

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE MEDICINA

DEPARTAMENTO DE SALUD PÚBLICA

SALUD PÚBLICA II

**LECTURAS BÁSICAS PARA EL APOYO
DE LAS UNIDADES DE ESTADÍSTICA
DESCRIPTIVA Y ANALÍTICA**

OCTUBRE 1996

I

Dr. Felix Morales Gonzalez

DIRECTORIO DEL DEPARTAMENTO

Dr. Leopoldo Vega Fraco
Jefe del Departamento

Dra. Alejandra Moreno Altamirano
Coordinadora de Investigación y Posgrado

Dr. Carlos Eslava Campos
Coordinador de Investigación del
Área Biomédica Básica

Dra. Isabel Cristina Morán Álvarez
Coordinadora del Primer Nivel de
Enseñanza

Dr. Juan José García García
Coordinador del Segundo Nivel de
Enseñanza

Ing. Alejandro Pérez Colín
Administrador

DIRECTORIO DEL DEPARTAMENTO

Dr. Leopoldo Vega Fraco
Jefe del Departamento

Dra. Alejandra Moreno Altamirano
Coordinadora de Investigación y Posgrado

Dr. Carlos Eslava Campos
Coordinador de Investigación del
Área Biomédica Básica

Dra. Isabel Cristina Morán Álvarez
Coordinadora del Primer Nivel de
Enseñanza

Dr. Juan José García García
Coordinador del Segundo Nivel de
Enseñanza

Ing. Alejandro Pérez Colín
Administrador

INTRODUCCIÓN

Con el propósito de facilitar el aprendizaje de la estadística en los estudiantes de medicina se ha preparado este material original que incluye, tanto el enfoque conceptual como el metodológico e instrumental, intencionalmente simplificado para facilitar su comprensión y aplicación.

El material que aquí se presenta corresponde a estadística descriptiva y analítica está basado en la solución de problemas como estrategia didáctica primordial responde a los ajustes previsto en la enseñanza de la estadística a fin de que el conocimiento de la metodología y su dominio fundamental permita su utilización, no sólo en el campo epidemiológico, sino en cualquier circunstancia que amerite la descripción o el análisis de un fenómeno estudiado con rigor científico.

Para facilitar su aplicación se incluye un plan de trabajo para cada una de las semanas con la especificación del material que se sugiere sea revisado en cada sesión.

**EL MATERIAL ORIGINAL FUE ELABORADO POR
DR. JESÚS REYNAGA OBREGON**

SELECCIÓN Y COMPILACIÓN

**DRA. ISABEL CRISTINA MORÁN ALVAREZ
COORDINADORA DEL PRIMER
NIVEL DE ENSEÑANZA**

**DR. JULIAN ALCALÁ RAMÍREZ
ASESOR DE LA COORDINACIÓN**

**LIC. ORALIA ACUÑA DAVILA
RESPONSABLE SALUD PÚBLICA II**

**ESPERANZA NÁPOLES GONZÁLEZ
RESPONSABLE SALUD PÚBLICA I**

**DOLORES HERNÁNDEZ GONZÁLEZ
SECRETARIA**

FACULTAD DE MEDICINA
DEPARTAMENTO DE SALUD PUBLICA

Materia: Salud Pública I I
Unidad temática: Estadística 1996-1997

Contenidos:

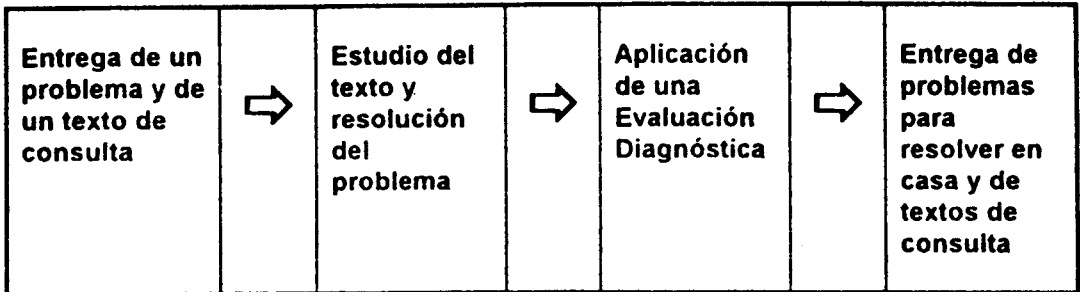
1. Relaciones entre la investigación científica y el método estadístico. Deducción de consecuencias verificables. Relación entre las consecuencias verificables y los diseños de investigación. Relaciones entre los diseños de investigación y los procedimientos estadísticos. Etapas del método estadístico. 33
2. Etapa de Recolección. Procedimientos de medición. Concepto de variables. Clasificación estadística de las variables. Escalas de modalidades y escalas de clases. 41-70
3. Etapa de Recuento. Procedimientos manuales y computarizados de recuento. 71
4. Etapa de Presentación: Selección, construcción e interpretación de cuadros estadísticos.
5. Etapa de Presentación: Selección, construcción e interpretación de gráficos estadísticos.
6. Etapa de Descripción: Selección, cálculo e interpretación de medidas de resumen para variables cualitativas. Medidas de uso común en salud pública.
7. Etapa de Descripción: Selección, cálculo e interpretación de medidas de resumen para variables cuantitativas en series simples.
8. Etapa de Descripción: Selección, cálculo e interpretación de medidas de resumen para variables cuantitativas en series agrupadas
9. Etapa de Descripción: Relación entre la desviación estándar y la curva normal. Características de la curva normal. Empleo de la tabla de áreas bajo la curva normal.
10. Etapa de Descripción: Cálculo e interpretación de medidas para la verificación del sesgo y la curtosis en series simples y agrupadas.
11. Etapa de Análisis: Relación entre diseños de investigación para la asociación de dos variables cualitativas o cuantitativas en un grupo, y las correspondientes pruebas de análisis estadístico.
12. Etapa de Análisis: Relación entre diseños de investigación para la comparación de dos grupos independientes o apareados, con variables cualitativas o cuantitativas, y las correspondientes pruebas de análisis estadístico.
13. Etapa de Análisis: Relación entre diseños de investigación para la comparación de tres o más grupos independientes o apareados, con variables cualitativas o cuantitativas, y las correspondientes pruebas de análisis estadístico.
1. Panorama del muestreo. Esquemas de muestreo. Fundamentos del muestreo aleatorio simple. Aplicaciones del muestreo en salud pública.

Salud Pública 1

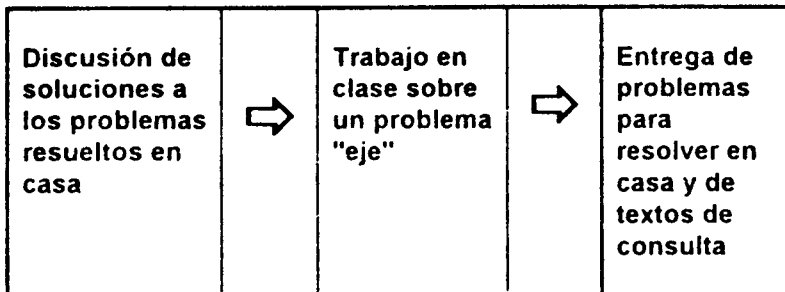
Unidad Temática Estadística 1993-1994

Esquema de trabajo

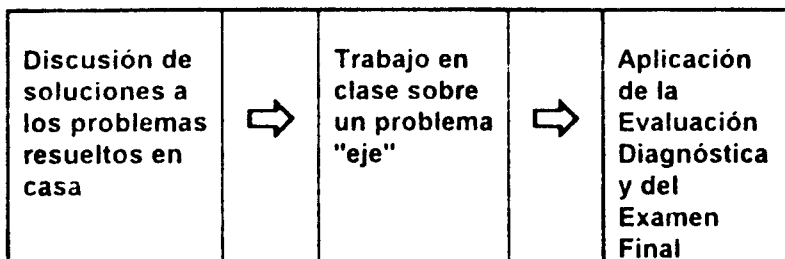
SESION 1



SESIONES SIGUIENTES



ULTIMA SESION



Plan de trabajo para la semana 1

Sesión 1

Identificar la importancia del manejo sistemático y ordenado de información constituye una primera etapa en el dominio de los métodos de la estadística aplicada.

A través de un problema práctico y cotidiano, el alumno debe reflexionar en la conveniencia de manejar información mediante el sentido común y la aplicación de operaciones aritméticas sencillas.

Durante esta sesión, el alumno utilizará los siguientes documentos:

- **Problemas: ¿Qué opina usted de esta información?** 1 página
- **Texto: ¿Padece de Insuficiencia Estadística?** 3 páginas

Con el propósito de medir el nivel de dominio de algunos procedimientos estadísticos estudiados con anterioridad, se aplicará al alumno una evaluación diagnóstica sin valor para calificación.

Finalmente, al término de la sesión, el alumno recibirá indicaciones para el empleo de los siguientes materiales:

- **Problema: La mortalidad de los médicos por cáncer pulmonar en relación con el hábito de fumar (versión 1)** 2 páginas
 - **Texto: La Investigación Científica y la Estadística** 10 páginas
-

Sesión 2

Resulta necesario adquirir habilidad para identificar las etapas del método científico para que, posteriormente, se puedan discutir las relaciones entre dicho método y los procedimientos de la estadística aplicada.

Para ello, en esta sesión se discutirán las soluciones planteadas por el alumno al problema *La mortalidad de los médicos por cáncer pulmonar en relación con el hábito de fumar (versión 1)* y se ampliarán conceptos relacionados con el método científico.

Al final de la sesión, el alumno recibirá indicaciones para el empleo de los siguientes materiales de trabajo para la primera sesión de la semana 2:

- **Problemas: ¿Son correctas las deducciones?** 1 página
 - **Problema: La rubeola y las malformaciones congénitas** 1 página
 - **Texto: Hipótesis y diseños de investigación (la necesidad de deducir consecuencias verificables)** 4 páginas
-

PROBLEMAS: ¿Qué opina usted de esta información?**Problema 1**

Se siguieron los precios de algunos artículos en una tienda de autoservicio a lo largo de 1991, 1992 y 1993. Se encontró lo siguiente:

En febrero de 1991 el frasco de 12 cápsulas Binotal 500 mgrs. costaba \$7,130; el kilo de melón chino costaba \$2,100 y el kilo de cecina de res costaba \$9,200

Hacia febrero de 1992 los precios de dichos artículos se encontraron en \$8,600, \$3,100 y \$13,500, respectivamente.

Para febrero de 1993 se encontró que el medicamento ya había llegado a los \$10,730, la fruta subió a \$4,500 y la cecina estaba en \$17,800.

En las siguientes líneas describa la información anterior de la manera más sintética y objetiva posible:

Problema 2

Ciento veinte personas acudieron a una comida de bodas. En contra de la tradición de preparar la comida en casa, los anfitriones compraron sopa de lata en la tienda de al lado y encargaron mole a una fonda recientemente establecida en la colonia y que prometía un esmerado servicio y un precio inmejorable.

De los ciento veinte invitados, ochenta y uno comieron sopa y el resto no. De los que comieron sopa sesenta y cuatro presentaron diarrea y el resto no. De los que no comieron sopa cinco presentaron diarrea y el resto no.

Por otra parte, de los ciento veinte asistentes a la comida, cincuenta y siete comieron mole y el resto no lo comieron. De los que comieron mole treinta y tres presentaron diarrea y el resto no. De los que no lo comieron treinta y seis enfermaron y el resto no.

En las siguientes líneas señale cuáles medidas de control tomaría usted y justifique breve y objetivamente su respuesta:

TEXTO: ¿Padece de Insuficiencia Estadística?

Introducción:

Si existiera tal enfermedad su definición sería la siguiente:

"Se denomina *Insuficiencia Estadística* a un padecimiento en el que existe un grave bloqueo de las capacidades para seleccionar, recoger, contar, presentar, resumir, analizar y comparar información cualitativa y cuantitativa.

Sus principales manifestaciones son: imposibilidad para entender literatura científica, dificultad para tomar decisiones objetivas e imposibilidad para participar en casi cualquier proyecto de investigación. Entre sus secuelas se encuentra el rezago en los estudios profesionales y la imposibilidad para realizar estudios de posgrado.

Las principales medidas preventivas contra la *Insuficiencia Estadística* consisten en la aplicación del sentido común y el uso de pequeñas dosis de aritmética, ambas orientadas por la utilización de un sencillo procedimiento llamado *Método Estadístico*.

Se sabe que los individuos menos propensos a padecer la enfermedad son aquellos que saben aceptar retos constructivos y que tienen un carácter perseverante".

Los siguientes párrafos consisten en una prueba auto suministrada para verificar si se es propenso a la enfermedad de referencia.

Si ante la pregunta que se plantea solo se puede responder con *no sé* o *voy a pensarlo* habrá que poner en práctica medidas de control urgentes.

En cambio si se responde de manera similar a la que se describe luego del siguiente relato el pronóstico es bueno.

Relato:

Un grupo de 89 alumnos de primer año de la Facultad de Medicina fue entrevistado en 1991 para averiguar tanto la orientación política predominante de sus padres como el tipo de educación que recibieron durante sus estudios de educación primaria.

Se encontró que en 32 casos la orientación política de los padres fue conservadora; de estos estudiantes siete recibieron educación de tipo flexible, diez de tipo moderada y 15 de tipo rígida.

En 30 estudiantes cuyos padres tenían orientación política de tipo moderada se encontró que nueve recibieron educación flexible, diez moderada y 11 rígida.

Finalmente, en el caso de los últimos 27 estudiantes, cuyos padres tenían orientación política de tipo liberal, se encontró que 14 recibieron educación de tipo flexible, ocho de tipo moderada y cinco de tipo rígida.

Planteamiento y desarrollo:

1o. Por una parte, se calificó a los estudiantes según la *orientación política de los padres* en alguna de las siguientes tres modalidades: *conservadora, moderada, liberal*; esto es, se midió una característica o variable de tipo cualitativo ya que la orientación política no es susceptible de cuantificar. Por otra parte, se observó en los estudiantes el *tipo de educación primaria recibida* y se hizo la siguiente clasificación: *flexible, moderada, rígida*; esto es, también se midió en los estudiantes otra característica o variable de tipo cualitativo.

En resumen, las características o variables seleccionadas y recogidas en este estudio fueron dos, siendo ambas de tipo cualitativo: *orientación política de los padres* y *tipo de educación primaria recibida*.

2o. El conteo de la información arrojó los siguientes resultados:

Total de estudiantes con padres con orientación política conservadora:	32	Total de estudiantes con padres con orientación política moderada:	30	Total de estudiantes con padres con orientación política liberal:	27
Estudiantes que recibieron educación primaria de tipo flexible:	7	Estudiantes que recibieron educación primaria de tipo flexible:	9	Estudiantes que recibieron educación primaria de tipo flexible:	14
Estudiantes que recibieron educación primaria de tipo moderada:	10	Estudiantes que recibieron educación primaria de tipo moderada:	10	Estudiantes que recibieron educación primaria de tipo moderada:	8
Estudiantes que recibieron educación primaria de tipo rígida	15	Estudiantes que recibieron educación primaria de tipo rígida	11	Estudiantes que recibieron educación primaria de tipo rígida	5

3o. Para comprender la información se puede presentar en un cuadro como el siguiente:

88 estudiantes de primer año, según orientación política de los padres y tipo de educación primaria recibida
Facultad de Medicina, 1991

Orientación política de los padres	Tipo de educación primaria recibida			Totales No. y (%)
	Flexible No. y (%)	Moderada No. y (%)	Rígida No. y (%)	
Conservadora	7 (22)	10 (31)	15 (47)	32 (100)
Moderada	9 (30)	10 (33)	11 (37)	30 (100)
Liberal	14 (52)	8 (30)	5 (18)	27 (100)

3o., 4o. y 5o. El resumen, análisis y comparación de la información pueden llevar a una conclusión como la siguiente:

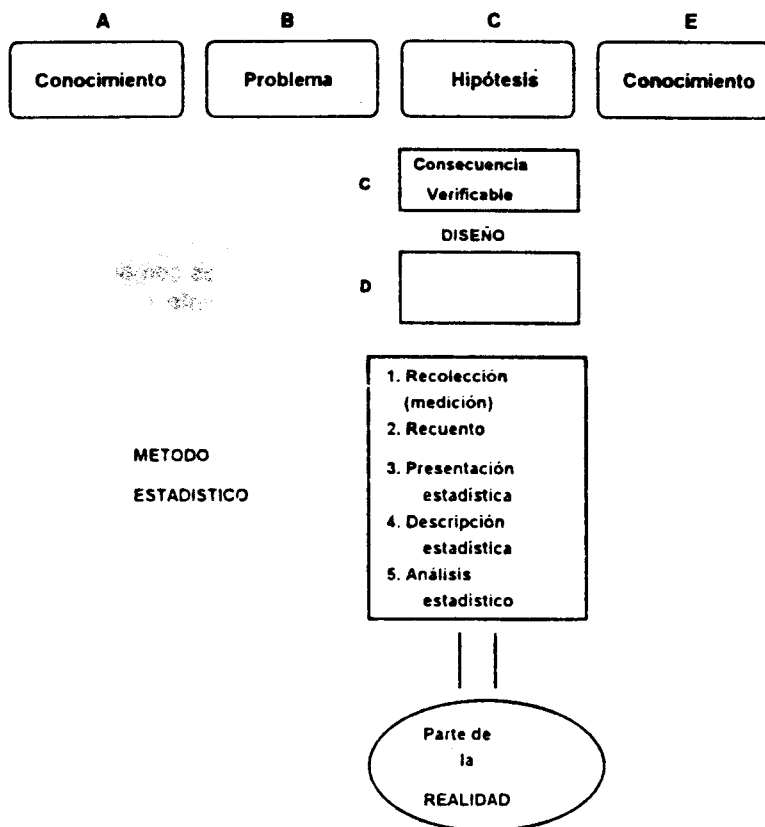
Parece existir asociación entre la orientación política de los padres y el tipo de educación primaria recibida por los 88 estudiantes de primer año de medicina.

Esta conclusión, que es de aplicarse principalmente a los 88 alumnos estudiados y no necesariamente refleja a todos los demás alumnos, se fundamenta en que:

- los hijos de padres con orientación política conservadora que recibieron educación primaria de tipo rígida (47 %) predominaron sobre los sometidos a los otros dos tipos de educación primaria ,
- en el caso de hijos de padres con orientación política liberal más de la mitad de ellos (52%) recibieron educación primaria de tipo flexible;
- finalmente, en el caso de los hijos de padres con orientación política moderada no pareció haber un predominio contundente de alguno de los tres tipos de educación primaria recibida.

PROBLEMA : La mortalidad de los médicos por cáncer pulmonar en relación con el hábito de fumar

A manera de recordatorio, en el siguiente esquema, con las letras A a la E se muestran las etapas del *Método Científico* y con los números 1. al 5. se muestran las etapas del *Método Estadístico* :



Con los mismo sistema de enumeración empleado en el esquema anterior, identifique las etapas del *Método Científico* que se encuentren presentes en el siguiente relato y anote las letras correspondientes en los recuadros de los párrafos (algunos recuadros pueden quedar vacíos).

Al revisar las estadísticas de enfermedades y defunciones por enfermedades pulmonares se encontró que casi todos los casos de enfisema pulmonar y varios de cáncer pulmonar tenían en muchas ocasiones el antecedente de tabaquismo crónico.

Sin saber el por qué de tal relación, en un estudio de numerosos casos de cáncer pulmonar se observó que alrededor del 70% de los enfermos eran antiguos fumadores

- Esto hizo pensar que el consumo de tabaco pudiera ser un factor favorecedor o causal del cáncer pulmonar.
- Si lo anterior fuera cierto, podría inferirse lo siguiente: a) Un grupo de fumadores de determinada edad, sexo y nivel socioeconómico presentaría un número mayor de muertes por cáncer pulmonar que un grupo de no fumadores de edad, sexo y nivel socioeconómico similares y b) La frecuencia de cáncer pulmonar sería mayor al aumentar el número de cigarrillos consumidos.
- Luego de discutir si convenía formar un grupo de médicos fumadores y seguirlo a lo largo de varios años o formar un grupo de médicos fumadores y uno de no fumadores, se resolvió hacer la investigación enviando un cuestionario a un único pero numeroso grupo de médicos del Reino Unido.
- De ellos, se solicitaron datos individuales sobre, edad y hábitos de fumar por medio del cuestionario de referencia. Se establecieron además conexiones con el Registro de Defunciones para asegurar la información sobre la mortalidad por cáncer pulmonar ocurrida en médicos.
- Los datos se mostraron solamente en forma de tablas.
- El resumen de la información obtenida se expresó en tasas de mortalidad por cada mil médicos.
- Los resultados mostraron que el riesgo de morir fue 14 veces mayor en fumadores que en no fumadores. También se observó que en los fumadores la frecuencia de éstas muertes aumentó cuando la cantidad de tabaco que se consumía era mayor.

TEXTO: La Investigación Científica y la Estadística (1)

1.0 INTRODUCCION

El presente material tiene como propósito facilitar el inicio de estudios de bioestadística dentro del campo de la salud. Su elaboración surgió al observar la existencia del problema de que en la mayoría de los textos de estadística que se encuentran a disposición de los estudiantes de medicina ocurre que, desde las primeras páginas, son presentadas diversas técnicas estadísticas sin preocuparse por señalar el lugar que éstas ocupan en la investigación científica.

Lo anterior ocasiona que el estudiante se sienta confundido sobre la importancia que pueda tener el conocimiento de la estadística para la investigación en las ciencias de la salud y que termine por conceder poco aprecio a esa importante disciplina o por suponer que la utilización abundante e indiscriminada de las técnicas estadísticas le facilitará la culminación exitosa de cualquier investigación.

Por ello, se pretende proporcionar en forma resumida una visión de las relaciones entre el método científico, la investigación y la estadística y facilitar al estudiante la construcción de un esquema mental propio de la ubicación de la metodología y procedimientos de la estadística dentro del proceso de la investigación.

Debe mencionarse que gran parte de las características de este material fueron consecuencia de las sugerencias del E.S.P. Jaime S. García Romero y de la doctora Blanca de Garay Gómez del Villar tras de haberlo probado con los alumnos de sus Cursos de Bioestadística para alumnos de medicina en la Facultad de Medicina de la UNAM.

Dr. Jesús Reynaga Obregón

1980

¹ Extracto del manuscrito original del documento *La investigación científica y la estadística* del Departamento de Medicina Social, Medicina Preventiva y Salud Pública, Facultad de Medicina, UNAM, México, 1980

2.0 LAS EXPLICACIONES DE LA CIENCIA

2.1. ¿Qué es el conocimiento científico?

Se entiende por conocimiento a la posesión de datos de los fenómenos materiales y espirituales.

El conocimiento científico se entiende como la posesión de explicaciones objetivas y confirmadas, pero siempre verificables, de los procesos existentes en el universo.

El conocimiento científico, al ser producto del razonamiento, constituye en esencia el reflejo mental producido por los procesos existentes y representa los enlaces reales entre los componentes de los fenómenos.

Considerando que los fenómenos existen independientemente de la forma en que las personas los conozcan o los imaginen, hay la posibilidad de verificar los datos referentes a dichos fenómenos.

A manera de ejemplo: el fenómeno de la desnutrición existe en el mundo y está condicionado mayormente por factores de tipo económico y escasamente por factores de tipo moral. Entonces independientemente de que sea un religioso, un sociólogo o un biólogo quien explique dicho fenómeno, la desnutrición se presenta condicionada en la realidad por los factores mencionados; lo que cambia es la forma de imaginar el fenómeno.

En este ejemplo el conocimiento científico referente al condicionamiento económico de la desnutrición ha sido confirmado y puede ser verificado cuantas veces sea necesario para complementarlo y relacionarlo con otros conocimientos.

No todo el conocimiento es científico; por ejemplo: la relación causa-efecto que algunas personas establecen entre la depresión de la fontanela y la diarrea que sufren los niños. Es claro que resulta imposible probar que la diarrea es causada por el hundimiento de la fontanela; sin embargo la observación de que estas manifestaciones se presentan casi siempre juntas ha generado un conocimiento de tipo no científico que se transmite por experiencia, generación tras generación.

Los animales también conocen características de la naturaleza y de algunos fenómenos que en ella se presentan, pero este conocimiento tampoco es científico.

2.2. ¿Cómo surge el conocimiento científico?

El conocimiento se origina en los procesos que ocurren en el universo, de tal manera que el universo y sus procesos constituyen la base para comprobar el conocimiento. Como ya se mencionó los fenómenos existen objetiva e independientemente de la forma en que los humanos los imaginen.

Con mayor especificidad, el conocimiento científico tiene origen en las diversas actividades que el hombre realiza; y el mismo ser humano forma parte del universo.

En tanto que los fenómenos del universo existen, el conocimiento se presenta como resultado de la actividad humana. Al igual que se han ido desarrollando y diversificando las actividades del ser humano a lo largo de la historia, también el conocimiento científico se ha ampliado,

diversificado y actualizado.

La gran variedad de actividades del ser humano procura la satisfacción de sus necesidades. Dicha satisfacción de necesidades requiere del trabajo para la elaboración de bienes y para la realización de servicios, además de su distribución entre los integrantes de la sociedad.

Las modalidades que adopta la producción y la distribución y las relaciones sociales que se establecen entre los hombres para realizar estos procesos son las que permiten definir las características de la sociedad; así mismo son las que permiten conocer el grado de desarrollo social y al mismo tiempo influyen en los propios cambios sociales.

De acuerdo a lo anterior, los cambios sociales que ocurren a causa de la satisfacción de las necesidades humanas obligan a la adquisición de conocimientos científicos sobre los fenómenos del mundo y éstos, a su vez, tienen influencia sobre el progreso social.

*Cuanto más avanza el dominio del hombre sobre el mundo, mayor resulta ser la productividad del trabajo humano; y, a la vez, el incremento de la productividad del trabajo acaba por provocar cambios en la organización social. Por su parte, los cambios sociales influyen poderosamente en el avance de la ciencia o, lo que es lo mismo, en el dominio del hombre sobre el mundo. Por tanto, la ciencia no existe por sí misma ni puede separarse de las otras actividades humanas, sino que es un producto de la vida social del hombre y, al mismo tiempo, ejerce una acción definida sobre la sociedad. De este modo, la ciencia sólo puede entenderse en función del desenvolvimiento histórico de la sociedad en su conjunto.*²

El conocimiento científico puede ser, en ocasiones, limitado, ya que depende de las condiciones en que ha sido logrado, Sin embargo, estas condiciones no son invariables. Por lo contrario, se modifican constantemente y cada nuevo conocimiento adquirido establece nuevas posibilidades para el mejoramiento de las propias condiciones en que ha de adquirirse más conocimiento.

*En consecuencia, los límites del conocimiento se ensanchan con el avance del conocimiento mismo.*³

Un ejemplo de lo mencionado está constituido por el enorme incremento de los conocimientos en cirugía surgidos a raíz de la necesidad de atender a los heridos en los conflictos armados. Dichos conocimientos han ido aplicándose para el tratamiento de padecimientos que no tiene como único origen las lesiones de guerra.

Ejemplos como el anterior pueden citarse en casi todos los campos de la actividad humana y resulta fácil comprobar como el surgimiento de conocimientos es producto de las actividades humanas y como ese mismo conocimiento ha provocado cambios en las actividades y procesos sociales.

² Farrington, B. La Ciencia Griega Buenos Aires, Hachette, 1957, p 25-27.

³ De Gortari, E. Introducción a la Lógica Dialéctica. México, Grijalbo 1979, p 18

2.3 ¿Cómo logra el ser humano el conocimiento?

Ya se ha mencionado que el conocimiento es la posesión de datos acerca de los fenómenos del universo. Esta posesión necesita de cuatro factores para realizarse: un sujeto pensante, el proceso psíquico del pensar, el ente o fenómeno pensado y la expresión de lo pensado.

Considerando los factores mencionados, se reitera la idea de que el pensamiento se refiere a la reflexión de los procesos del universo en la mente humana. Pero no todos los pensamientos son iguales o tienen la misma categoría.

La forma más simple del pensamiento es el concepto, el cual se refiere a las propiedades esenciales de un objeto sin afirmar o negar nada de él. Su expresión es el término (por ejemplo: sol, tranquilidad, once).

Otro tipo de razonamiento es el llamado analógico, el cual parte de proposiciones particulares para lograr otra proposición particular. Es decir, después de reconocer semejanzas comunes a dos o más objetos, se concluye la existencia de características comunes para otros objetos similares (por ejemplo: la úlcera duodenal fácilmente produce dolor epigástrico; la úlcera gástrica también produce dolor epigástrico frecuentemente; el cáncer gástrico tiene muchas manifestaciones similares a las de las úlceras duodenales y gástricas; entonces, el cáncer gástrico produce dolor epigástrico).

Es importante señalar que tanto el razonamiento deductivo como el inductivo y el analógico son formas o estructuras del pensamiento que se complementan entre sí. Dichas formas y las reglas o leyes que las fundamentan son el objeto de estudio de la lógica.

Por otra parte, los elementos, procesos o entes que se están conociendo, son el objeto de estudio de cada ciencia particular. Así, los fenómenos relacionados con la existencia de la vida en la tierra son el objeto de la biología y las relaciones que se establecen entre los grupos sociales son el objeto de la sociología.

2.4 ¿Qué es la Ciencia?

*Por ciencia se entiende la explicación objetiva y racional del universo. Como explicación, la ciencia describe las diversas formas en que se manifiestan los procesos existentes, distingue las fases sucesivas y coexistentes observadas en el desarrollo de los mismos procesos, desentraña sus enlaces internos y sus conexiones con otros procesos, pone al descubierto las acciones reciprocas entre los procesos y encuentra las condiciones y los medios necesarios para permitir la intervención humana en el curso de los propios procesos.*⁴

En forma sintética podría entenderse por ciencia al conjunto sistemático de conocimientos racionales, exactos, verificables y perfectibles que explican los procesos del universo.

2.5 Características de la Ciencia

La ciencia posee notables características; una primera es el control práctico que ella permite sobre la naturaleza. Como ejemplo pueden citarse las innumerables contribuciones que la ciencia ha hecho para la prevención de las enfermedades y la conservación de la salud. Sin embargo, no debe confundirse a la ciencia con los adelantos tecnológicos que de ella se

⁴ De Gortari, E.: Op. Cit., p.13

derivan; una cosa son los conocimientos de la ciencia con respecto a la relatividad y otra cosa es su aplicación en la fabricación de armas atómicas.

Un segundo rasgo sobresaliente de la ciencia destaca cuando se le considera como dirigida a la obtención de más conocimientos sistemáticos y confiables, los cuales son empleados para garantizar conclusiones válidas acerca de la forma y condiciones en que se presentan diversos fenómenos.

Una tercera característica fundamental de la ciencia es su posesión de un método. Esta es una de las características más permanentes y que garantiza la confiabilidad de sus conclusiones. El método de la ciencia está constituido por fases generales que, mediante normas lógicas, permite la adquisición del conocimiento científico.

3.0 EL METODO CIENTIFICO

3.1 ¿Qué es el método?

El método es considerado como la manera razonada de conducir los procesos del pensamiento con objeto de llegar a un resultado determinado y, preferentemente, al descubrimiento de la verdad⁵.

3.2 ¿Qué es el método científico?

De acuerdo a la definición anterior, el método científico persigue la solución de problemas mediante el empleo de los diversos tipos de razonamiento. Es en el método científico en donde los procesos del pensamiento encuentran su más adecuada expresión al través de enunciados y argumentos que explican los fenómenos del universo.

El método científico es todo un procedimiento formado por una secuencia lógica de actividades que procura descubrir las características de los fenómenos, las relaciones internas entre sus elementos y sus conexiones con otros fenómenos, mediante el raciocinio y la comprobación a través de la demostración y la verificación.

Cada clase de problemas obliga a la aplicación de diferentes técnicas de acuerdo a los fenómenos inherentes a una u otra ciencia.

No debe confundirse a la técnica como sinónimo de método. Una técnica es un conjunto de acciones aplicables a diversos fines; en cambio, el método es un proceso compuesto por varias fases que se aplican al ciclo entero de obtención del conocimiento. Por ejemplo, el diagnóstico de las enfermedades requiere de la aplicación del método clínico, en el cual se emplean diversas técnicas como son el interrogatorio, la inspección, la auscultación y la percusión.

Al igual que el conocimiento científico, el método científico ha evolucionado a través del tiempo, especialmente en lo referente al predominio que en él se ha observado de los diversos tipos de razonamientos. Recuérdese el enfoque deductivo empleado por Aristóteles para la obtención de conocimiento, la influencia de la inducción en la época de Bacon o las aplicaciones dialécticas de Marx a partir de los conceptos Hegelianos.

3.3 ¿Cómo es el método científico?

El método científico parte de conocimientos previos para llegar a conocimientos nuevos.

Existen conocimientos referentes a entes abstractos, ideales, que sólo existen en la mente humana; conocimientos que están representados por signos y relaciones entre ellos, por ejemplo los números y los símbolos que sirven para asociarlos en las operaciones aritméticas. Estos conocimientos son objeto de las ciencias formales⁶ (lógica matemática).

Por otro lado, existen conocimientos referentes a sucesos, procesos y objetos que existen fuera de la mente humana (peso, talla, volumen); estos conocimientos son objeto de las ciencias

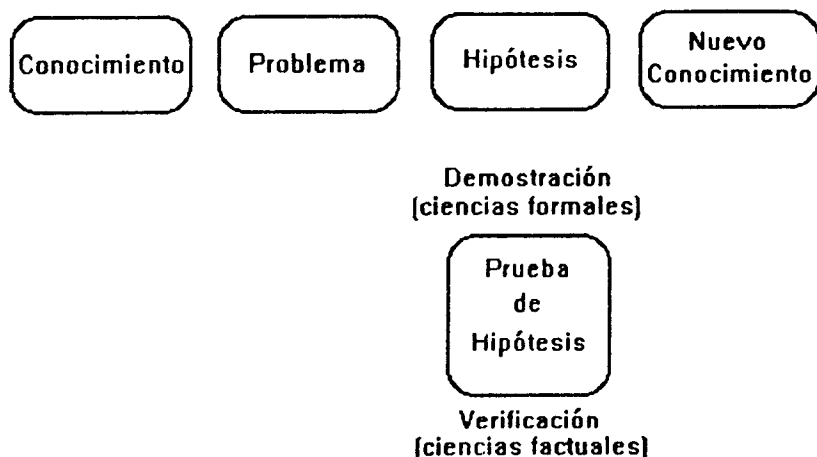
⁵ Diccionario Léxico Hispano, 4a. ed., México, W M. Jaskson, 1978.

⁶ Bunge, M.: La ciencia, su método y su filosofía. Buenos Aires, Siglo Veinte, 1975, p. 9-36

factuales⁷, como por ejemplo biología y física.

Ambos tipos de conocimientos y ciencias se diferencian para fines de comprensión, pero en la realidad se complementan y mezclan en la vida cotidiana y, como ya se señaló, dan lugar a la adquisición de conocimientos nuevos mediante la aplicación del método científico.

Con el propósito de buscar una más fácil descripción del método científico y relacionarlo con la investigación y la estadística, se inserta en seguida un esquema que auxiliará en explicaciones posteriores:



3.3.1 Los problemas científicos

Como ya se mencionó anteriormente, los fenómenos del universo y las actividades del ser humano dan origen al conocimiento. Las características del conocimiento dependen del grado de desarrollo social, pero estos no son estáticos ni limitados, sino que avanzan continuamente influenciándose mutuamente.

Ocurre que el conocimiento que en una época es adecuado para explicar los procesos que el hombre conoce llega a ser insuficiente para explicar algún fenómeno observado; el ser humano trata de encontrar las relaciones entre los elementos del fenómeno tal como está acostumbrado a hallarlas, pero existen discordancias entre dichas relaciones; requiere modificar alguna situación con los conocimientos que posee y no le es posible; el camino para lograr su meta presenta algún obstáculo y sus patrones fijos de conducta o su razonamiento cotidiano no le permiten salvar ese obstáculo; delibera y toma conciencia de la dificultad; surge entonces su problema.

En general, al enfrentarse el hombre a situaciones inhabituales para las que no tiene conocimientos suficientes surge una dificultad; si esa dificultad no se puede resolver con lo que se sabe o con las conductas instintivas, se está ante un problema.

Un problema científico se plantea basándose en los conocimientos científicos existentes y se estudia por medios científicos con el propósito fundamental de aumentar el conocimiento.

En todo problema científico pueden apreciarse tres aspectos fundamentales: el propio problema, es decir la explicación requerida; la pregunta; esto es, el proceso mental de entender que se

⁷ Bunge, M.: La ciencia, su método y su filosofía. Buenos Aires, Siglo Veinte, 1975, p. 9-36

desea saber algo; y finalmente, la expresión del problema, es decir el planteamiento oral o escrito en términos claramente definidos.

Algunos problemas lo son solamente en apariencia (pseudo problemas) o no lo son en absoluto; es posible que lo que ocurra es que se carezca del conocimiento o no se haya establecido contacto con él, aunque dicho conocimiento ya exista.

Para proceder al planteamiento de problemas se deben realizar diversas actividades entre las que se pueden citar: redacción adecuada y comprensible del problema, ubicación del problema dentro de teorías existentes, previsión de estrategias posibles para su resolución, descomposición en sus partes elementales, búsqueda de analogía con otros problemas conocidos, simplificación y simbolización en términos matemáticos y lógicos.

3.3.2 Las hipótesis científicas

Si la existencia en el universo fuera tan constante que sólo se tuvieran que usar los mismos conocimientos una y otra vez y todas las actividades debieran realizarse en forma automática, sólo se tendría que recurrir al uso de la memoria y de las habilidades adquiridas; pero la existencia de problemas obliga al ser humano a buscar explicaciones satisfactorias.

La adecuación de las explicaciones siempre tendrá una vigencia limitada ya que los procesos del universo son constantemente cambiantes.

Precisamente es la necesidad humana de poseer esas explicaciones la que da lugar a las hipótesis. Una hipótesis es un supuesto, una explicación probable, una anticipación; y como tal debe ser comprobada. Precisamente, esa característica de respuesta anticipada hace que la hipótesis oriente y conduzca al pensamiento y a la acción hacia ciertos campos que permitirán juzgar la veracidad o falsedad de las explicaciones.

*Una hipótesis es una suposición que permite establecer relaciones entre hechos. El valor de una hipótesis radica en su capacidad para establecer esas relaciones entre los hechos y explicar por que se producen.*⁸

La peculiaridad característica de la hipótesis reside en que sistematiza el conocimiento científico y forma un sistema de abstracciones. La hipótesis es un sistema de juicios, conceptos y razonamientos sintetizados y ordenados con el propósito de explicar el fenómeno que ha de investigarse.

La construcción de una hipótesis parte del conocimiento de los fenómenos de la naturaleza y la sociedad y del conocimiento ordenado en forma de leyes y teorías, sus explicaciones son comprobadas y los resultados de esa comprobación se emplean para formular y completar nuevas leyes o teorías.

La realidad existente es la base, el fin y el criterio para decidir la veracidad de una hipótesis.

Aunque pueda adoptar modalidades diferentes de expresión en las diferentes ciencias, la hipótesis es la guía y orientación por excelencia para captar y poseer el conocimiento que a cada ciencia corresponde.

Todas las teorías existentes han pasado por fases más o menos largas de hipótesis y, de hecho,

⁸ López Cano, J L : Método e hipótesis científica (parte 2) México, ANUIES, 1975, p. 29

la causa de la existencia de las teorías se explican por la existencia de hipótesis. Si los científicos no formularan hipótesis ni se dedicaran a probarlas, no existirían nuevas leyes ni teorías, inclusive, no aparecerían nuevas hipótesis.

Una hipótesis que ha sido probada, ya sea que se acepte o rechace, genera nuevo conocimiento que permite explicar fenómenos que no eran cabalmente comprendidos y solucionar problemas no superados. Este nuevo conocimiento determina cambios en la concepción que el hombre tenga del universo; lo cual, a su vez, acarrea cambios en las relaciones sociales que terminarán por volver a plantear nuevos problemas que habrán de requerir otras hipótesis; continuándose así, permanentemente, la conquista de conocimientos a través del método científico.

3.3.3. Las Consecuencias verificables de las hipótesis

Por la razón de tener como objeto de estudio a los hechos, en la ciencias factuales la comprobación de las hipótesis implica la necesidad de deducir la forma en que se encontrarán los hechos en el caso de que las hipótesis sean ciertas.

La deducción de los arreglos de hechos o los acontecimientos que habrán de observarse si la hipótesis es verdadera es una condición fundamental si, como es de esperarse, se desea proceder a probar dicha hipótesis y no quedarse en la etapa del puro planteamiento.

Por otra parte, la anticipación de los hallazgos que han de encontrarse, si la hipótesis es cierta, facilita al ser humano la labor de investigar científicamente los fenómenos de la naturaleza, pues el señalamiento de las consecuencias verificables que se han deducido, orienta hacia el tipo de información que ha de recogerse y los procedimientos que han de realizarse con ella.

A manera de ejemplo, considérese la siguiente hipótesis:

La mortalidad infantil depende más estrechamente del ingreso económico que de la existencia de servicios médicos de urgencia.

Como puede observarse no es claro el sentido de la dependencia entre la mortalidad infantil, el ingreso económico y los servicios médicos de urgencia. Además, la comprobación de esta hipótesis resulta de tal manera difícil o puede adoptar tal número de modalidades, que la comparación de los resultados de las investigaciones realizadas por diferentes investigadores sería poco menos que imposible.

Si a partir de la hipótesis mencionada se deduce alguna consecuencia tal como la siguiente:

Si la hipótesis es cierta, entonces: las poblaciones con igual número y tipo de servicios médicos de urgencia tendrán una más alta tasa de mortalidad infantil mientras más bajo sea el ingreso económico promedio

fácilmente podrá decidirse cuáles tipos de información habría que recoger en una investigación e, inclusive, cuál sería el manejo, presentación y descripción que habría que realizar con dicha información.

3.4. Resumen

Concretando, el método científico permite la adquisición de conocimiento racional y ordenado a partir de los siguientes elementos: insuficiencia del conocimiento, planteamiento y delimitación de un problema, elaboración de hipótesis y prueba de dicha hipótesis.

Un atractivo ejemplo del ciclo de adquisición de conocimiento, que muestra las etapas del método científico, está constituido por los descubrimientos del ginecólogo húngaro Ignaz Philipp Semmelweiss (1818-1865).⁹

Al terminar sus estudios en la Universidad de Viena, Semmelweiss ingresó como médico asistente en la sección de maternidad del Hospital General de la Ciudad. Rápidamente adquirió fama de extravagante y fastidioso por su insistente clamor ante el gran número de fallecimientos de las parturientas. Ante sus preguntas acerca de la causa de esas muertes, la respuesta que le daban sus superiores era siempre la misma: un miasma invisible atacaba a las mujeres.

Semmelweiss había observado que tan solo en el año 1846 se presentaron 451 muertes en la primera sección de maternidad, mientras que en la segunda sección sólo fallecieron 90 madres. Semmelweiss encontraba insuficiente una teoría en la que ocurría que un miasma se ensañaba en la primera sección de maternidad y era condescendiente con la segunda. Empezó a centrar su atención en las características del problema; siempre iniciaba su trabajo, rodeado de alumnos, practicando disecciones de los cadáveres de las parturientas fallecidas el día anterior. Luego, siempre acompañado de sus discípulos, realizaba la revisión de las pacientes de la primera sección. La segunda sección de maternidad, donde la muerte era menos asoladora, era atendida por parteras que jamás ponían el pie en el anfiteatro de anatomía.

Semmelweiss supuso que él mismo y sus alumnos acarreaban alguna sustancia o germen desde el anfiteatro hasta las parturientas. En 1847 dio órdenes de que cualquier persona que estuviera en el anfiteatro, realizando alguna disección, debía lavarse escrupulosamente las manos y desinfectarlas con cloruro de calcio. Tomadas estas medidas, la mortalidad debida a la fiebre puerperal que en mayo de aquel año era de 12.24 por ciento, llegó a desaparecer casi por completo al cabo de dos años mientras que en la segunda sección se mantenía estacionaria.

El conocimiento logrado por Semmelweiss, que de ninguna manera fue casual o por inspiración divina sino producto de una necesidad humana (disminuir un problema de la población), produjo importantes repercusiones en la atención médica de épocas posteriores; específicamente, inició los importantes trabajos sobre antisepsia y asepsia.

⁹ Tomado y resumido de: Margota, Roberto: Historia de la Medicina. México, Novaro, 1972, pp. 267-268.

PROBLEMAS: ¿Qué opina usted de esta información?**Problema 1**

Se siguieron los precios de algunos artículos en una tienda de autoservicio a lo largo de 1991, 1992 y 1993. Se encontró lo siguiente:

En febrero de 1991 el frasco de 12 cápsulas Binotal 500 mgrs. costaba \$7,130; el kilo de melón chino costaba \$2,100 y el kilo de cecina de res costaba \$9,200

Hacia febrero de 1992 los precios de dichos artículos se encontraron en \$8,600, \$3,100 y \$13,500, respectivamente.

Para febrero de 1993 se encontró que el medicamento ya había llegado a los \$10,730, la fruta subió a \$4,500 y la cecina estaba en \$17,800.

En las siguientes líneas describa la información anterior de la manera más sintética y objetiva posible:

Problema 2

Ciento veinte personas acudieron a una comida de bodas. En contra de la tradición de preparar la comida en casa, los anfitriones compraron sopa de lata en la tienda de al lado y encargaron mole a una fonda recientemente establecida en la colonia y que prometía un esmerado servicio y un precio inmejorable.

De los ciento veinte invitados, ochenta y uno comieron sopa y el resto no. De los que comieron sopa sesenta y cuatro presentaron diarrea y el resto no. De los que no comieron sopa cinco presentaron diarrea y el resto no.

Por otra parte, de los ciento veinte asistentes a la comida, cincuenta y siete comieron mole y el resto no lo comieron. De los que comieron mole treinta y tres presentaron diarrea y el resto no. De los que no lo comieron treinta y seis enfermaron y el resto no.

En las siguientes líneas señale cuáles medidas de control tomaría usted y justifique breve y objetivamente su respuesta:

Planteamiento y desarrollo:

1o. Por una parte, se calificó a los estudiantes según la *orientación política de los padres* en alguna de las siguientes tres modalidades: *conservadora, moderada, liberal*; esto es, se midió una característica o variable de tipo cualitativo ya que la orientación política no es susceptible de cuantificar. Por otra parte, se observó en los estudiantes el *tipo de educación primaria recibida* y se hizo la siguiente clasificación: *flexible, moderada, rígida*; esto es, también se midió en los estudiantes otra característica o variable de tipo cualitativo.

En resumen, las características o variables seleccionadas y recogidas en este estudio fueron dos, siendo ambas de tipo cualitativo: *orientación política de los padres* y *tipo de educación primaria recibida*.

2o. El conteo de la información arrojó los siguientes resultados:

Total de estudiantes con padres con orientación política conservadora:	32	Total de estudiantes con padres con orientación política moderada:	30	Total de estudiantes con padres con orientación política liberal:	27
Estudiantes que recibieron educación primaria de tipo flexible:	7	Estudiantes que recibieron educación primaria de tipo flexible:	9	Estudiantes que recibieron educación primaria de tipo flexible:	14
Estudiantes que recibieron educación primaria de tipo moderada:	10	Estudiantes que recibieron educación primaria de tipo moderada:	10	Estudiantes que recibieron educación primaria de tipo moderada:	8
Estudiantes que recibieron educación primaria de tipo rígida	15	Estudiantes que recibieron educación primaria de tipo rígida	11	Estudiantes que recibieron educación primaria de tipo rígida	5

3o. Para comprender la información se puede presentar en un cuadro como el siguiente:

88 estudiantes de primer año, según orientación política de los padres y tipo de educación primaria recibida
Facultad de Medicina, 1991

Orientación política de los padres	Tipo de educación primaria recibida			Totales No. y (%)
	Flexible No. y (%)	Moderada No. y (%)	Rígida No. y (%)	
Conservadora	7 (22)	10 (31)	15 (47)	32 (100)
Moderada	9 (30)	10 (33)	11 (37)	30 (100)
Liberal	14 (52)	8 (30)	5 (18)	27 (100)

TEXTO: ¿Padece de *Insuficiencia Estadística*?**Introducción:**

Si existiera tal enfermedad su definición sería la siguiente:

"Se denomina *Insuficiencia Estadística* a un padecimiento en el que existe un grave bloqueo de las capacidades para seleccionar, recoger, contar, presentar, resumir, analizar y comparar información cualitativa y cuantitativa.

Sus principales manifestaciones son: imposibilidad para entender literatura científica, dificultad para tomar decisiones objetivas e imposibilidad para participar en casi cualquier proyecto de investigación. Entre sus secuelas se encuentra el rezago en los estudios profesionales y la imposibilidad para realizar estudios de posgrado.

Las principales medidas preventivas contra la *Insuficiencia Estadística* consisten en la aplicación del sentido común y el uso de pequeñas dosis de aritmética, ambas orientadas por la utilización de un sencillo procedimiento llamado *Método Estadístico*.

Se sabe que los individuos menos propensos a padecer la enfermedad son aquellos que saben aceptar retos constructivos y que tienen un carácter perseverante".

Los siguientes párrafos consisten en una prueba auto suministrada para verificar si se es propenso a la enfermedad de referencia.

Si ante la pregunta que se plantea solo se puede responder con *no sé* o *voy a pensarlo* habrá que poner en práctica medidas de control urgentes.

En cambio si se responde de manera similar a la que se describe luego del siguiente relato el pronóstico es bueno.

Relato:

Un grupo de 89 alumnos de primer año de la Facultad de Medicina fue entrevistado en 1991 para averiguar tanto la orientación política predominante de sus padres como el tipo de educación que recibieron durante sus estudios de educación primaria.

Se encontró que en 32 casos la orientación política de los padres fue conservadora; de estos estudiantes siete recibieron educación de tipo flexible, diez de tipo moderada y 15 de tipo rígida.

En 30 estudiantes cuyos padres tenían orientación política de tipo moderada se encontró que nueve recibieron educación flexible, diez moderada y 11 rígida.

Finalmente, en el caso de los últimos 27 estudiantes, cuyos padres tenían orientación política de tipo liberal, se encontró que 14 recibieron educación de tipo flexible, ocho de tipo moderada y cinco de tipo rígida.

3o., 4o. y 5o. El resumen, análisis y comparación de la información pueden llevar a una conclusión como la siguiente:

Parece existir asociación entre la orientación política de los padres y el tipo de educación primaria recibida por los 88 estudiantes de primer año de medicina.

Esta conclusión, que es de aplicarse principalmente a los 88 alumnos estudiados y no necesariamente refleja a todos los demás alumnos, se fundamenta en que:

- los hijos de padres con orientación política conservadora que recibieron educación primaria de tipo rígida (47 %) predominaron sobre los sometidos a los otros dos tipos de educación primaria ,
- en el caso de hijos de padres con orientación política liberal más de la mitad de ellos (52%) recibieron educación primaria de tipo flexible;
- finalmente, en el caso de los hijos de padres con orientación política moderada no pareció haber un predominio contundente de alguno de los tres tipos de educación primaria recibida.

PROBLEMAS: ¿Son correctas las deducciones?

Para cada uno de los tres siguientes problemas se requiere identificar si las consecuencias verificables que se han deducido a partir de sus correspondientes hipótesis son correctas o incorrectas. Es posible que una, otra o ambas deducciones sean correctas o incorrectas.

Justifique sus respuestas.

PROBLEMA 1

Hipótesis general de la investigación: Debido a la distinta forma en que son tratados a lo largo del tiempo la gran mayoría de los * y la gran mayoría de los C, es probable que los * terminen siendo más altos hacia el final de su existencia.

Consecuencia No. 1: Si la hipótesis es cierta, entonces, al comparar un grupo de * con un grupo de C, estando ambos hacia el final de su existencia, se esperaría que el promedio de estatura de ambos grupos sea diferente.

Consecuencia No. 2: Si la hipótesis es cierta, entonces al estudiar el mecanismo de crecimiento de ambos grupos se esperaría encontrar que la sustancia aceleradora W-237 posee diferentes concentraciones a diversas horas del día.

Consecuencia No. 3: Si la hipótesis es cierta, entonces al tratar de una misma forma a lo largo del tiempo tanto a un grupo de * como a un grupo de C, se esperaría que los promedios de estatura de ambos grupos sean iguales hacia el final de su existencia.

PROBLEMA 2

Hipótesis general de la investigación: Es probable que la desnutrición infantil esté más intensamente influida por el nivel de ingresos familiares que por el nivel educativo familiar.

Consecuencia No. 1 Si la hipótesis es cierta, entonces, al comparar dos grupos niños: uno formado por desnutridos y otro formado por niños sin desnutrición se esperaría que en el grupo de desnutridos predominara un cierto nivel de ingresos familiar (tal vez bajo) y que los niveles educativos familiares fueran diversos y que en el grupo de niños sin desnutrición no hubiera predominio en sus familias de un determinado nivel de ingresos ni de un cierto nivel educativo.

Consecuencia No. 2 Si la hipótesis es cierta, entonces, al comparar dos grupos de niños: uno formado por desnutridos y otro formado por niños sin desnutrición se esperaría que en el grupo de desnutridos predominara la falta de alimentos de origen proteico y en el grupo de niños no desnutridos las dietas se encontrarán balanceadas.

PROBLEMA 3

Hipótesis general de la investigación: Es probable que el stress produzca aumento de peso en las glándulas suprarrenales de las ratas de experimentación.

Consecuencia No. 1 Si la Hipótesis es cierta, entonces, al analizar las concentraciones de adrenalina en sangre de un grupo de ratas, antes y después de someterlas a situaciones crónicas de stress, se esperaría encontrar que existen aumentos de esta sustancia y de sus análogos químicos.

Consecuencia No. 2 Si la hipótesis es cierta, entonces, al comprarar dos grupos de ratas: uno sometido a situaciones de stress y otro no, se esperaría encontrar que el peso promedio de las glándulas suprarrenales de uno y otro grupo sean diferentes.

PROBLEMA: La rubeola y las malformaciones congénitas

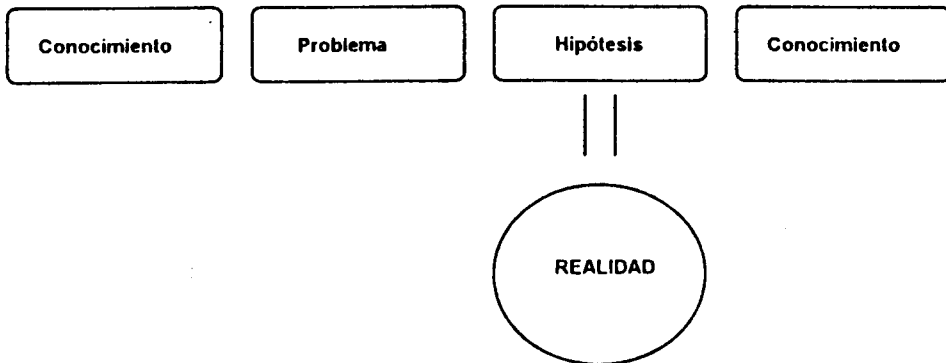
Selecione de entre los párrafos a) al f) los que correspondan a las siguientes etapas del Método Científico y anote la letra correspondiente:

- PLANTEAMIENTO DE HIPOTESIS
- DEDUCCION DE CONSECUENCIAS VERIFICABLES
- DISEÑO DE LA INVESTIGACION

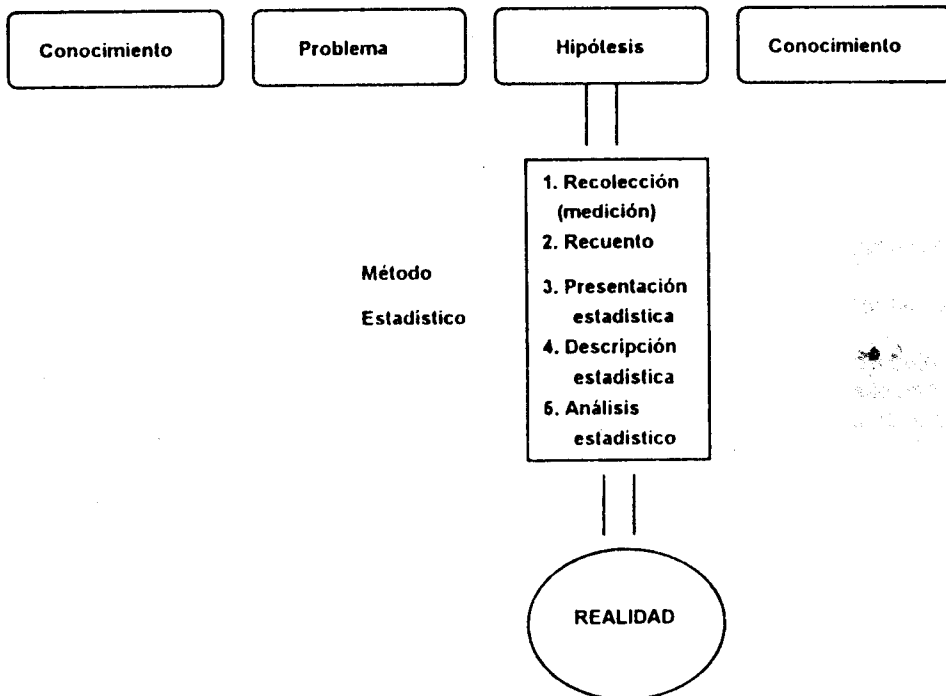
- a) Se elaboraron tablas de asociación para mostrar, simultáneamente, el antecedente positivo o negativo de rubeola durante el embarazo y la presencia o ausencia de malformaciones congénitas.
- b) Al final se concluyó que eran más frecuentes los casos de microcefalia y de defectos congénitos cardíacos entre los hijos de las madres que habían padecido rubeola.
- c) Se formaron dos grupos: uno de niños con microcefalia y defectos congénitos cardíacos y uno de niños con otro tipo de padecimientos presentes desde el nacimiento. En ambos grupos se averiguó el tipo de enfermedades padecidas por la madre durante la gestación.
- d) Se supuso que podría haber asociación entre la rubeola y la presencia de microcefalia y de defectos cardíacos congénitos.
- e) Para sintetizar la información se hicieron cálculos de proporciones.
- f) Si la hipótesis era cierta, se esperaba que la gran mayoría de los niños con microcefalia y malformaciones cardíacas congénitas procedieran de madres que habían padecido rubeola durante el embarazo.

TEXTO: Hipótesis y diseños de investigación (la necesidad de deducir consecuencias verificables)

En otros documentos ¹ ha sido discutido el ciclo de la investigación científica y el papel de la estadística en la comprobación de la hipótesis de investigación. Como punto de partida, se ha dicho que siendo las hipótesis construcciones conceptuales han de ser comprobadas de manera práctica en la realidad.



Por otra parte, se ha señalado que el método estadístico facilita y sistematiza el manejo de la información cualitativa y cuantitativa que permite la comprobación de la hipótesis de la investigación.



¹ *La Investigación Científica y la Estadística*. Departamento de Medicina Social, Medicina Preventiva y Salud Pública, Facultad de Medicina, UNAM. Serie Publicaciones Técnicas de Medicina Preventiva y Social, No. 8. Mexico, 1980

Sin embargo, resulta necesario insistir en la necesidad de efectuar *deducción*² de *consecuencias verificables*³ a partir de la hipótesis de investigación ya que a partir de tales consecuencias verificables se puede seleccionar el *diseño*⁴ de la investigación.

La importancia de que una investigación posea un diseño nítido consiste en que todos los procedimientos estadísticos que habrán de utilizarse están subordinados a tal diseño.

Por ejemplo: si el diseño consiste en la comparación simultánea de los promedios de tres o más grupos independientes, entonces la única prueba de análisis estadístico que corresponderá a tal diseño es la llamada de *análisis de varianza*.

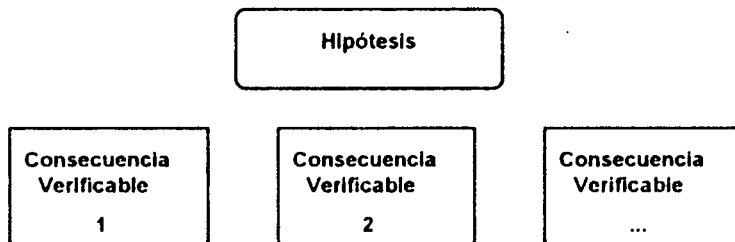
En cambio, si el diseño consiste en la comparación de las frecuencias de dos características cualitativas que se medirán en un único grupo de individuos, entonces la única prueba de análisis estadístico que estará indicada es la llamada *chi cuadrada*.

Las consecuencias verificables se refieren a efectos que se considera que habrán de presentarse en la realidad en caso de que la hipótesis de la investigación sea cierta.

En realidad, al ser anticipaciones producidas por deducciones originadas en una hipótesis, las consecuencias verificables vienen siendo, a su vez, también hipótesis.

Sin embargo, las consecuencias verificables suelen ser de tal nivel de especificidad y se acostumbra a que estén redactadas de manera tan práctica u operativa que resultan sumamente orientadoras para elegir estrategias para su comprobación.

Es totalmente factible que a partir de una sola hipótesis de investigación se pueda efectuar la deducción de más de una consecuencia verificable.



A manera de ejemplo supóngase la siguiente hipótesis general de una cierta investigación:

Es probable que el síndrome de inmunodeficiencia adquirida (SIDA) se transmita por vía sanguínea y no por vía cutánea.

Como puede observarse, la redacción de la hipótesis solo se refiere a conceptos y términos inmatrimales y, aunque señala posibles relaciones entre tales términos, no orienta a una determinada realidad en la cual efectuar comprobaciones.

² Deducción: Obtención de conclusiones particulares y específicas a partir de premisas generales.

³ Consecuencias verificables: efectos observables y medibles en una porción específica de la realidad.

⁴ Diseño: Forma y estrategia general de la investigación.

La deducción de consecuencias verificables en este caso consistiría en anticipar efectos en una determinada y específica parte de la realidad en el caso de que la hipótesis fuera cierta; por ejemplo:

Consecuencia verificable 1. Si la hipótesis es cierta, entonces se esperaría que en un grupo de familias de enfermos de SIDA, en la cuales se haya presentado contacto corporal frecuente entre el enfermo y sus familiares, hubiera igual prevalencia de la enfermedad que en un grupo de familias en las que no hubiera contacto corporal frecuente.

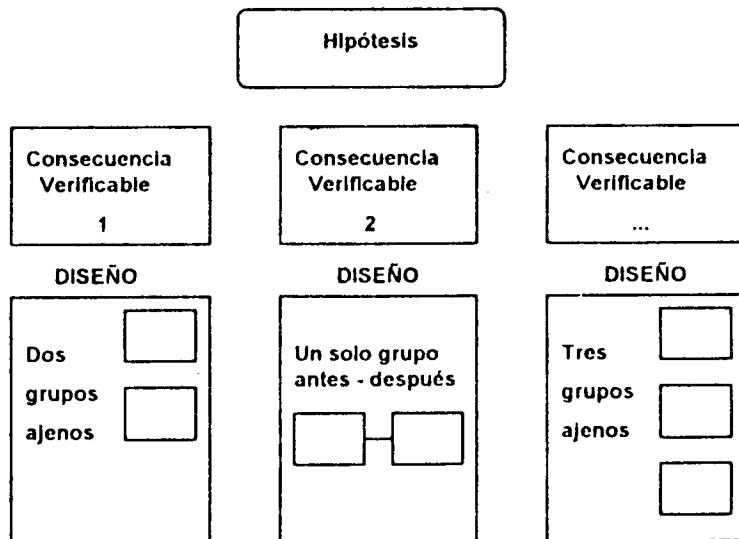
Consecuencia verificable 2. Si la hipótesis es cierta, entonces se esperaría que los familiares que hayan recibido donaciones sanguíneas de parientes con SIDA presenten tarde o temprano la enfermedad, independientemente de que no hayan tenido contacto corporal frecuente con el enfermo.

Como puede observarse, cuando solo se había leído la hipótesis general de investigación, pudieron ser imaginadas muchas y muy diversas maneras de efectuar la investigación. En cambio, una vez leídas y analizadas algunas consecuencias verificables deducidas de dicha hipótesis general uno puede anticipar estrategias bastante bien definidas para efectuar la investigación.

Por ejemplo, es claro que al tratar de verificar la primera consecuencia verificable se buscaría formar dos grupos de familias en las que exista un enfermo con SIDA: unas clasificadas como familias con contacto corporal frecuente y otras clasificadas como con contacto corporal infrecuente. Una vez clasificadas se compararía, en ambos grupos, la frecuencia de casos en el transcurso del tiempo.

Una vez que han quedado deducidas una o más consecuencias verificables, resulta relativamente fácil identificar los diseños de investigación pertinentes para comprobar cada una de tales consecuencias.

El diseño consiste en la forma y estrategia para efectuar la comprobación de una consecuencia verificable. En el diseño queda explicado el número de grupos que habrán de formarse, las características que habrán de medirse o recogerse e, inclusive, los criterios para incluir o excluir a los individuos dentro de cada grupo.

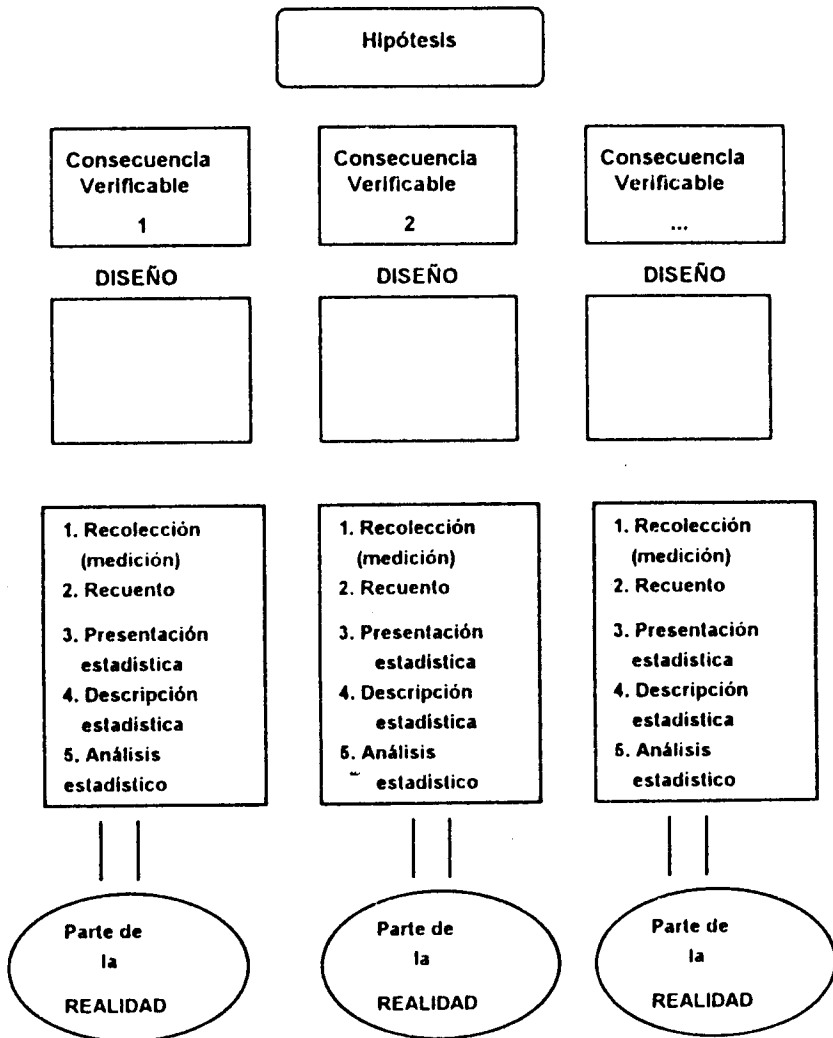


Si han sido deducidas correctamente una o más consecuencias verificables y, por ello mismo, han sido identificados cuidadosamente los diseños correspondientes, entonces las técnicas estadísticas a utilizar durante la o las investigaciones serán fácilmente elegibles.

Por todo lo anterior, un esquema más integral para representar la relación entre la hipótesis de investigación y los diseños de investigación necesariamente debe incluir a la deducción de consecuencias verificables.

Adicionalmente, ha de considerarse entonces que la correcta utilización de las técnicas del método estadístico solo ocurre una vez que se han identificado un diseño para cada una de las consecuencias verificables deducidas a partir de la hipótesis de investigación.

Así pues, un esquema más integral sería el siguiente:



Plan de trabajo para la semana 2

Sesión 1

La mayoría de los procedimientos estadísticos que han de aplicarse en la investigación médica debieran ser seleccionados con anticipación. Dicha anticipación requiere de la deducción de consecuencias **verificables** derivadas de la hipótesis de investigación. Esto ocurre así porque, una vez que se han definido las consecuencias verificables, se pueden seleccionar diseños de investigación pertinentes con lo cual no reviste mayor dificultad escoger y aplicar los procedimientos estadísticos correspondientes.

Para desarrollar habilidades en los temas anteriormente señalados, en esta sesión se discutirán las soluciones planteadas por el alumno al problema *¿Son correctas las deducciones?* y al problema *La rubeola y las malformaciones congénitas* y se ampliarán conceptos correspondientes a la relación entre las consecuencias verificables y los diseños de investigación.

Al final de la sesión, el alumno recibirá indicaciones para el empleo de los siguientes materiales de trabajo para la segunda sesión de la semana 2:

- **Problema: La mortalidad de los médicos por cáncer pulmonar en relación con el hábito de fumar (versión 2)** 2 páginas
 - **Texto: El método estadístico** 4 páginas
-

Sesión 2

El método estadístico es una secuencia lógica y estable de procedimientos para el manejo de la información en la investigación científica. En vista de que el método estadístico formará parte de los futuros esquemas de pensamiento y de acción del alumno, resulta necesario que se adquiera habilidad para aplicar dicha secuencia a lo largo de los estudios profesionales y en la práctica de la profesión.

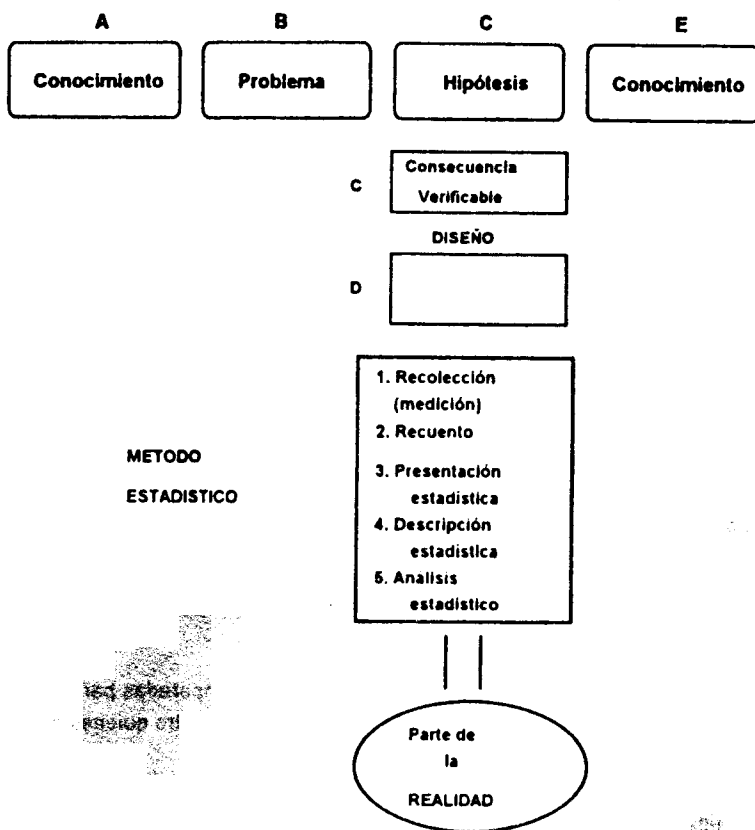
Para contribuir a ello, en esta sesión se discutirán las soluciones planteadas por el alumno al problema *La mortalidad de los médicos por cáncer pulmonar en relación con el hábito de fumar (versión 2)* y se ampliarán conceptos relacionados con el método estadístico.

Al final de la sesión, el alumno recibirá indicaciones para el empleo de los siguientes materiales de trabajo para la primera sesión de la semana 3:

- **Problemas: ¿Qué opina de la recolección?** 2 páginas
 - **Problemas: Construcción de escalas** 2 páginas
 - **Texto: La recolección como primera etapa del método estadístico** 5 páginas
-

PROBLEMA : La mortalidad de los médicos por cáncer pulmonar en relación con el hábito de fumar
(versión 2)

A manera de recordatorio, en el siguiente esquema, con las letras A a la E se muestran las etapas del *Método Científico* y con los números 1. al 5. se muestran las etapas del *Método Estadístico* :



Con los mismo sistema de enumeración empleado en el esquema anterior, identifique las etapas del *Método Estadístico* que se encuentren presentes en el siguiente relato y anote las letras correspondientes en los recuadros de los párrafos (algunos recuadros pueden quedar vacíos).

Al revisar las estadísticas de enfermedades y defunciones por enfermedades pulmonares se encontró que casi todos los casos de enfisema pulmonar y varios de cáncer pulmonar tenían en muchas ocasiones el antecedente de tabaquismo crónico.

Sin saber el por qué de tal relación, en un estudio de numerosos casos de cáncer pulmonar se observó que alrededor del 70% de los enfermos eran antiguos fumadores.

- Esto hizo pensar que el consumo de tabaco pudiera ser un factor favorecedor o causal del cáncer pulmonar.
- Si lo anterior fuera cierto, podría inferirse lo siguiente: a) Un grupo de fumadores de determinada edad, sexo y nivel socioeconómico presentaría un número mayor de muertes por cáncer pulmonar que un grupo de no fumadores de edad, sexo y nivel socioeconómico similares y b) La frecuencia de cáncer pulmonar sería mayor al aumentar el número de cigarrillos consumidos.
- Luego de discutir si convenía formar un grupo de médicos fumadores y seguirlo a lo largo de varios años o formar un grupo de médicos fumadores y uno de no fumadores, se resolvió hacer la investigación enviando un cuestionario a un único pero numeroso grupo de médicos del Reino Unido.
- De ellos, se solicitaron datos individuales sobre, edad y hábitos de fumar por medio del cuestionario de referencia. Dicho cuestionario fue puesto a prueba para comprobar que medía con precisión las diversas variables en estudio. Se establecieron además conexiones con el Registro de Defunciones para asegurar la información sobre la mortalidad por cáncer pulmonar que fuera ocurriendo en los médicos a quienes se incluyó en el estudio.
- Aunque el número de datos preguntados a cada médico era pequeño, la cantidad de médicos a quienes se envió el cuestionario sumaba varios miles. Por ello tuvo que hacerse uso de las que en esa época se consideraban modernas computadoras.
- Los datos se mostraron solamente en forma de cuadros o tablas.
- El resumen de la información obtenida se expresó en tasas de mortalidad por cada mil médicos.
- Las tasas de mortalidad por cada mil médicos fueron comparadas para averiguar si parecían existir diferencias significativas, desde el punto de vista estadístico entre quienes murieron por cáncer pulmonar según el antecedente de si fumaban o no.
- Los resultados mostraron que el riesgo de morir fue 14 veces mayor en fumadores que en no fumadores. También se observó que en los fumadores la frecuencia de éstas muertes aumentó cuando la cantidad de tabaco que se consumía era mayor.

TEXTO: El método estadístico

EL MÉTODO ESTADÍSTICO

El método estadístico consiste en una secuencia de procedimientos para el manejo de los datos cualitativos y cuantitativos de la investigación.

Dicho manejo de datos tiene por propósito la comprobación, en una parte de la realidad, de una o varias consecuencias verificables deducidas de la hipótesis general de la investigación.

Las características que adoptan los procedimientos propios del método estadístico dependen del diseño de investigación seleccionado para la comprobación de la consecuencia verificable en cuestión.

El método estadístico tiene las siguientes etapas:

1. **Recolección (medición)**
2. **Recuento (cómputo)**
3. **Presentación**
4. **Descripción**
5. **Análisis**

Tales etapas siempre se encuentran en el orden descrito y cada una de ellas consiste, de manera resumida, en lo siguiente:

1. Recolección (medición)

En esta etapa se recoge la información cualitativa y cuantitativa señalada en el diseño de la investigación. En vista de que los datos recogidos suelen tener diferentes magnitudes o intensidades en cada elemento observado (por ejemplo el peso o la talla de un grupo de personas), a dicha información o datos también se les conoce como variables.

Por lo anterior, puede decirse que esta etapa del método estadístico consiste en la medición de las variables.

La recolección o medición puede realizarse de diferentes maneras; a veces ocurre por simple observación y en otras ocasiones se requiere de complejos procedimientos de medición; en algunas ocasiones basta con una sola medición y en otras se requiere una serie de ellas a lo largo de amplios períodos de tiempo.

La calidad técnica de esta etapa es fundamental ya que de ella depende que se disponga de datos exactos y confiables en los cuales se fundamenten las conclusiones de toda la investigación.

Es tan grande la importancia de esta etapa que algunas clasificaciones de las investigaciones se basan en la forma en que ocurre la medición; por ejemplo si la información es recogida en una sola ocasión suele decirse que la investigación es transversal; en cambio, si la información es recogida a lo largo del tiempo se denomina longitudinal a la investigación.

En ocasiones, la recolección de la información debe ocurrir en grupos tan grandes de individuos que se hace impráctico tratar de abarcar a todos ellos; entonces es cuando se ponen en práctica procedimientos de muestreo. Tales procedimientos de muestreo están subordinados a la consecuencia verificable que se desea comprobar y al diseño de investigación seleccionado.

2. Recuento (cómputo)

En esta etapa del método estadístico la información recogida es sometida a revisión, clasificación y cómputo numérico.

A veces el recuento puede realizarse de manera muy simple, por ejemplo con rayas o palotes; en otras ocasiones se requiere el empleo de tarjetas con los datos y, en investigaciones con mucha información y muchos casos, puede requerirse el empleo de computadoras y programas especiales para el manejo de bases de datos.

En términos generales puede decirse que el recuento consiste en la cuantificación de la frecuencia con que aparecen las diversas características medidas en los elementos en estudio; por ejemplo: el número de personas de sexo femenino y el de personas de sexo masculino o el número de niños con peso menor a 3 kilos y el número de niños con peso igual o mayor a dicha cifra.

3. Presentación

En esta etapa del método estadístico se elaboran los cuadros y los gráficos que permiten una inspección precisa y rápida de los datos.

La elaboración de cuadros, que también suelen llamarse tablas, tiene por propósito acomodar los datos de manera que se pueda efectuar una revisión numérica precisa de los mismos.

La elaboración de gráficos tiene por propósito facilitar la inspección visual rápida de la información.

Casi siempre a cada cuadro con datos le puede corresponder una gráfica pertinente que represente la misma información.

Presentar la misma información tanto en un cuadro como en su correspondiente gráfico permite obtener una clara idea de la distribución de las frecuencias de las características estudiadas.

4. Descripción

En esta etapa la información es resumida en forma de medidas que permiten expresar de manera sintética las principales propiedades numéricas de grandes series o agrupamientos de datos.

La condensación de la información, en forma de medidas llamadas de resumen, tiene por propósito facilitar la comprensión global de las características fundamentales de los agrupamientos de datos.

Tales medidas de resumen, al ser comunicadas, permiten a los interlocutores evocar de una misma manera la esencia de los datos; por ejemplo, cuando alguien informa que el promedio de calificaciones de un grupo de alumnos es 9.6, en una escala que va del 0 al 10, la imagen que se transmite es la de un grupo con buen aprovechamiento escolar; igualmente, cuando se dice que el porcentaje de defunciones con una cierta técnica quirúrgica es de 80%, las personas que escuchan se imaginan que se trata de un procedimiento peligroso.

La información cualitativa, como el sexo, la ocupación o los tipos de enfermedades, requiere ser condensada a través de medidas de resumen diferentes a la que se usan para sintetizar la información cuantitativa o numérica como el peso, la talla o la concentración de glucosa.

Entre las principales medidas de resumen para sintetizar a los datos cualitativos se encuentran las razones, las proporciones y las tasas.

Entre las principales medidas para sintetizar los datos cuantitativos se encuentra la moda y la amplitud, la mediana y los percentiles y el promedio y la desviación estándar.

5

Análisis

En esta etapa, mediante fórmulas estadísticas y el uso de tablas específicamente diseñadas, se efectúa la comparación de las medidas de resumen previamente calculadas; por ejemplo, si antes se han calculado los promedios de peso de dos grupos de personas sometidas a diferentes dietas, el análisis estadístico de los datos consiste en la comparación de ambos promedios con el propósito de decidir si parece haber diferencias significativas entre tales promedios.

Existen procedimientos bien establecidos para la comparación de las medidas de resumen que se hayan calculado en la etapa de descripción. Tales procedimientos, conocidos como pruebas de análisis estadístico cuentan con sus fórmulas y procedimientos propios.

Cada prueba de análisis estadístico debe utilizarse siempre en función del tipo de diseño de investigación que se haya seleccionado para la comprobación de cada consecuencia verificable deducida a partir de la hipótesis general de investigación.

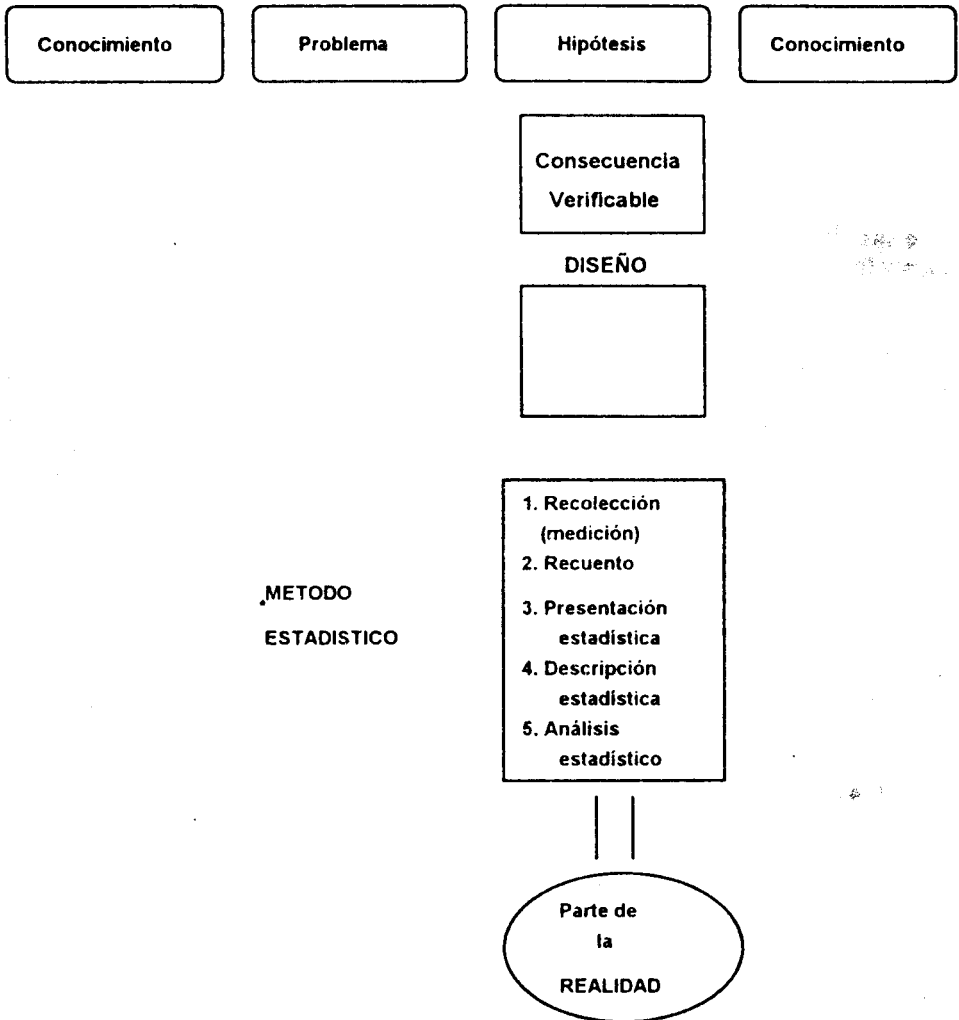
En los primeros párrafos se mencionó que el método estadístico es una secuencia ordenada de procedimientos para el manejo de los datos en las investigaciones. Luego de haber revisado las características y propósitos de las diversas etapas del método, parece quedar claro que tal secuencia guarda siempre el orden descrito.

Es lógico efectuar el recuento de la información (2a. etapa, Recuento) solo hasta que previamente fue recogida (1a. etapa, Recolección o Medición). Antes de condensar la información en la etapa de descripción (4a. etapa) siempre conviene haberla comprendido a través de la elaboración de cuadros y gráficos (3a. etapa Presentación).

Desde luego es impensable efectuar análisis estadístico (5a. etapa, Análisis) para comparar medidas de resumen si antes estas no han sido calculadas (4a. etapa, Descripción)

Por lo anterior, puede considerarse a la estadística como una disciplina que posee su propio método. Tal disciplina emplea conocimientos de otras ciencias como la lógica y la matemática y por ello se dice que la estadística es una forma razonable de emplear el sentido común y la aritmética para el manejo de los datos en las investigaciones.

En la siguiente ilustración se esquematiza la relación que existe entre la estadística y las etapas de la investigación científica:



PROBLEMAS: ¿QUE OPINA DE LA RECOLECCION?

1.

Un joven pasante de medicina fue contratado por una fábrica constructora de espejos para que estableciera programas integrales de prevención y control de enfermedades y accidentes en el trabajo, pues los propietarios de la empresa estaban preocupados por la abundante cantidad de heridos que ya se habían presentado en el curso de los cinco años que tenía en funcionamiento la fábrica. Aunque su consultorio estaba ubicado en la planta alta de la sección de corte y esmerilado de los espejos y tenía acceso directo a ella a través de un moderno elevador, el joven pasante siempre procuraba llegar a su oficina por una escalera exterior del edificio, pues le causaba molestia el intenso ruido producido por la maquinaria de corte y esmerilado.

Con el propósito de averiguar las razones de los numerosos accidentes, el pasante citaba diariamente a dos trabajadores a su consultorio y los interrogaba acerca de su edad, horario de entrada y salida, nivel educativo, sueldo y cursos de entrenamiento en el manejo de vidrio que hubieran tomado.

Luego de haber entrevistado a los 189 trabajadores de la fábrica, el pasante propuso a los dueños de la empresa que: a) solo contrataran a trabajadores mayores de 28 años y menores de 45, b) establecieran dos jornadas de descanso en el curso del día, c) incrementaran los sueldos de los trabajadores, y d) realizaran dos cursos anuales y obligatorios sobre prevención de accidentes.

Preguntas:

1. ¿Qué entrenamiento, además de los estudios regulares de medicina, debiera tener el responsable de un programa de prevención y control de enfermedades y accidentes en una fábrica de espejos?
2. ¿Que procedimiento de recolección de información utilizó el pasante? ¿Debió haber utilizado otro procedimiento?

2.

En contra de lo que se suponía, un investigador pensaba que no era cierto que los escolares aplicados en las escuelas primarias siempre se sientan en las primeras filas de asientos. El opinaba que esos lugares suelen ser ocupados por los niños que no oyen o ven bien.

Para comprobar lo anterior, compró una carta de optotipos como las que se usan en las ópticas y un viejo y ruidoso reloj de cuerda y clasificó la visión y audición tanto de los niños de las primeras filas como de los de las últimas.

Preguntas:

1. ¿Cuáles fueron los instrumentos de medición? ¿El investigador debió haber utilizado otros instrumentos? ¿Cuáles serían?
2. ¿Cuáles son los nombres de las tres características medidas por el investigador en los niños?

3.

En nuestro país se realizan censos de población cada 10 años (el último se realizó apenas en 1990). Tales censos se aplican en la inmensa mayoría de las casas y es raro que existan familias que no recuerden tal evento, pues la cobertura de los censos se ha incrementado década con década.

Interroga a sus familiares acerca de las preguntas que se les hicieron en el último censo y haga una lista de ellas. Una vez que disponga de la lista, califique a cada pregunta según si fue respondida con exactitud o no. En el caso de las preguntas para las que se otorgaron respuestas inexactas señale las posibles razones.

4.

Identificándose como estudiante que realiza una práctica escolar acuda a la óptica más cercana a su domicilio y averigüe lo siguiente:

- Tipo de preparación del optometrista
- Nombre de los instrumentos que emplea
- Tipo de registros utilizados para hacer sus anotaciones
- Nombre de las unidades en que se registran los resultados de los exámenes de agudeza visual
- Tipo de errores que pueden cometerse en la medición de la agudeza visual atribuibles a:
 - el optometrista
 - el paciente
 - el equipo
 - el procedimiento de medición
 - el registro de los resultados de la medición

PROBLEMAS: Construcción de escalas

A continuación se presentan varias escalas. En cada una de ellas existen uno más errores; identifíquelos y corrijalos.

1

Estudiantes según medio de transporte a su plantel
Público
Privado
Taxi

2

Estudiantes según tiempo de transporte
Prolongado
Regular
Breve
1 - 2 horas

3

Estudiantes según medio de transporte a su plantel
Camión
Metro
Taxi
Automóvil privado
A pie
Bicicleta
Motocicleta
Combinados
Público

4

Estudiantes según tiempo de transporte
0 a 29 minutos
30 a 59 minutos
60 a 89 minutos
90 a 119 minutos
120 o más minutos
más de una hora

5

Estudiantes según aprovechamiento escolar
Alerta
Somnoliento
Adormilado
Dormido

6

Recién nacidos según coloración de la piel
Pálido
Sonrosado
Amorotado
Rubio

7

Profesores según estado civil
Solteros
Viudos
Divorciados
En unión Libre

8

Alumnos según procedencia
Nacionales
Extranjeros
Guatemaltecos
Hondureños
Salvadoreños
Nicaragüenses
Costarricenses

9

Alumnos según aciertos en el examen final
MB
B
S
NA
NP
Reprobado

10

Alumnos según resultado en el examen final
0 - 19
20 - 39
40 - 59
60 - 89
85 y más
Con mención honorífica
Sin mención honorífica

TEXTO: La recolección como primera etapa del método estadístico

Introducción:

En documentos anteriores (1) se ha mencionado que la primera etapa del método estadístico es la de recolección. Desde el punto de vista estadístico, el término *recolección* se refiere a un conjunto de acciones que permiten disponer de la información cuantitativa y cualitativa estrictamente necesaria para comprobar las consecuencias verificables que se hayan deducido de una hipótesis.

El conjunto de acciones que permite poseer dicha información está orientado por el tipo de diseño de investigación que se haya seleccionado para comprobar a cada una de las consecuencias verificables.

A lo largo de este documento se utilizará como término equivalente de recolección al de *medición*, ya que lo que realmente se hace al recoger información es una especie de medición de la magnitud, intensidad o tipo de variación que tienen diversas características o atributos.

Como ejemplos de recolección de información cuantitativa se pueden mencionar: la medición del peso al nacer de un grupo de niños, la medición de la duración de las consultas médicas otorgadas por un grupo de médicos, la medición de la cantidad de glucosa sanguínea de un grupo de enfermos de diabetes, etc.

Como ejemplos de recolección de información cualitativa se pueden mencionar: la medición del grado de aceptación del embarazo en un grupo de adolescentes, la medición de la agilidad de un grupo de atletas, la medición del color de los ojos de un grupo de personas, etc.

Componentes de la recolección:**1. El observador o responsable de la medición**

En el caso de la investigación científica, el observador o responsable de la medición usualmente es el mismo investigador. En ocasiones, cuando la investigación es realizada por un grupo de investigadores, suele designarse a alguno de los integrantes como encargado de efectuar la medición de las características o atributos que servirán para comprobar las consecuencias verificables que se hayan deducido de la hipótesis de la investigación.

Tanto si es el propio investigador el que realiza la medición como si es alguno de los miembros de un grupo, la observación que realice debe buscar mediciones válidas y confiables. La impreparación del observador o su falta de cuidado puede provocar que las mediciones efectuadas no proporcionen información útil para el proceso de investigación.

¹ *La investigación Científica y la Estadística, Hipótesis y Diseños de Investigación, El Método Estadístico*

Muchas de las investigaciones requieren que el observador tenga una alta preparación en técnicas de medición específicas para determinado tipo de investigaciones; por ejemplo, la medición de actitudes requiere de una preparación que tal vez incluya estudios especializados en el campo de la psicología, la medición de alteraciones anatómicas puede requerir de preparación especializada en técnicas radiológicas modernas o la medición de características microscópicas puede requerir de entrenamiento en técnicas de microscopía electrónica.

2. Las características medidas o variables

Para fines estadísticos las características o atributos a los que se mide su magnitud, intensidad o valor se denominan variables ⁽²⁾. En algunos casos la característica que se mide admite una valoración cuantitativa de tipo numérico, como la estatura o el número de aciertos en un examen; en otras ocasiones la característica que se mide solo admite valoraciones cualitativas en términos de palabras, como el color de ojos o los oficios a que se dedica un grupo de personas.

Tanto si las variables recogidas son de tipo cuantitativo como si son de tipo cualitativo, deben haberse definido a la luz del diseño de investigación y deben servir precisamente para comprobar la consecuencia verificable que originó la selección de dicho diseño.

3. El procedimiento de medición

Los procedimientos de medición o recolección son tan variados que su enumeración sería vastísima. Algunos procedimientos consisten en la observación visual y el registro de lo observado; por ejemplo, si se deseara investigar la forma en que un grupo de amas de casa prepara los alimentos en su hogar, una forma sería observarlas y efectuar las anotaciones pertinentes. Otros procedimientos consisten en efectuar preguntas directas; por ejemplo si se deseara averiguar el tiempo que emplean en transportarse a su escuela los estudiantes de alguna licenciatura, una forma podría ser interrogar con precisión a una muestra de dichos estudiantes. En otras ocasiones los procedimientos pueden consistir en pedir que sea contestado un cuestionario, ya sea con preguntas abiertas o con preguntas con opciones.

En algunas disciplinas los procedimientos de recolección o medición son efectuados a través de dispositivos y técnicas que proporcionan valores de tipo numérico; por ejemplo, para recoger la magnitud de las estaturas de un grupo de niños puede emplearse desde una cinta métrica hasta un preciso escalímetro. En algunos casos, los procedimientos de medición llegan a adquirir niveles extremados de precisión; por ejemplo, para medir la duración de algunos fenómenos es posible emplear relojes que miden hasta millonésimas de segundo. En otros casos los procedimientos de recolección implican el empleo de sustancias que desencadenan reacciones biológicas o químicas que ponen en evidencia alguna característica o atributo que, al ser medida reflejan su real magnitud o intensidad.

² Ver una revisión más amplia del concepto y clasificación de variables en el documento de instrucción programada: *Variables 93*

Escalas de modalidades o clases

Antes de proceder a la recolección de la información es necesario que se hayan definido las escalas de modalidades y las escalas de clases que servirán para efectuar tanto el recuento de la información como para aplicar técnicas de presentación, de descripción y de análisis estadístico.

Una **escala de modalidades** como la del siguiente ejemplo sirve para clasificar y contar las frecuencias de una característica o atributo de tipo cualitativo:

Nombre de la característica a medir →

Escala de modalidades →

ESTADO CIVIL
Solteros
Casados
Viudos
Divorciados
En Unión Libre

Una **escala de clases (o de intervalos)** como la del siguiente ejemplo sirve para clasificar y contar las frecuencias de una característica o atributo de tipo cuantitativo:

Nombre de la característica a medir →

Escala de clases (o de intervalos) →

ESTATURA (cms)
140 a 149
150 a 159
160 a 169
170 a 179
180 a 189

Tanto uno como otro tipo de escalas, deben cumplir, entre otras, con las siguientes normas:

1. La escala debe ser exhaustiva; es decir, sus modalidades o clases deben abarcar a todo el tipo posible de mediciones que se vayan a efectuar. En los siguientes ejemplos se aprecian escalas que no son exhaustivas:

OPINION SOBRE LAS CONDICIONES DE LAS AULA
Muy buena
Regular
Mala
Muy mala

Falta la modalidad *Buena*

PESO DE UN GRUPO DE ADOLESCENTES (KGRS.)
30 a 34
35 a 39
40 a 44
50 a 54
55 a 59

Falta la clase *45 a 49*

2. La escala debe contener modalidades o clases mutuamente excluyentes. Lo anterior se refiere al hecho de que cualquier medición realizada pueda ser ubicada en una o en otra modalidad o clase sin confusiones o ambigüedades de ningún tipo. En los siguientes ejemplos se aprecian escalas con modalidades o clases que no son mutuamente excluyentes:

AGILIDAD MOSTRADA EN UN EVENTO ATLETICO
Estupenda
Excelente
Buena
Regular
Mala

Las modalidades *Estupenda* y *Excelente* se traslapan

PESO DE UN GRUPO DE ADOLESCENTES (KGRS.)
30 a 34
33 a 39
40 a 44
45 a 49
50 a 54

La clase *33 a 39* se traslapa con la de *30 a 34*; porque un individuo de *34* kgrs. podría ubicarse tanto en una como en otra clase.

3. La escala debe contener modalidades o clases pertenecientes a una única variable. Lo anterior significa que no deben mezclarse modalidades de o clases de una escala con las de otra, aunque sean sumamente parecidas. En los siguientes ejemplos se aprecian modalidades o clases que no pertenecen a una misma variable:

TIPO DE DIETA
Hipocalórica
Normocalórica
Hiperocalórica
Equilibrada

Las tres modalidades relativas a calorías pueden pertenecer a la escala de una sola característica y la modalidad *Equilibrada* puede pertenecer a otra escala perteneciente a otra diferente característica. Así, despejando la mezcla de escalas y modalidades, las escalas de dos diferentes características o variables pueden quedar de la siguiente manera:

Tipo de dieta según aporte calórico
Hipocalórica
Normocalórica
Hiperocalórica

Tipo de dieta según equilibrio de nutrientes
Equilibrada
Desequilibrada

Como puede observarse, aunque son muy parecidas, las características que se desea medir son diferentes y a cada una de ellas le corresponde su propia escala de modalidades.

Plan de trabajo para la semana 3

Sesión 1

Como primera etapa del método estadístico, la de *recolección* tiene una importancia fundamental. Esto ocurre porque en ella se realiza la medición de las variables que habrán de ser objeto de los restantes procedimientos estadísticos de recuento, presentación, descripción y análisis. Fácilmente puede comprenderse que si la calidad de los datos recogidos no es buena o es imperfecta, los resultados de cualquier estudio o investigación serán deficientes.

Para desarrollar una mejor comprensión de la importancia de esta etapa, en esta sesión se discutirán las soluciones planteadas por el alumno a los problemas *¿Qué opina de la recolección?* y *Construcción de escalas* y se ampliarán conceptos relacionados con el proceso de medición y con los componentes de la etapa de recolección.

Al final de la sesión, el alumno recibirá indicaciones para el empleo de los siguientes materiales de trabajo para la segunda sesión de la semana 3:

- **Problemas: Identificación básica y clasificación de variables** **2 páginas**
 - **Problemas: Identificación avanzada de variables** **2 páginas**
 - **Texto: Variables (documento de autoenseñanza)** **6 páginas**
-

Sesión 2

La esencia de la etapa de recolección del método estadístico es la medición de características o atributos en grupos de casos. Tales características o atributos, al tener valores o magnitudes diferentes en cada uno de los casos sometidos a medición, justifican la aplicación de procedimientos estadísticos para su presentación, descripción y análisis.

En vista de la necesidad de poseer habilidad para identificar, clasificar y definir variables, en esta sesión se discutirán las soluciones planteadas por el alumno a los problemas *Identificación básica y clasificación de variables* e *Identificación avanzada de variables*

Al final de la sesión, el alumno recibirá indicaciones para el empleo de los siguientes materiales de trabajo para la primera sesión de la semana 4:

- **Problemas: Recuento de información (Procedimiento de palotes)** **4 páginas**
 - **Problemas: Recuento de información (Procedimiento de tarjetas simples)** **5 páginas**
 - **Texto: Recuento de información** **5 páginas**
-

PROBLEMAS: Identificación básica y clasificación de variables

Primera parte: Clasifique a cada variable marcando con una cruz en la columna correspondiente

No.	Nombre de la variable y escala correspondiente	Cualitativas		Cuantitativas	
		Nominal	Ordinal	Discreta	Continua
1	Agilidad mostrada al saltar un charco:				
2	Sexo:				
3	Disminución de peso bajo un régimen de dieta:				
4	Grado militar:				
5	Escolaridad:				
6	Convulsiones:				

Segunda parte: Clasifique a cada variable marcando con una cruz en la columna correspondiente

No.	Nombre de la variable y escala correspondiente	Cualitativas		Cuantitativas	
		Nominal	Ordinal	Discreta	Continua
1	Agilidad mostrada al saltar un charco: Minima Regular Buena				
2	Sexo: Masculino Femenino				
3	Disminución de peso bajo un régimen de dieta (kgr. enteros): 01 - 05 06 - 10 11 - 15 16 - 20				
4	Disminución de peso bajo un régimen de dieta deficiente regular buena excelente				
5	Grado militar: Jefes Oficiales Tropa				

continua ...

6	Escolaridad: Ausente Presente				
7	Escolaridad: Analfabeta Sabe leer y/o escribir (sin haber asistido a la escuela) Primaria incompleta Primaria completa Educación media incompleta Educación media o mayor				
8	Escolaridad (años completos): 0 años 01-06 años 07-09 años 10-12 años 13-18 años				
9	Convulsiones: 1 - 3 4 - 6 7 - 9				
10	Convulsiones: Tónico-clónicas Tónicas				

PROBLEMAS: Identificación avanzada de variables.**Problema 1**

Un grupo de investigadores suponían que las situaciones de tensión (stress) podrían influir en el peso de las glándulas suprarrenales de animales de experimentación

- a) Desde el punto de vista metodológico, ¿cuál era la variable independiente (antecedente) incluida en la hipótesis?
- b) Desde el punto de vista metodológico, ¿cuál era la variable dependiente (consecuente) incluida en la hipótesis?
- c) Desde el punto de vista estadístico, ¿de qué tipo era la variable independiente?
- d) Desde el punto de vista estadístico, ¿de qué tipo era la variable dependiente?

Problema 2

El mismo grupo de investigadores consideraban que si su hipótesis era cierta, al formar dos grupos de ratas y exponer a uno a stress y al otro no, podría encontrarse que los pesos de las glándulas suprarrenales de las ratas de ambos grupos serían diferentes.

A continuación se muestran sus resultados:

Grupo sometido a stress		Grupo NO sometido a stress	
Rata No.	Peso de las glándulas (gr.)	Rata No.	Peso de las glándulas (gr.)
1	5.2	1	3.9
2	4.9	2	4.1
3	4.7	3	3.4
4	3.3	4	3.2
5	4.8	5	4.7

- a) Desde el punto de vista metodológico, ¿cuál era la variable independiente (antecedente) incluida en la consecuencia verificable deducida de la hipótesis?
- b) Desde el punto de vista metodológico, ¿cuál era la variable dependiente (consecuente) incluida en la consecuencia verificable deducida de la hipótesis?
- c) Desde el punto de vista estadístico, ¿de qué tipo era la variable independiente?
- d) Desde el punto de vista estadístico, ¿de qué tipo era la variable dependiente?

Problema 3

Con el propósito de simplificar la presentación de sus resultados en un congreso, los investigadores decidieron mostrar sus datos de la siguiente forma:

Grupo sometido a stress		Grupo NO sometido a stress	
Rata No.	Peso de las glándulas	Rata No.	Peso de las glándulas
1	alto	1	bajo
2	medio	2	medio
3	medio	3	bajo
4	bajo	4	bajo
5	medio	5	medio

- a) Desde el punto de vista metodológico, ¿cuál era la variable independiente (antecedente) incluida en la tabla?
- b) Desde el punto de vista metodológico, ¿cuál era la variable dependiente (consecuente) incluida en la tabla?
- c) Desde el punto de vista estadístico, ¿de qué tipo era la variable independiente?
- d) Desde el punto de vista estadístico, ¿de qué tipo era la variable dependiente?



TEXTO: Variables

(documento de autoenseñanza)

INTRODUCCIÓN: Características constantes y características variables

Usted sabe que a nivel del mar el agua siempre hierve a 100 grados centígrados. Esto ocurre invariablemente en cualquier época del año y en cualquier país.

Usted diría que esta característica del agua (hervir a una temperatura de 100 grados centígrados a nivel del mar) es una característica:

VARIABLE

CONSTANTE

(encierra en un círculo la opción correcta)

Es cierto, el hecho de que el agua hierva a 100 grados a nivel del mar es una característica **CONSTANTE**.

La característica que recién se ha mencionado es constante; pero existen otras características que pueden tomar diversos valores o magnitudes. Por ejemplo, el tamaño de los árboles no es constante; cuando son jóvenes tienen un tamaño pequeño y cuando son adultos tienen un gran tamaño.

Los escolares de un mismo grupo tienen diverso peso. Los integrantes de una familia tienen diversas edades.

Cuando usted desayuna ingiere cierta cantidad de proteínas. Usted opina que la cantidad de proteínas en su desayuno diario es una característica: (encierra en un círculo)

VARIABLE

CONSTANTE

Muy bien, si usted señaló que la cantidad de proteínas contenidas en su desayuno diario es variable se ha dado cuenta de que algunas características toman diversos valores, magnitudes o intensidades. Veamos otros ejemplos:

(en el lado derecho, anote si la característica señalada en el lado izquierdo es variable o es constante)

Característica	Tipo (variable o constante)
Tiempo empleado para resolver un examen	
Cantidad mensual de dinero gastado en diversiones	
Temperatura a la que se funde la plata al nivel del mar	
Escolaridad alcanzada por los habitantes de un país	

Si usted respondió que todas las características anteriores son **VARIABLES**, excepto la temperatura a la cual se funde la plata a nivel del mar, está en lo correcto.

CONCEPTO DE VARIABLES

Ahora que hemos entrado en materia, definamos el concepto de **VARIABLES**.

VARIABLES son las cualidades, rasgos, atributos, características o propiedades que toman diferentes valores, magnitudes o intensidades en un grupo de elementos.

En el trabajo estadístico es fundamental conocer con precisión el tipo de variables que han de manejarse, pues a cada tipo de variables le corresponde una diferente manera de ser presentada en forma de tablas o gráficos y de acuerdo al tipo de variable se empleará uno u otro procedimiento estadístico.

A continuación escriba con sus propias palabras el concepto de variable:

Cada variable puede tener, como ya se dijo, diferentes valores, magnitudes, intensidades o modalidades. Por ejemplo: la variable APROVECHAMIENTO ESCOLAR puede tener los siguientes valores: BUENO, REGULAR, MALO. La variable SEXO puede tener las siguientes modalidades: MASCULINO, FEMENINO.

Veamos qué tal se ha entendido lo anterior. Para las variables anotadas en la columna izquierda, consigne las diferentes modalidades o valores que ellas puedan adoptar:

TIPO DE SANGRE: -----, -----, -----, -----,
 FACTOR Rho: -----, -----,
 GRADO DE ENSEÑANZA
 SECUNDARIA: -----, -----, -----,

Esperamos que haya respondido de la siguiente manera:

TIPO DE SANGRE: A, B, AB, O
 FACTOR Rho: +, -
 GRADO DE ENSEÑANZA
 SECUNDARIA: Primero, Segundo, Tercero

En algunos casos las modalidades o magnitudes que adopta una variable pueden ser anotadas en forma gruesa o en forma muy detallada; por ejemplo, la variable HIGIENE PERSONAL puede adoptar los siguientes valores: ADECUADA, INADECUADA. Para la misma variable, HIGIENE PERSONAL, pueden consignarse otras modalidades más detalladas como BUENA, REGULAR, MALA; aún más, pueden anotarse como valores más detallados los siguientes: EXCELENTE, MUY BUENA, BUENA, REGULAR, MALA, MUY MALA, PÉSIMA.

CLASIFICACIÓN DE LAS VARIABLES: cualitativas y cuantitativas

Atendiendo a la forma en que las variables se pueden observar, medir o captar se pueden clasificar en dos principales grupos: CUALITATIVAS Y CUANTITATIVAS.

Son variables CUALITATIVAS las que se refieren a características o atributos que solamente se describen con palabras; por ejemplo la variable SEXO solamente se describe en términos de palabras; sus modalidades son: MASCULINO, FEMENINO; la variable PRESENCIA DE ENFERMEDAD se describe en términos de SI, NO.

Son variables CUANTITATIVAS las que son susceptibles de describir en términos numéricos; por ejemplo: la variable NUMERO DE HIJOS se puede describir a través de las siguientes modalidades o magnitudes: 1 hijo, 2 hijos, 3 hijos, 4 hijos, etc.; la variable ESTATURA AL NACER puede tener los valores 47 cm., 48 cm., 49 cm., 50 cm., etc.

Practiquemos un poco. Marque con una cruz si la variable de la columna del lado izquierdo es CUALITATIVA o CUANTITATIVA:

Variable	Cualitativa	Cuantitativa
Peso al nacer		
Estado Civil		
Material de construcción de la vivienda		

Usted debió responder:

- que el peso al nacer es una variable cuantitativa (se mide en número de gramos),
- que el estado civil es una variable cualitativa (existen las modalidades soltero, casado, viudo, divorciado, unión libre)
- y que la variable material de la vivienda es cualitativa (existen viviendas de cartón, ladrillo, madera adobe, etc.)

Es importante que usted se de cuenta de que las modalidades de las variables cuantitativas pueden expresarse como si fueran cualitativas; por ejemplo, la variable INTEGRANTES DE FAMILIA será cuantitativa cuando se exprese en los siguientes términos: 2 integrantes, 3 integrantes, 4 integrantes, 5 integrantes; pero si la misma variable, INTEGRANTES DE FAMILIA, se expresa en los siguientes términos: PADRE, MADRE, HIJOS, se convierte en una variable de tipo cualitativa.

Para los procedimientos estadísticos es necesario definir con anticipación tanto la clasificación como las modalidades de las variables que se han de manejar.

SUBDIVISIÓN DE LAS VARIABLES CUALITATIVAS: nominales y ordinales

Como ya se mencionó anteriormente, las variables CUALITATIVAS poseen modalidades que solamente se expresan a través de palabras, pero algunas de dichas modalidades dan una idea de jerarquía entre ellas y otras no.

A manera de ejemplo obsérvese la diferencia entre dos variables CUALITATIVAS y sus respectivas modalidades:

NACIONALIDAD: Mexicano, Peruano, Argentino, etc.

LUGAR OCUPADO EN UNA COMPETENCIA: Primero, Segundo, Tercero, etc.

Como podrá observarse existen dentro de las variables cualitativas algunas que adoptan modalidades que no tienen punto de comparación, como el tipo de enfermedad, profesión, etc. A estas variables se les conoce como variables CUALITATIVAS NOMINALES.

Así, la variable NACIONALIDAD es una variable CUALITATIVA NOMINAL.

En el caso de la variable LUGAR OCUPADO EN UNA COMPETENCIA se puede observar jerarquía u orden en las modalidades, pues quien obtuvo el primer lugar llegó antes del segundo y el tercero; a su vez el segundo llegó después del primero pero antes del tercero. En este caso se trata de una variable de tipo CUALITATIVO ORDINAL.

Veamos si está claramente entendida la subdivisión de las variables CUALITATIVAS. A continuación identifique, marcando con una cruz, el tipo a que corresponden las siguientes variables cualitativas.

Variable CUALITATIVA	Nominal	Ordinal
Profesión		
Area de Trabajo		
Color de ojos		
Partido político preferido		
Lugar ocupado al nacer		
Agilidad		
Intensidad de un esfuerzo		

Usted debió haber identificado a las primeras cuatro variables cualitativas como NOMINALES y a las últimas tres como ORDINALES.

SUBDIVISIÓN DE LAS VARIABLES CUANTITATIVAS: discontinuas y continuas

También las variables susceptibles de describir en términos cuantificables, es decir las variables CUANTITATIVAS, pueden ser subdivididas en dos tipos: DISCONTINUAS o CONTINUAS.

Algunas variables CUANTITATIVAS tienen modalidades que solo se pueden expresar en números enteros; por ejemplo la variable TOTAL DE HIJOS solo se expresa en términos de 1 hijo, 2 hijos, etc., y no es posible hablar de 1.5 hijos o de 1 1/2 hijos. A estas variables CUANTITATIVAS se les denomina DISCONTINUAS o también DISCRETAS.

Otras variables CUANTITATIVAS sí son susceptibles de expresar en términos de fracciones; por ejemplo la variable ESTATURA se puede medir en metros, centímetros, milímetros, etc. y así es posible hablar de alguna persona que mida 1.756 mts. A este último tipo de variables CUANTITATIVAS se les denomina CONTINUAS.

Veamos si ha quedado clara la subdivisión de las variables CUANTITATIVAS. A continuación identifique, marcando con una cruz, el tipo a que corresponden las siguientes variables CUANTITATIVAS:

Variable CUANTITATIVA	Discontinua	Continua
Número de habitaciones de un grupo de casas		
Intentos de suicidio en un grupo de personas		
Automóviles vendidos por mes en varias agencias		
Tiempo empleado por un grupo para resolver un examen		
Peso de un grupo de niños		
Estatura de un grupo de empleados		

Usted debió haber identificado a las tres primeras variables como CUANTITATIVAS DISCONTINUAS y a las tres últimas como CUANTITATIVAS CONTINUAS.

EJERCICIO INTEGRAL

Ahora que usted ya conoce las características de las variables y su clasificación, le pediremos un esfuerzo adicional para reunir los conocimientos que ha adquirido durante su lectura.

Observe en el siguiente relato cómo se han identificado las variables y cómo se han clasificado, para que posteriormente usted realice un ejercicio semejante:

En un estudio sobre la Lepra se observó a un grupo de pacientes. Interesaba identificarlos según su sexo, su edad, el tiempo de evolución de la enfermedad y el tipo de medicamento empleado.

Variable Identificada	Tipo de variable			
	Cualitativa Nominal	Cualitativa Ordinal	Cuantitativa Discontinua	Cuantitativa Continua
Sexo	X			
Edad				X
Tiempo de evolución				X
Tipo de medicamento	X			

Como podrá usted observar, en el relato anterior se han identificado dos variables cualitativas nominales y dos variables cuantitativas continuas.

Ahora identifique usted las variables en el siguiente relato y anótelas, marcando su clasificación en el cuadro correspondiente:

En una población rural se realizó una investigación para estudiar la edad de los habitantes, también se observó el total de los hijos en cada familia, otro factor investigado fue la escolaridad de las madres de familia. También se estudió el tipo de alimentación de las familias.

Variable Identificada	Tipo de variable			
	Cualitativa Nominal	Cualitativa Ordinal	Cuantitativa Discontinua	Cuantitativa Continua

Usted debió haber identificado las siguientes variables:

Variable Identificada	Tipo de variable			
	Cualitativa Nominal	Cualitativa Ordinal	Cuantitativa Discontinua	Cuantitativa Continua
Edad de los habitantes				X
Número de hijos por familia			X	
Escolaridad de la madres		X		
Tipo de alimentación de las familias	X			

Si usted contestó de acuerdo al cuadro que le presentamos lo felicitamos; creemos que ahora usted ya es capaz de identificar variables y clasificarlas con fines de realizar procedimientos estadísticos.

OTRAS CLASIFICACIONES DE LAS VARIABLES:

Prácticamente todo el texto se ha referido a una clasificación de las variables que tiene uso cotidiano en los procesos de tipo estadístico; sin embargo existen otras clasificaciones de utilidad en la investigación.

Una clasificación de uso frecuente, denominada clasificación metodológica, divide a las variables en Independientes y Dependientes.

Las variables independientes son antecedentes a las variables dependientes. A su vez, las variables dependientes se consideran consecuentes de las independientes.

Aunque no son sinónimos exactos desde el punto de vista científico, en el siguiente cuadro se plantean varios términos que tienen utilización semejante:

Variables Independientes	Variables Dependientes
Antecedentes	Consecuentes
causa	efecto
origen	resultado
razón	producto

La clasificación metodológica de las variables cobra importancia cuando en una investigación se procede a definir el diseño de la misma. Por ejemplo, si se desea averiguar cual de dos hipoglucemiantes, "A" o "B", produce una mejor disminución de los niveles de glucosa en sangre, se puede clasificar a las variables de la siguiente manera:

Variable Independiente	Variable Dependiente
Tipo de hipoglucemiante ("A" o "B")	Nivel de Glucosa en sangre

PROBLEMAS

Recuento de Información (Procedimiento de palotes)

A una comida de bodas asistieron 120 personas. Se ofrecieron tres alimentos: sopa, arroz y barbacoa. Algunas horas después de la comida varias personas presentaron diarrea. Los datos se presentan en forma de un listado con los datos de cada una de las 120 personas.

1. A través del procedimiento de palotes efectúe un primer recuento o cómputo de la información auxiliándose de una tabla de trabajo como la siguiente:

		Palotes o rayas	Total
Comió sopa	S		
Comió sopa	N		
Comió arroz	S		
Comió arroz	N		
Comió barbacoa	S		
Comió barbacoa	N		
Tuvo diarrea	S		
Tuvo diarrea	N		

Caso #	Comió Sopa	Comió arroz	Comió barbacoa	Tuvo diarrea	Caso #	Comió Sopa	Comió arroz	Comió barbacoa	Tuvo diarrea
1	S	N	S	S	31	N	S	S	N
2	N	S	S	N	32	S	N	N	S
3	S	N	S	N	33	S	N	N	S
4	S	N	N	S	34	S	N	N	S
5	S	N	S	S	35	S	S	N	S
6	N	S	N	N	36	S	S	N	S
7	S	S	S	S	37	N	N	S	N
8	S	S	S	S	38	S	N	S	N
9	S	N	S	N	39	S	N	S	N
10	N	S	N	S	40	S	S	S	S
11	S	S	S	S	41	N	N	N	N
12	S	S	S	S	42	N	S	S	N
13	N	S	S	N	43	S	S	S	S
14	S	S	S	S	44	N	N	S	N
15	S	N	N	S	45	N	N	S	N
16	N	N	N	N	46	S	N	S	N
17	S	S	S	S	47	N	S	S	N
18	S	S	S	S	48	S	N	N	S
19	S	N	S	S	49	S	S	S	N
20	S	N	N	S	50	N	S	S	N
21	S	N	N	S	51	N	N	S	N
22	S	N	N	S	52	N	S	S	N
23	S	N	N	S	53	N	N	S	N
24	S	N	N	S	54	S	S	S	S
25	S	S	S	S	55	N	N	N	N
26	N	N	S	S	56	N	N	S	N
27	S	N	N	S	57	S	S	S	N
28	N	S	S	N	58	S	N	S	S
29	S	S	N	S	59	N	N	S	N
30	S	S	N	S	60	S	S	N	N

Caso #	Comió Sopa	Comió arroz	Comió barbacoa	Tuvo diarrea	Caso #	Comió Sopa	Comió arroz	Comió barbacoa	Tuvo diarrea
61	S	N	S	S	91	N	S	S	N
62	N	S	S	S	92	S	N	N	S
63	S	N	S	N	93	S	N	N	S
64	S	N	N	S	94	S	N	N	S
65	S	N	S	S	95	S	S	N	S
66	N	S	N	N	96	S	S	N	S
67	S	S	S	S	97	S	N	S	N
68	S	S	S	S	98	S	N	S	N
69	S	N	S	N	99	S	N	S	N
70	N	S	N	N	100	S	S	S	N
71	S	S	S	S	101	N	N	N	N
72	S	S	S	S	102	N	S	S	N
73	N	S	S	N	103	S	S	S	S
74	S	N	S	S	104	N	N	S	N
75	S	N	N	S	105	N	N	S	N
76	N	N	N	N	106	S	S	S	N
77	S	S	S	S	107	N	S	S	S
78	S	S	S	S	108	S	N	N	S
79	S	N	S	S	109	S	S	N	S
80	S	N	N	S	110	S	S	S	N
81	S	N	N	S	111	N	N	S	N
82	S	S	S	S	112	S	N	N	S
83	S	N	N	S	113	N	N	S	N
84	S	N	N	S	114	S	S	S	S
85	N	S	S	N	115	N	N	N	N
86	N	N	S	S	116	S	N	N	S
87	S	N	N	S	117	S	S	S	N
88	N	S	S	N	118	S	N	S	S
89	S	S	N	S	119	N	N	S	N
90	S	S	N	S	120	S	S	N	N

2. A través del procedimiento de palotes efectúe un segundo recuento o cómputo de la información auxiliándose de tablas de trabajo como las siguientes:

		Palotes o rayas	Total
Comió sopa	S		
tuvo diarrea	S		
Comió sopa	S		
tuvo diarrea	N		
Comió sopa	N		
tuvo diarrea	S		
Comió sopa	N		
tuvo diarrea	N		

		Palotes o rayas	Total
Comió arroz	S		
tuvo diarrea	S		
Comió arroz	S		
tuvo diarrea	N		
Comió arroz	N		
tuvo diarrea	S		
Comió arroz	N		
tuvo diarrea	N		

		Palotes o rayas	Total
Comió barbacoa	S		
tuvo diarrea	S		
Comió barbacoa	S		
tuvo diarrea	N		
Comió barbacoa	N		
tuvo diarrea	S		
Comió barbacoa	N		
tuvo diarrea	N		

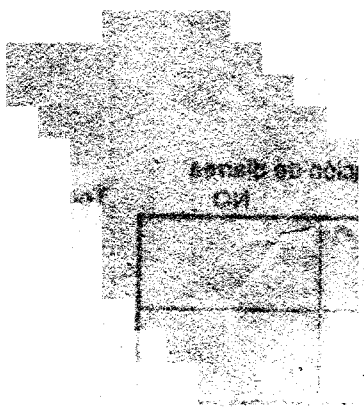
3. Una vez que haya efectuado el recuento o cómputo de la información llene las siguientes tablas auxiliares de trabajo.

Ingestión de sopa	Presentación de diarrea		Totales
	SI	NO	
SI			
NO			

Ingestión de arroz	Presentación de diarrea		Totales
	SI	NO	
SI			
NO			

Ingestión de barbacoa	Presentación de diarrea		Totales
	SI	NO	
SI			
NO			

4. Emita una opinión con respecto al tipo de asociación que se observa entre la ingestión de cada alimento y la presentación de diarrea



PROBLEMAS**Recuento de Información (Procedimiento de tarjetas simples)**

A una comida de bodas asistieron 120 personas de ambos sexos y diferentes edades. Se ofrecieron tres alimentos: sopa, arroz y barbacoa. Algunas horas después de la comida varias personas presentaron diarrea.

Los datos referentes a cada una de las 120 personas se presentan en el listado de las páginas siguientes.

1. Usando como modelo la tarjeta que se muestra a continuación, traslade los datos del listado a 120 tarjetas y efectúe el recuento requerido para llenar las seis tablas auxiliares de trabajo que se muestran más adelante

edad	sexo	horas entre comida y diarrea
	caso #	
barbacoa	arroz	sopa

Grupo de Edad	Tuvo diarrea		Totales
	SI	NO	
menos de 24			
25 a 44			
45 y más			

Horas entre Comida y diarrea	Tuvo diarrea		Totales
	SI	NO	
0			
1 a 6			
7 a 12			
13 y más			

Sexo	Tuvo diarrea		Totales
	SI	NO	
Femenino			
Masculino			

Comió sopa	Tuvo diarrea		Totales
	SI	NO	
SI			
NO			

Comió arroz	Tuvo diarrea		Totales
	SI	NO	
SI			
NO			

Comió barbacoa	Tuvo diarrea		Totales
	SI	NO	
SI			
NO			

2. Identifique a cada una de las siete variables recogidas y las escalas de clases o modalidades que se construyeron para cada variable.

3. Emita una opinión en relación con la posible asociación entre las modalidades de la variable *presentación de diarrea* y las clases o modalidades de las otras seis variables.

Caso #	Edad	Sexo	Horas	Comió sopa	Comió arroz	Comió barbacoa	Tuvo diarrea
1	43	M	8	S	N	S	S
2	51	M	0	N	S	S	N
3	28	F	0	S	N	S	N
4	29	M	12	S	N	N	S
5	36	M	7	S	N	S	S
6	39	M	0	N	S	N	N
7	61	M	9	S	S	S	S
8	57	F	11	S	S	S	S
9	48	M	0	S	N	S	N
10	46	F	10	N	S	N	S
11	54	M	12	S	S	S	S
12	42	F	7	S	S	S	S
13	21	M	0	N	S	S	N
14	19	F	9	S	S	S	S
15	28	M	12	S	N	N	S
16	23	F	0	N	N	N	N
17	29	M	11	S	S	S	S
18	28	M	8	S	S	S	S
19	26	M	13	S	N	S	S
20	27	M	11	S	N	N	S
21	21	F	6	S	N	N	S
22	20	M	5	S	N	N	S
23	19	F	10	S	N	N	S
24	47	M	15	S	N	N	S
25	40	F	8	S	S	S	S
26	28	F	11	N	N	S	S
27	36	M	9	S	N	N	S
28	35	F	0	N	S	S	N
29	10	F	6	S	S	N	S
30	8	M	6	S	S	N	S
31	28	M	0	N	S	S	N
32	16	F	7	S	N	N	S
33	15	F	8	S	N	N	S
34	17	F	13	S	N	N	S
35	46	F	14	S	S	N	S
36	64	F	7	S	S	N	S
37	48	M	0	N	N	S	N
38	39	F	0	S	N	S	N
39	52	M	0	S	N	S	N
40	51	M	9	S	S	S	S

Caso #	Edad	Sexo	Horas	Comió sopa	Comió arroz	Comió barbacoa	Tuvo diarrea
41	63	F	0	N	N	N	N
42	45	F	0	N	S	S	N
43	68	M	6	S	S	S	S
44	48	F	0	N	N	S	N
45	49	F	0	N	N	S	N
46	42	M	0	S	N	S	N
47	61	M	0	N	S	S	N
48	66	F	4	S	N	N	S
49	35	F	0	S	S	S	N
50	68	M	0	N	S	S	N
51	38	F	0	N	N	S	N
52	27	F	0	N	S	S	N
53	34	M	0	N	N	S	N
54	41	M	7	S	S	S	S
55	67	F	0	N	N	N	N
56	45	F	0	N	N	S	N
57	46	M	0	S	S	S	N
58	53	M	6	S	N	S	S
59	29	F	0	N	N	S	N
60	45	M	0	S	S	N	N
61	44	M	8	S	N	S	S
62	55	F	8	N	S	S	S
63	28	F	0	S	N	S	N
64	29	F	12	S	N	N	S
65	36	M	7	S	N	S	S
66	39	M	0	N	S	N	N
67	61	F	5	S	S	S	S
68	57	F	14	S	S	S	S
69	48	M	0	S	N	S	N
70	32	M	0	N	S	N	N
71	54	M	12	S	S	S	S
72	42	F	7	S	S	S	S
73	21	M	0	N	S	S	N
74	19	M	12	S	N	S	S
75	28	M	12	S	N	N	S
76	23	F	0	N	N	N	N
77	29	F	10	S	S	S	S
78	28	M	8	S	S	S	S
79	29	M	10	S	N	S	S
80	27	M	11	S	N	N	S

Caso #	Edad	Sexo	Horas	Comió sopa	Comió arroz	Comió barbacoa	Tuvo diarrea
81	21	F	6	S	N	N	S
82	17	M	7	S	S	S	S
83	19	F	10	S	N	N	S
84	47	M	15	S	N	N	S
85	24	M	0	N	S	S	N
86	28	F	11	N	N	S	S
87	36	M	9	S	N	N	S
88	35	F	0	N	S	S	N
89	16	M	16	S	S	N	S
90	28	M	6	S	S	N	S
91	28	M	0	N	S	S	N
92	36	F	7	S	N	N	S
93	25	M	18	S	N	N	S
94	17	F	13	S	N	N	S
95	46	F	14	S	S	N	S
96	44	F	7	S	S	N	S
97	28	F	0	S	N	S	N
98	39	F	0	S	N	S	N
99	52	M	0	S	N	S	N
100	51	M	9	S	S	S	S
101	63	M	0	N	N	N	N
102	45	F	0	N	S	S	N
103	28	M	6	S	S	S	S
104	48	F	0	N	N	S	N
105	49	F	0	N	N	S	N
106	32	M	0	S	S	S	N
107	61	M	5	N	S	S	S
108	26	M	14	S	N	N	S
109	25	F	8	S	S	N	S
110	68	M	0	S	S	S	N
111	38	F	0	N	N	S	N
112	27	F	6	S	N	N	S
113	34	F	0	N	N	S	N
114	41	M	7	S	S	S	S
115	37	F	0	N	N	N	N
116	45	F	9	S	N	N	S
117	26	M	0	S	S	S	N
118	53	M	6	S	N	S	S
119	19	F	0	N	N	S	N
120	35	M	0	S	S	N	N

TEXTO: Recuento de Información ¹**1. Sistema de recuento por medio de palotes:**

Para contar a través de este sistema se requiere de una hoja de trabajo que muestre las posibles modalidades o clases de cada variable investigada y que tenga un espacio adjunto en donde se anotará un palote (/) o raya por cada elemento que posea determinada modalidad o valor. Conviene hacer agrupaciones de cuatro rayas diagonales y una raya horizontal (###) que representan grupos de cinco casos para facilitar la suma final.

Ejemplo: A partir de los instrumentos de recolección (o sea, los formularios) de una investigación se elaboró la siguiente hoja de trabajo para efectuar el cómputo de datos a través del sistema de palotes:

VARIABLE	MODALIDADES O CLASES	NUMERO DE CASOS	TOTALES
SEXO	Femenino	### III III III	19
	Masculino	### III III I	16
INGRESO FAMILIAR	0 - 3000	### III III	15
	3001 - 9000	### III	10
	9001 - 15000	### I	6
	15001 y más	###	4
ESTADO CIVIL	Soltero	### III III III	20
	Casado	###	5
	Divorciado	###	5
	Viudo	I	1
	Unión libre	###	4

Con el sistema de palotes también es posible realizar el cómputo de acuerdo a dos o más criterios; por ejemplo, véase la siguiente hoja de trabajo que resultó después de revisar los instrumentos de recolección:

VARIABLES	MODALIDADES O CLASES DE LA 1a. VARIABLE	MODALIDADES O CLASES DE LA 2a. VARIABLE	NUMERO DE CASOS	TOTALES
ESTADO CIVIL Y SEXO	Soltero	Femenino	### III	8
		Masculino	### III II	12
	Casado	Femenino	###	5
		Masculino		0
	Divorciado	Femenino	II	2
		Masculino	###	3
	Viudo	Femenino	I	1
		Masculino	I	1
	Unión libre	Femenino	###	3
		Masculino		0

Como puede observarse, los datos para el cómputo de la información se toman directamente del instrumento de recolección y se anotan en una hoja de trabajo como la mostrada.

El sistema de palotes tiene ciertas desventajas. En primer lugar, es impráctico si se estudia un gran número de casos; especialmente si se manejan demasiadas características a la vez. En segundo lugar, se presta a frecuentes errores, ya que no se puede verificar si se anotó o no un determinado caso. Cualquier error en el resultado final requiere la realización de todo el trabajo nuevamente.

¹ Extracto del documento: RECUESTO DE INFORMACION de J. García, J. Reynaga, B. de Garay Documento 139 de la serie EAS; Depto. de Medicina Social, Medicina Preventiva y Salud Pública. Facultad de Medicina, UNAM, mayo de 1980

Este sistema de recuento solo debe emplearse para menos de 100 casos con 3 a 5 variables a lo sumo.

2. Sistema de recuento por medio de tarjetas simples:

Este sistema de recuento consiste en transcribir los valores y modalidades de las variables estudiadas que se encuentran registradas en los formularios en tarjetas de cartulina (usualmente de 12.5 cms. X 7.5 cms.) y, a partir de éstas, efectuar el cómputo.

Para esto, se requiere que los valores o modalidades tengan asignado un código, preferentemente desde el asentamiento de los datos hecho en el instrumento en que se hayan recogido. Dichos códigos serán transcritos a la tarjeta de cartulina en los sitios establecidos para ello, a partir de una tarjeta modelo o matriz; por tanto, cada formulario da lugar a una tarjeta que contendrá la misma información ya codificada.

Ejemplo: Obsérvese el caso de un instrumento de recolección ya contestado y con los códigos ya asignados a los valores de las diferentes modalidades o clases a que pertenece el caso:

INSTRUMENTO DE RECOLECCION

1	Alumno No.			19
2	Preparatoria de origen	Privada (P)		B
		CCH UNAM (C)		
		Bachilleres (B)		
		Prepa UNAM (U)		
		Vocacional IPN (VI)		
3	Calificación 1er. examen	No presento (NP)		S
		No aprobado (NA)		
		Suficiente (S)		
		Bien (B)		
		Muy bien (MB)		
4	Calificación 2o. examen	No presento (NP)		B
		No aprobado (NA)		
		Suficiente (S)		
		Bien (B)		
		Muy bien (MB)		
5	Sexo	Femenino (F)		M
		Masculino (M)		
6	Edad	17 - 21 años (1)		2
		22 - 26 años (2)		
		27 y más (3)		
7	Otras ocupaciones	Otra carrera (C)		S
		Trabaja (T)		
		Solo esta carrera (S)		
8	Hrs. de estudio en casa	0 - 1 horas (1)		1
		2 - 3 horas (2)		
		4 o más horas (3)		
9	Lugar de estudio	Biblioteca (B)		P
		Casa propia (P)		
		Casa de amigos (A)		
		Otros lugares (O)		

Obsérvese la tarjeta a que da lugar el anterior formulario:

B	S	B
P	19	M
1	S	2

Como puede verse, se han ocupado el centro, las esquinas y las partes medias de cada borde de la tarjeta para anotar los códigos contenidos en el formulario. Los sitios donde se anota cada código se determinan previamente mediante la elaboración de una tarjeta modelo.

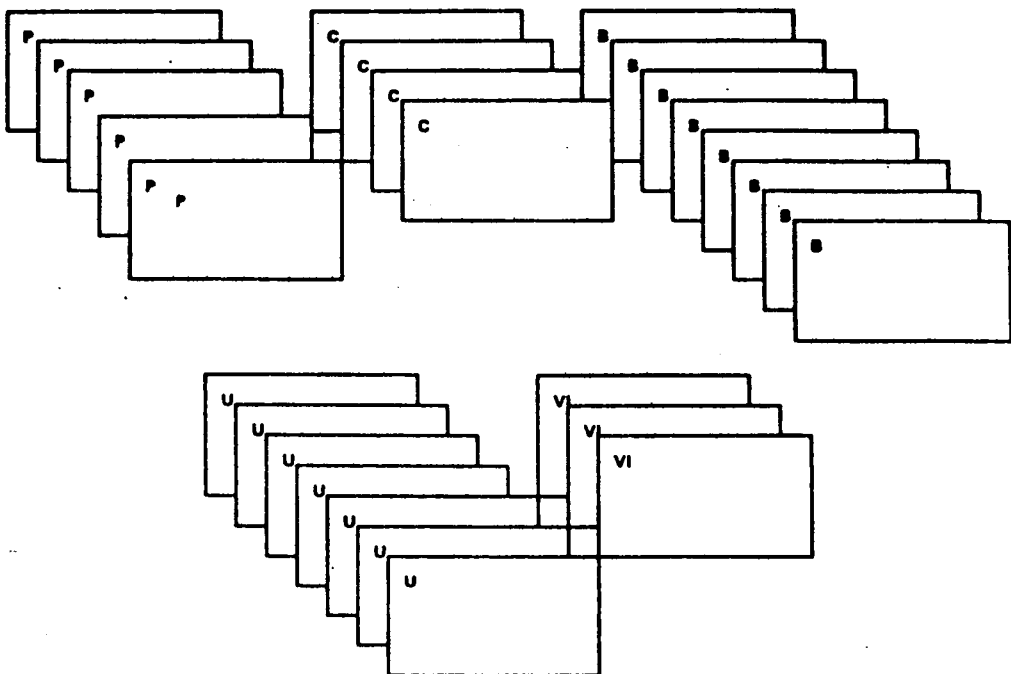
En el ejemplo citado, la tarjeta modelo fue elaborada con el siguiente formato:

TARJETA MODELO

Prepa de origen	Calificación 1er. examen	Calificación 2o. examen
Lugar de estudio	<u>Número del alumno</u>	Sexo
Tiempo de estudio	Otra ocupación	Edad

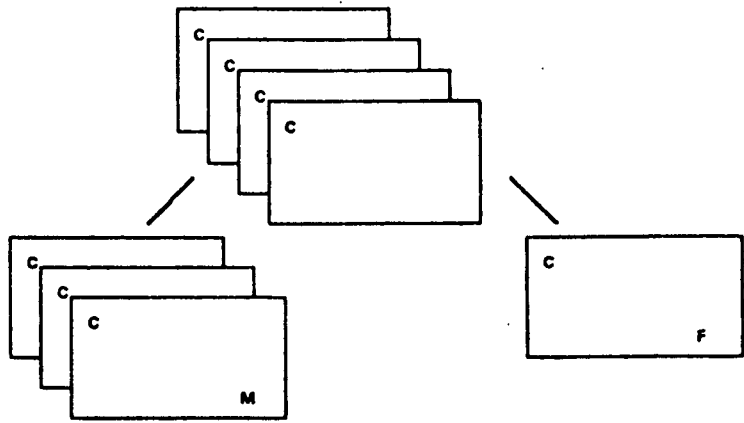
Para proceder al cómputo, simplemente se separan las tarjetas en paquetes diferentes, de acuerdo a las diversas modalidades o clases de una variable, y se cuentan las tarjetas resultantes.

Por ejemplo, si se deseara contar a los individuos según su escuela de origen se procedería a separar las tarjetas formando paquetes de la siguiente manera:



En caso necesario, cuando se desea realizar el cómputo de individuos clasificados bajo dos criterios, se procede a una nueva separación de los grupos de tarjetas resultantes de la primera clasificación para proceder a su cómputo.

Por ejemplo, para el caso que está sirviendo de ilustración, si se deseara conocer cuántos de los individuos que proceden del Colegio de Ciencias y Humanidades son hombres y cuántas son mujeres (esto es: si se deseará conocer a los individuos según escuela de origen y sexo simultáneamente) se procedería según el siguiente esquema:



El sistema de recuento por medio de tarjetas simples es útil para contar con comodidad hasta 500 casos que contenga cada uno a lo sumo ocho variables.

3. Sistema de recuento por medio de programas para computadoras:

Junto con la aparición de las computadoras de tipo personal se ha puesto a disposición del público general una serie de programas (software) que anteriormente solo estaban al alcance de los expertos.

Un tipo de programas de fácil empleo está constituido por los denominados *manejadores de bases de datos*. Entre tales programas destaca uno que se ha constituido en un estándar para el manejo de grandes volúmenes de información denominado dBASE III Plus.

En esencia, los programas manejadores de bases de datos, tales como dBASE III Plus, actúan como eficientes sistemas de archivo. Ellos permiten, de manera rápida y sencilla almacenar, revisar y modificar la información, ordenar los datos en diferentes formas y efectuar recuentos con suma facilidad.

Un sistema de recuento que emplee programas para computadoras de tipo personal permite contar miles de casos con cientos de variables casi sin ningún esfuerzo como no sea el de capturar cuidadosamente los datos.

Plan de trabajo para la semana 4

Sesión 1

El **recuento** de la información constituye la segunda etapa del método estadístico. Los procedimientos que se ponen en práctica en esta etapa tienen por propósito identificar las frecuencias de las modalidades o de las clases de las diversas variables que se estén estudiando en cada uno de los elementos investigados. Si durante la etapa de recuento se utilizan procedimientos sistemáticos para contar se pueden detectar diversas inconsistencias en la anotación de los datos previamente recogidos; por ello a esta etapa también se le conoce como de *revisión y recuento*.

Aunque es deseable que el recuento se haga por medio de programas de computadora (por ejemplo programas para manejo de bases de datos), puede ocurrir que no se tenga acceso inmediato a tal equipo o que el pequeño volumen de los casos y el escaso número de variables haga impráctico el uso de la computadora y de los programas de cómputo correspondientes. Por otra parte, el uso de los programas computacionales para el manejo de bases de datos requiere que el usuario tenga al menos una mínima experiencia en el recuento con procedimientos manuales.

Con el propósito de comprobar si el alumno ha adquirido habilidad para el recuento a través de procedimientos manuales, en esta sesión se discutirán las soluciones planteadas por el alumno a los problemas *Recuento de información (Procedimiento de palotes)* y *Recuento de información (Procedimiento de tarjetas simples)* y se discutirán conceptos relacionados con los procedimientos computarizados para el manejo de bases de datos.

Al final de la sesión, el alumno recibirá indicaciones para el empleo de los siguientes materiales de trabajo para la segunda sesión de la semana 4:

- **Problemas: Presentación de variables cualitativas** **2 páginas**
 - **Texto: La presentación estadística (cuadros y gráficos)** **8 páginas**
-

Sesión 2

El proverbio *una imagen vale más que mil palabras* es perfectamente aplicable a la tercera etapa del método estadístico: **la presentación**. En esta etapa se procede a la construcción de los cuadros y gráficos que dan inicio a la comprensión de los datos previamente recogidos y contados.

El entendimiento, a la vez global y detallado, de la forma en que se distribuyen las frecuencias de las modalidades o de las clases de las diversas variables que se estén estudiando en cada uno de los elementos investigados se facilita si se elaboran e interpretan tanto los cuadros con la información como sus correspondientes gráficos.

En esta sesión se discutirán las soluciones planteadas por el alumno a los problemas *Presentación de variables cualitativas*.

Al final de la sesión, el alumno recibirá indicaciones para el empleo del siguiente material de trabajo para la primera sesión de la semana 5:

- **Problemas: Presentación de variables cuantitativas** **páginas**
-

PROBLEMAS: Presentación de variables cualitativas

1. En cada uno de los siguientes cuadros existen dos o más errores. Identifíquelos y corrijalos:

**Estado Civil
Hospital Francés
1992**

Estado Civil	Número	%
Solteros	46	22.2
Casados	84	40.6
Viudos	25	12.1
Divorciados	16	7.7
Unión Libre	22	10.6
Pareja	14	6.8
Total	407	100.0

**Pacientes atendidos en Urgencias
según tipo de cirugía
Hospital Francés
1992**

Tipo de paciente	Número	%
Radical	226	23.3
Conservadora	452	66.7
Total	678	100.0

**Escolares según estado
de higiene oral**

Estado de higiene oral	Número	%
Excelente	38	2.75
Bueno	124	8.96
Regular	246	17.77
Malo	356	25.72
Pésimo	620	44.80
Total	1,384	100.00

**Alumnos según calificación
en examen final
Materia de Anatomía
1992**

Calificación	Número	%
MB	12	13.3
B	32	35.6
S	14	15.6
NA	8	8.9
NP	2	2.2
Pésimo	22	24.4

2. Una vez que haya corregido los cuadros anteriores, elabore los gráficos que corresponden a cada uno de los cuadros.

3. Para cada uno de los siguientes relatos elabore un cuadro y su correspondiente gráfico:

- a) Se entrevistó a 128 embarazadas con edades de 14 a 18 años para calificar su opinión con respecto a la conveniencia de la lactancia materna. De las entrevistadas, 48 respondieron que era muy conveniente, 24 que era conveniente, 16 que era inconveniente y el resto manifestó que no tenían ninguna opinión al respecto.

- b) Un grupo de 89 alumnos de primer año de la Facultad de Medicina fue entrevistado en 1991 para averiguar la orientación política predominante de sus padres

Se encontró que en 32 casos la orientación política de los padres fue conservadora, en 30 estudiantes los padres tenían orientación política de tipo moderada y en los últimos 27 estudiantes sus padres tenían orientación política de tipo liberal.

- c) Un grupo de 89 alumnos de primer año de la Facultad de Medicina fue entrevistado en 1991 para averiguar tanto la orientación política predominante de sus padres como el tipo de educación que recibieron durante sus estudios de educación primaria.

Se encontró que en 32 casos la orientación política de los padres fue conservadora; de estos estudiantes siete recibieron educación de tipo flexible, diez de tipo moderada y 15 de tipo rígida.

En 30 estudiantes cuyos padres tenían orientación política de tipo moderada se encontró que nueve recibieron educación flexible, diez moderada y 11 rígida.

Finalmente, en el caso de los últimos 27 estudiantes, cuyos padres tenían orientación política de tipo liberal, se encontró que 14 recibieron educación de tipo flexible, ocho de tipo moderada y cinco de tipo rígida.

TEXTO: La presentación estadística (cuadros y gráficos)

En el método estadístico, luego de las etapas de recolección y de recuento, suele desarrollarse la etapa denominada *presentación*. En ella se elaboran y describen los datos de los estudios o investigaciones a través de cuadros y gráficos.

A través de los siguientes ejemplos se podrá apreciar la importancia de esta etapa.

Supóngase que se lee o se escucha la siguiente descripción: *De 120 trabajadores de una fábrica de gabinetes metálicos que fueron seguidos a lo largo de la década 1981-1990, en la población de San Garabato, 81 estaban expuestos regularmente a ruido intenso; de estos, casi tres cuartas partes presentaron diversos grados de sordera al cabo del período de observación. Por otra parte, de 39 trabajadores que solo se exponían esporádicamente al ruido, la gran mayoría casi no presentaban manifestaciones de sordera al final del período de observación.*

Salvo pocas excepciones, la gran mayoría de las personas no logra captar de primera intención la magnitud de las cifras y el modo en que se distribuyen las variables a las que se está haciendo referencia. Es casi seguro que una importante proporción de las personas tendrá que releer la descripción dos o más veces para emitir una opinión respecto a la información leída o escuchada.

En cambio, al disponer de un cuadro en el que se presenta dicha información, la comprensión de los datos es casi automática y las magnitudes numéricas se comparan de manera sumamente ágil. Obsérvese el efecto que se logra con la presentación tabular de la información:

**Trabajadores de una fábrica de gabinetes metálicos
según exposición al ruido y presencia de sordera**

San Garabato
1981-1990

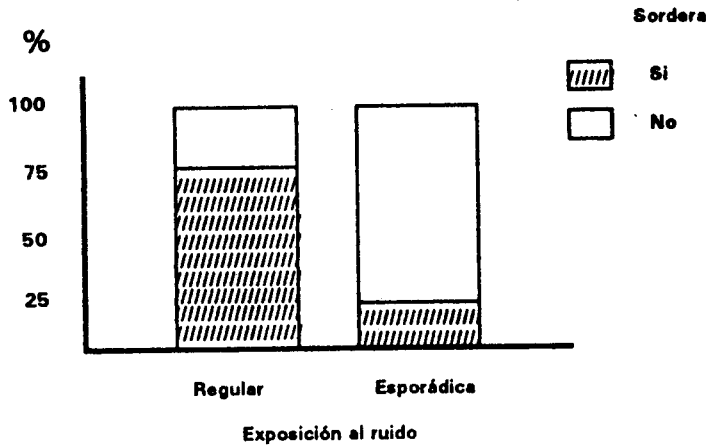
Exposición al ruido	Presencia de sordera		Totales
	SI	NO	
Regular	60 (74%)	21 (26%)	81 (100%)
Esporádica	7 (18%)	32 (82%)	39 (100%)

Si bien los cuadros permiten efectuar comparaciones precisas de tipo numérico, también se hace necesario disponer de una forma de presentación que facilite la revisión visual rápida de las características esenciales de los datos.

Obsérvese cómo el gráfico que corresponde a la tabla anterior propicia una fácil comprensión de la información con la sola inspección visual:

**Trabajadores de una fábrica de gabinetes metálicos
según exposición al ruido y presencia de sordera**

San Garabato
1981-1990



Como ya se dijo antes, el principal propósito de la etapa de presentación es facilitar la comprensión rápida y práctica de la información recién contada. Por lo anterior, tanto la elaboración de cuadros como la elaboración de gráficos debe cumplir con normas mínimas que promuevan la facilidad en la descripción de la información en ellos contenida.

En el caso de los cuadros pueden mencionarse como normas mínimas las siguientes:

- Cada cuadro debe tener un título completo que responda a las preguntas
 - ¿Qué se está presentando?,
 - ¿Cómo se está presentando?,
 - ¿De dónde provienen los datos? y
 - ¿De cuándo son los datos?
- Las variables deben tener un nombre exacto que esté acorde con la escala de clases o modalidades que se hayan definido para dichas variables.
- La escala de cada variable debe tener modalidades o clases que sean exhaustivas y mutuamente excluyentes.
- En vista de que los cuadros tienen por propósito presentar las características numéricas de los datos, los cálculos deben ser exactos.

En el caso de los gráficos pueden mencionarse como normas mínimas las siguientes:

- Cada gráfico debe tener un título completo que responda a las preguntas

*¿Qué se está presentando?,
¿Cómo se está presentando?,
¿De donde provienen los datos? y
¿De cuándo son los datos?*

- Para no distorsionar la apreciación visual de las variables que se estén presentando, la relación entre el eje vertical (ordenada) y el eje horizontal (abscisa) de los gráficos debe ser de 1 a 1.5; por ejemplo: si la ordenada mide 3 centímetros, la ordenada medirá 4.5
- La escala de cada variable debe tener modalidades o clases que sean exhaustivas y mutuamente excluyentes.
- En vista de que los gráficos tienen por propósito agilizar la captación visual de las características esenciales de los datos sin distorsiones de ningún tipo, debe evitarse el uso de colores o de efectos tridimensionales de volumen:

En las siguientes páginas se muestran los cuadros y los gráficos pertinentes para la presentación de:

- Una sola variable cualitativa nominal u ordinal
- Una sola variable cuantitativa discreta
- Una sola variable cuantitativa continua
- Dos variables cualitativas simultáneamente (nominales u ordinales)
- Dos variables cuantitativas simultáneamente (discretas o continuas)

En cada caso se encuentran las indicaciones técnicas que deben tomarse en cuenta para la construcción tanto de los cuadros como de los gráficos correspondientes. Aunque el listado anterior no abarca a toda la amplia gama de cuadros y gráficos que podrían construirse, se considera que los que se incluyen forman el arsenal básico de cualquier profesional.

Presentación (cuadros y gráficos)
UNA SOLA VARIABLE CUALITATIVA NOMINAL U ORDINAL

TITULO COMPLETO →
 QUE →
 COMO →
 DONDE →
 CUANDO →

PROFESORES SEGUN
ESTADO CIVIL
 UNIVERSIDAD NUEVO MUNDO
1986

NOMBRE EXACTO DE LA VARIABLE →

ESTADO CIVIL	Numero	%
SOLTEROS	384	25.5
CASADOS	542	35.9
DIVORCIADOS	266	17.6
VIUDOS	194	12.9
UNION LIBRE	122	8.1
T O T A L :	1.508	100.0

TABLA DE DISTRIBUCION
DE FRECUENCIAS

MODALIDADES MUTUAMENTE EXCLUYENTES Y EXHAUSTIVAS →

FRECUENCIAS RELATIVAS

↑
FRECUENCIAS ABSOLUTAS

TITULO COMPLETO →

QUE →
 COMO →
 DONDE →
 CUANDO →

PROFESORES SEGUN
ESTADO CIVIL
 UNIVERSIDAD NUEVO MUNDO
1986

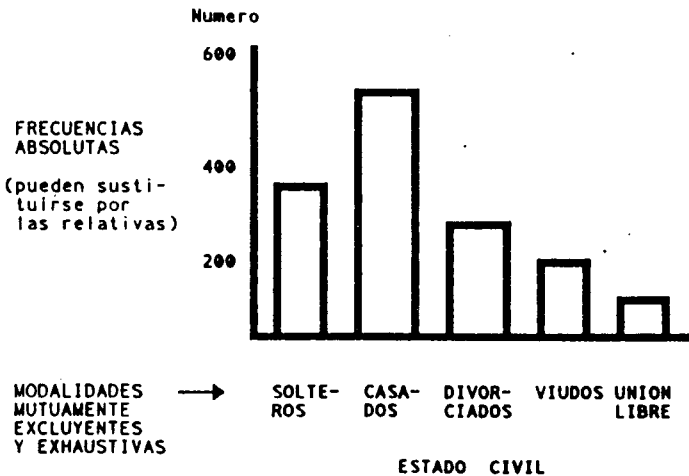


GRAFICO DE BARRAS
SEPARADAS

MODALIDADES MUTUAMENTE EXCLUYENTES Y EXHAUSTIVAS →

NOMBRE EXACTO DE LA VARIABLE →

**Presentación (cuadros y gráficos):
UNA SOLA VARIABLE CUANTITATIVA DISCRETA**

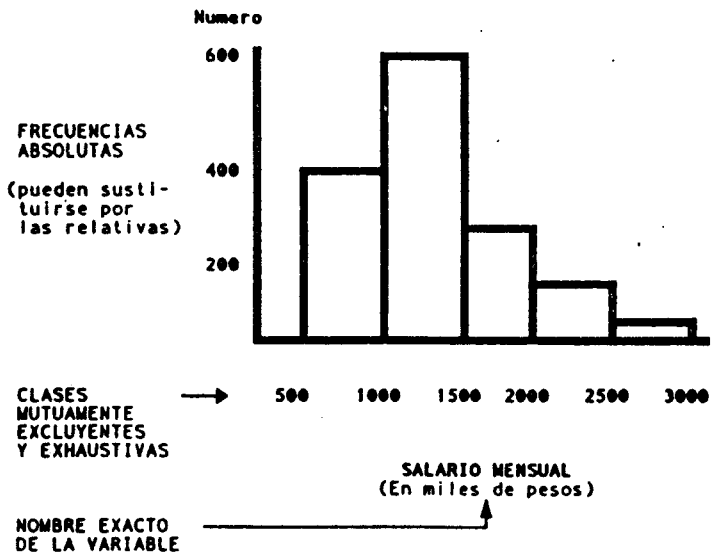
TITULO COMPLETO →
 QUE COMO → PROFESORES SEGUN SALARIO
 DONDE CUANDO → UNIVERSIDAD NUEVO MUNDO 1986

HOMBRE EXACTO DE LA VARIABLE →	SALARIO MENSUAL *	Numero	%
	500 - 999	400	26.6
	1000 - 1499	600	39.8
CLASES MUTUAMENTE EXCLUYENTES Y EXHAUSTIVAS →	1500 - 1999	299	19.8
	2000 - 2499	149	9.9
	2500 - 2999	60	3.9

TABLA DE DISTRIBUCION DE FRECUENCIAS

T O T A L : 1,508 100.0 FRECUENCIAS RELATIVAS
 * En miles de pesos ↑ ↑ FRECUENCIAS ABSOLUTAS

TITULO COMPLETO →
 QUE COMO → PROFESORES SEGUN SALARIO
 DONDE CUANDO → UNIVERSIDAD NUEVO MUNDO 1986



HISTOGRAMA

Presentación (cuadros y gráficos) UNA SOLA VARIABLE CUANTITATIVA CONTINUA

TITULO COMPLETO →
 QUE →
 COMO →
 DONDE →
 CUANDO →

PROFESORES SEGUN PESO
 UNIVERSIDAD NUEVO MUNDO 1986

NOMBRE EXACTO DE LA VARIABLE →

PESO *	Numero	%
40.0 - 49.9	152	10.0
50.0 - 59.9	302	20.0
60.0 - 69.9	608	40.3
70.0 - 79.9	298	19.8
80.0 - 89.9	148	9.9
TOTAL :	1.508	100.0

TABLA DE DISTRIBUCION DE FRECUENCIAS

CLASES MUTUAMENTE EXCLUYENTES, EXHAUSTIVAS Y CONTINUAS →

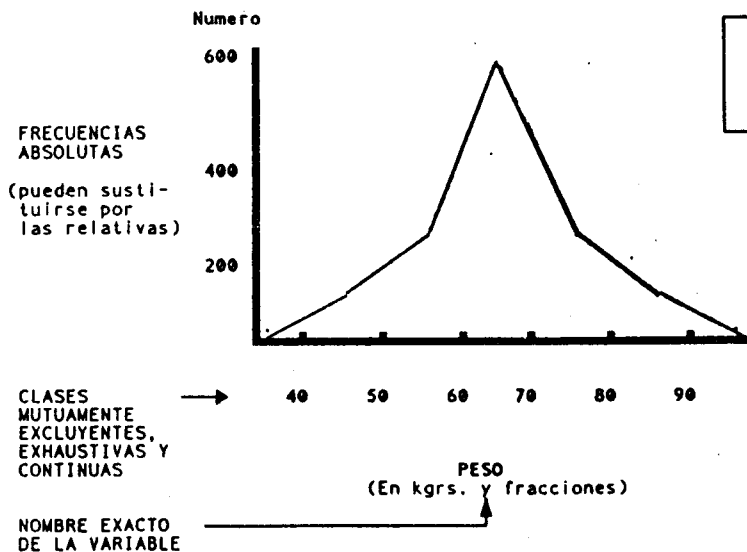
* En kgrs y fracciones

FRECUENCIAS RELATIVAS

FRECUENCIAS ABSOLUTAS

TITULO COMPLETO →
 QUE →
 COMO →
 DONDE →
 CUANDO →

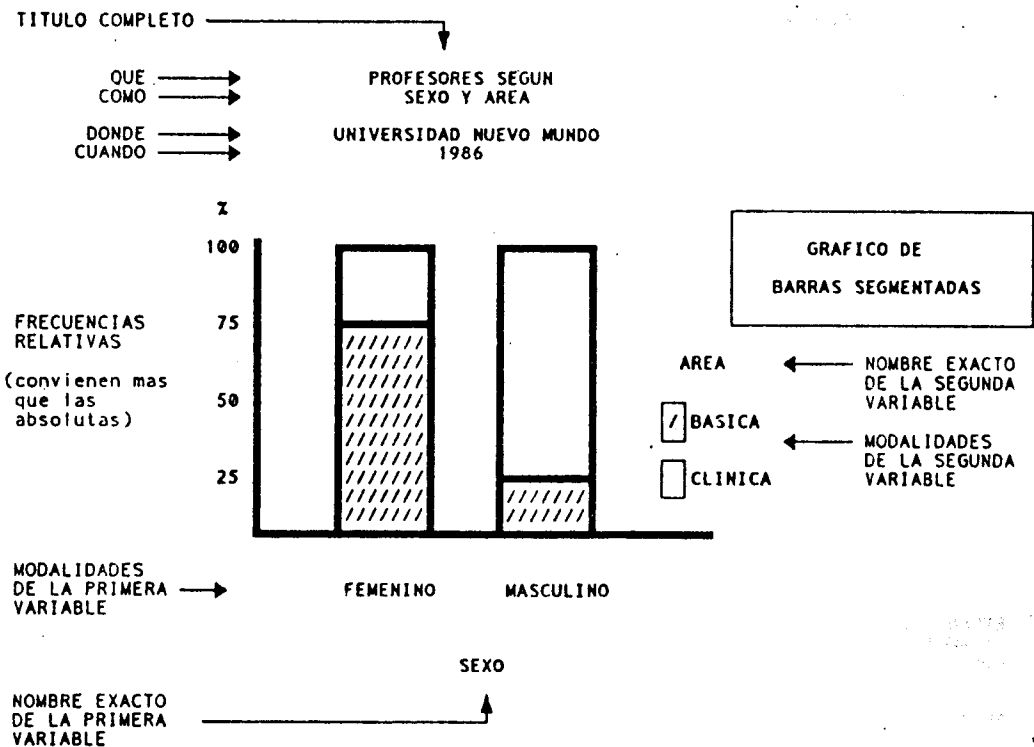
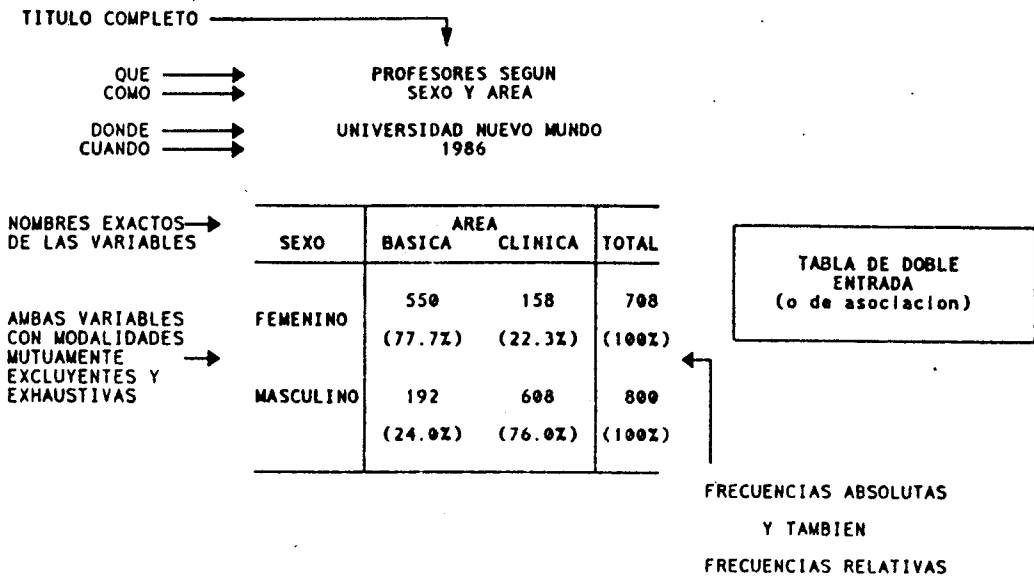
PROFESORES SEGUN PESO
 UNIVERSIDAD NUEVO MUNDO 1986



POLIGONO DE FRECUENCIAS

Presentación (cuadros y gráficos)

DOS VARIABLES CUALITATIVAS SIMULTANEAMENTE (nominales u ordinales)



Presentación (cuadros y gráficos)

DOS VARIABLES CUANTITATIVAS SIMULTANEAMENTE (discretas o continuas)

TITULO COMPLETO →

QUE COMO → SIETE PROFESORES SEGUN ANTIGUEDAD Y SUELDO

DONDE CUANDO → UNIVERSIDAD NUEVO MUNDO 1986

NOMBRES EXACTOS DE LAS VARIABLES →

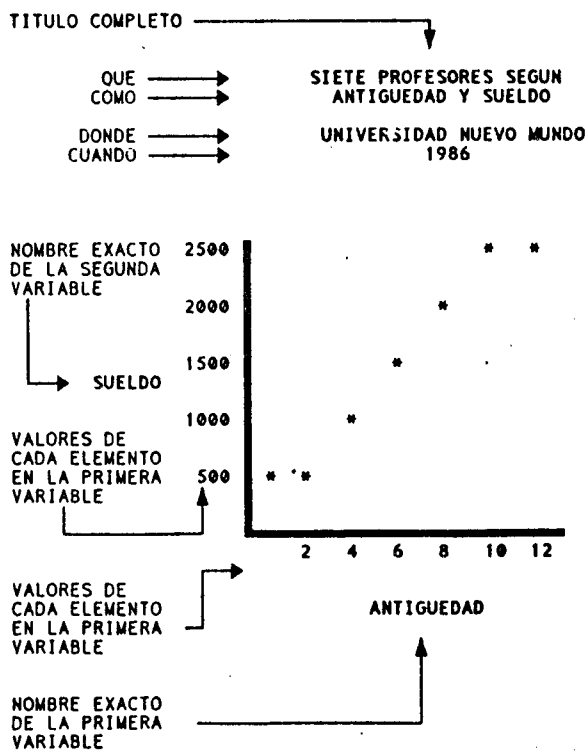
PROFESOR	ANTIGUEDAD ¹	SUELDO ²
1	4	1 000
2	2	500
3	8	2 000
4	6	1 500
5	10	2 500
6	1	500
7	12	2 500

CADA ELEMENTO CON SU PAR DE VALORES EN AMBAS VARIABLES →

¹ EN AÑOS

² EN MILES DE PESOS

LISTA SIMPLE DE PARES DE DATOS PARA DOS VARIABLES CORRESPONDIENTES A CADA ELEMENTO



PROBLEMAS: Presentación de variables cuantitativas

1. En cada uno de los siguientes cuadros existen dos o más errores. Identifíquelos y corrijalos:

**Pacientes epilépticos según
Número de convulsiones diarias
Hospital Francés
1992**

Estado Civil	Número	%
1 a 3	6	21.4
4 a 6	14	50.0
6 a 10	5	17.9
10 a 12	3	10.7
Total	38	100.0

**Familias según
número de hijos
Colonia Rosedal**

Número de hijos	Número de familias	%
1 a 2	232	29.5
3 a 4	452	57.5
4 a 6	78	9.9
7 a 8	24	3.1
Total	786	100.0

**Alumnos según aciertos
en examen final
Materia de Anatomía
1992**

Número de aciertos	Número de alumnos	%
0 a 20	38	2.75
21 a 40	124	8.96
41 a 80	246	17.77
61 a 80	356	25.72
81 a 100	620	44.80
Total	1,384	100.00

**Alumnos según porcentaje de
aciertos en examen final
Materia de**

Porcentaje de aciertos	Número de alumnos	%
0.0 a 20.9	38	42.75
21.0 a 40.9	124	8.96
41.9 a 60.9	246	17.77
61.0 a 80.9	356	25.72
81.0 a 100.0	620	44.80
Total	1,384	100.00

2. Una vez que haya corregido los cuadros anteriores, elabore los gráficos que corresponden a cada uno de los cuadros.

3. Para cada uno de los siguientes relatos elabore un cuadro y su correspondiente gráfico:

- a) Durante el año de 1991, en el Centro de Salud "Dr. Luis Pasteur", se midió el aumento de peso desde el principio de la gestación hasta el final de la misma en 128 embarazadas. De las gestantes, 68 aumentaron entre 6 y 8 kilos; 24 aumentaron entre 9 y 10 kilos; 16 aumentaron entre 11 y 12 kilos y el resto aumentó entre 13 y 14 kilos.
- b) Durante el año de 1991, en el Centro de Salud "Dr. Luis Pasteur", se midió el aumento de peso desde el principio de la gestación hasta el final de la misma en 128 embarazadas. De las gestantes, 68 aumentaron entre 6.000 y 8.999 kilos; 24 aumentaron entre 9.000 y 10.999 kilos; 16 aumentaron entre 11.000 y 12.999 kilos y el resto aumentó entre 13.000 y 14.999 kilos.
- c) Durante el año de 1987, en el Hospital Cosmos, se midió la duración de las intervenciones quirúrgicas.

Luego de efectuar el recuento, se encontró lo siguiente:

32 operaciones duraron entre 30.00 y 59.99 minutos
48 operaciones duraron entre 60.00 y 89.99 minutos
96 intervenciones duraron entre 90 y 119.99 minutos
48 intervenciones duraron entre 120 y 149.99 minutos
32 intervenciones duraron entre 150 y 179.99 minutos

Sesión 1

Aunque pudieran parecer más complicados los procedimientos para construir cuadros y gráficos para la presentación de variables cuantitativas que para las variables cualitativas, es posible que el alumno haya comprobado que, si se siguen los principios básicos de sencillez y pertinencia, la elaboración de dichos cuadros y gráficos puede ser igualmente simple y sistemática.

Para verificar lo anterior, en esta sesión se discutirán las soluciones planteadas por el alumno al problema *Presentación de variables cuantitativas*.

Al final de la sesión, el alumno recibirá indicaciones para el empleo de los siguientes materiales de trabajo para la segunda sesión de la semana 5:

- **Problema: Crítica a la presentación estadística en artículos de investigación** 1 página
 - **Texto: La presentación estadística (cuadros y gráficos) << ya utilizado en las sesiones anteriores >>** 8 páginas
-

Sesión 2

En esta sesión el alumno discutirá, luego de haberlo comprobado, que no siempre lo publicado es perfecto y a veces ni siquiera bueno. Para verificar si se ha adquirido capacidad crítica para interpretar y valorar cuadros y gráficos elaborados por otros, en la presente sesión se revisarán los resultados de las actividades señaladas en el problema *Crítica a la presentación estadística en artículos de investigación*.

Con la anterior actividad se dejará a la etapa del método estadístico denominada *Presentación de la Información* para ingresar a una cuarta etapa del método conocida como *Descripción Estadística*.

Esta etapa consiste, fundamentalmente, en la condensación o síntesis de la información con el propósito de comunicar, en cifras, la esencia o médula de los datos.

Describir estadísticamente a la información significa calcular e interpretar medidas de resumen tales como el promedio o los porcentajes. Una persona bien adiestrada en el cálculo e interpretación de medidas de resumen es capaz de evocar o imaginar grandes fragmentos de la realidad solo con escuchar el valor de alguna de las medidas de resumen; por ejemplo, cualquiera que escuche que un grupo de alumnos de licenciatura tiene un peso promedio de 94 kilos, de inmediato traerá a su mente la imagen de un grupo de estudiantes muy corpulentos.

En vista de la importancia que tiene la adquisición de habilidad para el cálculo e interpretación de medidas de resumen, al final de esta sesión se discutirá un esquema relacionado con las medidas de resumen para la descripción estadística tanto de las variables de tipo cualitativo como de las de tipo cuantitativo y se darán indicaciones para el empleo del siguiente material de trabajo para la primera sesión de la semana 6:

- **Problemas: Cálculo e interpretación de medidas de resumen para variables cualitativas: razones, proporciones y tasas** 2 páginas
 - **Texto: Medidas de resumen para variables cualitativas: Razones, proporciones y tasas** 4 páginas
-

PROBLEMAS: Crítica a la presentación estadística en artículos de investigación

1. A partir de artículos publicados en las revistas *Salud Pública de México* y *American Journal Of Epidemiology*, localice gráficos de los tipos que se mencionan posteriormente; describa la información que en ellos se presenta y trate de encontrar defectos técnicos a tales gráficos.

- Barras simples
- Histograma
- Polígono de frecuencias
- Barras segmentadas
- Correlación

REYNAGA, JESÚS
1993
Salud Pública de México

PROBLEMAS:**Cálculo e interpretación de medidas de resumen para variables cualitativas: razones, proporciones y tasas**

1. En un distrito escolar urbano se encontró que existía una población escolar a nivel primaria de 123,456 niños. Para uso de tales niños se encontraron disponibles 1,380 sanitarios y 820 bebederos de agua.

Por otra parte, en un distrito escolar rural, se encontró una población escolar de primaria de 34,244 niños. En las escuelas de dicho distrito se encontraron 179 sanitarios y 54 bebederos de agua.

Con la información anterior calcule e interprete las siguientes razones para los dos distritos escolares:

- razón niños / sanitario
- razón niños / bebedero

2. En un país en desarrollo ocurrieron 432,614 defunciones a lo largo de un año. De tales defunciones las correspondientes a niños menores de un año fueron 125,315.

Por otra parte, en un país desarrollado, ocurrieron 1' 347,915 defunciones en total. De tales defunciones 114,304 correspondieron a niños menores de un año.

Con la información anterior calcule e interprete la proporción de las defunciones de menores de un año en relación con el total de las defunciones. Compare la importancia de la mortalidad de los niños menores de un año en cada país.

3. En un estudio sobre deserción escolar a nivel de licenciatura se encontró que de 894 estudiantes que ingresaron a la carrera de química 148 abandonaron los estudios durante el primer año de la carrera. Por otra parte, de 1,311 alumnos que ingresaron a la carrera de medicina 118 abandonaron los estudios durante el primer año de la carrera.

Calcule y compare las tasas de deserción en primer año para ambas licenciaturas.

4. En un estudio efectuado en 1987 sobre el personal académico de instituciones públicas y privadas de educación superior se encontró que había 894 hombres y 1,814 mujeres en una universidad pública mientras que en una universidad privada de la misma región había 633 hombres y 274 mujeres.

En la universidad pública se encontró que de los hombres 296 poseían grados académicos superiores al de licenciatura y de las mujeres 524 poseían grados análogos. Con respecto a la universidad privada, se encontró que 92 hombres y 38 mujeres poseían grados académicos superiores al de licenciatura.

En cuanto a las solicitudes de permiso para efectuar estudios de posgrado en el último ciclo lectivo, se encontró que 9 hombres y 21 mujeres de la universidad pública pidieron tales permisos mientras que en la universidad privada las cifras correspondientes fueron de 2 hombres y 6 mujeres.

Con la información anterior calcule y describa para cada universidad lo siguiente:

- razones hombres / mujeres
- proporciones de académicos con estudios de posgrado
- tasas de solicitud para efectuar estudios de posgrado.

5. Para su propio grupo de compañeros de clase, calcule las siguientes medidas de resumen para variables cualitativas:

- razón estudiantes nacidos en el D.F. / estudiantes nacidos fuera del D.F.
- proporción de estudiantes que usan anteojos
- tasa de catarro en el mes anterior

Averigue los datos necesarios y efectue el cálculo de las mismas medidas de resumen por lo menos para otros dos grupos de alumnos.

TEXTO : **Medidas de Resumen para Variables Cualitativas:
Razones, Proporciones y Tasas**

INTRODUCCION

La cuarta etapa del método estadístico se denomina *descripción*; en ella se realizan procedimientos de cálculo numérico que producen cifras o medidas que condensan o resumen en sí mismas una gran cantidad de información.

La condensación numérica de los datos da lugar a medidas que, debidamente analizadas y entendidas, permiten evocar o imaginar las características fundamentales de toda la información que dio origen a dichas medidas.

La ventaja principal del cálculo de las medidas de resumen consiste en que a través de ellas se facilita la comunicación de las peculiaridades esenciales de grandes volúmenes de datos.

Por ejemplo: si una persona se quisiera informar acerca de las enfermedades que dieron origen a los fallecimientos ocurridos en una ciudad podría efectuar una lectura de todos y cada uno de los certificados de defunción de las personas fallecidas; sin embargo, luego de la lectura de 40 o 50 certificados, ya habría perdido el interés por la información o quizá ya habría caído en una total confusión al tratar de recordar la causa que condujo a la muerte a cada uno de los difuntos.

En cambio, si se hubieran calculado las medidas de resumen llamadas *tasas* para las cinco primeras causas de muerte y esta información estuviera disponible, tales medidas de resumen permitirían imaginar las características básicas de la mortalidad en la ciudad de referencia.

En el caso de las *variables de tipo cualitativo* la condensación de la información produce los siguientes tres principales tipos de medidas de resumen:

- Razones
- Proporciones
- Tasas

RAZONES

Definición: Una razón es una medida de resumen para variables cualitativas que consiste en la comparación, a través de una división, entre dos conjuntos o grupos de elementos de diferente o igual naturaleza.

Ejemplo: Se desea condensar la siguiente información: en un hospital rural existen 18 residentes que se encargan de la atención de 126 pacientes; por otra parte, en un hospital urbano existen 8 residentes que se encargan de atender a 32 pacientes.

Procedimiento: de acuerdo a la definición de razón deben dividirse dos grupos de elementos. En este caso, para el hospital rural se divide el número de pacientes entre el número de residentes:

$$\text{Razón pacientes/residentes en hospital rural} = \frac{126}{18} = 7$$

En cuanto al hospital urbano, se divide también el número de pacientes entre el número de residentes de dicho hospital:

$$\text{Razón pacientes/residentes en hospital urbano} = \frac{32}{8} = 4$$

Interpretación:

Para el caso del hospital rural: Existen siete pacientes a atender por cada residente.

Para el caso del hospital urbano: Existen cuatro pacientes a atender por cada residente.

En conclusión: parece haber mayor cantidad de trabajo para los residentes del hospital rural.

PROPORCIONES

Definición: Una proporción es una medida de resumen para variables cualitativas que consiste en la comparación, a través de una división, entre un subconjunto y el conjunto al que pertenece.

Ejemplo: Se desea condensar la información siguiente: en un país subdesarrollado ocurrieron 422,350 defunciones en total; de ellas 124,352 pertenecieron a niños que aún no cumplían el año de edad.

Por otra parte, en un país industrializado ocurrieron un total de 1' 721,215 defunciones de las cuales 206,876 se presentaron en niños menores de un año de edad.

Procedimiento: de acuerdo a la definición de proporción deben dividirse un subconjunto entre el conjunto al que pertenece.

Así, para el país subdesarrollado:

$$\text{proporción de muertes de menores de un año en país subdesarrollado} = \frac{124,352}{422,350} = 0.29$$

Por otra parte, para el país industrializado:

$$\text{proporción de muertes de menores de un año en país industrializado} = \frac{206,876}{1'721,215} = 0.12$$

Interpretación:

Para el caso del país subdesarrollado: La importancia del subconjunto de muertes de menores de un año, en relación con el conjunto total de muertes en el país es de 0.29, lo que equivale a casi la tercera parte de todas las defunciones.

Para el caso del país industrializado: La importancia del subconjunto de muertes de menores de un año, en relación con el conjunto total de muertes en el país es de 0.12, lo que equivale a aproximadamente una octava parte de todas las defunciones.

En conclusión: la importancia de las defunciones de niños menores de un año es mucho mayor en el país subdesarrollado que en el país industrializado.

Notas: Conviene mencionar que, en ocasiones, las proporciones son multiplicadas por el número 100 y con ello se habla de los *porcentajes*. Así, para el ejemplo, podría haberse hablado de que los porcentajes de defunciones de menores de un año en el país subdesarrollado y en el país industrializado eran de 29% y de 12%, respectivamente.

TASAS

Definición: Una tasa es una medida de resumen para variables cualitativas que consiste en la comparación, a través de una división, entre el número de veces que ocurre un cierto tipo de eventos y la población en la que *puede ocurrir* dicho tipo de eventos. Usualmente el resultado de tal división consiste en una cifra fraccionaria menor a uno; por ello, el resultado de la división suele ser multiplicado por alguna constante que sea múltiplo del número 10.

Ejemplo: En una ciudad, a lo largo del año 1991, ocurrieron 345 defunciones por cáncer de próstata. Dicha ciudad tenía una población total de 2' 453, 310 habitantes. De ellos, 1' 210, 425 eran del sexo masculino. Se desea condensar la información de tal manera que los cálculos produzcan una medida de resumen que permita imaginar o evocar la magnitud del riesgo que existe para los habitantes de tal ciudad de fallecer por cáncer de próstata.

Procedimiento: de acuerdo a la definición de tasa debe dividirse el evento entre la población en la cual dicho evento puede ocurrir. El resultado debe multiplicarse por un múltiplo del número 10.

$$\text{tasa de defunciones por cáncer de próstata} = \frac{345}{1,210,425} \times 100,000 = 28.50 \approx 29$$

Interpretación: En la ciudad estudiada, a lo largo del año de referencia, fallecieron 29 de cada 100,000 habitantes del sexo masculino por cáncer de próstata.

Notas: En *primer lugar* obsérvese que la división del número de eventos, en este caso las defunciones por cáncer de próstata, fue efectuada entre el número de habitantes del sexo masculino y no entre el total de la población, ya que solo los hombres pueden padecer tal patología.

En *segundo lugar* nótese que la constante utilizada para multiplicar el resultado de la división anterior fue el número 100,000. Tal cifra fue escogida porque con ella se obtiene un resultado final que incluye uno o dos dígitos enteros (es usual que las tasas de mortalidad específica tengan como constante para multiplicar el resultado de la división al número 100,000)

En *tercer lugar* conviene destacar la importancia del uso de tasas para condensar información de tipo cualitativo. Supóngase que interesa comparar la mortalidad por enfermedades del corazón entre los jóvenes de 15 a 24 años edad y los ancianos de 65 a 74 años y que se utilizan los siguientes datos:

Grupo poblacional	Defunciones por enfermedades del corazón	Total de defunciones por todas las causas	Proporción de defunciones por enfermedades del corazón en relación con las defunciones por todas las causas
Jóvenes	12,763	48,999	.26
Ancianos	11,425	306,025	.04

Si se optara por comparar las *proporciones* de defunciones por enfermedades del corazón de ambos grupos poblacionales podría quedar la impresión de que las enfermedades del corazón son un problema mucho más grave en el grupo de jóvenes que en el de ancianos; esta falsa imagen se produce porque no se está tomando en cuenta la cantidad de personas que componen a cada grupo poblacional; es decir, no se está tomando en consideración el volumen de las poblaciones en que pueden ocurrir ese tipo de eventos.

En cambio obsérvese lo que ocurre cuando se usan los siguientes datos, necesarios para el cálculo de tasas:

Grupo poblacional	Defunciones por enfermedades del corazón	Número de habitantes que forman el grupo poblacional	Tasa de defunciones por enfermedades del corazón
Jóvenes	12,763	23'900,749	53.4 defunciones por cada 100,000 habitantes de ese grupo
Ancianos	11,425	6'368,450	179.4 defunciones por cada 100,000 habitantes de ese grupo

Con los últimos resultados ya se puede efectuar una comparación correcta al decir que por cada 100,000 jóvenes solo fallecen 53.4 de ellos por enfermedades del corazón; en cambio por cada 100,000 ancianos fallecen 179.4 por la misma causa. Lo anterior significa que tales padecimientos constituyen un riesgo mucho más importante como causa de mortalidad en el grupo de ancianos que en el grupo de jóvenes.

En *cuarto lugar*, aunque las tasas son calculadas con datos que pertenecen a periodos ya concluidos, se considera que tales tasas sirven para anticipar la probabilidad o el riesgo de ocurrencia de un evento en una población para periodos inmediatos a aquel para el que se efectuaron los cálculos.

Plan de trabajo para la semana 6

Sesión 1

En esta sesión serán discutidos los resultados obtenidos para los problemas *Cálculo e interpretación de medidas de resumen para variables cualitativas: razones, proporciones y tasas.*

A través de la discusión de los trabajos se espera que el alumno aprecie la importancia de la condensación de masas de datos en cifras sintéticas con el propósito de comunicar la esencia de la información, en este caso de tipo cualitativa.

Al final de la sesión, el alumno recibirá indicaciones para el empleo de los siguientes materiales de trabajo para la segunda sesión de la semana 6:

- **Problemas: Cálculo e interpretación de medidas de resumen para variables cualitativas: tasas de uso común en salud pública** **1 página**
 - **Textos:** Los que recomiende el profesor y dos útiles materiales de lectura publicados por el Departamento de Salud Pública de la Facultad de medicina de la UNAM: *TASAS Y RAZONES DE USO HABITUAL EN SALUD PUBLICA* (Documento EAS-0017) y *FORMULAS DE ALGUNAS TASAS DE USO FRECUENTE* (documento EAS-0018)
-

Sesión 2

Gran parte del trabajo de la salud pública requiere de la disposición de indicadores que reflejen el estado de salud de la población. En su gran mayoría, tales indicadores de salud están constituidos por tasas.

En vista de que el alumno, al egresar de la carrera, se verá involucrado mayor o menormente en acciones de salud pública, conviene que comprenda que el cálculo e interpretación de tasas requiere de una preparación estadística básica.

Por lo anterior, en esta sesión se discutirán los resultados de sus indagaciones orientadas por la resolución de los problemas *Cálculo e interpretación de medidas de resumen para variables cualitativas: tasas de uso común en salud pública.*

Al final de la sesión, el alumno recibirá indicaciones para el empleo de los siguientes materiales de trabajo para la primera sesión de la semana 7:

- **Problemas: Medidas de Resumen para variables cuantitativas en series simples de valores: Moda y Amplitud, Mediana y Percentiles** **2 páginas**
 - **Texto: Medidas de Resumen para variables cuantitativas en series simples de valores: Moda y Amplitud, Mediana y Percentiles** **2 páginas**
-

PROBLEMAS: Cálculo e interpretación de medidas de resumen para variables cualitativas: tasas de uso común en salud pública (1)

1. Averigüe y compare las siguientes tasas para México, para un país más desarrollado y para un país menos desarrollado (en lo posible para un mismo año):

- Tasa de mortalidad general
- Tasas de mortalidad para las tres principales causas de defunción
- Tasa de mortalidad infantil
- Tasa de natalidad

2. Averigüe y compare las tasas de incidencia para tres diferentes enfermedades en México.

3. Averigüe y compare las tasas de incidencia de una misma enfermedad en al menos tres diferentes regiones de México.

4. Averigüe y compare las tasas de incidencia de una misma enfermedad en al menos tres diferentes épocas de México.

5. Averigüe y compare las diferencias entre la tasa de incidencia y la tasa de prevalencia de la diabetes y del catarro.

6. Averigüe y compare las diferencias entre las tasas de letalidad de la lepra y la de la rabia humana.

¹ Dos útiles materiales de lectura están publicados por el Departamento de Salud Pública de la Facultad de Medicina de la UNAM: *TASAS Y RAZONES DE USO HABITUAL EN SALUD PUBLICA* (Documento EAS-0017) y *FORMULAS DE ALGUNAS TASAS DE USO FRECUENTE* (documento EAS-0018)

PROBLEMAS:

**Medidas de Resumen para Variables Cuantitativas
en Series Simples de Valores:
Moda y Amplitud, Mediana y Percentiles**

Problema 1. Luego de criar durante dos años a un grupo de 20 cerdos en una granja modelo que funcionaba bajo el régimen de sociedad cooperativa, en condiciones de estricta higiene animal, se sacrificaron y observaron sus cerebros en búsqueda de quistes de cisticercos. Los hallazgos se compararon con otro grupo de 20 cerdos criados en los patios de las casas de la misma localidad.

CERDOS PROVENIENTES DE LA GRANJA MODELO

Cerdo No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Quistes	1	3	0	2	3	1	2	2	1	4	1	0	1	1	2	1	0	2	4	1

CERDOS PROVENIENTES DE PATIOS DE CASAS

Cerdo No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Quistes	7	4	2	4	5	1	0	2	2	9	2	3	4	4	1	6	4	4	3	1

PREGUNTAS:

- Para cada grupo identifique la moda y emita una opinión con respecto a los resultados.
- Para cada grupo calcule la amplitud y responda a la siguiente cuestión: ¿Cuál grupo fue más homogéneo y cuál más heterogéneo en cuanto al número de quistes?
- Para cada grupo identifique la mediana (o percentil 50) y emita una opinión con respecto a los resultados.
- Para cada grupo identifique los percentiles 25 y 75 y llene la siguiente tabla:

	Valor mínimo	Percentil 25	Percentil 75	Valor máximo
Grupo "Granja"	1	5	15	4
Grupo "Pacios"	1	5	15	9

e) Emita una opinión en relación con los hallazgos en términos de porcentajes de cada grupo; por ejemplo: "mientras que en el grupo Granja el 25% de los cerdos tuvo 1 o menos quistes cerebrales por animal, en el grupo Pacios el 25 % de los cerdos tuvo 2 o menos quistes cerebrales por cada animal"

No olvide discutir los hallazgos para el 50% de los animales encontrados entre los percentiles 25 y 75 de cada uno de los grupos. También, opine sobre la homogeneidad de los grupos.

Problema 2. Se identificó a una muestra de 20 madres de familia de una localidad y se les preguntó acerca del número de veces que habían asistido en el transcurso del año anterior al centro de salud para recibir pláticas de orientación sobre nutrición infantil. En dicha visita se

aprovechó para pesar y medir a los hijos de las entrevistadas y para dejarles unos atractivos folletos para la anotación de los pesos y tallas infantiles y una serie de recetas de comidas para niños.

Transcurrido un año de la entrevista se volvió a visitar a las mismas 20 madres de familia y se les volvió a hacer la misma pregunta del año anterior. A continuación se muestran los resultados de ambas visitas:

PRIMERA ENTREVISTA

Madre No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Visitas al Centro	1	3	0	2	3	1	2	2	1	4	1	0	1	1	2	1	0	2	4	1

SEGUNDA ENTREVISTA

Madre No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Visitas al Centro	7	4	2	4	5	1	0	2	2	9	2	3	4	4	1	6	4	4	3	1

PREGUNTAS:

- Para cada entrevista identifique la moda y emita una opinión con respecto a los resultados.
- Para cada entrevista calcule la amplitud y responda a la siguiente cuestión: ¿En cuál entrevista se encontró más homogeneidad y en cuál más heterogeneidad en cuanto al número de visitas?
- Para cada entrevista identifique la mediana (o percentil 50) y emita una opinión con respecto a los resultados.
- Para cada entrevista identifique los percentiles 25 y 75 y llene la siguiente tabla:

	Valor mínimo	Percentil 25	Percentil 75	Valor máximo
Primera entrevista	1	6.4	6.4	4
Segunda entrevista	1	13.6	13.6	9

- Emita una opinión en relación con los hallazgos en términos de porcentajes en cada entrevista; por ejemplo: *mientras que en la primera entrevista el 25% de las madres manifestó haber asistido 1 o menos veces al centro de salud, en la segunda entrevista el 25% de las madres señaló que habían visitado el centro de salud 2 o menos ocasiones.*

No olvide discutir los hallazgos para el 50% de las madres encontradas entre los percentiles 25 y 75 de cada uno de las dos entrevistas. También, opine sobre la homogeneidad de las visitas en ambas entrevistas.

TEXTO:**Medidas de Resumen para Variables Cuantitativas
en Series Simples de Valores:
Moda y Amplitud, Mediana y Percentiles**

La siguiente serie simple de valores se utilizará como ejemplo para ilustrar el cálculo e interpretación de las medidas de resumen:

**PESO EN KILOGRAMOS DE UN GRUPO DE 20 NIÑOS
DE UN AÑO DE EDAD**

9.1	9.4	8.9	9.6	10.5	8.8	9.4	9.2	9.0	8.1
9.3	8.8	9.5	9.7	9.2	9.4	9.6	9.0	9.4	9.8

MODA:

Definición: Es el valor que en una serie se repite con mayor frecuencia.

Procedimiento: Simplemente buscar si existe un valor que se repita con mayor frecuencia que los demás. En este caso el valor 9.4 se repite en cuatro ocasiones.

Interpretación: "El peso más frecuente en el grupo de niños fue 9.4 kilogramos".

AMPLITUD: O RANGO

Definición: Es la diferencia entre el valor mayor y el valor menor de una serie.

Procedimiento: Encontrar, por sustracción o resta, la diferencia entre el valor más grande de la serie (X_{\max}) y el valor más pequeño (X_{\min}). En esta serie el resultado equivale a $10.5 - 8.1 = 2.4$.

Interpretación: "La diferencia de peso entre el niño más pesado y el niño más ligero fue de 2.4 kilogramos".

MEDIANA (O PERCENTIL 50) :

Definición: En una serie de valores ordenados de menor a mayor, o viceversa, es aquel valor que divide en dos partes de igual tamaño a toda la serie.

Procedimiento: Ordenar la serie y localizar el valor que la divida en dos partes de igual tamaño, de tal manera que en una parte quede el 50% de los datos y en la otra el 50% restante.

8.1	8.8	8.8	8.9	9.0	9.0	9.1	9.2	9.2	9.3
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

9.4	9.4	9.4	9.4	9.5	9.6	9.6	9.7	9.8	10.5
-----	-----	-----	-----	-----	-----	-----	-----	-----	------

Usualmente este valor se ubica en la posición $(N + 1) / 2$.

En este caso, a la posición $(N + 1) / 2$ le corresponde el lugar $(20 + 1) / 2 = 10.5$ avo.

Como tal lugar (10.5avo) se encuentra vacío, se considera que en ese sitio se encontraría un valor que fuera el promedio de los que están colocados en las posiciones 10a. y 11a.; es decir, correspondería al promedio de los valores 9.3 y 9.4, que equivale a un valor de 9.35.

Interpretación: "La mitad de los niños tuvieron un peso igual o menor que 9.35 kilogramos y la otra mitad pesaron 9.35 o más kilogramos"

PERCENTILES (Pp) :

Definición: En una serie de valores ordenados, de menor a mayor o viceversa, es aquel valor que divide en dos partes porcentualmente complementarias a toda la serie. Por ejemplo: el percentil 40 divide a la serie en una parte que contiene al 40 % de los valores iguales o inferiores a él y, simultáneamente, en otra parte que contiene al 60% de los valores de la serie iguales o mayores a dicho percentil.

Procedimiento: Ordenar la serie y localizar el valor que la divide en los porcentajes complementarios deseados.

Por ejemplo: para encontrar el valor del percentil 25 debe localizarse a aquel que deje a una cuarta parte de los valores con menor o igual magnitud a él y a las tres cuartas partes restantes de los valores con magnitud más grande o iguales a él.

8.1	8.8	8.8	8.9	9.0	9.0	9.1	9.2	9.2	9.3
9.4	9.4	9.4	9.4	9.5	9.6	9.6	9.7	9.8	10.5

En esta serie el primer valor 9.0 ocupa una posición tal que hasta él se encuentra el 25% de los casos y, simultáneamente, desde él se encuentra el 75% restante de los casos.

Usualmente cualquier percentil se ubica en la posición $(N \times P) / 100$. Para el ejemplo del percentil 25, a la posición $(N \times 25) / 100$ le corresponde el lugar $(20 \times 25) / 100 = 5o$. Como ya se observó, la quinta posición está ocupada por el primer valor 9.0 antes ya identificado.

Interpretación: (Válida para el Percentil 25 o P25) : " El 25% de los niños tuvieron un peso de 9.0 kilogramos o menor y el 75% restante tuvo peso de 9.0 kilogramos ó mayor".

Plan de trabajo para la semana 7

Sesión 1

En esta sesión serán discutidos los resultados obtenidos previamente por los alumnos para los problemas: *Medidas de Resumen para variables cuantitativas en series simples de valores: Moda y Amplitud, Mediana y Percentiles*

Al igual que en el caso de las medidas de resumen para variables cualitativas, con las variables de tipo cuantitativo deberá apreciarse la importancia de la condensación de masas de datos en cifras sintéticas con el propósito de comunicar la esencia de la información, en este caso de tipo cuantitativa y, en particular, para series simples de datos.

Al final de la sesión, el alumno recibirá indicaciones para el empleo de los siguientes materiales de trabajo para la segunda sesión de la semana 7:

- **Problemas: Medidas de Resumen para Variables Cuantitativas en Series Simples de Valores: Promedio y Desviación Estándar** 1 página
 - **Texto: Medidas de Resumen para Variables Cuantitativas en Series Simples de Valores: Promedio y Desviación Estándar** 3 páginas
-

Sesión 2

Aunque casi todo mundo aprendió desde muy temprano la forma en que se calcula el promedio y ha utilizado abundantemente esta medida de resumen con propósitos de síntesis, por ejemplo para evaluar su propio desempeño escolar, casi nadie ha reflexionado en el significado real de esta medida y mucho menos en su acompañante inseparable: la desviación estándar.

Para juzgar el avance conceptual y metodológico del alumno en relación con estas dos importantes medidas de resumen, en esta sesión se discutirán los resultados obtenidos en la resolución de los Problemas *Medidas de Resumen para Variables Cuantitativas en series simples de valores: promedio y desviación estándar*

Al final de la sesión, el alumno recibirá indicaciones para el empleo de los siguientes materiales de trabajo para la primera sesión de la semana 8:

- **Problemas: Cuestión de Curvas** 2 páginas
 - **Texto: Relación entre la Desviación Estándar y la Curva Normal** 7 páginas
-

PROBLEMAS:**Medidas de Resumen para Variables Cuantitativas
en Series Simples de Valores:
Promedio y Desviación Estándar**

Problema 1. Se estudió el tiempo que tardaban dos medicamentos en hacer efecto para controlar el dolor de cabeza. Para ello, se formaron dos grupos de pacientes diagnosticados con dicho problema y a uno de los grupos se les dio el medicamento **A** y a otro el medicamento **B**.

A continuación se muestran los resultados:

Tiempo para quitar el dolor de cabeza
(minutos)

Medicamento **A**

3.5	5.5	6.5	15.0	45.5	6.0	5.5	4.0	8.5	10.0
-----	-----	-----	------	------	-----	-----	-----	-----	------

Tiempo para quitar el dolor de cabeza
(minutos)

Medicamento **B**

19.5	18.5	18.5	21.0	20.5	20.0	19.0	21.0	21.5	20.0
------	------	------	------	------	------	------	------	------	------

PREGUNTAS:

1. Si a todos los pacientes que tomaron el medicamento **A** se les quitara el dolor de cabeza al mismo tiempo, ¿a los cuantos minutos ocurriría el alivio? Si a todos los pacientes que tomaron el medicamento **B** se les quitara el dolor de cabeza al mismo tiempo, ¿a los cuantos minutos ocurriría el alivio?
2. ¿Cual medicamento produjo mejoría de manera más homogénea?, ¿Por qué?, ¿Cual medicamento preferiría tomar usted?, ¿Por qué?

TEXTO:**Medidas de Resumen para Variables Cuantitativas
en Series Simples de Valores:
Promedio y Desviación Estándar**

La siguiente serie simple de valores se utilizará como ejemplo para ilustrar el cálculo e interpretación de las medidas de resumen:

**PESO EN KILOGRAMOS DE UN GRUPO DE 20 NIÑOS
DE UN AÑO DE EDAD**

9.1	9.4	8.9	9.6	10.5	8.8	9.4	9.2	9.0	8.1
9.3	8.8	9.5	9.7	9.2	9.4	9.6	9.0	9.4	9.8

PROMEDIO: *Media*

Definición: Es el valor que tendrían todos los datos de una serie numérica si ellos fueran de igual valor.

Fórmula:

$$\bar{x} = \frac{\sum x}{n}$$

Procedimiento: Sumar todos los valores y dividir tal suma entre el número de valores que componen a la serie simple. En este caso la suma es: 185.7 y, entonces, el promedio vale:

$$185.7 / 20 = 9.285 \text{ Kilogramos}$$

Interpretación: "Si todos los niños tuvieran peso igual, éste sería de 9.285 kilogramos".

DESVIACION ESTANDAR ¹

Definición: Es la raíz cuadrada de la varianza. A su vez, la varianza equivale al promedio de las desviaciones o diferencias cuadráticas de cada valor de una serie con respecto al promedio de dicha serie.

¹ Junto con esta sección se recomienda estudiar documentos y realizar ejercicios relacionados con la *curva normal*

Fórmula:

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

Procedimiento:

Paso 1. Obtener el promedio de la serie de valores. En este caso, por tratarse de una serie

simple de valores se usa la fórmula anteriormente revisada: $\bar{x} = \frac{\sum x}{n}$

Paso 2. Calcular la desviación o diferencia de cada valor en relación con el promedio de la serie; es decir obtener una serie de valores $(x - \bar{x})$

Paso 3. Elevar al cuadrado cada una de las anteriores desviaciones; es decir obtener una serie de valores $(x - \bar{x})^2$

Paso 4. Efectuar la suma de desviaciones cuadráticas; es decir obtener el valor $\sum (x - \bar{x})^2$

Paso 5. Dividir la suma anterior entre el número de valores; es decir: obtener el promedio de

desviaciones cuadráticas o varianza: $\frac{\sum (x - \bar{x})^2}{n}$

Paso 6. Obtener la raíz cuadrada del anterior promedio; es decir obtener la desviación estándar:

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

Así, los cálculos para los pesos de los 20 niños son los siguientes:

Para el paso 1: El promedio, ya calculado en párrafos anteriores vale: 9.285 kgrs.

Para los pasos 2, 3 y 4 se recomienda utilizar una tabla auxiliar como la siguiente para efectuar los cálculos:

Cada uno de los valores X_i	Desviación de cada valor con respecto al promedio $(X_i - \bar{X})$ (paso 2)	Elevación al cuadrado de cada una de las desviaciones $(X_i - \bar{X})^2$ (paso 3)
9.1	-0.185	0.034
9.4	0.115	0.113
8.9	-0.385	0.148
.	.	.
.	.	.
9.8	0.515	0.265
		4.446 (paso 4)

\bar{X} 9.285

Para el paso 5: El promedio de desviaciones cuadráticas, o varianza, vale entonces:

$$4.446 / 20 = 0.222$$

Para el paso 6: Finalmente la desviación estándar, que equivale a la raíz cuadrada de la varianza equivale a la raíz cuadrada de 0.222 o sea: 0.471 kgrs.

Interpretación: La interpretación está condicionada a la suposición de que los valores tienen una distribución semejante a la de la *curva normal*².

Dicha interpretación puede ser realizada en múltiples sentidos ya que se sabe que el 68.27% de los valores de una serie que se distribuye como la curva normal están agrupados alrededor del promedio si a éste se le resta una vez y también se le suma una vez el valor calculado para la desviación estándar.

Para el ejemplo, puede decirse entonces que el 68.27% de los niños tuvieron pesos que fluctuaron desde 8.814 kgrs (es decir: 9.285 kgrs menos 0.471 kgrs) hasta 9.756 kgrs. (es decir: 9.285 más 0.471 kgrs.).

Por otra parte, como se sabe que la curva normal tiene una área que equivale a un total de 100%, entonces también puede decirse que hubo un 31.73% de los niños que pesaron menos de 8.814 kgrs. o que pesaron más de 9.756 kgrs. (ya que si a 100% se le resta 68.27% quedan 31.73%).

² Ver documentos relacionados con la *curva normal*

PROBLEMAS:**Cuestión de curvas****Relato**

Se practicó cuantificación de glucosa en sangre a 2,100 adultos jóvenes, aparentemente sanos.

Se les encontró un promedio de glucosa de 80 mgrs. por cada 100 ml. de sangre. El valor de la desviación estándar fue de 5 mgrs./100 ml.

Por lo anterior, de inmediato se supo que el .6826, es decir el 68.26 %, de las personas tenían cifras de glucosa que fluctuaban entre 75 y 85 mgrs. por cada 100 ml.

Preguntas 1 y 2

¿Cómo se supo que el .6826 de las personas tenían cifras de glucosa que fluctuaban entre 75 y 85 mgrs./ 100 ml.? ¿A cuántas personas correspondió tal proporción?

Tip: Recuerde las proporciones de área bajo la curva que hay desde -1 hasta +1 desviaciones estándar alrededor del promedio

Preguntas 3 y 4

¿Qué proporción de personas tuvo 85 y más mgrs. de glucosa por cada 100 ml. de sangre? ¿A cuántas personas correspondió tal proporción?

Tip: Recuerde la proporción de área bajo la curva que hay desde +1 desviaciones estándar en adelante

Preguntas 5 y 6

¿Qué proporción de personas tuvo 75 y menos mgrs. de glucosa por cada 100 ml. de sangre? ¿A cuántas personas correspondió tal proporción?

Tip: Recuerde la proporción de área bajo la curva que hay antes de -1 desviaciones estándar

Preguntas 7 y 8

¿Qué proporción de personas tuvo 82 y más mgrs. de glucosa? ¿A cuántas personas correspondió tal proporción?

Tip: Efectue la transformación de su X valor de glucosa a un Z valor en la abscisa de la curva normal mediante la fórmula

$$z = \frac{x - \bar{x}}{s} \text{ y utilice dicho valor } Z \text{ para ubicarse en la tabla de áreas bajo la curva}$$

Preguntas 9 y 10

¿Qué proporción de personas tuvo 82 y menos mgrs. de glucosa? ¿A cuántas personas correspondió tal proporción?

Tip: Recuerde que toda el área bajo la curva vale 1.0 y que en la pregunta 7 ya calculó usted la proporción de personas con 82 y más mgrs. de glucosa.

Preguntas 11 y 12

¿Qué proporción de personas tuvo 71 y más mgrs. de glucosa? ¿A cuántas personas correspondió tal proporción?

Tip: Efectue la transformación de su X valor de glucosa a un Z valor en la abscisa de la curva normal mediante la fórmula

$z = \frac{x - \bar{x}}{s}$ y utilice dicho valor Z para ubicarse en la tabla de áreas bajo la curva

Preguntas 13 y 14

¿Qué proporción de personas tuvo 71 y menos mgrs. de glucosa? ¿A cuántas personas correspondió tal proporción?

Tip: Recuerde que toda el área bajo la curva vale 1.0 y que en la pregunta 11 ya calculó usted la proporción de personas con 71 y más mgrs. de glucosa.

Preguntas 15 a 20

¡Solo para mentalidades avanzadas! ¡Esta pregunta vale por seis!: ¿A partir de que cifra de glucosa se encontraba el 10 % superior de las personas?

Tip número 1:

En primer lugar tendrá que buscar en la tabla de áreas bajo la curva cuál es el valor z a la derecha del cual queda una área de .10 y que a su izquierda deja una área que equivale al .90 del área total bajo la curva.

En segundo lugar, ya disponiendo del valor Z , tendrá que

pasar de la fórmula $z = \frac{x - \bar{x}}{s}$ hacia la fórmula

$x = (zs) + \bar{x}$ para encontrar el X valor de glucosa buscado

Explicación algebraica:

Primero: $z = \frac{x - \bar{x}}{s}$

Segundo: $zs = x - \bar{x}$

Tercero:

$$(zs) + \bar{x} = x ; \text{ por tanto: } x = (zs) + \bar{x}$$

Tip número 2:

Consiga la tabla de distribución de frecuencias que contenga los valores de las 2,100 personas y calcule el percentil 90 mediante la fórmula para series agrupadas.

TEXTO: Relación entre la Desviación Estándar y la Curva Normal

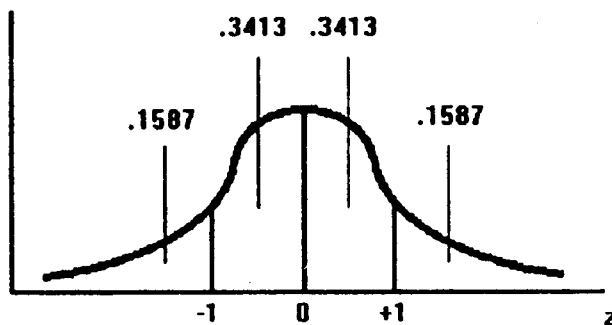
Cuando se calcula la desviación estándar para una serie de datos no siempre es evidente el significado del resultado obtenido y menos lo es aún si no se compara con la desviación estándar de otra serie diferente de datos.

Para muchas personas podría tener significado que le dijeran que el promedio de peso de un grupo de 300 personas fue de 80 kilos pues, si se acuerda de la definición del promedio, imaginaría que si todos los individuos tuvieran el mismo peso este sería de 80 kilos; sin embargo para quienes no tienen conocimiento de las características básicas del modelo de la curva normal podría carecer de significado que le mencionaran que la desviación estándar del peso de las mismas personas fue de 5 kilos.

Interpretar la desviación estándar y comprender cabalmente lo que ella significa en relación con los datos que se están manejando solo es posible a la luz del conocimiento del modelo de la curva normal.

PROPIEDADES PRINCIPALES DE LA CURVA NORMAL

1. La curva normal es un polígono de frecuencias en forma de campana para el que están calculadas sus áreas en función de los diversos valores del eje horizontal o abscisa.



2. En la abscisa se encuentran valores de tipo cuantitativo continuo, genéricamente denominados valores Z , cuyas magnitudes teóricamente pueden ir, de izquierda a derecha desde $-\infty$ hasta $+\infty$ (desde menos infinito hasta más infinito).

3. El *promedio* de todos los valores z de la abscisa equivale a cero, pues la mitad son negativos y la mitad son positivos. En el sitio de la abscisa que corresponde al cero, es decir al promedio, se encuentra la parte más alta de la curva. En este sitio también se encuentra la mediana de todos los valores z de la abscisa, pues el 50% de ellos está antes del cero y el 50% restante se encuentra después.

4. La curva es simétrica alrededor del promedio; esto es, hay una mitad izquierda que es reflejo de la mitad derecha.

5. En la abscisa existen segmentos unitarios de igual longitud y de tamaño 1. Los segmentos a la izquierda del promedio tienen signo negativo y los segmentos a la derecha del promedio

tienen signo positivo. Tales segmentos, denominados *desviaciones estandar* pueden dividirse en fracciones infinitamente pequeñas y continuas.

6. La curva es asintótica; es decir, sus extremos teóricamente nunca tocan a la abscisa. Por ello, la longitud de la abscisa podría ser infinitamente larga; sin embargo se acostumbra graficar solo hasta la distancia de tres segmentos a la izquierda y a la derecha del promedio.

7. Toda el área bajo la curva vale 1. Por lo anterior el área a la izquierda del promedio vale .5 y el área a la derecha del promedio vale también .5

8. El área que se encuentra sobre el segmento de la abscisa que va desde el promedio hasta el valor z de +1 vale .3413; por simetría, el area que se encuentra sobre el segmento que va desde el promedio hasta el valor z de -1 de la abscisa también vale .3413

Por lo anterior el área que se encuentra por arriba del amplio segmento que va desde el valor z de -1 hasta el valor z de +1 equivale a .6826; es decir a la suma de .3413 mas .3413

9. El área que se encuentra sobre el segmento de la abscisa que van más allá del valor z de +1 vale .1587; por simetría, el area que se encuentra sobre el segmento que va más alla (hacia menos infinito) del valor z de -1 de la abscisa también vale .1587

10. Para cualquier segmento de la abscisa, y aún para fracciones de segmento, se encuentran calculadas las áreas correspondientes en tablas como la siguiente:

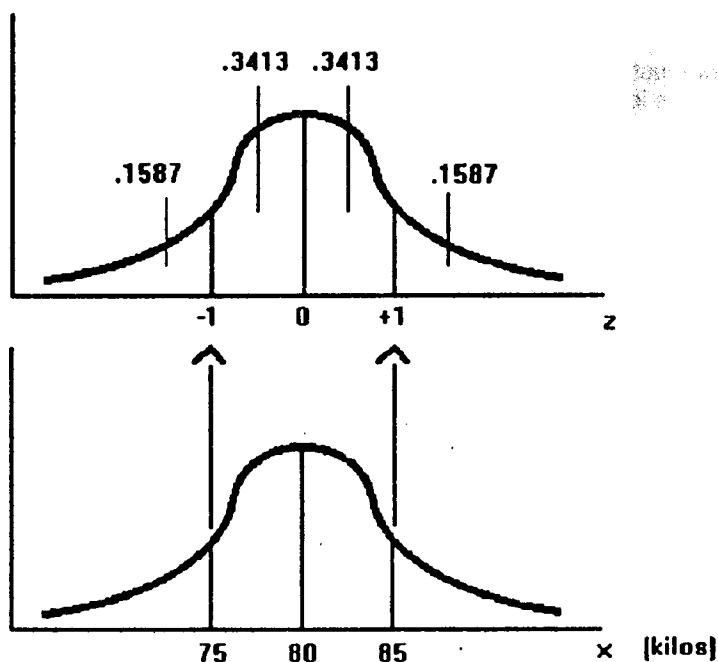
(A) Valor z	(B) Area entre el promedio y el valor z	(C) Area más allá del valor z
0.00	.0000	.5000
0.25	.0987	.4013
0.50	.1915	.3085
0.75	.2734	.2266
1.00	.3413	.1587
1.25	.3944	.1056
1.50	.4332	.0668
1.65	.4505	.0495
1.75	.4599	.0401
1.96	.4750	.0250
2.00	.4772	.0228
2.58	.4950	.0050

APROVECHAMIENTO DE LAS PROPIEDADES DE LA CURVA NORMAL PARA LA INTERPRETACION DE LA DESVIACION ESTANDAR

Al principio de este documento se comentó que sin conocer las características básicas del modelo de la curva normal podría carecer de significado que se mencionara que el valor de la desviación estándar del peso de 300 personas fue de 5 kilos.

Una vez que se han comprendido las propiedades principales de la curva normal es posible entender el significado del valor de la desviación estándar del peso de las 300 personas si se hacen suposiciones como las siguientes:

Suponiendo que al graficar el peso de los 300 individuos con un polígono de frecuencias, el gráfico resultante fuera muy parecido al modelo de la curva normal como se muestra en la siguiente ilustración:



entonces podría decirse que:

- el área bajo la curva de valores de peso que contiene a los individuos vale 300 de manera semejante a la propiedad del modelo de la curva normal de que su área vale 1;
- a la izquierda del promedio existen 150 individuos y a la derecha del promedio existen los otros 150;

- así como en la curva normal existe una área de .3413 sobre el segmento que va desde el valor z de 0 hasta el valor z de + 1 en la curva de valores x (es decir kilos de peso) habrá el .3413 de 300 o sea que habrá 102 personas sobre el segmento que va desde el valor x de 80 kilos hasta el valor x de 85 kilos;
- de acuerdo al párrafo anterior, habra 204 personas con pesos que van desde 75 hasta 85 kilos;
- al igual que en la curva normal existe simetría alrededor del promedio, se puede considerar que en la curva de valores de peso habrá 102 personas sobre el segmento que va desde 80 kilos hasta 75 kilos de peso;
- en la curva de valores peso habrá un .1587 de las 300 personas; es decir 48 personas, con pesos de 85 y más kilos;
- de manera semejante a la curva normal, por simetría habrá un .1587 de las 300 personas; es decir 48 personas, con pesos de 75 y menos kilos.

Como puede apreciarse, una vez que se conocen las características del modelo de la curva normal, la interpretación del resultado de la desviación estándar que se haya calculado para una serie de datos es mucho más fácil y brinda una gran cantidad de información sobre la manera en que se distribuyen los valores.

Para confirmar que la comprensión del significado de la desviación estándar brinda una importante cantidad de información obsérvese el siguiente ejemplo:

Relato: Se aplicó un mismo examen escrito a dos grupos de 90 alumnos cada uno. En un caso se imprimió el examen en hojas de color amarillo paja y en otro caso en hojas de color marrón. Se midió con cronómetro el tiempo, en minutos y fracciones, que tardaron los alumnos en entregar el examen y se calculó el promedio y la desviación estándar para ambos grupos obteniéndose los siguientes resultados:

Grupo	Promedio	Desviación Estándar
Color Paja	45'	5'
Color Marrón	45'	15'

Algunas interpretaciones a partir de los valores de la desviación estándar:

- Los alumnos a quienes se aplicó el examen impreso en hojas color paja entregaron el examen en tiempos más homogéneos, pues el .6826 de ellos (es decir 61 alumnos) lo entregaron entre 40 y 50 minutos luego de haberlo iniciado.
- Los alumnos a quienes se aplicó el examen impreso en hojas color marrón entregaron el examen en tiempos más heterogéneos, pues el .6826 de ellos (es decir 61 alumnos) lo entregaron entre 30 y 60 minutos luego de haberlo iniciado.
- En el grupo paja el .1587 más lento de los alumnos (es decir: 14) entregaron su examen luego de 50 minutos, mientras que en el grupo marrón la misma cantidad de alumnos lo hizo luego de 60 minutos.

TRANSFORMACIÓN DE VALORES x A VALORES z ; USO DE LA TABLA DE ÁREAS BAJO LA CURVA

En la sección anterior se ha visto que hay correspondencia entre las áreas de la curva normal y las de la serie de datos cuantitativos continuos que se esté manejando siempre y cuando se haya comprobado que esta última, al ser graficada con un polígono de frecuencias, muestra un parecido razonable con el perfil de la curva normal.

Tal correspondencia ha permitido solamente mencionar las áreas que se encuentran sobre segmentos completos de la abscisa; es decir, solamente se ha hecho mención de áreas por arriba o más allá de desviaciones estándar enteras.

Sin embargo, cómo podría responderse a la pregunta *¿cuántos alumnos de cada grupo tardaron 47 o más minutos en entregar su examen?*

En este caso se aprecia que no hay coincidencia entre el valor z de $+1$ y el valor x de 47 minutos y por ello deja de ser útil el método de comparación analógica de los gráficos que se utilizó en páginas anteriores.

La respuesta estriba en el uso de una fórmula para transformar cualquier valor x en su correspondiente valor z y en hacer uso de la tabla de áreas bajo la curva normal.

Una vez que se han calculado tanto el promedio como la desviación estándar para una serie de datos cuantitativos continuos, el valor z que, en la abscisa de la curva normal corresponde a un determinado valor x de la abscisa de los datos que se están manejando, se encuentra con la fórmula:

$$z = \frac{x - \bar{x}}{s}$$

Para responder a la pregunta recientemente planteada de *¿cuántos alumnos de cada grupo tardaron 47 o más minutos en entregar su examen?* se hacen las siguientes sustituciones:

Para el grupo al que se aplicó el examen en hojas color paja se tiene que

$$\bar{x} = 45' \text{ y } s = 5';$$

el valor z que se desea conocer es el correspondiente a un valor x de 47;

$$\text{entonces: } z = \frac{47 - 45}{5} = \frac{2}{5} = .4$$

El valor z obtenido, en este caso .4 debe localizarse en la columna A de la tabla de áreas bajo la curva ¹. Una vez localizado tal valor, se busca en la columna C cuál es el área que en la curva normal se encuentra más allá de dicho valor z ; en este caso es de .3446.

¹ Usar la tabla más detallada de áreas bajo la curva normal que se encuentra como anexo de este documento

Como el área encontrada (.3446) es una proporción del área total, entonces la misma proporción se aplica al total de alumnos del grupo para saber cuantos tardaron más de 47 minutos en entregar el examen.

Así, luego de efectuar la operación $.3446 \times 90 = 31$, puede responderse a la pregunta con el señalamiento de que hubo en este grupo 31 alumnos que tardaron 47 o más minutos en entregar su examen. Desde luego, al conocer las propiedades básicas de la curva normal, también se puede decir que hubo 59 alumnos que tradaron 47 o menos minutos en entregar su examen.

Por otra parte, para el grupo al que se aplicó el examen en hojas color marrón se tiene que

$$\bar{x} = 45' \text{ y } s = 15';$$

el valor z que se desea conocer es el correspondiente a un valor x de 47;

$$\text{entonces: } z = \frac{47 - 45}{15} = \frac{2}{15} = .13$$

El valor z obtenido, en este caso .13 debe localizarse en la columna A de la tabla de áreas bajo la curva ². Una vez localizado tal valor, se busca en la columna C cuál es el área que en la curva normal se encuentra más allá de dicho valor z ; en este caso es de .3446.

Como el área encontrada (.4483) es una proporción del área total, entonces la misma proporción se aplica al total de alumnos del grupo para saber cuantos tardaron más de 47 minutos en entregar el examen.

Así, luego de efectuar la operación $.4483 \times 90 = 40$, puede responderse a la pregunta con el señalamiento de que hubo en este grupo 40 alumnos que tardaron 47 o más minutos en entregar su examen. Desde luego, al conocer las propiedades básicas de la curva normal, también se puede decir que hubo 50 alumnos que tradaron 47 o menos minutos en entregar su examen.

² Usar la tabla más detallada de áreas bajo la curva normal que se encuentra como anexo de este documento

TABLA DE AREAS BAJO LA CURVA NORMAL
(versión condensada)

$$Z = \frac{X - \bar{X}}{S}$$

(A) Valor z	(B) Area entre el promedio y el valor z	(C) Area más allá del valor z	(A) Valor z	(B) Area entre el promedio y el valor z	(C) Area más allá del valor z
-------------------	--	--	-------------------	--	-------------------------------------

0.00	.0000	.5000
0.05	.0199	.4801
0.10	.0398	.4602
0.15	.0596	.4404
0.20	.0793	.4207
0.25	.0987	.4013
0.30	.1179	.3821
0.35	.1368	.3632
0.40	.1554	.3446
0.45	.1736	.3264
0.50	.1915	.3085
0.55	.2088	.2912
0.60	.2257	.2743
0.65	.2422	.2578
0.70	.2580	.2420
0.75	.2734	.2266
0.80	.2881	.2119
0.85	.3023	.1977
0.90	.3159	.1841
0.95	.3289	.1711
1.00	.3413	.1587
1.05	.3531	.1469
1.10	.3643	.1357
1.15	.3749	.1251
1.20	.3849	.1151
1.25	.3944	.1056
1.30	.4032	.0968
1.35	.4115	.0885
1.40	.4192	.0808
1.45	.4265	.0735
1.50	.4332	.0668
1.55	.4394	.0606
1.60	.4452	.0548
1.65	.4505	.0495
1.70	.4554	.0446

1.75	.4599	.0401
1.80	.4641	.0359
1.85	.4678	.0322
1.90	.4713	.0287
1.95	.4744	.0256
1.96	.4750	.0250
2.00	.4772	.0228
2.05	.4798	.0202
2.10	.4821	.0179
2.15	.4842	.0158
2.20	.4861	.0139
2.22	.4868	.0132
2.27	.4884	.0116
2.32	.4898	.0102
2.37	.4911	.0089
2.42	.4922	.0078
2.47	.4932	.0068
2.52	.4941	.0059
2.57	.4949	.0051
2.58	.4950	.0050
2.62	.4956	.0044
2.67	.4962	.0038
2.72	.4967	.0033
2.77	.4972	.0028
2.79	.4974	.0026
2.84	.4977	.0023
2.89	.4981	.0019
2.94	.4984	.0016
2.99	.4986	.0014
3.04	.4988	.0012
3.09	.4990	.0010
3.14	.4992	.0008
3.19	.4993	.0007
3.24	.4994	.0006
3.45	.4997	.0003

Plan de trabajo para la semana 8

Sesión 1

A través de la discusión de las soluciones obtenidas en la resolución de los problemas *Cuestión de Curvas* debe quedar aclarada la relación entre la desviación estándar y el modelo de la curva normal, ya que la comprensión de tal relación es esencial para poder efectuar correctamente la interpretación de la desviación estándar.

Adicionalmente, con la discusión de resultados habrá de verificarse la adquisición de habilidad para el empleo de tablas de áreas bajo la curva normal y su aprovechamiento en la solución de problemas que tengan que ver con la interpretación de la curva normal.

Al final de la sesión, el alumno recibirá indicaciones para el empleo de los siguientes materiales de trabajo para la segunda sesión de la semana 8:

- **Problemas: Aplicación de criterios de selección de las medidas de resumen para variables cuantitativas** **3 páginas**
 - **Texto: Criterios para la selección de las medidas de resumen con que se describen las variables cuantitativas** **2 páginas**
-

Sesión 2

Casi seguramente a estas alturas del trabajo con las medidas de resumen para variables cuantitativas en series simples ya se tiene buena capacidad de cálculo e interpretación; sin embargo aún es necesario adquirir criterios para la selección de las medidas pertinentes para cada situación específica.

Por lo anterior, en esta sesión se discutirán los resultados a los Problemas *Aplicación de criterios de selección de las medidas de resumen para variables cuantitativas*

Al final de la sesión, el alumno recibirá indicaciones para el empleo de los siguientes materiales de trabajo para la primera sesión de la semana 9:

- **Problemas: Cálculo e interpretación del sesgo y la curtosis en series simples de valores** **2 páginas**
 - **Texto: Cálculo e interpretación del sesgo y de la curtosis de datos cuantitativos continuos en series simples mediante el método de momentos** **3 páginas**
-

PROBLEMAS: **Aplicación de criterios de selección de las medidas de resumen para variables cuantitativas**

Problema 1. Un grupo de investigadores estaba interesado en identificar cuál era la cifra más frecuente en el número de hijos por familia en un barrio acomodado de la ciudad; es decir, estaban interesados en saber qué tantos hijos estaba de moda tener entre dichas familias.

También les interesaba saber qué tan grande era la diferencia entre la familia que tenía mas hijos y la que menos tenía.

Luego de una encuesta encontraron los siguientes datos:

Familia	1	2	3	4	5	6	7	8	9	10	11	12	13
Hijos	5	1	2	4	2	3	1	2	2	3	2	1	2

Familia	14	15	16	17	18	19	20	21	22	23	24	25	26
Hijos	2	5	1	2	2	4	1	6	2	2	4	3	3

Cuestiones:

- Siendo la variable *número de hijos por familia* de tipo cuantitativa, ¿es discreta o es continua?
- De acuerdo a su respuesta anterior ¿cuáles podrían ser las medidas de resumen para describir sintéticamente la información recogida en la encuesta?
- Atendiendo a su respuesta anterior, y agregando como elemento de decisión el propósito que tenían los investigadores, ¿cuál sería la medida de posición y cuál la medida de dispersión que debieron haber seleccionado?
- De acuerdo a su respuesta anterior, realice el cálculo e interpretación de la única medida de posición y de la única medida de dispersión indicadas para la descripción estadística de la información recogida por los investigadores.

Problema 2. En una etapa más avanzada de su proyecto, los investigadores del problema anterior ampliaron su encuesta y recogieron más información. Sin embargo, a estas alturas de la investigación, sus propósitos ya habían cambiado y ahora eran:

- identificar el número de hijos por familia por detrás y por delante del cual se encontraba la mitad de las familias,
- identificar el número de hijos por familia por *detrás* del cual se encontraba solo una cuarta parte de las familias,
- identificar el número de hijos por familia por *delante* del cual solo se encontraba una cuarta parte de las familias.

A continuación se muestran todos los datos que hasta el momento ya poseían los investigadores:

Familia	1	2	3	4	5	6	7	8	9	10	11	12	13
Hijos	5	1	2	4	2	3	1	2	2	3	2	1	2

Familia	14	15	16	17	18	19	20	21	22	23	24	25	26
Hijos	2	5	1	2	2	4	1	6	2	2	4	3	3

Familia	27	28	29	30	31	32	33	34	35	36	37	38	39
Hijos	7	5	2	2	3	3	4	1	1	2	3	2	2

Familia	40	41	42	43	44	45	46	47	48	49	50	51	52
Hijos	5	1	2	3	4	2	2	9	2	4	2	1	8

Cuestiones:

- Siendo la variable *número de hijos por familia* de tipo cuantitativa, ¿es discreta o es continua?
- De acuerdo a su respuesta anterior y considerando que se ignoraran los propósitos actuales del grupo de investigadores, ¿cuáles podrían ser las medidas de resumen para describir sintéticamente la información completa recogida en la encuesta ampliada?
- Atendiendo a su respuesta anterior, y agregando como elemento de decisión los nuevos propósitos que ahora tenían los investigadores, ¿cuál sería la medida de posición y cuáles las medidas de dispersión que debieron haber seleccionado?
- De acuerdo a su respuesta anterior, realice el cálculo e interpretación de la única medida de posición y de las únicas medidas de dispersión indicadas para la descripción estadística de la información recogida por los investigadores.

Problema 3. Los valores de la serie simple que a continuación se muestran son de tipo cuantitativo continuo. Represente a cada uno de ellos con un cuadrado en la cuadrícula que se muestra posteriormente, acomodándolos de manera semejante a la construcción de un histograma.

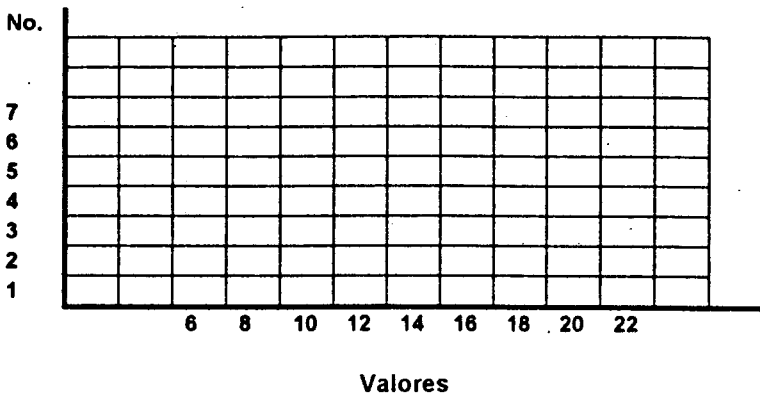
Comente si la distribución de valores parece ser semejante a la de la curva normal.

6.00	8.00	10.00	12.00	14.00	16.00	18.00	20.00	22.00
------	------	-------	-------	-------	-------	-------	-------	-------

14.00	14.00	14.00	14.00	14.00	12.00	12.00	12.00	16.00
-------	-------	-------	-------	-------	-------	-------	-------	-------

16.00	16.00	10.00	10.00	18.00	18.00	20.00	22.00	8.00
-------	-------	-------	-------	-------	-------	-------	-------	------

6.00	14.00							
------	-------	--	--	--	--	--	--	--

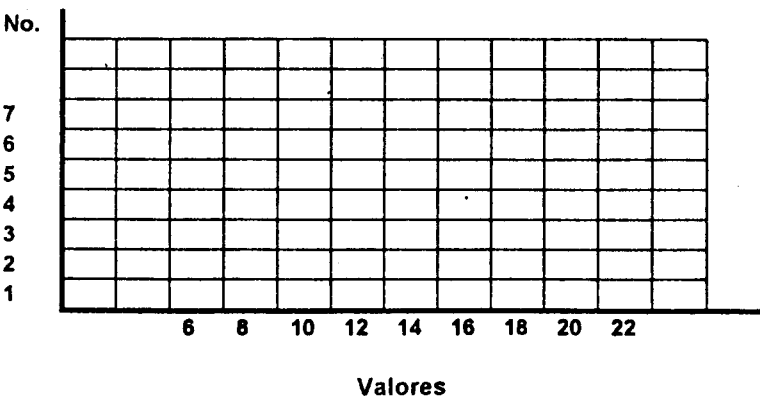


En caso de que a usted le parezca que la imagen resultante tiene un perfil parecido al de la curva normal, calcule e interprete las medidas de resumen pertinentes.

Problema 4. Los valores de la serie simple que a continuación se muestran son de tipo cuantitativo continuo. Represente a cada uno de ellos con un cuadrado en la cuadrícula que se muestra posteriormente, acomodándolos de manera semejante a la construcción de un histograma.

Comente si la distribución de valores parece ser semejante a la de la curva normal.

6.00	8.00	10.00	12.00	14.00	16.00	18.00	20.00	22.00
8.00	8.00	10.00	10.00	10.00	10.00	12.00	12.00	16.00
16.00	16.00	10.00	10.00	18.00	18.00	20.00	22.00	8.00
6.00	14.00							



En caso de que a usted le parezca que la imagen resultante no tiene un perfil parecido al de la curva normal, calcule e interprete las medidas de resumen pertinentes.

**TEXTO: Criterios para la selección de las medidas de resumen
conque se describen las variables cuantitativas:**

Los comentarios relacionados con los criterios para la selección de las medidas de resumen pertinentes para la descripción estadística de las variables de tipo cuantitativo se realizarán en función de los cuatro tipos de situación que se ilustran en el siguiente esquema:

Tipo de Variable Cuantitativa	Moda y Amplitud	Mediana y Percentiles	Promedio y Desviación Estándar
Discreta	<input checked="" type="checkbox"/> SITUACION 1	<input checked="" type="checkbox"/> SITUACION 2	
Continua con distribución NO semejante a la de la curva normal		<input checked="" type="checkbox"/> SITUACION 3	
Continua con distribución semejante a la de la curva normal			<input checked="" type="checkbox"/> SITUACION 4

Situación 1 y Situación 2: Las variables cuantitativas *discretas* (como el número de hijos, el número de convulsiones, el número de habitaciones en las viviendas) no admiten de entrada el uso del promedio y la desviación estándar porque el cálculo de estas medidas de resumen suele producir cifras con enteros y fracciones. Es obvio que carece de sentido hablar de fracciones de hijo, de fracciones de convulsión o de fracciones de habitación.

Por lo anterior solo cabe calcular, por un lado, la moda y la amplitud o, por otro lado, la mediana y algunos percentiles.

Decidirse por la moda y la amplitud depende de si el propósito de la descripción estadística es expresar el valor cuantitativo discreto que se presentó con mayor frecuencia y mencionar la diferencia entre el valor más grande y el más pequeño.

Decidirse por la mediana y diversos percentiles depende de si el propósito de la descripción estadística es identificar valores por detrás y por delante de los cuales se encuentran segmentos porcentualmente complementarios de la serie ordenada de valores.

Situación 3 y Situación 4: Aunque es factible calcular de primera intención el promedio y la desviación estándar a las variables cuantitativas *continuas*, tales medidas de resumen solo deben utilizarse cuando se ha verificado que la distribución de los valores es bastante semejante a la del modelo de la *curva normal*.

En medicina existen variables cuantitativas continuas de origen biológico, como las concentraciones de diversas sustancias en sangre, que, al ser graficadas a través de un polígono de frecuencias, muestran un perfil muy semejante al de la curva normal y que, por ello, admiten el empleo del promedio y de la desviación estándar, pues, justamente, tales medidas de resumen constituyen parámetros fundamentales de dicho modelo (1).

Pero también en medicina existen variables de origen social y económico, como el ingreso económico de las familias o el tiempo de estudios, cuyo polígono de frecuencias está sumamente sesgado y difiere radicalmente del perfil de la curva normal.

Además de convenir la elaboración y verificación del perfil del polígono de frecuencias para determinar si la distribución de una variable cuantitativa continua se parece al perfil de la curva normal, existen procedimientos numéricos más precisos para tal comprobación. Entre los principales procedimientos numéricos destaca el cálculo del sesgo y de la curtosis de la serie de valores a través del método de momentos.

¹ Considérese que cualquier cálculo de áreas bajo la curva normal implica el empleo del promedio y de la desviación estándar de manera obligada.

PROBLEMAS:**Cálculo e interpretación del sesgo y la curtosis
en series simples de valores****Problema 1**

Los valores de la serie simple que a continuación se muestran son de tipo cuantitativo continuo.

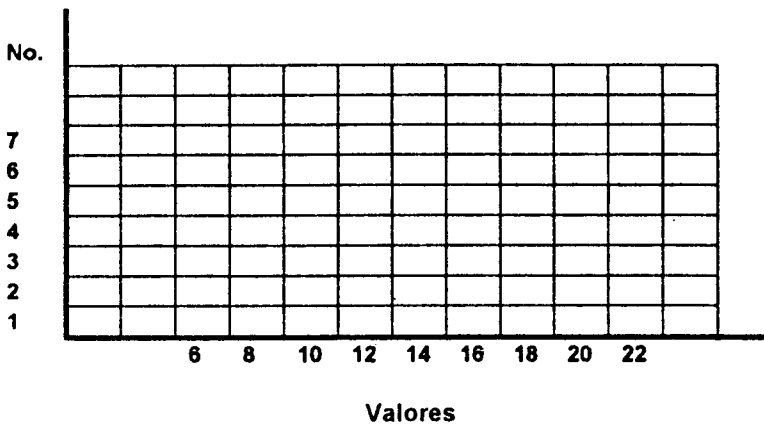
a) Represente a cada uno de ellos con un cuadrado en la cuadrícula que se muestra posteriormente, acomodándolos de manera semejante a la construcción de un histograma.

6.00	8.00	10.00	12.00	14.00	16.00	18.00	20.00	22.00
------	------	-------	-------	-------	-------	-------	-------	-------

14.00	14.00	14.00	14.00	14.00	12.00	12.00	12.00	16.00
-------	-------	-------	-------	-------	-------	-------	-------	-------

16.00	16.00	10.00	10.00	18.00	18.00	20.00	22.00	8.00
-------	-------	-------	-------	-------	-------	-------	-------	------

6.00	14.00							
------	-------	--	--	--	--	--	--	--



b) Comente si la distribución de valores parece ser semejante a la de la curva normal.

c) Efectúe el cálculo de la moda, la mediana y el promedio de la anterior serie de valores y comente sus hallazgos.

d) Efectue el cálculo del sesgo y la curtosis de la anterior serie de valores a través del método de momentos y decida si su distribución es semejante a la de la curva normal.

Problema 2

Un grupo de investigadores deseaba comparar el peso que al año de edad habían alcanzado dos diferentes grupos de niños que al nacer habían tenido pesos que fluctuaban entre 2.950 y 3.050 kgrs.

Luego de confirmar que se hallaban ante una variable cuantitativa continua, se les ocurrió que lo primero que había que hacer era resumir el peso de uno y otros grupos mediante el cálculo de sendos promedios y desviaciones estándar para, más adelante, efectuar una prueba de análisis estadístico que les permitiera decidir si había diferencias significativas entre tales promedios.

Cuando apenas habían empezado a realizar sus cálculos alguien les preguntó si era prudente usar como medida de resumen al promedio y a la desviación estándar o si sería mejor usar a la mediana y a diversos percentiles.

Los investigadores no supieron responder y su desconcierto aumentó cuando se les preguntó si, ya después de haber resumido los datos, procederían a utilizar la clásica prueba de comparación de dos promedios independientes a través de la distribución t de Student o si sería mejor usar la prueba U de Mann-Withney.

a) Con el propósito de resolver las cuestiones, calcule para ambos grupos el sesgo y la curtosis a través del método de momentos.

PESO EN KILOGRAMOS DE 20 NIÑOS
DE UN AÑO DE EDAD
GRUPO "A"

9.1	9.4	8.9	9.6	10.5	8.8	9.4	9.2	9.0	8.1
9.3	8.8	9.5	9.7	9.2	9.4	9.6	9.0	9.4	9.8

PESO EN KILOGRAMOS DE 19 NIÑOS
DE UN AÑO DE EDAD
GRUPO "B"

6.9	9.3	8.9	10.7	10.5	8.0	9.4	9.2	9.0	7.1
9.3	8.8	9.7	9.7	9.9	9.8	9.6	9.0	9.8	

b) De acuerdo a los resultados obtenidos con el cálculo del sesgo y la curtosis para uno y otro grupo, seleccione, calcule e interprete las medidas de resumen pertinentes para cada uno de los dos grupos.

TEXTO: Cálculo e interpretación del sesgo y de la curtosis de datos cuantitativos continuos en series simples mediante el método de momentos

1 Se dice que una distribución de valores cuantitativos continuos tiene semejanza a la curva normal si su sesgo, calculado a través del método de momentos, vale entre - 0.5 y + 0.5 , lo cual se simboliza de la siguiente forma :

$$- 0.5 < a_3 < + 0.5$$

y su curtosis, también calculada a través del método de momentos, vale entre 2 y 4, lo cual se simboliza de la siguiente forma :

$$2 < a_4 < 4$$

2 Las fórmulas para calcular el sesgo y la curtosis, a través del método de momentos, son las siguientes:

SESGO
$$a_3 = \frac{m_3}{\left(\sqrt{m_2} \right)^3}$$

CURTOSIS
$$a_4 = \frac{m_4}{\left(m_2 \right)^2}$$

3 El cálculo de momentos para *series simples* de datos cuantitativos continuos se hace con las fórmulas siguientes:

MOMENTOS EN SERIES SIMPLES

Momento 2

$$m_2 = \frac{\sum \left(x - \bar{x} \right)^2}{n}$$

Momento 3

$$m_3 = \frac{\sum \left(x - \bar{x} \right)^3}{n}$$

Momento 4

$$m_4 = \frac{\sum \left(x - \bar{x} \right)^4}{n}$$

4 En el caso de las series simples de valores, conviene efectuar el cálculo de los momentos a través de una tabla auxiliar de trabajo como la del siguiente ejemplo:

Ejemplo con una serie simple de valores:

PESO EN KILOGRAMOS DE UN GRUPO DE 20 NIÑOS
DE UN AÑO DE EDAD

9.1	9.4	8.9	9.6	10.5	8.8	9.4	9.2	9.0	8.1
9.3	8.8	9.5	9.7	9.2	9.4	9.6	9.0	9.4	9.8

El promedio vale: 9.285 kgrs.

Cada uno de los valores	Desviación de cada valor con respecto al promedio	Elevación al cuadrado de cada una de las desviaciones	Elevación al cubo de cada una de las desviaciones	Elevación a la cuarta de cada una de las desviaciones
X_i	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$	$(X_i - \bar{X})^3$	$(X_i - \bar{X})^4$
9.1	-0.185	0.034	-0.006	0.001
9.4	0.115	0.013	0.002	0.000
8.9	-0.385	0.148	-0.057	0.022
9.6	0.315	0.099	0.031	0.010
10.5	1.215	1.476	1.794	2.179
8.8	-0.485	0.235	-0.114	0.055
9.4	0.115	0.013	0.002	0.000
9.2	-0.085	0.007	-0.001	0.000
9	-0.285	0.081	-0.023	0.007
8.1	-1.185	1.404	-1.664	1.972
9.3	0.015	0.000	0.000	0.000
8.8	-0.485	0.235	-0.114	0.055
9.5	0.215	0.046	0.010	0.002
9.7	0.415	0.172	0.071	0.030
9.2	-0.085	0.007	-0.001	0.000
9.4	0.115	0.013	0.002	0.000
9.6	0.315	0.099	0.031	0.010
9	-0.285	0.081	-0.023	0.007
9.4	0.115	0.013	0.002	0.000
9.8	0.515	0.265	0.137	0.070
Sumas	0.000	4.446	0.077	4.421

Sustituyendo en las fórmulas para el cálculo de momentos en series simples se tiene:

$$m_2 = \frac{\sum \langle x - \bar{x} \rangle^2}{n} \quad m_2 = \frac{4.446}{20} = 0.222 \quad m_3 = \frac{\sum \langle x - \bar{x} \rangle^3}{n} \quad m_3 = \frac{0.077}{20} = 0.004$$

$$m_4 = \frac{\sum \langle x - \bar{x} \rangle^4}{n} \quad m_4 = \frac{4.421}{20} = 0.221$$

Finalmente, usando los valores calculados para los momentos y sustituyendo para las fórmulas de sesgo y curtosis en series simples, se tiene:

$$\text{SESGO } a_3 = \frac{m_3}{\left(\sqrt{m_2}\right)^3}$$

$$a_3 = \frac{0.004}{\left(\sqrt{0.222}\right)^3} = \frac{0.004}{(0.471)^3} = \frac{0.004}{0.105} = 0.038$$

$$\text{CURTOSIS } a_4 = \frac{m_4}{\left(m_2\right)^2}$$

$$a_4 = \frac{0.221}{\left(0.222\right)^2} = \frac{0.221}{0.049} = 4.484$$

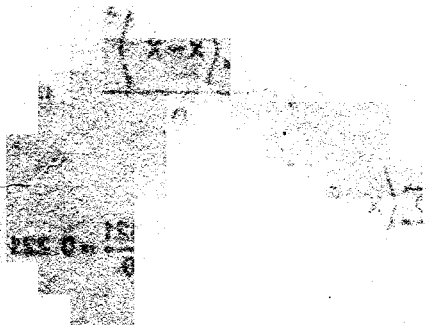
Interpretación de los resultados:

En vista de que el sesgo calculado se encuentra en el intervalo que va desde - 0.5 hasta + 0.5 puede decirse que la distribución de los pesos de los 20 niños es semejante en simetría a la de la curva normal.

Sin embargo, en vista de que la curtosis calculada está fuera del intervalo que va desde 2 hasta 4 no puede decirse que el grado de apuntamiento o aplanamiento de los pesos de los 20 niños sea semejante a la de la curva normal.

En resumen, la serie es simétrica como la curva normal pero más apuntada o elevada que ésta.

Por lo anterior, no deben utilizarse como medidas de resumen para describir a esta serie ni al promedio ni a la desviación estándar y, en su lugar, se debieran utilizar la mediana y diversos percentiles.



Plan de trabajo para la semana 9

Sesión 1

A través de la discusión de las soluciones obtenidas en la resolución de los problemas **Cálculo e interpretación del sesgo y la curtosis en series simples de valores**, en esta sesión se confirmará que los resultados de los cálculos del sesgo y de la curtosis tienen una representación gráfica específica según los valores de dichos resultados. También se verificará la adquisición de habilidad para aplicar las fórmulas para el cálculo y para la interpretación de los resultados.

Al final de la sesión, el alumno recibirá indicaciones para el empleo de los siguientes materiales de trabajo para la segunda sesión de la semana 9:

- | | | |
|---------------------------------|--|------------------------|
| •Problemas: | Elección de medidas de resumen para datos cuantitativos continuos | 1 página |
| •Instructivo y diskette: | Programa NORMAL para el cálculo de momentos, sesgo y curtosis | 1 página
1 diskette |
-

Sesión 2

Casi seguramente el alumno habrá encontrado que el cálculo del sesgo y de la curtosis brinda muy importante información acerca de la forma en que se distribuye una serie de datos cuantitativos continuos, pero habrá encontrado que los cálculos, aunque necesarios, suelen ser bastante laboriosos.

Por lo anterior, en esta sesión se discutirán las soluciones obtenidas para los Problemas *Elección de medidas de resumen para datos cuantitativos continuos*, que debieron haberse obtenido mediante la utilización del programa para computadora denominado **NORMAL**.

Al final de la sesión, el alumno recibirá indicaciones para el empleo de los siguientes materiales de trabajo para la primera sesión de la semana 10:

- | | | |
|--------------------|---|-----------|
| •Problemas: | Medidas de Resumen para Variables Cuantitativas en Series Agrupadas de Valores:
Moda y Amplitud, Mediana y Percentiles | 1 página |
| •Texto: | Medidas de Resumen para Variables Cuantitativas en Series Agrupadas de Valores:
Moda y Amplitud, Mediana y Percentiles | 5 páginas |
-

**PROBLEMAS: Elección de medidas de resumen para datos
cuantitativos continuos**

1. Aunque sabían que el ingreso económico familiar promedio puede considerarse como una variable de tipo cuantitativo continuo, un grupo de economistas interesados en la influencia del ingreso en la presencia de desnutrición, tenían serias dudas en relación con la conveniencia de resumir sus resultados utilizando el promedio y la desviación estándar.

Con el propósito de resolver la cuestión, calcule el sesgo y la curtosis de los siguientes datos recogidos por los economistas y, en base a los resultados, calcule la mediana junto con los percentiles 25 y 75 o el promedio junto con la desviación estándar:

Familia número	Ingreso promedio mensual	Familia número	Ingreso promedio mensual	Familia número	Ingreso promedio mensual
1	1 678	6	2 754	11	1 815
2	1 925	7	3 428	12	2 267
3	1 842	8	1 824	13	1 799
4	6 210	9	1 600	14	1 846
5	5 300	10	2 925	15	1 833

2. A continuación se presentan los pesos alcanzados a los 12 años de edad por dos grupos de niñas provenientes de dos distintas regiones del país. Identifique, a través del cálculo del sesgo y de la curtosis, si el peso de las niñas de una y otra región observa una distribución semejante a la de la curva normal:

**Niñas de
región rural**

27.200
31.500
28.400
28.300
26.900
30.400
34.800
35.600
29.100
28.200
27.900
39.100
28.600
28.600
34.500
28.200

**Niñas de
región urbana**

32.500
31.300
32.400
30.500
33.700
35.700
32.600
35.100
32.500
29.600
35.600
32.500
30.400
32.500
32.550
28.100
32.450
32.450
32.500
32.550

**INSTRUCTIVO:
curtosis****Programa *NORMAL* para el cálculo de momentos, sesgo y**

Introducción: Los cálculos del sesgo y de la curtosis de las series simples o agrupadas de datos cuantitativos continuos constituyen una actividad de la mayor importancia pues dependiendo de los resultados de tales cálculos se tomarán decisiones relacionadas con la descripción y el análisis estadístico de los datos.

Las decisiones pueden ser tan sencillas pero tan trascendentes como optar por escoger resumir los datos con el promedio y la desviación estándar o con la mediana y los percentiles o tan complejas, e igualmente trascendentes, como optar por pruebas de análisis estadístico de tipo paramétrico o de tipo no paramétrico.

Lamentablemente, el cálculo del sesgo y de la curtosis no es una práctica común en muchos estudios. La falta de esta importantísima verificación en los datos está influida por lo laborioso de los cálculos y por la escasez de programas computacionales de fácil acceso.

Descripción: El programa *NORMAL* es una rutina para el cálculo de los momentos, el promedio y la desviación estándar de series numéricas, simples y agrupadas, para utilizarse en computadoras compatibles con IBM y está escrito en lenguaje *BASIC*.

Entre sus ventajas se cuentan: 1) facilidad de manejo, 2) no se requiere el uso de archivos grabados con programas manejadores de bases de datos, 3) no requiere ser instalado en el disco duro de la computadora, pues puede funcionar desde el mismo disco flexible en que está almacenado, 4) puede utilizarse en computadoras con procesadores 8088, 286, 386 o 486.

Entre sus limitaciones se encuentran: 1) no se pueden grabar los datos trabajados, 2) no se pueden grabar los resultados de los cálculos (sin embargo sí se pueden imprimir los resultados, pues estos se presentan en una sola pantalla).

Utilización: 1o. Una vez encendida la computadora, instalarse en la unidad manejadora de disco que contiene al disco con el programa *NORMAL*, mediante la orden

a: [ENTER] o b: [ENTER] (1)

2o. Teclar *NORMAL* [ENTER]

3o. seguir las instrucciones que aparecen en cada pantalla del programa.

4o. Al terminar de usar el programa la computadora mantiene en su memoria al lenguaje *BASIC*. Para regresar al sistema operativo teclar:

SYSTEM [ENTER]

¹ la instrucción [ENTER] se refiere a la acción de oprimir la tecla con el símbolo ↵

**INSTRUCTIVO:
curtosis****Programa *NORMAL* para el cálculo de momentos, sesgo y**

Introducción: Los cálculos del sesgo y de la curtosis de las series simples o agrupadas de datos cuantitativos continuos constituyen una actividad de la mayor importancia pues dependiendo de los resultados de tales cálculos se tomarán decisiones relacionadas con la descripción y el análisis estadístico de los datos.

Las decisiones pueden ser tan sencillas pero tan trascendentes como optar por escoger resumir los datos con el promedio y la desviación estándar o con la mediana y los percentiles o tan complejas, e igualmente trascendentes, como optar por pruebas de análisis estadístico de tipo paramétrico o de tipo no paramétrico.

Lamentablemente, el cálculo del sesgo y de la curtosis no es una práctica común en muchos estudios. La falta de esta importantísima verificación en los datos está influida por lo laborioso de los cálculos y por la escasez de programas computacionales de fácil acceso.

Descripción: El programa *NORMAL* es una rutina para el cálculo de los momentos, el promedio y la desviación estándar de series numéricas, simples y agrupadas, para utilizarse en computadoras compatibles con IBM y está escrito en lenguaje *BASIC*.

Entre sus ventajas se cuentan: 1) facilidad de manejo, 2) no se requiere el uso de archivos grabados con programas manejadores de bases de datos, 3) no requiere ser instalado en el disco duro de la computadora, pues puede funcionar desde el mismo disco flexible en que está almacenado, 4) puede utilizarse en computadoras con procesadores 8088, 286, 386 o 486.

Entre sus limitaciones se encuentran: 1) no se pueden grabar los datos trabajados, 2) no se pueden grabar los resultados de los cálculos (sin embargo sí se pueden imprimir los resultados, pues estos se presentan en una sola pantalla).

Utilización: 1o. Una vez encendida la computadora, instalarse en la unidad manejadora de disco que contiene al disco con el programa *NORMAL*, mediante la orden

a: [ENTER] o b: [ENTER] (1)

2o. Teclar NORMAL [ENTER]

3o. seguir las instrucciones que aparecen en cada pantalla del programa.

4o. Al terminar de usar el programa la computadora mantiene en su memoria al lenguaje *BASIC*. Para regresar al sistema operativo teclar:

SYSTEM [ENTER]

1 la instrucción [ENTER] se refiere a la acción de oprimir la tecla con el símbolo ↵

PROBLEMAS:**Medidas de Resumen para Variables Cuantitativas
en Series Agrupadas de Valores:
Moda y Amplitud, Mediana y Percentiles**

1. Calcule e interprete la moda, amplitud, mediana y percentiles 10, 25, 75 y 90 para la siguiente serie agrupada de valores:

**Peso (en grs) de un grupo de recién nacidos vivos
País industrializado**

(A) PESO	(B) FRECUENCIA	(C) %	(D) FRECUENCIA ACUMULADA	(E) % ACUMULADO
2100 - 2299	31	3		
2300 - 2499	134	12		
2500 - 2699	358	33		
2700 - 2899	326	30		
2900 - 3099	143	13		
3100 - 3299	43	4		
3300 - 3499	30	3		
3500 - 3699	21	2		
3700 - 3899	11	1		
TOTAL	1,097	100		

2. Calcule e interprete la moda, amplitud, mediana y percentiles 10, 25, 75 y 90 para la siguiente serie agrupada de valores y compare los resultados con los del problema 1

**Peso (en grs) de un grupo de recién nacidos vivos
País subdesarrollado**

(A) PESO	(B) FRECUENCIA	(C) %	(D) FRECUENCIA ACUMULADA	(E) % ACUMULADO
2100 - 2299	131			
2300 - 2499	134			
2500 - 2699	258			
2700 - 2899	76			
2900 - 3099	43			
3100 - 3299	30			
3300 - 3499	12			
3500 - 3699	8			
3700 - 3899	2			
TOTAL		100		

TEXTO: **Medidas de Resumen para Variables Cuantitativas
en Series Agrupadas de Valores:
Moda y Amplitud, Mediana y Percentiles**

La siguiente serie agrupada de valores se utilizará como ejemplo para ilustrar el cálculo e interpretación de las medidas de resumen:

**NIVELES DE COLESTEROL EN SUERO
EN 1,097 VARONES DE 40 - 59 AÑOS**

(A) COLESTEROL EN SUERO (mg/100 ml.)	(B) FRECUENCIA	(C) %	(D) FRECUENCIA ACUMULADA	(E) % ACUMULADO
119.5 - 159.5	31	3	31	3
159.5 - 199.5	134	12	165	15
199.5 - 239.5	358	32	523	47
239.5 - 279.5	326	30	849	77
279.5 - 319.5	143	13	992	90
319.5 - 359.5	43	4	1035	94
359.5 - 399.5	30	3	1065	97
399.5 - 439.5	21	2	1086	99
439.5 - 479.5	11	1	1097	100
TOTAL	1,097	100		

MODA:

Definición: Es el valor que en una serie se repite con mayor frecuencia.

Procedimiento: En primer lugar identificar la clase o intervalo con mayor frecuencia (Clase Modal) y en segundo lugar utilizar la siguiente fórmula:

$$Mo = L.inf. + \left\{ \left(\frac{d_1}{d_1 + d_2} \right) W \right\}$$

Donde:

L.inf. = Límite inferior de la clase modal

d_1 = Diferencia entre la frecuencia de la clase modal y la clase anterior

d_2 = Diferencia entre la frecuencia de la clase modal y la clase posterior

W = Amplitud de la clase modal

Para el caso de la tabla de valores de colesterol:

Clase modal: 199.5 - 239.5 (porque en esta clase está la mayor frecuencia: 358 personas)

$$Mo = 199.5 + \left\{ \left(\frac{224}{224 + 32} \right) 40 \right\} \quad Mo = 199.5 + \{ (0.875) 40 \} \quad Mo = 199.5 + 35$$

$$Mo = 234.5 \text{ mgrs/100 ml}$$

Interpretación: "El valor de colesterol más frecuente en el grupo de 1,097 varones fue de 234.5 mgrs/100 ml".

AMPLITUD:

Definición: Es la diferencia entre el mayor centro de clase y el menor centro de clase de una serie agrupada. La amplitud suele simbolizarse por las siglas Am.

Procedimiento: Encontrar, por sustracción o resta, la diferencia entre el centro de clase más grande de la serie ($X' \text{ max}$) y el centro de clase más pequeño ($X' \text{ min}$).

En la serie agrupada de valores de colesterol el centro de clase más grande es 459.5 (que resulta de sumar los límites superior e inferior de la última clase y dividir entre dos); por otra parte, el centro de la primera clase vale $(119.5 + 159.5) / 2 = 139.5$

Por lo tanto, la amplitud vale $Am = X' \text{ max} - X' \text{ min} = 459.5 - 139.5 = 320$ mgrs de colesterol

Interpretación: "La diferencia entre el mayor y el menor valor de colesterol fue de 320 mgrs/100 ml."

MEDIANA (O PERCENTIL 50) :

Definición: En una serie de valores agrupados en clases o intervalos, es aquel valor que divide en dos partes de igual tamaño a toda la serie; dicho de otra manera, es el valor por detrás del cual queda un 50% de los valores y por delante del cual queda el 50% restante.

Procedimiento: En primer lugar, analizando una columna con porcentajes acumulados (como la columna E de la tabla de valores de colesterol que se muestra al principio de este documento), identificar la clase en la que se acumula el 50% de las observaciones (*identificación de la clase que contiene a la mediana*).

Posteriormente, aplicar la siguiente fórmula genérica válida para cualquier percentil (recordar que la mediana también se denomina percentil 50):

$$P_p = L.\text{inf.} + \left\{ \left(\frac{\frac{n(p)}{100} - FA}{fp} \right) W \right\}$$

Donde:

- P_p = Percentil a calcular
- $L.\text{inf}$ = Límite inferior de la clase que contiene a la mediana
- n = número total de valores de la serie
- p = percentil buscado (en este caso el 50)
- FA = frecuencia acumulada (columna D) hasta la clase anterior a la que contiene a la mediana
- fp = frecuencia simple de la clase que contiene a la mediana
- W = ancho de la clase que contiene a la mediana

Para el ejemplo de los 1,097 valores de colesterol, se tiene que la clase que contiene a la mediana es la que tiene como límite inferior a 239.5 y como límite superior a 279.5, porque en la columna de porcentajes acumulados (E) se observa que existe un 78% de los valores hasta tal clase. Por otra parte, en la clase anterior (199.5 - 239.5) apenas se había acumulado un 48% de todos los valores. Así pues, en la clase 239.5 - 279.5 necesariamente estará el valor que deja al 50% de los valores de la serie por detrás de él.

Sustituyendo en la fórmula, se tiene:

$$Md = P_{50} = 239.5 + \left\{ \left(\frac{\frac{1,097 (50)}{100} - 523}{326} \right) 40 \right\} \quad Md = P_{50} = 239.5 + \left\{ \left(\frac{548.5 - 523}{326} \right) 40 \right\}$$

$$Md = P_{50} = 239.5 + \{ (0.078) 40 \}$$

$$Md = P_{50} = 239.5 + 3.12 \quad Md = P_{50} = 242.62$$

Interpretación: "La mitad de los 1,097 varones de 40 a 59 años tuvieron valores de colesterol iguales o menores que 242.62 mgrs/100 ml y el 50 % restante tuvo valores iguales o superiores a dicha cifra".

PERCENTILES (Pp) :

Definición: En una serie agrupada de valores, es aquel valor que divide en dos partes *porcentualmente complementarias* a toda la serie. Por ejemplo: el percentil 40 divide a la serie en una parte que contiene al 40 % de los valores iguales o inferiores a él y simultáneamente, en otra parte que contiene al 60% de los valores de la serie iguales o mayores a dicho percentil.

Procedimiento: En primer lugar, analizando una columna con porcentajes acumulados (como la columna E de la tabla de valores de colesterol que se muestra al principio de este documento), identificar la clase en la que se acumula el porcentaje de las observaciones correspondientes al percentil que se desea calcular (*identificación de la clase que contiene al percentil*).

Posteriormente, aplicar la siguiente fórmula genérica válida para cualquier percentil:

$$P_p = L.\text{inf.} + \left\{ \left(\frac{\frac{n(p) - FA}{100}}{fp} \right) W \right\}$$

Donde:

P_p = Percentil a calcular

$L.\text{inf.}$ = Límite inferior de la clase que contiene al percentil

n = número total de valores de la serie

p = percentil buscado

FA = frecuencia acumulada (columna D) hasta la clase anterior a la que contiene al percentil

fp = frecuencia simple de la clase que contiene al percentil

W = ancho de la clase que contiene al percentil

Para el ejemplo de los 1,097 valores de colesterol, y suponiendo que se desea calcular el percentil 40 (P_{40}) se tiene que la clase que contiene al percentil 40 es la que tiene como límite inferior a 199.5 y como límite superior a 239.5, porque en la columna de porcentajes acumulados (E) se observa que existe un 48% de los valores hasta tal clase. Por otra parte, en la clase anterior (159.5 - 199.5) apenas se había acumulado un 15% de todos los valores. Así pues, en la clase 199.5 - 239.5 necesariamente estará el valor que deja al 40% de los valores de la serie por detrás de él.

Sustituyendo en la fórmula, se tiene:

$$P_p = 199.5 + \left\{ \left(\frac{1,097 (40) - 165}{358} \right) 40 \right\}$$

$$P_p = 199.5 + \left\{ \left(\frac{438.8 - 165}{358} \right) 40 \right\}$$

$$P_{40} = 199.5 + \{ (0.764) 40 \}$$

$$P_{40} = 199.5 + 30.56 \quad P_{40} = 230.06$$

Interpretación: "El 40% de los 1,097 varones de 40 a 59 años tuvieron valores de colesterol iguales o menores que 230.06 mgrs/100 ml y el 60 % restante tuvo valores iguales o superiores a dicha cifra".

Plan de trabajo para la semana 10

Sesión 1

Aunque ya se ha obtenido capacidad para el cálculo e interpretación de la moda y la amplitud, la mediana y los percentiles y el promedio y la desviación estándar en series simples de datos, aún se requiere poseer un buen nivel de habilidad para el empleo de las fórmulas para tales medidas en series agrupadas de datos. Por ello, a través de la discusión de las soluciones obtenidas en la resolución de los problemas *Medidas de Resumen para Variables Cuantitativas en Series Agrupadas de Valores: Moda y Amplitud, Mediana y Percentiles*, en esta sesión se confirmará que lo único que se requiere es una buena calculadora y tiempo para aplicar las fórmulas correspondientes.

Al final de la sesión, el alumno recibirá indicaciones para el empleo de los siguientes materiales de trabajo para la segunda sesión de la semana 10:

- | | | |
|--------------------|---|-----------------|
| •Problemas: | Medidas de Resumen para Variables Cuantitativas en Series Agrupadas de Valores:
Promedio y Desviación Estándar | 1 página |
|--------------------|---|-----------------|

 - | | | |
|----------------|---|------------------|
| •Texto: | Medidas de Resumen para Variables Cuantitativas en Series Agrupadas de Valores:
Promedio y Desviación Estándar | 4 páginas |
|----------------|---|------------------|
-

Sesión 2

Casi seguramente el alumno habrá comprobado que, aunque parecían formidables, las fórmulas para el cálculo de medidas de resumen en series agrupadas de datos pueden ser aplicadas fácilmente si se dispone de tablas auxiliares de trabajo para la realización de los cálculos intermedios. En esta sesión se discutirán los resultados que se hayan obtenido al resolver el problema *Medidas de Resumen para Variables Cuantitativas en Series Agrupadas de Valores: Promedio y Desviación Estándar*.

Al final de la sesión, el alumno recibirá indicaciones para el empleo de los siguientes materiales de trabajo para la primera sesión de la semana 11:

- | | | |
|--------------------|---|-----------------|
| •Problemas: | Aplicación de criterios de selección de las medidas de resumen para variables cuantitativas (series agrupadas) | 1 página |
|--------------------|---|-----------------|

 - | | | |
|----------------|---|------------------|
| •Texto: | Criterios para la selección de las medidas de resumen con que se describen las variables cuantitativas | 2 páginas |
|----------------|---|------------------|
-

PROBLEMAS:**Medidas de Resumen para Variables Cuantitativas
en Series Agrupadas de Valores:
Promedio y Desviación Estándar**

1. A continuación se muestran las edades de tres grupos de alumnos de primer ingreso a la licenciatura según preparatoria de origen. Para cada uno de ellos calcule el promedio y la desviación estándar; interprete por separado y luego compare los resultados.

Edad de los alumnos de Preparatoria Privada		Edad de los alumnos de Preparatoria UNAM		Edad de los alumnos de Bachillerato CCH	
Clases de edad	Número	Clases de edad	Número	Clases de edad	Número
18.0 - 18.4	17	17.5 - 17.9	4	17.5 - 17.9	4
18.5 - 18.9	11	18.0 - 18.4	3	18.0 - 18.4	4
Total	28	18.5 - 18.9	9	18.5 - 18.9	8
		19.0 - 19.4	5	19.0 - 19.4	12
		19.5 - 19.9	5	19.5 - 19.9	0
		20.0 - 20.4	0	20.0 - 20.4	0
		20.5 - 20.9	1	20.5 - 20.9	0
		21.0 - 21.4	0	21.0 - 21.4	2
		21.5 - 21.9	0	21.5 - 21.9	0
		22.0 - 22.4	0	22.0 - 22.4	0
		22.5 - 22.9	1	22.5 - 22.9	1
		Total	28	23.0 - 23.4	1
				Total	32

2. A continuación se muestran la ganancia de peso desde el nacimiento hasta el año de edad de dos grupos de niños. Calcule el promedio y la desviación estándar para ambos grupos.

Grupo sin orientación
nutricional continua

Ganancia de peso	Número de niños
11 a 12	7
9 a 10	5
7 a 8	12
5 a 6	34
3 a 4	49

Total 107

Grupo con orientación
nutricional continua

Ganancia de peso	Número de niños
11 a 12	3
9 a 10	15
7 a 8	53
5 a 6	30
3 a 4	2

Total 103

TEXTO:**Medidas de Resumen para Variables Cuantitativas
en Series Agrupadas de Valores:
Promedio y Desviación Estándar**

La siguiente serie agrupada de valores se utilizará como ejemplo para ilustrar el cálculo e interpretación de las medidas de resumen:

**NIVELES DE COLESTEROL EN SUERO
EN 1,097 VARONES DE 40 - 59 AÑOS**

(A) COLESTEROL EN SUERO (mg/100 ml.)	(B) FRECUENCIA	(C) %	(D) FRECUENCIA ACUMULADA	(E) % ACUMULADO
119.5 - 159.5	31	3	31	3
159.5 - 199.5	134	12	165	15
199.5 - 239.5	358	32	523	47
239.5 - 279.5	326	30	849	77
279.5 - 319.5	143	13	992	90
319.5 - 359.5	43	4	1035	94
359.5 - 399.5	30	3	1065	97
399.5 - 439.5	21	2	1086	99
439.5 - 479.5	11	1	1097	100
TOTAL	1,097	100		

PROMEDIO :

Definición: Es el valor que tendrían todos los datos de una serie numérica, ya sea en una serie simple o en una serie agrupada, si ellos fueran de igual valor.

Fórmula para series agrupadas:

$$\bar{x} = \frac{f \cdot x'}{n}$$

Procedimiento: Sumar todos productos de la multiplicación de cada frecuencia por su correspondiente centro de clase y dividir tal suma entre el número de valores que componen a la serie agrupada. Para efectuar con mayor facilidad el procedimiento es muy útil el empleo de una tabla auxiliar de trabajo como la que se ejemplifica a continuación:

Clases o Intervalos COLESTEROL EN SUERO (mg/100 ml.)	frecuencia <i>f</i>	Centros de cada clase <i>x'</i>	Productos <i>f x'</i>
119.5 - 159.5	31	139.5	4,324.5
159.5 - 199.5	134	179.5	24,053.0
199.5 - 239.5	358	219.5	78,581.0
239.5 - 279.5	326	259.5	84,597.0
279.5 - 319.5	143	299.5	42,828.5
319.5 - 359.5	43	339.5	14,598.5
359.5 - 399.5	30	379.5	11,385.0
399.5 - 439.5	21	419.5	8,809.5
439.5 - 479.5	11	459.5	5,054.5
TOTAL	1,097		274,231.5

Sustituyendo en la fórmula para el cálculo del promedio en series agrupadas:

$$\bar{x} = \frac{\sum f x}{n} \qquad \bar{x} = \frac{274,231.5}{1,097} = 249.98$$

Interpretación: "Si todos los individuos tuvieran la misma concentración de colesterol esta sería de 249.98 mgrs/100 ml".

DESVIACION ESTANDAR (en series agrupadas)

Definición: Es la raíz cuadrada de la varianza. A su vez, la varianza equivale a la división de una suma de productos entre el número de valores de la serie; es decir equivale a un promedio.

Tal promedio resulta de dividir la suma de productos de las frecuencias de cada clase multiplicadas por las desviaciones o diferencias cuadráticas de cada centro de clase con respecto al promedio de la serie agrupada entre el número de valores de la serie.

Fórmula para series agrupadas:

$$s = \sqrt{\frac{\sum f (x' - \bar{x})^2}{n}}$$

Procedimiento:

Paso 1. Obtener el promedio de la serie agrupada de valores (utilizando la fórmula para el cálculo del promedio en series agrupadas),

Paso 2. Calcular la desviación o diferencia de cada centro de clase en relación con el promedio de la serie,

Paso 3. Elevar al cuadrado cada una de las anteriores desviaciones; es decir, obtener diferencias cuadráticas,

Paso 4. Multiplicar la frecuencia de cada clase por su correspondiente diferencia cuadrática; es decir, obtener productos

Paso 5. Efectuar la suma de los anteriores productos,

Paso 6. Dividir la suma de productos entre el número de valores; es decir, obtener el promedio de desviaciones cuadráticas o varianza,

Paso 7. Obtener la raíz cuadrada de la varianza..

Con el propósito de sistematizar los cálculos, se recomienda utilizar una tabla auxiliar de trabajo como la que se muestra a continuación:

Clases o intervalos	Frecuencias	Centros de cada clase	Diferencias de cada centro en relación al promedio	Diferencias cuadráticas	Productos
	f	x'	$(x' - \bar{x})$	$(x' - \bar{x})^2$	$f(x' - \bar{x})^2$
TOTAL					

Así, para el caso de los 1,097 valores de colesterol se tienen los siguientes cálculos:

Paso 1. El promedio, calculado en una sección anterior de este documento, vale: 249.98

Clases o intervalos	Frecuencias	Centros de cada clase	Diferencias de cada centro en relación al promedio	Diferencias cuadráticas	Productos
COLESTEROL EN SUERO (mg/100 ml.)	f	x'	$(x' - \bar{x})$	$(x' - \bar{x})^2$	$f(x' - \bar{x})^2$
			(PASO 2)	(PASO 3)	(PASO 4)
119.5 - 159.5	31	139.5	-110.34	12,174.92	377,422.38
159.5 - 199.5	134	179.5	-70.34	4,947.72	662,993.89
199.5 - 239.5	358	219.5	-30.34	920.52	329,544.58
239.5 - 279.5	326	259.5	9.66	93.32	30,420.89
279.5 - 319.5	143	299.5	49.66	2,466.12	352,654.53
319.5 - 359.5	43	339.5	89.66	8,038.92	345,673.37
359.5 - 399.5	30	379.5	129.66	16,811.72	504,351.47
399.5 - 439.5	21	419.5	169.66	28,784.52	604,474.83
439.5 - 479.5	11	459.5	209.66	43,957.32	483,530.47
TOTAL	1,097				3,691,066.41
					(PASO 5)

Pasos 6 y 7

$$s = \sqrt{\frac{\sum f(x' - \bar{x})^2}{n}} \quad s = \sqrt{\frac{3,691,066.41}{1,097}} \quad s = \sqrt{3,364.69} \quad s = 58.00 \text{ mgrs./100 ml.}$$

Interpretación: La interpretación, que está condicionada a la suposición de que los valores tienen una distribución semejante a la de la *curva normal*¹, puede ser realizada en múltiples sentidos ya que se sabe que el 68.27% de los valores de una serie que se distribuye como la curva normal están agrupados alrededor del promedio si a éste se le resta una vez y también se le suma una vez el valor calculado para la desviación estándar.

Para el ejemplo, puede decirse que el 68.27% de los 1,097 varones tuvieron concentraciones de colesterol que fluctuaron desde 191.88 mgrs./100 ml (es decir: 249.88 mgrs./100 ml. menos 58.00 mgrs./100 ml.) hasta 307.88 mgrs./100 ml. (es decir: 249.88 mgrs./100 ml. más 58.00 mgrs./100 ml.).

Por otra parte, como se sabe que la curva normal tiene una área que equivale a un total de 100%, entonces también puede decirse que hubo un 31.73% de los varones que tuvieron concentraciones de colesterol menores o cuando mucho iguales a 191.88 mgrs./100 ml. o que tuvieron concentraciones iguales o mayores a 307.88 mgrs./100 ml. (ya que si a 100% se le resta 68.27% quedan 31.73%).

¹ Ver documentos relacionados con la *curva normal*

PROBLEMA:**Aplicación de criterios de selección de las medidas de resumen para variables cuantitativas (series agrupadas)**

Un grupo de pediatras sabía que el peso al nacer está influido por múltiples factores entre los cuales destacan los de tipo socioeconómico. Para comprobar lo anterior decidieron revisar los registros de nacimiento de dos grupos de niños; un grupo proveniente de un país europeo y otro proveniente de un país africano. A continuación se presentan los datos que consiguieron:

**Peso (en grs) de un grupo de recién nacidos vivos
País europeo**

(A) PESO	(B) FRECUENCIA	(C) %
2100 - 2299	13	1.1
2300 - 2499	32	2.8
2500 - 2699	134	11.6
2700 - 2899	223	19.3
2900 - 3099	358	31.0
3100 - 3299	219	18.9
3300 - 3499	136	11.8
3500 - 3699	29	2.5
3700 - 3899	12	1.0
TOTAL	1,156	100.0

**Peso (en grs) de un grupo de recién nacidos vivos
País africano**

(A) PESO	(B) FRECUENCIA	(C) %
2100 - 2299	131	18.9
2300 - 2499	134	19.3
2500 - 2699	258	37.2
2700 - 2899	76	10.9
2900 - 3099	43	6.2
3100 - 3299	30	4.3
3300 - 3499	12	1.7
3500 - 3699	8	1.2
3700 - 3899	2	0.3
TOTAL	694	100.0

Los pediatras también sabían que por ser el peso al nacer una variable de tipo cuantitativo continuo tal vez podrían resumir sus datos con el promedio y con la desviación estándar; sin embargo, actuando cautelosamente decidieron no precipitarse a hacer los cálculos y primero pensaron en revisar gráficamente el perfil de la distribución del peso al nacer en ambos grupos.

a) Elabore un polígono de frecuencias para cada grupo de niños.

b) En caso de que el perfil de alguno de los polígonos NO se parezca al de la curva normal, calcule para cada serie de datos los percentiles 10, 50 y 90 y describa las diferencias.

TEXTO: Criterios para la selección de las medidas de resumen
con que se describen las variables cuantitativas:

Los comentarios relacionados con los criterios para la selección de las medidas de resumen pertinentes para la descripción estadística de las variables de tipo cuantitativo se realizarán en función de los cuatro tipos de situación que se ilustran en el siguiente esquema:

Tipo de Variable Cuantitativa	Moda y Amplitud	Mediana y Percentiles	Promedio y Desviación Estándar
Discreta	☑ SITUACION 1	☑ SITUACION 2	
Continua con distribución NO semejante a la de la curva normal		☑ SITUACION 3	
Continua con distribución semejante a la de la curva normal			☑ SITUACION 4

Situación 1 y Situación 2: Las variables cuantitativas *discretas* (como el número de hijos, el número de convulsiones, el número de habitaciones en las viviendas) no admiten de entrada el uso del promedio y la desviación estándar porque el cálculo de estas medidas de resumen suele producir cifras con enteros y fracciones. Es obvio que carece de sentido hablar de fracciones de hijo, de fracciones de convulsión o de fracciones de habitación.

Por lo anterior solo cabe calcular, por un lado, la moda y la amplitud o, por otro lado, la mediana y algunos percentiles.

Decidirse por la moda y la amplitud depende de si el propósito de la descripción estadística es expresar el valor cuantitativo discreto que se presentó con mayor frecuencia y mencionar la diferencia entre el valor más grande y el más pequeño.

Decidirse por la mediana y diversos percentiles depende de si el propósito de la descripción estadística es identificar valores por detrás y por delante de los cuales se encuentran segmentos porcentualmente complementarios de la serie ordenada de valores.

Situación 3 y Situación 4: Aunque es factible calcular de primera intención el promedio y la desviación estándar a las variables cuantitativas *continuas*, tales medidas de resumen solo deben utilizarse cuando se ha verificado que la distribución de los valores es bastante semejante a la del modelo de la *curva normal*.

En medicina existen variables cuantitativas continuas de origen biológico, como las concentraciones de diversas sustancias en sangre, que, al ser graficadas a través de un polígono de frecuencias, muestran un perfil muy semejante al de la curva normal y que, por ello, admiten el empleo del promedio y de la desviación estándar, pues, justamente, tales medidas de resumen constituyen parámetros fundamentales de dicho modelo (1).

Pero también en medicina existen variables de origen social y económico, como el ingreso económico de las familias o el tiempo de estudios, cuyo polígono de frecuencias está sumamente sesgado y difiere radicalmente del perfil de la curva normal.

Además de convenir la elaboración y verificación del perfil del polígono de frecuencias para determinar si la distribución de una variable cuantitativa continua se parece al perfil de la curva normal, existen procedimientos numéricos más precisos para tal comprobación. Entre los principales procedimientos numéricos destaca el cálculo del sesgo y de la curtosis de la serie de valores a través del método de momentos.

¹ Considérese que cualquier cálculo de áreas bajo la curva normal implica el empleo del promedio y de la desviación estándar de manera obligada.

Plan de trabajo para la semana 11

Sesión 1

A través de la discusión de las soluciones de los problemas *Aplicación de criterios de selección de las medidas de resumen para variables cuantitativas (series agrupadas)*, en esta sesión se volverán a practicar tales criterios, ya antes utilizados en el caso de las series simples.

Al final de la sesión, el alumno recibirá indicaciones para el empleo de los siguientes materiales de trabajo para la segunda sesión de la semana 11:

- | | | |
|-------------|---|----------|
| •Problemas: | Cálculo e interpretación del sesgo y la curtosis en series agrupadas de valores | 1 página |
| •Texto: | Cálculo e interpretación del sesgo y de la curtosis de datos cuantitativos continuos en series agrupadas mediante el método de momentos | 3 |
| páginas | | |
-

Sesión 2

En esta sesión se discutirán los resultados que se hayan obtenido al resolver el problema *Calculo e interpretación del sesgo y la curtosis en series agrupadas de valores*. El alumno habrá comprobado que la única forma de concluir los cálculos sin errores es trabajando con tablas auxiliares de trabajo y que, aunque los cálculos del sesgo y la curtosis son laboriosos, la información obtenida es fundamental para decidir si las series agrupadas de datos continuos son semejantes en su distribución a la curva normal.

Al final de la sesión, el alumno recibirá indicaciones para el empleo de los siguientes materiales de trabajo para la primera sesión de la semana 12:

- | | | |
|---------------|--|----------|
| •Problema: | ¿Las distribuciones se parecen a la normal? | 1 página |
| •Instructivo: | Programa <i>NORMAL</i> para el cálculo de momentos, sesgo y curtosis | 1 página |
-

PROBLEMA:**Cálculo e interpretación del sesgo y la curtosis
en series agrupadas de valores**

Un grupo de pediatras sabía que el peso al nacer está influido por múltiples factores entre los cuales destacan los de tipo socioeconómico. Para comprobar lo anterior decidieron revisar los registros de nacimiento de dos grupos de niños; un grupo proveniente de un país europeo y otro proveniente de un país africano. A continuación se presentan los datos que consiguieron:

Peso (en grs) de un grupo de recién nacidos vivos
Pais europeo

(A) PESO	(B) FRECUENCIA	(C) %
2100 - 2299	13	1.1
2300 - 2499	32	2.8
2500 - 2699	134	11.6
2700 - 2899	223	19.3
2900 - 3099	358	31.0
3100 - 3299	219	18.9
3300 - 3499	136	11.8
3500 - 3699	29	2.5
3700 - 3899	12	1.0
TOTAL	1,156	100.0

Peso (en grs) de un grupo de recién nacidos vivos
Pais africano

(A) PESO	(B) FRECUENCIA	(C) %
2100 - 2299	131	18.9
2300 - 2499	134	19.3
2500 - 2699	258	37.2
2700 - 2899	76	10.9
2900 - 3099	43	6.2
3100 - 3299	30	4.3
3300 - 3499	12	1.7
3500 - 3699	8	1.2
3700 - 3899	2	0.3
TOTAL	694	100.0

a) Con el propósito de verificar si podrían usar el promedio y la desviación estándar o deberían utilizar a la mediana y a algunos percentiles para resumir sus datos cuantitativos continuos, calcule el sesgo y la curtosis con el método de momentos para los pesos de los niños de ambos países.

TEXTO:**Cálculo e interpretación del sesgo y de la curtosis de datos cuantitativos continuos en series agrupadas mediante el método de momentos**

1 Se dice que una distribución de valores cuantitativos continuos tiene semejanza a la curva normal si su sesgo, calculado a través del método de momentos, vale entre -0.5 y $+0.5$, lo cual se simboliza de la siguiente forma :

$$-0.5 < a_3 < +0.5$$

y su curtosis, también calculada a través del método de momentos, vale entre 2 y 4 , lo cual se simboliza de la siguiente forma :

$$2 < a_4 < 4$$

2 Las fórmulas para calcular el sesgo y la curtosis, a través del método de momentos, son las siguientes:

$$\text{SESGO} \quad a_3 = \frac{m_3}{(\sqrt{m_2})^3}$$

$$\text{CURTOSIS} \quad a_4 = \frac{m_4}{(m_2)^2}$$

3 El cálculo de momentos para *series agrupadas* de datos cuantitativos continuos se hace con las fórmulas siguientes:

MOMENTOS EN SERIES AGRUPADAS

$$\text{Momento 2} \quad m_2 = \frac{\sum f (x' - \bar{x})^2}{n}$$

$$\text{Momento 3} \quad m_3 = \frac{\sum f (x' - \bar{x})^3}{n}$$

$$\text{Momento 4} \quad m_4 = \frac{\sum f (x' - \bar{x})^4}{n}$$

En el caso de las series agrupadas de valores, conviene efectuar el cálculo de los momentos y del sesgo y la curtosis a través de una tabla auxiliar de trabajo como la del siguiente ejemplo:

Ejemplo con una serie agrupada de valores:

NIVELES DE COLESTEROL EN SUERO
EN 1,097 VARONES DE 40 - 59 AÑOS
(mgrs/100 ml)

Clases		frecuen- cias	Centros de clase	$x' - \bar{x}$	$(x' - \bar{x})^2$	$(x' - \bar{x})^3$	$(x' - \bar{x})^4$	$f(x' - \bar{x})^2$	$f(x' - \bar{x})^3$	$f(x' - \bar{x})^4$
Límite inferior	Límite superior									
119.50	159.50	31	139.50	110.48	12,205.83	1,348,500.14	148,982,295.75	378,380.74	-41,803,504.42	4,618,451,168.36
159.50	199.50	134	179.50	-70.48	4,967.43	-350,104.49	24,675,364.78	665,635.67	-46,914,002.28	3,306,498,880.37
199.50	239.50	358	219.50	-30.48	929.03	-28,316.85	863,097.48	332,592.88	-10,137,431.08	308,988,899.32
239.50	279.50	326	259.50	9.52	90.63	862.80	8,213.87	29,545.51	281,273.26	2,677,721.43
279.50	319.50	143	299.50	49.52	2,452.23	121,434.45	6,013,433.93	350,668.95	17,365,126.27	859,921,052.66
319.50	359.50	43	339.50	89.52	8,013.83	717,398.10	64,221,477.68	344,594.71	30,848,118.19	2,761,523,540.24
359.50	399.50	30	379.50	129.52	16,775.43	2,172,753.75	281,415,065.11	503,262.91	65,182,612.36	8,442,451,953.16
399.50	439.50	21	419.50	169.52	28,737.03	4,871,501.39	825,816,916.21	603,477.64	102,301,529.26	17,342,155,240.42
439.50	479.50	11	459.50	209.52	43,898.63	9,197,641.04	1,927,089,751.00	482,884.93	101,174,051.46	21,197,987,260.95
Sumas		1097						3,691,043.95	218,297,773.02	58,840,655,716.90

Sustituyendo en las fórmulas para el cálculo de momentos en series agrupadas se tiene:

$$m_2 = \frac{\sum f(x' - \bar{x})^2}{n} = \frac{3,691,043.95}{1,097} = 3,364.67$$

$$m_3 = \frac{\sum f(x' - \bar{x})^3}{n} = \frac{218,297,773.02}{1,097} = 198,995.24$$

$$m_4 = \frac{\sum f(x' - \bar{x})^4}{n}$$

$$m_4 = \frac{58,840,655,716.90}{1,097} = 53,637,790.08$$

Finalmente, usando los valores calculados para los momentos y sustituyendo para las fórmulas de sesgo y curtosis en series agrupadas, se tiene:

$$\text{SESGO } a_3 = \frac{m_3}{(\sqrt{m_2})^3}$$

$$a_3 = \frac{198,995.24}{(\sqrt{3,364.67})^3} = \frac{198,995.24}{(58.01)^3} = \frac{198,995.24}{195,170.29} = 1.02$$

$$\text{CURTOSIS } a_4 = \frac{m_4}{(m_2)^2}$$

$$a_4 = \frac{53,637,790.08}{(3,364.67)^2} = \frac{53,637,790.08}{11,321,004} = 4.74$$

Interpretación de los resultados:

En vista de que el sesgo calculado no se encuentra en el intervalo que va desde - 0.5 hasta + 0.5 puede decirse que la distribución de los 1,097 valores de colesterol no es semejante en simetría a la de la curva normal.

Igualmente, en vista de que la curtosis calculada está fuera del intervalo que va desde 2 hasta 4 no puede decirse que el grado de apuntamiento o aplanamiento de los 1,097 valores de colesterol sea semejante a la de la curva normal.

En resumen, la serie no es simétrica como la curva normal ya que tiene una pendiente prolongada hacia la derecha y es más apuntada o elevada que ella ¹.

Por lo anterior, aunque los datos son de tipo cuantitativo continuo, no deben utilizarse como medidas de resumen para describir a esta serie agrupada ni al promedio ni a la desviación estándar y, en su lugar, se debieran utilizar la mediana y diversos percentiles.

¹ El lector debiera efectuar una comprobación visual de estos resultados elaborando el polígono de frecuencias correspondiente a los datos.

PROBLEMA:**¿Las distribuciones se parecen a la normal?**

A continuación se presenta el peso de tres grupos de recién nacidos.

Para cada uno de ellos elabore un polígono de frecuencias y calcule el sesgo y la curtosis (preferentemente a través del programa *Normal* para computadora)

Comente los resultados.

Peso (en grs) de un grupo de recién nacidos vivos
Pais europeo

(A) PESO	(B) FRECUENCIA	(C) %
2100 - 2499	14	1.2
2500 - 2899	65	5.6
2900 - 3299	577	49.4
3300 - 3699	367	31.4
3700 - 4099	145	12.4
TOTAL	1,168	100.0

Peso (en grs) de un grupo de recién nacidos vivos
Pais africano

(A) PESO	(B) FRECUENCIA	(C) %
2100 - 2499	265	38.2
2500 - 2899	334	48.1
2900 - 3299	73	10.5
3300 - 3699	20	2.9
3700 - 4099	2	0.3
TOTAL	694	100.0

Peso (en grs) de un grupo de recién nacidos vivos
Pais latinoamericano

(A) PESO	(B) FRECUENCIA	(C) %
2100 - 2499	83	6.9
2500 - 2899	293	24.3
2900 - 3299	462	38.4
3300 - 3699	287	23.8
3700 - 4099	79	6.6
TOTAL	1,204	100.0

INSTRUCTIVO: Programa *NORMAL* para el cálculo de momentos, sesgo y curtosis

Introducción: Los cálculos del sesgo y de la curtosis de las series simples o agrupadas de datos cuantitativos continuos constituyen una actividad de la mayor importancia pues dependiendo de los resultados de tales cálculos se tomarán decisiones relacionadas con la descripción y el análisis estadístico de los datos.

Las decisiones pueden ser tan sencillas pero tan trascendentes como optar por escoger resumir los datos con el promedio y la desviación estándar o con la mediana y los percentiles o tan complejas, e igualmente trascendentes, como optar por pruebas de análisis estadístico de tipo paramétrico o de tipo no paramétrico.

Lamentablemente, el cálculo del sesgo y de la curtosis no es una práctica común en muchos estudios. La falta de esta importantísima verificación en los datos está influida por lo laborioso de los cálculos y por la escasez de programas computacionales de fácil acceso.

Descripción: El programa *NORMAL* es una rutina para el cálculo de los momentos, el promedio y la desviación estándar de series numéricas, simples y agrupadas, para utilizarse en computadoras compatibles con IBM y está escrito en lenguaje *BASIC*.

Entre sus ventajas se cuentan: 1) facilidad de manejo, 2) no se requiere el uso de archivos grabados con programas manejadores de bases de datos, 3) no requiere ser instalado en el disco duro de la computadora, pues puede funcionar desde el mismo disco flexible en que está almacenado, 4) puede utilizarse en computadoras con procesadores 8088, 286, 386 o 486.

Entre sus limitaciones se encuentran: 1) no se pueden grabar los datos trabajados, 2) no se pueden grabar los resultados de los cálculos (sin embargo sí se pueden imprimir los resultados, pues estos se presentan en una sola pantalla).

Utilización: 1o. Una vez encendida la computadora, instalarse en la unidad manejadora de disco que contiene el disco con el programa *NORMAL*, mediante la orden

a: [ENTER] o b: [ENTER] (1)

2o. Teclar NORMAL [ENTER]

3o. seguir las instrucciones que aparecen en cada pantalla del programa.

4o. Al terminar de usar el programa la computadora mantiene en su memoria al lenguaje *BASIC*. Para regresar al sistema operativo teclar:

SYSTEM [ENTER]

¹ la instrucción [ENTER] se refiere a la acción de oprimir la tecla con el símbolo ↵

Plan de trabajo para la semana 12**Sesión 1**

En esta sesión se revisarán las soluciones al problema *¿Las distribuciones se parecen a la normal?*. El alumno habrá apreciado la facilidad para realizar cálculos en series agrupadas de datos si se utiliza el programa de cómputo **NORMAL**.

Al final de la sesión, el alumno recibirá indicaciones para el empleo de los siguientes materiales de trabajo para la segunda sesión de la semana 12:

- | | | |
|-------------|--|-----------|
| •Problemas: | Identificación de medidas de resumen en artículos de investigación | 2 páginas |
|-------------|--|-----------|
-

Sesión 2

En esta sesión se discutirán los resultados del problema práctico *Identificación de medidas de resumen en artículos de investigación*.

Al final de la sesión, el alumno recibirá indicaciones para el empleo de los siguientes materiales de trabajo para la primera sesión de la semana 13:

- | | | |
|------------|--|-----------|
| •Problema: | ¿Qué tan intensamente se asocian estas dos variables cualitativas? | 3 páginas |
| •Texto: | Medidas de resumen para cuantificar la asociación entre dos variables cualitativas | 6 páginas |
-

PROBLEMAS:**Identificación de medidas de resumen
en artículos de investigación**

1. A partir de artículos publicados en las revistas *Salud Pública de México* y *American Journal Of Epidemiology*, identifique párrafos en que se mencionen, implícita o explícitamente, las siguientes medidas de resumen y llene un formato como el que se muestra:

**Medida de
Resumen****Párrafo en que se menciona a la
medida de resumen****Interpretación**

Ejemplo: Proporción	Ejemplo: <i>It was found that of 63 patients provisionally diagnosed as suffering from salpingitis, 53 were in fact suffering from that condition and of 937 patients admitted with provisional diagnosis of appendicitis, ectopic pregnancy, ovarian tumour, etc. 9 were in fact suffering from salpingitis.</i>	Ejemplo: La proporción de casos de salpingitis entre quienes fueron provisionalmente diagnosticadas con tal padecimiento fue de $53 / 63 = 0.841$ mientras que la proporción de casos de salpingitis entre quienes fueron provisionalmente diagnosticadas con otro padecimiento fue de $9 / 937 = 0.009$ Por lo tanto, la importancia del primer subconjunto en relación con el conjunto al que pertenece es mayor que en el segundo caso.
Ejemplo: Desviación estándar	Ejemplo: <i>Se encontró que la desviación estándar de la duración de las consultas a pacientes diabéticos fue de 12.5 minutos</i>	Ejemplo: El 68.28% de las consultas a pacientes diabéticos duraron entre 12.5 minutos mas hasta 12.5 minutos menos alrededor del promedio.

Proporción (Proportion)		
Razón (Ratio)		
Tasa (Rate)		

Moda (Mode)		
Amplitud (Range)		
Mediana (Median)		
Percentiles (Percentiles)		
Promedio (Mean)		
Desviación Estándar (Standard Deviation)		

PROBLEMAS:

¿ Qué tan intensamente se asocian estas dos variables cualitativas ?

En las siguientes páginas se encuentran anotados los datos de 120 personas que asistieron a una comida. En la comida se ofrecieron los siguientes alimentos: sopa, arroz y mole. Algunos de los invitados presentaron diarrea.

1. Con algún procedimiento de recuento tal como el de tarjetas simples, efectue el cómputo de los datos y construya, para cada alimento, un cuadro como el siguiente:

Casos de diarrea según ingestión de _____ .

Ingestión de _____	Presentación de diarrea		Totales
	SI	NO	
SI			
NO			
Totales			120

2. Una vez elaborados los cuadros, describa la intensidad de la asociación de la ingestión de cada alimento con la presentación de diarrea a través del cálculo de chi cuadrada.

Núm. de caso	Comió SOPA	Comió ARROZ	Comió MOLE	Tuvo DIARREA	Núm. de caso	Comió SOPA	Comió ARROZ	Comió MOLE	Tuvo DIARREA
1	S	N	N	S	31	N	S	N	N
2	N	S	N	N	32	S	N	S	S
3	S	N	N	N	33	S	N	S	S
4	S	N	S	S	34	S	N	S	S
5	S	N	N	S	35	S	S	N	S
6	N	S	S	N	36	S	S	N	S
7	S	S	S	S	37	N	N	N	N
8	S	S	S	S	38	S	N	N	N
9	S	N	S	N	39	S	N	S	N
10	N	S	S	S	40	S	S	N	S
11	S	S	N	S	41	N	N	S	N
12	S	S	N	S	42	N	S	N	N
13	N	S	N	N	43	S	S	N	S
14	S	S	N	S	44	N	N	N	N
15	S	N	S	S	45	N	N	N	N
16	N	N	S	N	46	S	N	S	N
17	S	S	N	S	47	N	S	N	N
18	S	S	N	S	48	S	N	S	S
19	S	N	S	S	49	S	S	N	N
20	S	N	S	S	50	N	S	N	N
21	S	N	N	S	51	N	N	N	N
22	S	N	N	S	52	N	S	N	N
23	S	N	S	S	53	N	N	S	N
24	S	N	S	S	54	S	S	N	S
25	S	S	S	S	55	N	N	S	N
26	N	N	N	S	56	N	N	N	N
27	S	N	S	S	57	S	S	S	N
28	N	S	N	N	58	S	N	S	S
29	S	S	N	S	59	N	N	N	N
30	S	S	S	S	60	S	S	S	N

Núm. de caso	Comió SOPA	Comió ARROZ	Comió MOLE	Tuvo DIARREA	Núm. de caso	Comió SOPA	Comió ARROZ	Comió MOLE	Tuvo DIARREA
61	S	N	N	S	91	N	S	N	N
62	N	S	N	S	92	S	N	S	S
63	S	N	N	N	93	S	N	S	S
64	S	N	S	S	94	S	N	S	S
65	S	N	N	S	95	S	S	N	S
66	N	S	S	N	96	S	S	N	S
67	S	S	N	S	97	S	N	N	N
68	S	S	S	S	98	S	N	N	N
69	S	N	S	N	99	S	N	S	N
70	N	S	S	N	100	S	S	N	S
71	S	S	N	S	101	N	N	S	N
72	S	S	N	S	102	N	S	S	N
73	N	S	N	N	103	S	S	N	S
74	S	N	N	S	104	N	N	S	N
75	S	N	S	S	105	N	N	N	N
76	N	N	S	N	106	S	S	S	N
77	S	S	N	S	107	N	S	N	S
78	S	S	N	S	108	S	N	S	S
79	S	N	S	S	109	S	S	S	S
80	S	N	S	S	110	S	S	N	N
81	S	N	N	S	111	N	N	N	N
82	S	S	N	S	112	S	N	S	S
83	S	N	S	S	113	N	N	S	N
84	S	N	S	S	114	S	S	N	S
85	N	S	S	N	115	N	N	S	N
86	N	N	N	S	116	S	N	S	S
87	S	N	S	S	117	S	S	S	N
88	N	S	N	N	118	S	N	S	S
89	S	S	N	S	119	N	N	N	N
90	S	S	S	S	120	S	S	S	N

TEXTO: Medidas de resumen para cuantificar la asociación entre dos variables cualitativas:

χ^2 (chi cuadrada) en tablas de 2 x 2

Una manera práctica de resumir la magnitud de la asociación entre dos variables cualitativas consiste en el cálculo de chi cuadrada. El procedimiento de cálculo consiste en los siguientes tres pasos:

1. Acomodación de los datos en una tabla de contingencia de dos renglones por dos columnas.
2. Cálculo de frecuencias esperadas.
3. Cálculo e interpretación de chi cuadrada (χ^2)

1. Acomodación de los datos en una tabla de contingencia de dos renglones por dos columnas.

Supóngase que en un grupo de 100 niños fueron medidas simultáneamente dos variables cualitativas de tipo nominal: *exposición crónica al plomo* y *desempeño escolar* y que cada variable tuvo la siguiente escala:

Exposición crónica al plomo
Si
No

Desempeño escolar deficiente
Si
No

A continuación se muestra un listado parcial de los datos:

Niño No.	Exposición crónica al plomo	Desempeño escolar deficiente
1	Si	Si
2	No	Si
3	Si	No
.	.	.
.	.	.
99	Si	Si
100	No	No

A través de un procedimiento de recuento de información tal como el de palotes o el de tarjetas simples se cuentan las frecuencias de casos para acomodarlos en una tabla como la siguiente:

Casos de desempeño escolar deficiente según exposición crónica al plomo

Exposición crónica	Desempeño escolar deficiente		
	Si	No	
SI	42	8	50
NO	8	42	50
	50	50	100

La tabla, por contar con dos renglones y dos columnas para los datos, tiene cuatro celdillas en las que se anotan las frecuencias de casos. Ello permite apreciar la distribución que tuvieron tales casos.

A la distribución real de los casos se denomina *distribución observada* y a cada una de las frecuencias contenidas en las celdillas se denomina *frecuencia observada*. A cada frecuencia observada se le simboliza con f_o

En el supuesto de que no existiera ninguna asociación entre la exposición crónica al plomo y el desempeño escolar deficiente, las frecuencias que se esperarían encontrar en cada una de las celdillas serían las siguientes:

Exposición crónica	Desempeño deficiente		
	SI	NO	
SI	25	25	50
NO	25	25	50
	50	50	100

A la distribución de los casos, tal como se encontrarían hipotéticamente si no existiera ninguna asociación entre las dos variables se denomina *distribución esperada* y a cada una de las frecuencias contenidas en las celdillas se denomina *frecuencia esperada*. A cada frecuencia esperada se le simboliza con f_e .

2. Cálculo de frecuencias esperadas (f_e)

Una vez que se dispone en cada celdilla de los datos realmente observados (es decir de las frecuencias observadas f_o), se debe efectuar el cálculo de las correspondientes frecuencias esperadas f_e bajo la hipótesis de no asociación. El cálculo de la frecuencia esperada para cada una de las celdillas utiliza la siguiente fórmula:

$$f_e = \frac{(t_{mr})(t_{mc})}{tt}$$

donde: f_e = frecuencia esperada para una celdilla determinada
 t_{mr} = total marginal del renglón de dicha celdilla
 t_{mc} = total marginal de la columna de la misma celdilla
 tt = total de casos de toda la tabla

obsérvese cómo se efectúa el cálculo de la frecuencia esperada f_e para la celdilla superior izquierda que, de momento, solo contiene a la frecuencia observada f_o :

Casos de desempeño escolar deficiente según exposición crónica al plomo

Exposición crónica	Desempeño deficiente		
	SI	NO	
SI	42	8	50
NO	8	42	50
	50	50	100

t_{mr} de la celdilla = 50;

t_{mc} de la celdilla = 50

tt o gran total = 100

$$f_e \text{ para la celdilla} = \frac{(t_{mr})(t_{mc})}{tt} = \frac{(50)(50)}{100} = \frac{(2\ 500)}{100} = 25$$

De la misma manera se procede al cálculo de las frecuencias esperadas para cada una de las demás celdillas. Las frecuencias esperadas se anotan entre paréntesis junto a la correspondiente frecuencia observada de la siguiente manera:

**Casos de desempeño escolar deficiente
según exposición crónica al plomo**

Exposición crónica	Desempeño deficiente		
	SI	NO	
SI	42 (25)	8 (25)	50
NO	8 (25)	42 (25)	50
	50	50	100

3. Cálculo e interpretación de chi cuadrada (χ^2)

Una vez que ya se dispone de la tabla que contiene tanto a las frecuencias observadas como a las frecuencias esperadas se utiliza la fórmula para el cálculo del valor de chi cuadrada:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

para el caso que se está ejemplificando, los cálculos son como sigue:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(42 - 25)^2}{25} + \frac{(8 - 25)^2}{25} + \frac{(8 - 25)^2}{25} + \frac{(42 - 25)^2}{25} = 46.24$$

4. INTERPRETACION

El resultado 46.24 expresa qué tan diferente es la distribución de frecuencias observadas con respecto a la distribución de frecuencias esperadas.

Debe recordarse que la distribución de frecuencias esperadas representa a las frecuencias esperadas f_e que debieran encontrarse si no existiera ninguna asociación entre las dos variables; es decir, entre la exposición crónica al plomo y el desempeño escolar deficiente.

Por lo anterior, si las frecuencias observadas f_o hubieran sido iguales a las frecuencias esperadas f_e la diferencia entre ellas hubiera sido de cero. En consecuencia, el hecho de que no existiera diferencia alguna entre las frecuencias observadas y las frecuencias esperadas habría

provocado que en cada paso de los cálculos de la fórmula de chi cuadrada se hubieran obtenido valores de cero y que la suma final hubiera resultado cero.

Por lo anterior, cuando se encuentran resultados de cero o muy cercanos a esta cifra, se dice que no existe asociación entre las dos variables que se están estudiando. Por lo contrario, cuando se encuentran resultados muy lejanos al cero, se dice que existe asociación intensa entre una y otra variable.

Para el ejemplo que se ha desarrollado puede decirse que existe intensa asociación entre la exposición crónica al plomo y el desempeño escolar deficiente porque el valor calculado de chi cuadrada está muy alejado de cero.

En ocasiones, los resultados obtenidos son escasamente diferentes a cero y por no ser tan grandes causan la impresión de ser debidos solamente al azar. Obsérvese el siguiente caso:

Frecuencias Observadas

29	21
21	29

Frecuencias Esperadas

25	25
25	25



Ambas Frecuencias

29 (25)	21 (25)
21 (25)	29 (25)

Cálculo de chi cuadrada:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(29 - 25)^2}{25} + \frac{(21 - 25)^2}{25} + \frac{(21 - 25)^2}{25} + \frac{(29 - 25)^2}{25} = 2.56$$

En sentido estricto, el resultado no es cero y por ello tendría que decirse que, al no parecerse ambas distribuciones de frecuencias, existe asociación entre ambas variables. Sin embargo el resultado no parece ser tan grande y surge la duda de si conviene o no declarar que sí hay asociación.

Para resolver la anterior cuestión debe acudir al uso de una tabla de valores críticos de chi cuadrada que marca los valores límite para declarar si existe asociación significativa entre una y otra variable. A continuación se muestra un extracto de tal tabla:

Grados de libertad (g.l.)	Nivel de confianza .05	Nivel de confianza .01
1	3.84	6.64
2	5.99	9.21
3	7.82	11.35
4	9.49	13.28

La primera columna a la extrema izquierda de la tabla, con el encabezado *Grados de libertad*, se refiere a un valor que resulta de combinar el número de columnas con el número de renglones que tenga la tabla de contingencia conque se haya trabajado bajo la siguiente fórmula general:

$$\text{g.l.} = (\text{columnas menos } 1) (\text{renglones menos } 1).$$

En el caso de la tabla de contingencia que se está usando, la fórmula lleva a los siguientes cálculos : $\text{g.l.} = (2 - 1) (2 - 1) = (1) (1) = 1$; es decir que la tabla de contingencia tiene 1 grado de libertad.

Las siguientes dos columnas de la tabla de valores críticos de chi cuadrada se refieren a dos de los diversos niveles de confianza empleados para decidir si la asociación encontrada entre las dos variables puede estar dada o no por azar o por efecto de una verdadera asociación entre ambas variables.

Si se ha escogido el nivel de confianza de .05 y considerando que la tabla que contiene los datos tiene un grado de libertad, entonces se encuentra un *valor crítico* de 3.84; dicho valor no es excedido por el *valor calculado* para chi cuadrada y por ello se concluye que *con un riesgo de error de 5% puede afirmarse que no parece existir asociación entre las dos variables, ya que el valor calculado para chi cuadrada no excedió al valor crítico en el nivel de confianza de .05*

Módulo Análisis Estadístico para alumnos de Salud Pública II

ESTADISTICA 1996-1997

TEMAS	CONTENIDOS
Introducción 1	Bases para la selección de los procedimientos de análisis estadístico: El diseño de investigación. Las variables y su nivel de medición. Sesgo y curtosis y su relación con la curva normal; cálculo a través del método de momentos e interpretación en series simples y agrupadas. Un programa sencillo en microcomputadora para el cálculo de sesgo y curtosis en series simples o agrupadas.
2	Bondad de ajuste de la distribución de una variable medida en un solo grupo. Análisis del ajuste de una distribución observada con una distribución esperada mediante chi cuadrada
3	Asociación de dos variables cualitativas medidas en un solo grupo Análisis de tablas de contingencia de 2 x 2 y de n x n mediante chi cuadrada.
4	Asociación de dos variables cuantitativas medidas en un solo grupo Análisis de correlación lineal simple por medio de la prueba de hipótesis sobre el coeficiente de Pearson
5	Asociación de dos variables cuantitativas medidas en un solo grupo Análisis de correlación lineal simple por medio de la prueba de Spearman
6	Comparación de dos grupos Comparación de proporciones
7	Comparación de dos grupos Prueba de la probabilidad exacta de Fisher
8	Comparación de dos grupos Comparación de los promedios de dos grupos independientes
9	Comparación de dos grupos Prueba U de Mann-Whitney
10	Comparación de dos grupos Comparación de los promedios de dos grupos apareados
11	Comparación de dos grupos Prueba de Wilcoxon
12	Panorama e introducción al Muestreo

Opcional:

12	Comparación de tres o más grupos Análisis de varianza totalmente aleatorio
13	Comparación de tres o más grupos Análisis de varianza de Kruskal-Wallis
14	Comparación de tres o más grupos Análisis de varianza en diseño estratificado
15	Comparación de tres o más grupos Análisis de varianza en bloques
16	Comparación de tres o más grupos Análisis de varianza de Friedman

Problemas: **Evaluación de la bondad de ajuste de la distribución de una variable medida en un grupo**

(Diseño de un grupo y una variable)

Problema 1.

Un antropólogo sabía que la población de su país tenía una distribución de grupos sanguíneos de la siguiente forma:

Tipo A	Tipo B	Tipo AB	Tipo O
15%	7%	8%	70 %

Al estudiar a una aislada población de la costa descubrió que los rasgos étnicos de los pobladores eran semejantes a los de los pobladores de otro continente. En una muestra de 174 miembros de dicho lugar encontró la siguiente distribución de grupos sanguíneos:

Tipo A	Tipo B	Tipo AB	Tipo O
70 personas	22 personas	24 personas	58 personas

¿Pudieran provenir los habitantes de la citada población del otro continente?

Problema 2.

El administrador de un hospital suponía que los ingresos al servicio de cirugía provenían *equitativamente* de tres fuentes: a) consulta externa, b) urgencias o c) traslados desde otros hospitales.

Al estudiar el origen de los internamientos en cirugía en los últimos tres años encontró lo siguiente:

Origen de los ingresos al servicio de cirugía
y cantidad de los mismos

Consulta externa	Urgencias	Traslados
735	684	795

¿Qué tan razonable era la suposición del administrador del hospital?

Texto: **Evaluación de la bondad de ajuste de la distribución de una variable medida en un grupo**

(Diseño de un grupo y una variable)

Relato:

Un grupo de cirujanos observó que los resultados de la aplicación de una nueva técnica quirúrgica a 255 pacientes fueron los siguientes:

invalidez total:	invalidez parcial:	funcionamiento normal:	mejoría funcional:
31 casos	45 casos	73 casos	106 casos

Con la técnica quirúrgica antigua los resultados en una muy amplia serie de pacientes operados habían sido los siguientes:

invalidez total:	invalidez parcial:	funcionamiento normal:	mejoría funcional:
25.10%	30.87%	29.22%	14.81%

¿Los resultados observados con la nueva técnica quirúrgica aplicada a los 255 pacientes difieren o son semejantes a los que se esperarían con la técnica antigua?

Secuencia de resolución:

1. Problema:

Se ignora si el resultado de la aplicación de la nueva técnica quirúrgica produce resultados semejantes a los de la técnica quirúrgica antigua.

2. Planteamiento de la hipótesis de investigación:

Es probable que los resultados de la aplicación de la nueva técnica quirúrgica difieran de los de la técnica antigua.

3. Deducción de una o más consecuencias verificables:

Si la hipótesis es cierta, entonces, se esperaría que existan diferencias entre las frecuencias de pacientes ubicados en las diversas modalidades de resultados de la técnica quirúrgica nueva en comparación con las frecuencias que se encontrarían en las mismas modalidades si se hubiera aplicado la técnica quirúrgica antigua.

4. Diseño (s) para verificar (cada una de) la (s) consecuencia (s) :

Un solo grupo. Verificación de la similitud de una distribución de frecuencias observada con una distribución de frecuencias esperada (o teórica).

5. Clasificación metodológica y estadística de las variables contenidas en el diseño:

Metodológicamente existe una constante que es antecedente de una variable. Tal constante es la aplicación de una técnica quirúrgica que solo tiene una modalidad: NUEVA; estadísticamente no es susceptible de clasificación.

La variable consecuente o dependiente es el tipo de resultado de la técnica quirúrgica; sus modalidades son:

Resultado de la aplicación
de la técnica quirúrgica

Invalidez total
Invalidez parcial
Funcionamiento normal
Mejoría funcional

Estadísticamente, tal variable se clasifica como cualitativa ordinal.

6. Selección de la prueba de análisis estadístico indicada para el diseño y el tipo de variables:

Prueba de Bondad de Ajuste de la una distribución de frecuencias observada con relación a una distribución de frecuencias teórica o esperada a través de chi cuadrada.

7. Procedimiento para realizar la prueba:**Primero:**

Planteamiento de hipótesis estadísticas:

$$H_0: F_o = F_e$$

$$H_a: F_o \neq F_e$$

Segundo:

Disposición de ambas distribuciones de frecuencias. Para obtener la distribución de frecuencias teórica o esperada se aplican los porcentajes de los resultados de la técnica quirúrgica antigua al total de pacientes:

	invalidez total:	invalidez parcial:	funcionamiento normal:	mejoría funcional:
Distribución de frecuencias observada (Fo)	31 casos	45 casos	73 casos	106 casos
Distribución de frecuencias esperada (Fe)	25.10% de 255 = 64 casos	30.87% de 255 = 79 casos	29.22% de 255 = 74 casos	14.81% de 255 = 38 casos

Tercero:

Cálculo del valor χ^2 mediante la fórmula:

$$\chi^2 = \sum \frac{(o - e)^2}{e}$$

donde:

o = frecuencia observada en una modalidad

e = frecuencia esperada en la misma modalidad

	invalidez total:	invalidez parcial:	funcionamiento normal:	mejoría funcional:
(Fo)	31 casos	45 casos	73 casos	106 casos
(Fe)	64 casos	79 casos	74 casos	38 casos
$\frac{(o - e)^2}{e}$	$\frac{(31 - 64)^2}{64} = 17.02$	$\frac{(45 - 79)^2}{79} = 14.63$	$\frac{(73 - 74)^2}{74} = 0.01$	$\frac{(106 - 38)^2}{38} = 121.68$

$$\chi^2 = \sum \frac{(o-e)^2}{e} = 17.02 + 14.63 + 0.01 + 121.68 = 153.34$$

Cuarto:

Comparación de χ^2 calculada con un valor tabular (crítico) de χ^2 y conclusión en relación con las hipótesis estadísticas:

Extracto de la tabla de

Valores críticos de χ^2

G.L.	Niveles de Significancia	
	0.05	0.01
1	3.84	6.63
2	5.99	9.21
3	7.81	11.34
4	9.49	13.28
.		
100	124.34	135.81

La columna G.L., que significa Grados de Libertad, se refiere, específicamente en esta prueba, al número de modalidades menos uno.

Como el número de modalidades fue de cuatro, entonces el renglón a considerar es el de 3 Grados de Libertad; al cruzar dicho renglón con las columnas de los niveles de significancia de .05 y de .01 se observan los valores críticos de 7.81 y de 11.34 respectivamente.

En vista de que el valor de χ^2 calculada rebasa a ambos valores críticos de χ^2 , entonces se puede rechazar la Hipótesis Estadística Nula $H_0: F_o = F_e$ con una $p < 0.01$

8. Conclusión en relación con la consecuencia verificada:

Parecen existir diferencias entre las frecuencias de pacientes ubicados en las diversas modalidades de resultados de la técnica quirúrgica nueva en comparación con las frecuencias que se encontrarían en las mismas modalidades si se hubiera aplicado la técnica quirúrgica antigua, al menos en el caso de los 255 casos estudiados.

**PROBLEMA
EJEMPLAR****Evaluación de la bondad de ajuste de la
distribución de una variable medida en un grupo****(Diseño de un grupo y una variable)**

Un Profesor de anatomía había observado, a lo largo de sus muchos años de impartir la materia, que usualmente sus alumnos se distribuían, en cuanto a los resultados finales del curso, de la siguiente manera:

Muy Bien	Bien	Suficiente	No acreditó	No presentó
15%	40%	20%	20%	5%

En el año de 1993 se aplicaron muy rigurosos exámenes de ingreso a la carrera de medicina y se le aseguró al profesor que los nuevos alumnos de su materia tendrían un rendimiento superior, en comparación con las anteriores generaciones. A continuación se muestran los resultados en el curso de los 75 alumnos de la generación de 1993.

Muy Bien	Bien	Suficiente	No acreditó	No presentó
21 alumnos	24 alumnos	15 alumnos	10 alumnos	5 alumnos

¿Parece ser diferente la generación de 1993 a las de años anteriores en cuanto a su desempeño en la materia de anatomía?

ETAPAS PARA LA RESOLUCIÓN DEL PROBLEMA:

A. Tenga a la mano, para consulta, el texto *Evaluación de la bondad de ajuste de la distribución de una variable medida en un grupo*

B. Asegúrese de que comprende el concepto de variable, de que es capaz de clasificar variables y, en particular, que puede identificar las modalidades y frecuencias de variables de tipo cualitativo. Ejercítase en la aplicación de porcentajes a un total para obtener frecuencias parciales

1. Redacte el problema de la investigación

2. Redacte la hipótesis de la investigación

3. Opine si la siguiente consecuencia verificable, deducida de la hipótesis, le parece pertinente:

Si la hipótesis es cierta, entonces, al comparar las frecuencias de calificaciones observadas en la generación de 1993 con las frecuencias de calificaciones que teóricamente habrían obtenido si se les aplicaran los porcentajes de las generaciones anteriores, se esperaría que existieran diferencias estadísticamente significativas entre ambas distribuciones de frecuencias.

4. Opine si el diseño de la investigación corresponde al estudio de un único grupo a quien se le midió una variable cualitativa (específicamente a quien se le midieron las frecuencias de individuos en cada modalidad de tal variable cualitativa)

5. Decida si la variable en estudio es realmente cualitativa e identifique, en tal caso, las modalidades y las frecuencias de individuos en cada modalidad.
6. Mencione el nombre de la prueba de análisis estadístico que corresponde a esta investigación.
- 7 (a). Plantée, simbólicamente, las hipótesis estadísticas nula y alterna.
- 7 (b). Usando una tabla auxiliar de trabajo como la siguiente y observando el procedimiento mostrado en el texto de consulta, calcule el valor de chi cuadrada:

	Muy Bien	Bien	Suficiente	No acreditó	No presentó	Total
Frecuencias realmente observadas (Fo)						75 alumnos
Frecuencias esperadas al aplicar porcentajes de anteriores generaciones (Fe)						75 alumnos
$\frac{(o - e)^2}{e}$						

- 7 (c). Una vez calculado el valor de chi cuadrada y mediante el uso de la tabla correspondiente (tal como se muestra en el texto de consulta) decida si rechaza la hipótesis estadística nula.
8. Redacte una conclusión acerca de la consecuencia verificada que se dedujo a partir de la hipótesis de la investigación.

Problemas

Identifique la posible asociación entre cada uno de los antecedentes y la consecuencia

Ingestión de sopa	Presentación de diarrea		Totales
	SI	NO	
SI	64	17	81
NO	5	34	39
			120

Ingestión de arroz	Presentación de diarrea		Totales
	SI	NO	
SI	33	23	56
NO	36	28	64
			120

Ingestión de barbacoa	Presentación de diarrea		Totales
	SI	NO	
SI	34	40	74
NO	35	11	46
			120

Cálculo e interpretación de *Chi cuadrada* (χ^2) en los estudios Transversales

1. INTRODUCCION

Los estudios transversales suelen tener entre sus principales propósitos la indagación de la posible **asociación** entre dos variables usualmente de tipo cualitativas; por ejemplo: cuando interesa averiguar si uno u otro tipo de alimento puede estar asociado con un brote de diarrea en una población o si el tipo de orientación política de los padres influye en el tipo de escuela en que inscriben a sus hijos.

Los estudios transversales requieren la clasificación de un grupo de acuerdo a las modalidades que adopten ambas variables.

A través de un ejemplo, en la siguiente ilustración se esquematiza la variante más común de los estudios transversales:

Casos de deficiente desempeño escolar
según exposición crónica al plomo

Exposición crónica	Desempeño deficiente		
	SI	NO	
SI	42	8	50
NO	8	42	50
	50	50	100

A la distribución de los casos, tal como se encontraron en la realidad se denomina *distribución observada* y a cada una de las frecuencias contenidas en las celdillas se denomina *frecuencia observada*. A cada frecuencia observada se le simboliza con *f_o*

En el supuesto de que no existiera ninguna asociación entre la exposición crónica al plomo y el desempeño escolar deficiente las frecuencias que se esperarían encontrar en cada una de las celdillas serían las siguientes:

Exposición crónica	Desempeño deficiente		
	SI	NO	
SI	25	25	50
NO	25	25	50
	50	50	100

A la distribución de los casos, tal como se encontrarían hipotéticamente si no existiera ninguna asociación entre las dos variables se denomina *distribución esperada* y a cada una de las frecuencias contenidas en las celdillas se denomina *frecuencia esperada*. A cada frecuencia esperada se le simboliza con f_e .

2. DISPOSICION DE LOS DATOS

Una vez que se dispone en cada celdilla de los datos realmente observados (es decir de las frecuencias observadas f_o), se debe efectuar el cálculo de las correspondientes frecuencias esperadas f_e bajo la hipótesis de no asociación. El cálculo de la frecuencia esperada para cada una de las celdillas utiliza la siguiente fórmula:

$$f_e = \frac{(t_{mr})(t_{mc})}{tt}$$

donde: f_e = frecuencia esperada para una celdilla determinada
 t_{mr} = total marginal del renglón de dicha celdilla
 t_{mc} = total marginal de la columna de la misma celdilla
 tt = total de casos de toda la tabla

obsérvese cómo se efectúa el cálculo de la frecuencia esperada f_e para la celdilla superior izquierda que, de momento, solo contiene a la frecuencia observada f_o :

Casos de deficiente desempeño escolar
según exposición crónica al plomo

Exposición crónica	Desempeño deficiente		
	SI	NO	
SI	42	8	50
NO	8	42	50
	50	50	100

t_{mr} de la celdilla = 50;

t_{mc} de la celdilla = 50

gran total = 100

$$f_e \text{ para la celdilla} = \frac{(t_{mr})(t_{mc})}{tt} = \frac{(50)(50)}{100} = \frac{(2\ 500)}{100} = 25$$

De la misma manera se procede al cálculo de las frecuencias esperadas para cada una de las demás celdillas. Las frecuencias esperadas se anotan entre paréntesis junto a la correspondiente frecuencia observada de la siguiente manera:

**Casos de deficiente desempeño escolar
según exposición crónica al plomo**

Exposición crónica	Desempeño deficiente		
	SI	NO	
SI	42 (25)	8 (25)	50
NO	8 (25)	42 (25)	50
	50	50	100

3. PROCEDIMIENTO DE CALCULO DE CHI CUADRADA (χ^2)

Una vez que ya se dispone de la tabla que contiene tanto las frecuencias observadas como las frecuencias esperadas se utiliza la fórmula para el cálculo del valor de Chi cuadrada (χ^2):

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

para el caso que se está ejemplificando, los cálculos son como sigue:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(42 - 25)^2}{25} + \frac{(8 - 25)^2}{25} + \frac{(8 - 25)^2}{25} + \frac{(42 - 25)^2}{25} = 46.24$$

4. INTERPRETACION

El resultado 46.24 expresa qué tan diferente es la distribución de frecuencias observadas con respecto a la distribución de frecuencias esperadas.

Debe recordarse que la distribución de frecuencias esperadas representa a las frecuencias esperadas f_e que debieran encontrarse si no existiera ninguna asociación entre las dos variables; es decir, entre la exposición crónica al plomo y el desempeño escolar deficiente.

Por lo anterior, si las frecuencias observadas f_o hubieran sido iguales a las frecuencias esperadas f_e la diferencia entre ellas hubiera sido de cero. En consecuencia, el hecho de que no existiera diferencia alguna entre las frecuencias observadas y las frecuencias esperadas habría provocado

que en cada paso de los cálculos de la fórmula de chi cuadrada se hubieran obtenido valores de cero y que la suma final hubiera resultado de cero.

Por lo anterior, cuando se encuentran resultados de cero o muy cercanos a esta cifra, se dice que no existe asociación entre las dos variables que se están estudiando. Por lo contrario, cuando se encuentran resultados muy lejanos al cero, se dice que existe asociación intensa entre una y otra variable.

Para el ejemplo que se ha desarrollado puede decirse que existe intensa asociación entre la exposición crónica al plomo y el desempeño escolar deficiente porque el valor calculado de chi cuadrada está muy alejado de cero.

En ocasiones, los resultados obtenidos son escasamente diferentes a cero y por no ser tan grandes causan la impresión de ser debidos solamente al azar. Obsérvese el siguiente caso:

Frecuencias Observadas	Frecuencias Esperadas								
<table border="1" style="width: 100%; text-align: center;"> <tr><td>29</td><td>21</td></tr> <tr><td>21</td><td>29</td></tr> </table>	29	21	21	29	<table border="1" style="width: 100%; text-align: center;"> <tr><td>25</td><td>25</td></tr> <tr><td>25</td><td>25</td></tr> </table>	25	25	25	25
29	21								
21	29								
25	25								
25	25								
⇓	⇓								
Ambas Frecuencias									
<table border="1" style="width: 100%; text-align: center;"> <tr><td>29 (25)</td><td>21 (25)</td></tr> <tr><td>21 (25)</td><td>29 (25)</td></tr> </table>		29 (25)	21 (25)	21 (25)	29 (25)				
29 (25)	21 (25)								
21 (25)	29 (25)								

Cálculo de chi cuadrada:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(29 - 25)^2}{25} + \frac{(21 - 25)^2}{25} + \frac{(21 - 25)^2}{25} + \frac{(29 - 25)^2}{25} = 2.56$$

En sentido estricto, el resultado no es cero y por ello tendría que decirse que, al no parecerse ambas distribuciones de frecuencias, existe asociación entre ambas variables. Sin embargo el resultado no parece ser tan grande y surge la duda de si conviene o no declarar que sí hay asociación.

Para resolver la anterior cuestión debe acudir al uso de una tabla de valores críticos de chi cuadrada que marca los valores límite para declarar si existe asociación significativa entre el antecedente y el efecto. A continuación se muestra un extracto de tal tabla:

Grados de libertad (g.l.)	Nivel de confianza .05	Nivel de confianza .01
1	3.84	6.64
2	5.99	9.21
3	7.82	11.35
4	9.49	13.28

La primera columna a la extrema izquierda de la tabla, con el encabezado *Grados de libertad*, se refiere a un valor que resulta de combinar el número de columnas con el número de renglones que tenga la tabla de contingencia conque se haya trabajado bajo la siguiente fórmula general:

$$g.l. = (\text{columnas menos } 1) (\text{renglones menos } 1).$$

En el caso de la tabla de contingencia que se está usando, la fórmula lleva a los siguientes cálculos : $g.l. = (2 - 1) (2 - 1) = (1) (1) = 1$; es decir que la tabla de contingencia tiene 1 grado de libertad.

Las siguientes dos columnas de la tabla de valores críticos de chi cuadrada se refieren a dos de los diversos niveles de confianza empleados para decidir si la asociación encontrada entre las dos variables puede estar dada o no por azar o por efecto de una verdadera asociación entre ambas variables.

Si se ha escogido el nivel de confianza de .05 y considerando que la tabla que contiene los datos tiene un grado de libertad, entonces se encuentra un *valor crítico* de 3.84; dicho valor no es excedido por el *valor calculado* para chi cuadrada y por ello se concluye que *con un riesgo de error de 5% puede afirmarse que no parece existir asociación entre las dos variables ya que el valor calculado para chi cuadrada no excedió al valor crítico en el nivel de confianza de .05*

**PROBLEMA
EJEMPLAR**

**Asociación de dos variables cualitativas
Chi cuadrada en tablas de 2 x 2 y n x n**

(Diseño de un grupo y dos variables cualitativas)

Un grupo de 89 alumnos de primer año de la Facultad de Medicina fue entrevistado en 1991 para averiguar tanto la orientación política predominante de sus padres como el tipo de educación que recibieron durante sus estudios de educación primaria.

Por una parte, se calificó a los estudiantes según la *orientación política de los padres* en alguna de las siguientes tres modalidades: *conservadora, moderada, liberal*; esto es, se midió una característica o variable de tipo cualitativo ya que la orientación política no es susceptible de cuantificar.

Por otra parte, se observó en los estudiantes el *tipo de educación primaria recibida* y se hizo la siguiente clasificación: *flexible, moderada, rígida*; esto es, también se midió en los estudiantes otra característica o variable de tipo cualitativo.

En resumen, las características o variables seleccionadas y recogidas en este estudio de un solo grupo fueron dos, siendo ambas de tipo cualitativo: *orientación política de los padres* y *tipo de educación primaria recibida*.

El conteo de la información arrojó los siguientes resultados:

89 estudiantes de primer año, según orientación política
de los padres y tipo de educación primaria recibida
Facultad de Medicina, 1991

Orientación política de los padres	Tipo de educación primaria recibida			Totales No. y (%)
	Flexible No. y (%)	Moderada No. y (%)	Rígida No. y (%)	
Conservadora	7 (22)	10 (31)	15 (47)	32 (100)
Moderada	9 (30)	10 (33)	11 (37)	30 (100)
Liberal	14 (52)	8 (30)	5 (18)	27 (100)

Descripción Estadística:

- los hijos de padres con orientación política **conservadora** que recibieron educación primaria de tipo **rígida** (47 %) predominaron sobre los sometidos a los otros dos tipos de educación primaria ,

- en el caso de hijos de padres con orientación política liberal más de la mitad de ellos (52%) recibieron educación primaria de tipo flexible;
- finalmente, en el caso de los hijos de padres con orientación política moderada no pareció haber un predominio contundente de alguno de los tres tipos de educación primaria recibida, pues los porcentajes correspondientes fueron 30, 33 y 37 respectivamente.

¿Parece estar asociada la orientación política de los padres con el tipo de educación primaria recibida?

ETAPAS PARA LA RESOLUCIÓN DEL PROBLEMA:

A. Tenga a la mano, para consulta, el texto *Cálculo e interpretación de Chi cuadrada en los estudios transversales*

B. Asegúrese de que comprende el concepto de variable, de que es capaz de clasificar variables y, en particular, que puede identificar las modalidades y frecuencias de variables de tipo cualitativo.

1. Redacte el problema de la investigación
2. Redacte la hipótesis de la investigación
3. Opine si la siguiente consecuencia verificable, deducida de la hipótesis, le parece pertinente:

Si la hipótesis es cierta, entonces, al evaluar las frecuencias observadas en las diferentes modalidades de ambas variables, se esperaría que existiera una tendencia que muestre asociación estadísticamente significativa entre ambas variables.

4. Opine si el diseño de la investigación corresponde al estudio de un único grupo a quien se le midieron dos variables cualitativas (específicamente a quien se le midieron las frecuencias de individuos en cada modalidad de tales variables cualitativas)
5. Decida si las variables en estudio son realmente cualitativas e identifique, en tal caso, sus modalidades y las frecuencias de individuos en cada modalidad.
6. Mencione el nombre de la prueba de análisis estadístico que corresponde a esta investigación.
- 7 (a). Plantée, simbólicamente, las hipótesis estadísticas nula y alterna.

7 (b). Usando una tabla auxiliar de trabajo como la siguiente y observando el procedimiento mostrado en el texto de consulta, calcule para cada celdilla los valores f_e y los valores $\frac{(f_o - f_c)^2}{f_c}$. Posteriormente, calcule el valor de chicuadrada.

VALORES $\frac{(f_o - f_c)^2}{f_c}$

Orientación política de los padres	Tipo de educación primaria recibida			
	Flexible	Moderada	Rígida	
Conservadora				
Moderada				
Liberal				

7 (c). Una vez calculado el valor de chi cuadrada y mediante el uso de la tabla correspondiente (tal como se muestra en el texto de consulta) decida si rechaza la hipótesis estadística nula.

8. Redacte una conclusión acerca de la consecuencia verificada que se dedujo a partir de la hipótesis de la investigación.

**Diseño para la correlación
de 2 variables medidas en
un solo grupo**

GRUPO UNICO

Variable Variable
Cuantitativa Cuantitativa
X Y

Valores numericos	Valores numericos
X 1	Y 1
X 2	Y 2
X 3	Y 3
X 4	Y 4
.	.
.	.
.	.
.	.
X n	Y n

Pruebas Estadísticas:

- Prueba de significancia del coeficiente de correlación r de Pearson
(si ambas variables tienen distribución normal)
- Prueba de Spearman
(si las variables NO tienen distribución normal o son discretas)

**Diseño para la correlación
de 2 variables medidas en
un solo grupo**

<< EJEMPLO >>

GRUPO DE ALUMNOS

Calificación Calificación
en física en matemáticas

X 1	Y 1
X 2	Y 2
X 3	Y 3
X 4	Y 4
.	.
.	.
.	.
.	.
X n	Y n

Pruebas Estadísticas:

- Prueba de significancia del coeficiente de correlación r de Pearson
(si ambas variables tienen distribución normal)
- Prueba de Spearman
(si las variables NO tienen distribución normal o son discretas)

**Diseño para la correlación
de 2 variables medidas en
un solo grupo**

GRUPO UNICO

Variable Cualitativa X	Variable Cualitativa Y	
	Modalidad Y ₁	Modalidad Y ₂
Modalidad X ₁	Frecuencia X ₁ Y ₁	Frecuencia X ₁ Y ₂
Modalidad x ₂	Frecuencia X ₂ Y ₁	Frecuencia X ₂ Y ₂

TABLA DE CONTINGENCIA

Puede ser de mas
de dos columnas
y de mas de dos
renglones

Pruebas Estadísticas:

- Dependen de la hipótesis de investigación.
- Son aplicables las pruebas propias para tablas de contingencia; entre otras:
 - Chi Cuadrada
 - Prueba exacta de Fisher
 - Coefficiente r de correlación tetracórica
 - Coefficiente Phi
 - Q de Yule
 - Coefficiente C de contingencia
 - Razon de momios

**Diseño para la correlación
de 2 variables medidas en
un solo grupo**

<<< EJEMPLO >>>

GRUPO UNICO

INGESTION DE SOPA	PRESENTACION DE DIARREA	
	SI	NO
SI	Cantidad de personas que comieron sopa y enfermaron	Cantidad de personas que comieron sopa y no enfermaron
NO	Cantidad de personas que no comieron sopa y enfermaron	Cantidad de personas que no comieron sopa y no enfermaron

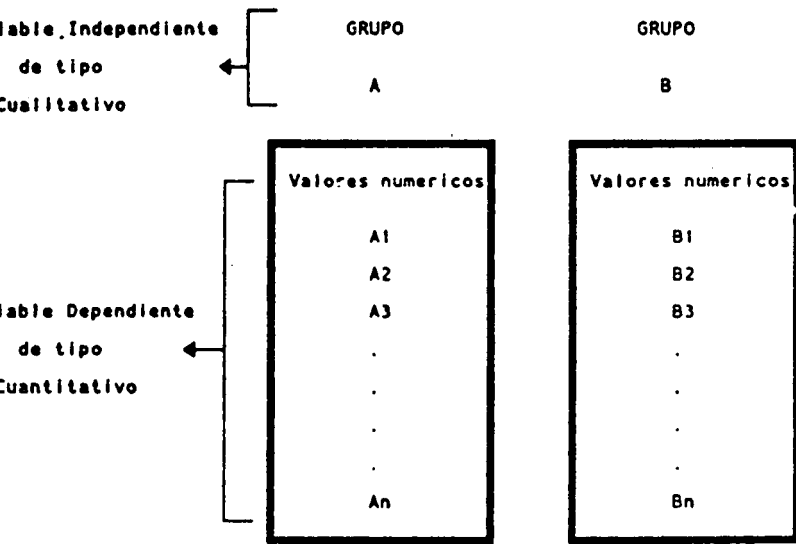
TABLA DE CONTINGENCIA

Puede ser de mas
de dos columnas
y de mas de dos
renglones

Pruebas Estadísticas:

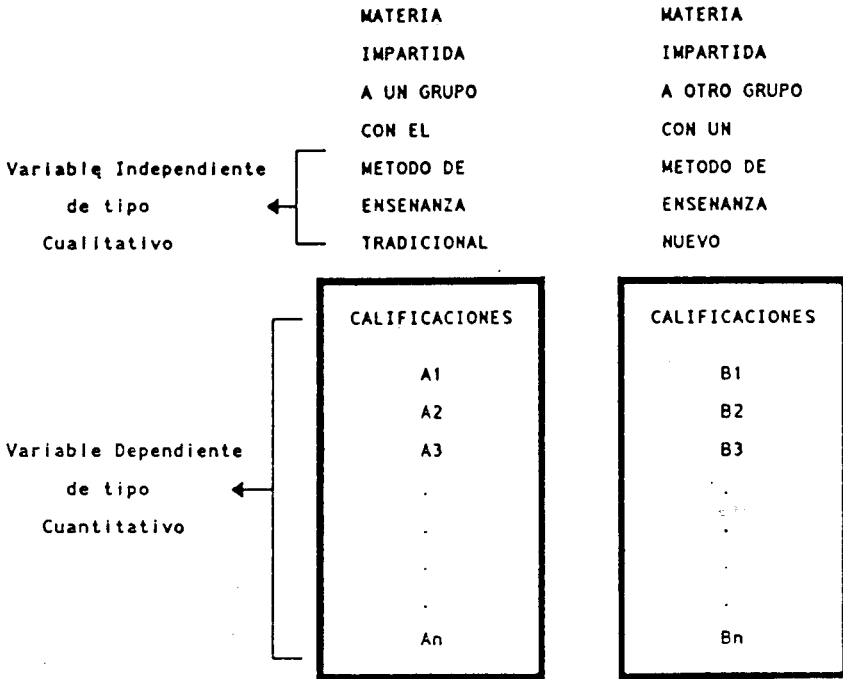
- Dependen de la hipótesis de investigación.
- Son aplicables las pruebas propias para tablas de contingencia; entre otras:
 - Chi Cuadrada
 - Prueba exacta de Fisher
 - Coefficiente r de correlación tetracórica
 - Coefficiente Phi
 - Q de Yule
 - Coefficiente C de contingencia
 - Razon de momios

Diseño para la comparación de 2 grupos diferentes



Diseño para la comparación de 2 grupos diferentes

<< EJEMPLO >>

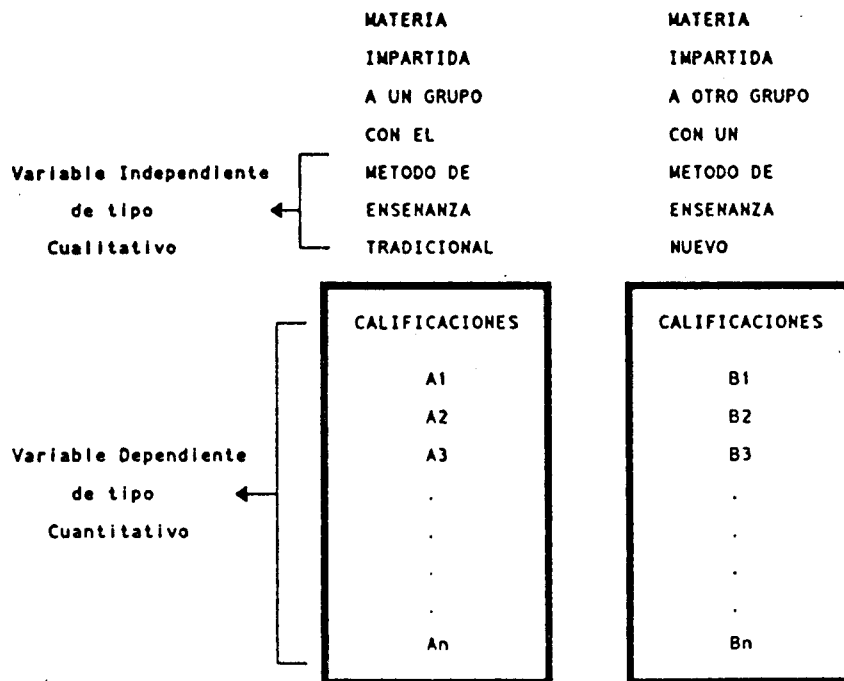


Pruebas Estadísticas:

- Comparación de 2 promedios a través de la distribución t de Student (si la variable dependiente tiene distribución normal)
- Prueba de Mann-Whitney (si la variable dependiente NO tiene distribución normal o es discreta)

**Diseño para la comparación
de 2 grupos diferentes**

<< EJEMPLO >>

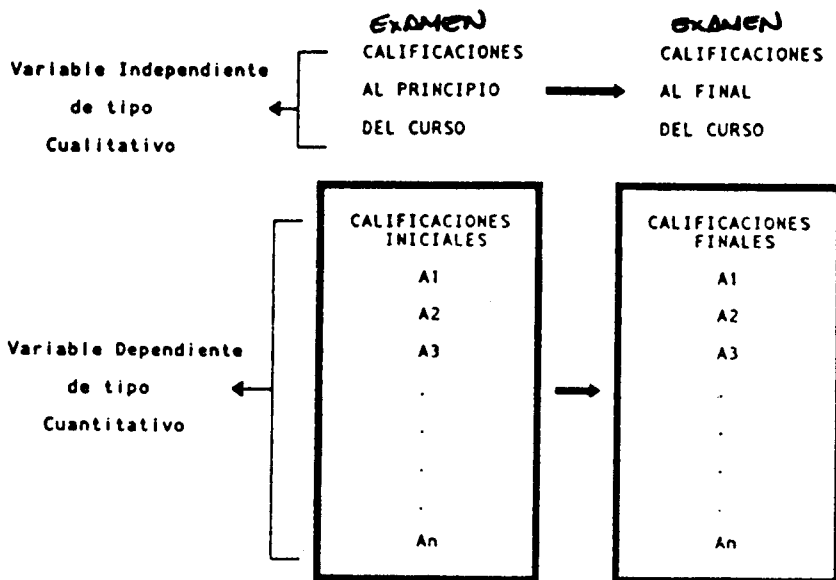


Pruebas Estadísticas:

- Comparación de 2 promedios a través de la distribución t de Student (si la variable dependiente tiene distribución normal)
- Prueba de Mann-Whitney (si la variable dependiente NO tiene distribución normal o es discreta)

Diseño para la comparación
de 1 grupo antes-después
o en dos situaciones diferentes

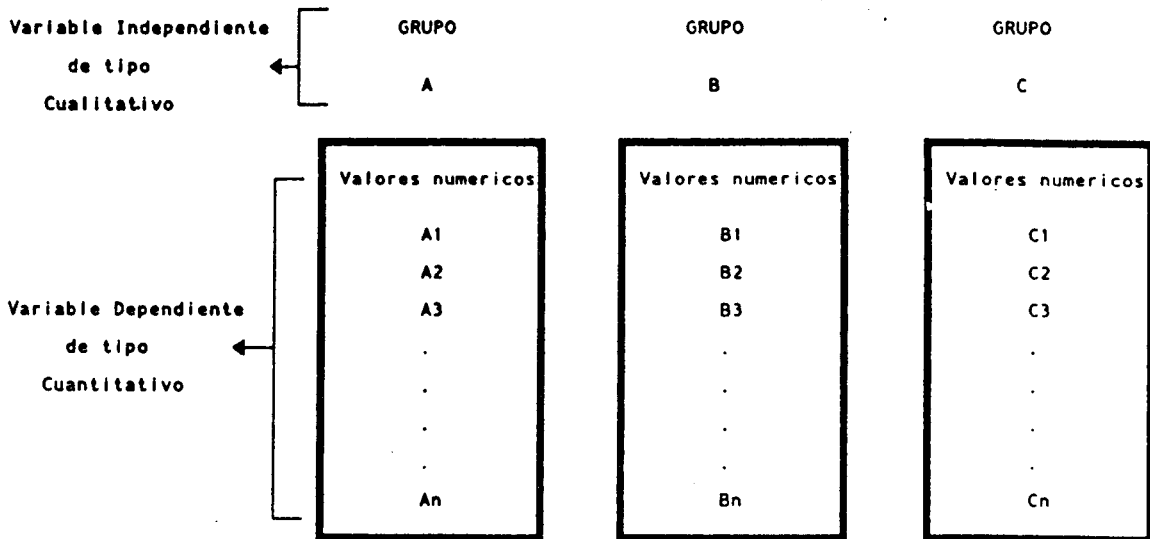
<< EJEMPLO >>



Pruebas Estadísticas:

- Comparación de 2 promedios a través de la distribución t de Student (si la variable dependiente tiene distribución normal)
- Prueba de Wilcoxon (si la variable dependiente NO tiene distribución normal o es discreta)

**Diseño para la comparación
de 3 o mas grupos diferentes**

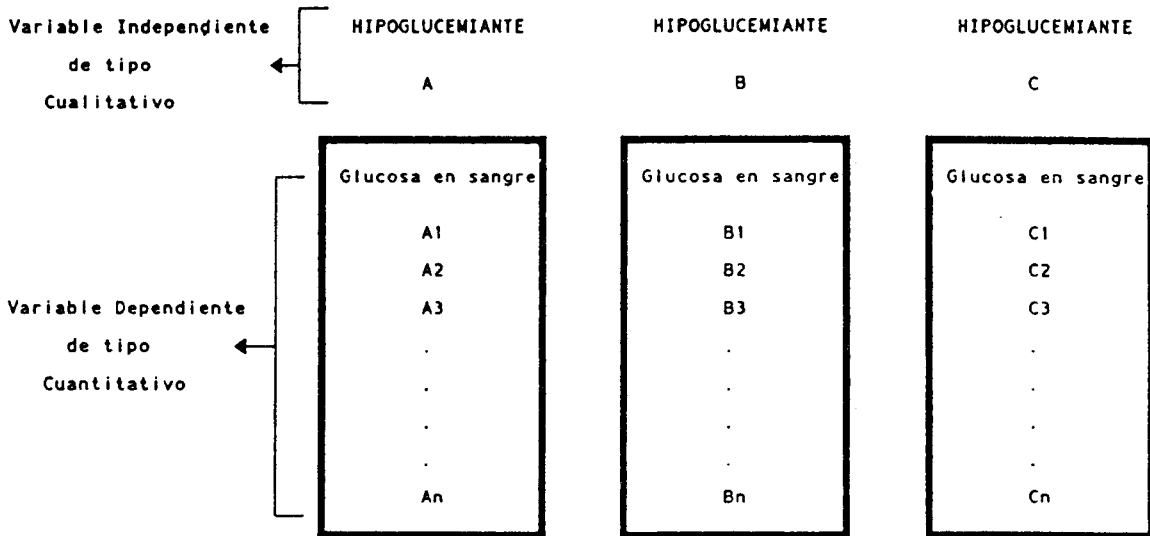


Pruebas Estadísticas:

- Comparación de 3 o mas promedios a través del diseño del Analisis de varianza (ANOVA) totalmente al azar (si la variable dependiente tiene distribución normal)
- Prueba de Kruskal-Wallis (si la variable dependiente NO tiene distribución normal o es discreta)

**Diseño para la comparación
de 3 o más grupos diferentes**

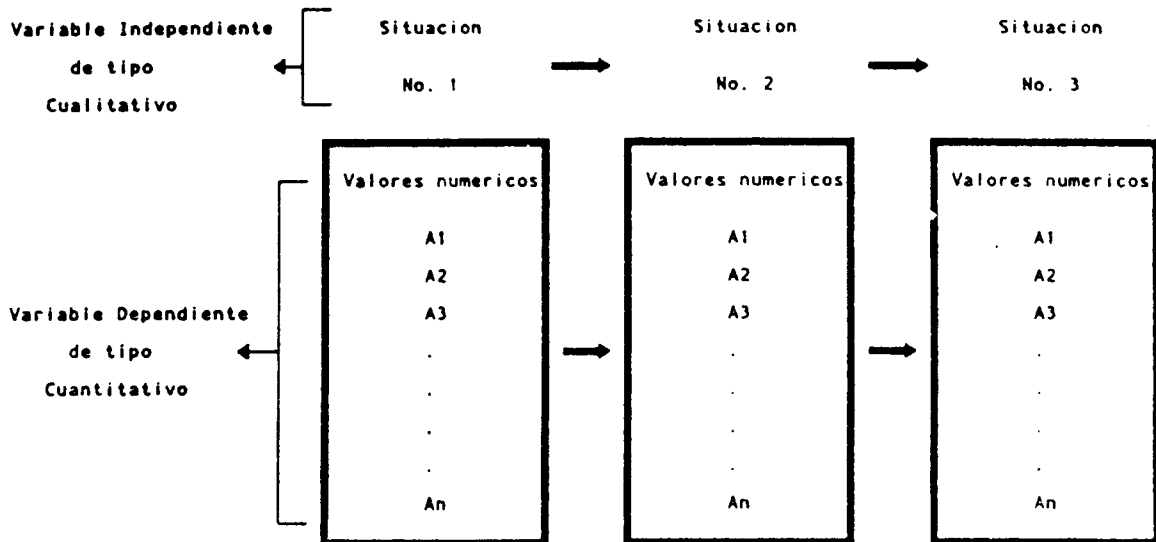
<<< EJEMPLO >>>



Pruebas Estadísticas:

- Comparación de 3 o más promedios a través del diseño del Analisis de varianza (ANOVA) totalmente al azar (si la variable dependiente tiene distribución normal)
- Prueba de Kruskal-Wallis (si la variable dependiente NO tiene distribución normal o es discreta)

Diseño para la comparación de 3 o mas grupos apareados o de un solo grupo en tres o mas situaciones diferentes



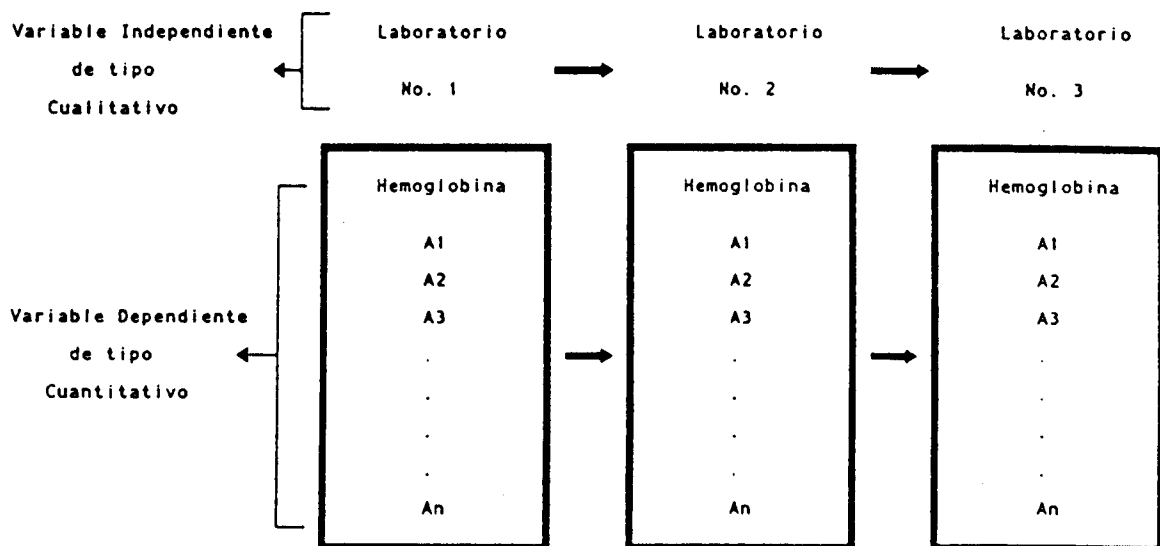
Pruebas Estadísticas:

- Comparación de 3 o mas promedios a través del diseño del Analisis de varianza (ANOVA) en bloques simples (si la variable dependiente tiene distribución normal)
- Prueba de Friedman (si la variable dependiente NO tiene distribución normal o es discreta)

Diseño para la comparación
de 3 o mas grupos apareados
o de un solo grupo en tres o
mas situaciones diferentes

<<< Ejemplo >>>

"Un grupo de muestras de sangre
analizadas en tres laboratorios"



Pruebas Estadísticas:

- Comparación de 3 o mas promedios a través del diseño del Analisis de varianza (ANOVA) en bloques simples (si la variable dependiente tiene distribución normal)
- Prueba de Friedman (si la variable dependiente NO tiene distribución normal o es discreta)

Análisis estadístico de la asociación de dos variables cuantitativas continuas:

Prueba de Hipótesis sobre el coeficiente de correlación de Pearson ¹

Jesús Reynaga Obregón

El coeficiente de correlación de Pearson permite identificar la forma en que se asocian dos variables cuantitativas continuas que hayan sido medidas a cada sujeto de un único grupo.

Sujetos	Valor de la primera variable (x)	Valor de la segunda variable (y)
1	x_1	y_1
2	x_2	y_2
3	x_3	y_3
n	x_n	y_n

La fórmula para el cálculo del coeficiente de correlación de Pearson es la siguiente:

$$r = \frac{n \sum xy - [(\sum x)(\sum y)]}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$$

El coeficiente de correlación de Pearson puede tener valores que varían entre - 1 hasta + 1, pasando por el cero.

Cuando el valor resultante es cercano a +1 se dice que ambas variables se asocian directamente de manera muy estrecha. Cuando el valor resultante es cercano a - 1 se dice que ambas variables se asocian inversamente de manera muy estrecha. Cuando el valor resultante es cercano a cero se dice que ambas variables no tienen asociación.

¹ Se presupone que el lector comprende las indicaciones, cálculo e interpretación del coeficiente de correlación de Pearson desde la perspectiva de la estadística descriptiva antes de utilizar este documento.

En vista de que un valor de cero indica la ausencia de correlación entre ambas variables, dicho valor, el cero, se constituye en el valor asignado a la hipótesis estadística nula H_0 .

La prueba de análisis estadístico sobre el coeficiente de correlación de Pearson exige el cumplimiento de al menos dos condiciones: a) la distribución de ambas variables debe ser semejante a la de la curva normal y b) las varianzas de ambas variables deben ser homogéneas. En el caso de que no se cumplan una o ambas condiciones, la prueba alterna de análisis estadístico es la de *Spearman*.

Para ilustrar el procedimiento de la prueba de análisis estadístico sobre el coeficiente de correlación de Pearson se usarán los siguientes datos:

**Puntuaciones de un grupo de 10 individuos
en pruebas de ansiedad y de depresión**

Individuo	Puntuación en prueba de ansiedad (x)	Puntuación en prueba de depresión (y)	x^2	y^2	xy
1	8.2	6.4	67.24	40.96	52.48
2	7.9	5.8	62.41	33.64	45.82
3	6.3	4.9	39.69	24.01	30.87
4	9.1	7.2	82.81	51.84	65.52
5	5.4	3.9	29.16	15.21	21.06
6	10.3	9.3	106.09	86.49	95.79
7	4.8	3.1	23.04	9.61	14.88
8	6.5	4.2	42.25	17.64	27.30
9	8.3	7.1	68.89	50.41	58.93
10	7.5	5.3	56.25	28.09	39.75
Sumatorias	74.3	57.2	577.83	357.9	452.4

Promedio	7.4	5.7
Desviación estándar	1.69	1.85
Varianza	2.86	3.41

Primer paso, Identificación del diseño:

Se trata de un solo grupo de individuos a quienes se midieron un par de variables, ambas de tipo cuantitativo continuo.

Segundo paso, Verificación de que se satisfacen las condiciones de la prueba de análisis estadístico:

Para la verificación de que ambas variables se distribuyen de manera semejante a la de la curva normal se realizó el cálculo del sesgo y de la curtosis a través del método de momentos y se encontraron los siguientes resultados que muestran que existe la semejanza requerida:

Variable cuantitativa continua	Sesgo	Curtosis
Puntuación de ansiedad	0.0017	2.1635
Puntuación de depresión	0.4443	2.5100

Para la verificación de que las varianzas de ambas variables son semejantes (homoscedásticas)

se calculó el valor $F_{max} = \frac{\text{Varianza mayor}}{\text{Varianza menor}} = \frac{3.41}{2.86} = 1.19$; tal valor resultó menor que el valor

crítico de 3.44 de la tabla de la distribución F para un nivel de significancia de 0.05, con 8 grados de libertad tanto para el numerador como para el denominador (ver tabla F al final del documento).

En vista de lo anterior, no se pudo rechazar la hipótesis de que las varianzas eran diferentes; por tanto se aceptó que las varianzas de ambas variables eran homogéneas, situación que constituye otro de los requisitos de la prueba que se está revisando.

Tercer paso, Planteamiento de hipótesis estadísticas:

Hipótesis estadística nula **Ho: r = 0**

Hipótesis estadística alterna **Ha: r ≠ 0**

Cuarto paso, cálculo del coeficiente de correlación:

Para el cálculo se empleó la fórmula anteriormente mostrada:

$$r = \frac{n \sum xy - [(\sum x)(\sum y)]}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$$

Los cálculos condujeron al siguiente resultado:

$$r = \frac{(10)(452.4) - [(74.3)(57.2)]}{\sqrt{[(10)(577.83) - (74.3)^2][10)(357.9) - (57.2)^2]}} = +0.97$$

Tal resultado, descriptivamente hablando, indica una intensa asociación de tipo directo entre ambas variables; es decir, las más altas puntuaciones en una de las variables correspondieron a las más altas puntuaciones en la otra y, viceversa, las más bajas puntuaciones en una variable correspondieron a las más bajas puntuaciones en la otra.

Quinto paso, comparación del valor calculado con un valor crítico tabular y evaluación de las hipótesis estadísticas:

En la tabla de valores críticos se localizaron los grados de libertad para el tamaño del grupo mediante la fórmula $G.L. = n - 2 = 10 - 2 = 8$

Los valores críticos para los niveles de significancia de 0.05 y de 0.01 fueron de 0.632 y de 0.765, respectivamente. Como el valor calculado de 0.97 supera a ambos valores críticos, entonces se rechazó la Hipótesis estadística nula **Ho: r = 0**

Sexto paso, conclusión en términos del problema de investigación:

Al menos para el grupo estudiado, puede considerarse que existe asociación directa entre las puntuaciones obtenidas por los sujetos en las escalas de ansiedad y de depresión; lo anterior se señala con una $p < 0.01$

**Tabla de valores críticos del coeficiente de correlación de Pearson
a los niveles de significancia de 0.05 y de 0.01**

Grados de libertad	0.05	0.01
1	0.997	0.9999
2	0.950	0.990
3	0.878	0.959
4	0.811	0.917
5	0.754	0.874
6	0.707	0.834
7	0.666	0.798
8	0.632	0.765
9	0.602	0.735
10	0.576	0.708
11	0.553	0.684
12	0.532	0.661
13	0.514	0.641
14	0.497	0.623
15	0.482	0.606
16	0.468	0.590
17	0.456	0.575
18	0.444	0.561
19	0.433	0.549
20	0.423	0.537
21	0.413	0.526
22	0.404	0.515
23	0.396	0.505
24	0.388	0.496
25	0.381	0.487
26	0.374	0.479
27	0.367	0.471
28	0.361	0.463
29	0.355	0.456
30	0.349	0.449
32	0.339	0.436
34	0.329	0.424

Grados de libertad	0.05	0.01
35	0.325	0.418
36	0.320	0.413
38	0.312	0.403
40	0.304	0.393
42	0.297	0.384
44	0.291	0.376
45	0.288	0.372
46	0.284	0.368
48	0.279	0.361
50	0.273	0.354
55	0.261	0.338
60	0.250	0.325
65	0.241	0.313
70	0.232	0.302
75	0.224	0.292
80	0.217	0.283
85	0.211	0.275
90	0.205	0.267
95	0.200	0.260
100	0.195	0.254
125	0.174	0.228
150	0.159	0.208
175	0.148	0.193
200	0.138	0.181
300	0.113	0.148
400	0.098	0.128
500	0.088	0.115
1,000	0.062	0.081

Problema:

Efectúe prueba de análisis estadístico sobre el coeficiente de correlación de Pearson para los siguientes datos.

Siga el procedimiento mencionado en el texto anterior aún cuando no se cumplan las condiciones de la prueba; sin embargo, señálelo en caso de ocurrir.

Valores de glucosa y de colesterol de diez estudiantes de entre 17 y 25 años

Estudiante	Glucosa	Colesterol
1	69	195
2	71	201
3	65	194
4	80	232
5	85	233
6	92	238
7	124	268
8	84	237
9	73	208
10	145	271

SOLUCION A UN PROBLEMA DE:

Análisis estadístico de la asociación de
dos variables cuantitativas continuas:

Prueba de Hipótesis sobre el
coeficiente de correlación de Pearson¹

Jesús Reynaga Obregón

Problema:

Efectúe prueba de análisis estadístico sobre el coeficiente de correlación de Pearson para los siguientes datos.

Siga el procedimiento mencionado en el texto correspondiente aún cuando no se cumplan las condiciones de la prueba; sin embargo, señálelo en caso de ocurrir.

Valores de glucosa y de colesterol de diez estudiantes de entre 17 y 25 años

Estudiante	Glucosa	Colesterol
1	69	195
2	71	201
3	65	194
4	80	232
5	85	233
6	92	238
7	124	268
8	84	237
9	73	208
10	145	271
\bar{x}	88.8	227.7
s	25.92	27.98

¹ Se presupone que el lector comprende las indicaciones, cálculo e interpretación del coeficiente de correlación de Pearson desde la perspectiva de la estadística descriptiva antes de utilizar este documento.

Primer paso, Identificación del diseño:

Se trata de un solo grupo de individuos a quienes se midieron un par de variables, ambas de tipo cuantitativo continuo. El objetivo del análisis es evaluar la intensidad de la asociación de ambas variables y verificar si el coeficiente de correlación difiere de cero.

Segundo paso, Verificación de que se satisfacen las condiciones de la prueba de análisis estadístico:

Para la verificación de que ambas variables se distribuyen de manera semejante a la de la curva normal se realizó el cálculo del sesgo y de la curtosis a través del método de momentos y se encontraron los siguientes resultados que muestran que **NO** existía la semejanza requerida; sin embargo, de acuerdo a las indicaciones, de todas formas se procedió a la prueba de análisis estadístico como si existiera la semejanza mencionada (*ver anexos con los cálculos de sesgo y curtosis para ambas variables*):

Variable cuantitativa continua	Sesgo	Curtosis
Concentración de glucosa	1.27	3.32
Concentración de colesterol	0.26	1.87

Para la verificación de que las varianzas de ambas variables eran semejantes (homoscedásticas)

se calculó el valor $F_{\max} = \frac{\text{Varianza mayor}}{\text{Varianza menor}} = \frac{s^2_{\text{mayor}}}{s^2_{\text{menor}}} = \frac{(27.98)^2}{(25.92)^2} = 1.17$; tal valor resultó menor

que el valor crítico de 3.44 de la tabla de la distribución F para un nivel de significancia de 0.05, con 8 grados de libertad tanto para el numerador como para el denominador (ver tablas de valores críticos de F).

En vista de lo anterior, no se pudo rechazar la hipótesis de que las varianzas eran diferentes; por tanto se aceptó que las varianzas de ambas variables eran homogéneas, situación que constituye otro de los requisitos de la prueba que se está revisando.

Tercer paso, Planteamiento de hipótesis estadísticas:

Hipótesis estadística nula **Ho: r = 0**

Hipótesis estadística alterna **Ha: r ≠ 0**

Cuarto paso, cálculo del coeficiente de correlación:

Para el cálculo se empleó la fórmula:

$$r = \frac{n \sum xy - [(\sum x)(\sum y)]}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$$

Para disponer de los valores incluidos en la fórmula se empleó una tabla auxiliar de trabajo como la siguiente:

Estudiante	Glucosa x	Colesterol y	x ²	y ²	xy
1	69	195	4,761	38,025	13,455
2	71	201	5,041	40,401	14,271
3	65	194	4,225	37,636	12,610
4	80	232	6,400	53,824	18,560
5	85	233	7,225	54,289	19,805
6	92	238	8,464	56,644	21,896
7	124	268	15,376	71,824	33,232
8	84	237	7,056	56,169	19,908
9	73	208	5,329	43,264	15,184
10	145	271	21,025	73,441	39,295
Sumatorias	888	2,277	84,902	525,517	208,216

Los cálculos condujeron al siguiente resultado:

$$r = \frac{(10)(208,216) - [(888)(2,277)]}{\sqrt{[(10)(84,902) - (888)^2][10(525,517) - (2,277)^2]}} = +0.92$$

Tal resultado, descriptivamente hablando, indica una intensa asociación de tipo directo entre ambas variables; es decir, las más altas puntuaciones en una de las variables correspondieron a las más altas puntuaciones en la otra y, viceversa, las más bajas puntuaciones en una variable correspondieron a las más bajas puntuaciones en la otra.

Quinto paso, comparación del valor calculado con un valor crítico tabular y evaluación de las hipótesis estadísticas:

En la tabla de valores críticos se localizaron los grados de libertad para el tamaño del grupo mediante la fórmula $G.L. = n - 2 = 10 - 2 = 8$

Los valores críticos para los niveles de significancia de 0.05 y de 0.01 fueron de 0.632 y de 0.765, respectivamente. Como el valor calculado de 0.92 supera a ambos valores críticos, entonces se rechazó la Hipótesis estadística nula

Ho: r = 0

Sexto paso, conclusión en términos del problema de investigación:

Puede considerarse que existe asociación directa entre las concentraciones de glucosa y las de colesterol encontradas en el grupo de diez estudiantes; lo anterior se señala con una $p < 0.01$

Verificación de semejanza a la distribución normal de los valores de glucosa

Glucosa $\bar{x} = 88.8$

x	$(x - \bar{x})$	$(x - \bar{x})^2$	$(x - \bar{x})^3$	$(x - \bar{x})^4$
69	-19.8	392.04	-7762.39	153,695.36
71	-17.8	316.84	-5639.75	100,387.59
65	-23.8	566.44	-13481.27	320,854.27
80	-8.8	77.44	-681.47	5,996.95
85	-3.8	14.44	-54.87	208.51
92	3.2	10.24	32.77	104.86
124	35.2	1,239.04	43614.21	1,535,220.12
84	-4.8	23.04	-110.59	530.84
73	-15.8	249.64	-3944.31	62,320.13
145	56.2	3,158.44	177504.33	9,975,743.23
totales:	0.00	6,047.60	189,476.64	12,155,061.87

$$m_2 = \frac{\sum (x - \bar{x})^2}{n} = \frac{6,047.60}{10} = 604.76$$

$$m_3 = \frac{\sum (x - \bar{x})^3}{n} = \frac{189,476.64}{10} = 18,947.66$$

$$m_4 = \frac{\sum (x - \bar{x})^4}{n} = \frac{12,155,061.87}{10} = 1,215,506.18$$

$$\text{sesgo} = a_3 = \frac{m_3}{(\sqrt{m_2})^3} = \frac{18,947.66}{(\sqrt{604.76})^3} = 1.27$$

$$\text{curtosis} = a_4 = \frac{m_4}{(m_2)^2} = \frac{1,215,506.18}{(604.76)^2} = 3.32$$

En vista de que el sesgo se encuentra fuera del intervalo de -0.5 a +0.5 se considera que la distribución de los valores de glucosa no es semejante a la de la curva normal.

Verificación de semejanza a la distribución normal de los valores de colesterol

Colesterol $\bar{y} = 227.7$

\bar{y}	$(y - \bar{y})$	$(y - \bar{y})^2$	$(y - \bar{y})^3$	$(y - \bar{y})^4$
195	-32.7	1069.29	-34965.78	1,143,381.10
201	-26.7	712.89	-19034.16	508,212.15
194	-33.7	1135.69	-38272.75	1,289,791.78
232	4.3	18.49	79.51	341.88
233	5.3	28.09	148.88	789.05
238	10.3	106.09	1092.73	11,255.09
268	40.3	1624.09	65450.83	2,637,668.33
237	9.3	86.49	804.36	7,480.52
208	-19.7	388.09	-7645.37	150,613.85
271	43.3	1874.89	81182.74	3,515,212.51
totales:	0.00	7,044.10	48,840.96	9,264,746.26

$$m_2 = \frac{\sum (y - \bar{y})^2}{n} = \frac{7,044.10}{10} = 704.41$$

$$m_3 = \frac{\sum (y - \bar{y})^3}{n} = \frac{48,840.96}{10} = 4,884.10$$

$$m_4 = \frac{\sum (y - \bar{y})^4}{n} = \frac{9,264,746.26}{10} = 926,474.63$$

$$\text{sesgo} = a_3 = \frac{m_3}{(\sqrt{m_2})^3} = \frac{4884.10}{(\sqrt{704.41})^3} = 0.26$$

$$\text{curtosis} = a_4 = \frac{m_4}{(m_2)^2} = \frac{926,474.63}{(704.41)^2} = 1.87$$

En vista de que la curtosis está fuera del intervalo de 2 a 4 se considera que la distribución de los valores de colesterol no es semejante a la de la curva normal.

Texto: Evaluación de la intensidad de la asociación entre dos variables cuantitativas mediante la prueba de Spearman

Introducción:

Tradicionalmente, la evaluación de la intensidad con que se asocian dos variables cuantitativas, medidas en un solo grupo de individuos, se ha venido efectuando a través del coeficiente r de correlación de Pearson. Sin embargo, en muchas ocasiones el uso de dicho procedimiento no ha estado totalmente acorde con la naturaleza de los datos; tal hecho ocurre porque los supuestos en que descansa tal técnica no siempre son cabalmente verificados.

El uso de la técnica de Pearson requiere que las dos variables hayan sido medidas hasta un nivel cuantitativo continuo y que la distribución de ambas sea semejante a la de la curva normal; es decir, que tengan sesgo y curtosis parecidos a los de dicha distribución teórica.

Cuando las variables, siendo cuantitativas continuas, no tienen una distribución semejante a la de la curva normal, o cuando una o ambas son de tipo discreto, la técnica idónea para evaluar la intensidad de la asociación es la descrita por Spearman.

Para ilustrar el procedimiento de cálculo del coeficiente de correlación r_s de Spearman se seguirá un ejemplo breve.

Relato:

Un investigador suponía que el desempeño de los alumnos de la carrera de medicina en materias afines y sinérgicas debiera ser semejante. Para comprobar lo anterior, aplicó dos exámenes a un grupo de diez alumnos. Su pretensión, al aplicar un examen de anatomía y otro de embriología, era averiguar si coincidían las puntuaciones bajas o altas en una materia con las puntuaciones bajas o altas en la otra.

A continuación se muestran los resultados que consisten en números enteros de aciertos en una y otra materia:

Alumno	Calificación en Anatomía	Calificación en Embriología
1	65	74
2	72	61
3	75	69
4	82	90
5	50	51
6	95	79
7	87	95
8	53	52
9	83	77
10	64	63

Secuencia:

1. Problema:

Se ignora si el desempeño escolar de los alumnos de la carrera de medicina es semejante en materias afines y sinérgicas.

2. Planteamiento de la hipótesis de investigación:

Es probable que el desempeño escolar de los alumnos de la carrera de medicina sea semejante en materias afines y sinérgicas.

3. Deducción de una o más consecuencias verificables:

Si la hipótesis es cierta, entonces, al aplicar a un solo grupo de alumnos un examen de anatomía y un examen de embriología, se esperaría que los alumnos que obtengan bajas puntuaciones en una materia también obtengan bajas puntuaciones en la otra y que los alumnos que obtengan altas puntuaciones en una materia también obtengan bajas puntuaciones en la otra. Dicho de otra forma: se esperaría que exista una asociación positiva entre el número de aciertos obtenidos por el grupo de alumnos en uno y otro examen.

4. Diseño (s) para verificar (cada una de) la (s) consecuencia (s) :

Un solo grupo. Medición de dos variables cuantitativas a cada individuo con el propósito de evaluar la intensidad de la asociación de ambas.

5. Clasificación metodológica y estadística de las variables contenidas en el diseño:

Metodológicamente no es menester que una variable sea antecedente a la otra y puede ser que ambas sean consecuentes a otra variable; en este caso no parece haber secuencia entre los aciertos obtenidos en uno y otro examen.

Estadísticamente ambas variables (número de aciertos en uno y otro examen) son de tipo cuantitativo discreto; por ello, no es admisible ni el uso de promedios ni el cálculo de sesgo y curtosis.

6. Selección de la prueba de análisis estadístico indicada para el diseño y el tipo de variables:

Prueba de Spearman (para el mismo diseño, si ambas variables fueran cuantitativas continuas con distribución semejante a la de la curva normal, se utilizaría la prueba de Pearson).

7. Procedimiento para realizar la prueba:**Primero:****Planteamiento de hipótesis estadísticas:**

$H_0: r_s = 0$ (un coeficiente de correlación de cero equivale a la ausencia de asociación)

$H_a: r_s \neq 0$

Segundo:

Asignación de rangos a los valores de cada una de las variables por separado y cálculo de las diferencias de ambos rangos y de sus correspondientes diferencias cuadráticas:

Alumno	Rangos de las Calificaciones en Anatomía	Rangos de las Calificaciones en Embriología	diferencia d	d ²
1	4	6	- 2	4
2	5	3	2	4
3	6	5	1	1
4	7	9	- 2	4
5	1	1	0	0
6	10	8	2	4
7	9	10	- 1	1
8	2	2	0	0
9	8	7	1	1
10	3	4	- 1	1
				20

Tercero:

Cálculo del coeficiente r_s mediante la fórmula:

$$r_s = 1 - \frac{6 \sum d^2}{n(n-1)(n+1)}$$

$$r_s = 1 - \frac{6(20)}{10(10-1)(10+1)} = 1 - \frac{120}{990} = 0.88$$

Cuarto:

Comparación de r_s calculado con r_s crítico y conclusión en relación con las hipótesis estadísticas:

Valores críticos para el coeficiente r_s

N	Niveles de Significancia	
	0.05	0.01
4	1.000	
5	0.900	1.000
6	0.829	0.943
7	0.714	0.893
8	0.643	0.833
9	0.600	0.783
10	0.564	0.746
12	0.506	0.712
14	0.456	0.645
16	0.425	0.601
18	0.399	0.564
20	0.377	0.534
22	0.359	0.508
24	0.343	0.485
26	0.329	0.465
28	0.317	0.448
30	0.306	0.432

r_s calculado (0.88) rebasa tanto al valor crítico correspondiente al nivel de significancia de 0.05 (0.564) como al de 0.01 (0.746)

Por lo anterior, puede rechazarse (¹) a la $H_0: r_s = 0$ con una $p < 0.01$

8. Conclusión en relación con la consecuencia verificada:

Parece existir asociación de tipo positivo entre el número de aciertos obtenidos por el grupo de alumnos en los exámenes de anatomía y de embriología.

9. Conclusión en relación con la hipótesis de investigación:

La conclusión obtenida al comprobar la consecuencia anterior apoya a la hipótesis de que el desempeño de los alumnos de la carrera de medicina en materias afines y sinérgicas puede ser semejante. Esto se afirma, al menos, para el caso del grupo de diez alumnos estudiados y tomando en cuenta las peculiaridades del estudio.

(¹) A diferencia de otras pruebas no paramétricas, en esta el rechazo de la hipótesis nula se da cuando el valor calculado excede al valor crítico de la tabla correspondiente.

SOLUCION A UN PROBLEMA DE:

Análisis estadístico de la asociación de dos variables cuantitativas CONTINUAS que NO se distribuyen de manera semejante a la de la curva normal o de dos variables cuantitativas DISCRETAS:

Prueba de Hipótesis sobre el coeficiente de correlación de SPEARMAN¹:

Jesús Reynaga Obregón

Problema:

Efectúe prueba de análisis estadístico sobre el coeficiente de correlación de **Spearman para los siguientes datos.**

Verifique que no se cumplen alguna o varias de las las condiciones de la prueba clásica de Pearson

Valores de glucosa y de colesterol de diez estudiantes de entre 17 y 25 años

Estudiante	Glucosa	Colesterol
1	69	195
2	71	201
3	65	194
4	80	232
5	85	233
6	92	238
7	124	268
8	84	237
9	73	208
10	145	271

250 19 8

¹ Se presupone que el lector comprende las indicaciones, cálculo e interpretación del coeficiente de correlación de Pearson desde la perspectiva de la estadística descriptiva antes de utilizar este documento.

Primer paso, Identificación del diseño:

Se trata de un solo grupo de individuos a quienes se midieron un par de variables, ambas de tipo cuantitativo continuo. El objetivo del análisis es evaluar la intensidad de la asociación de ambas variables y verificar si el coeficiente de correlación difiere de cero.

Segundo paso, Verificación de que se satisfacen o no las condiciones de la prueba clásica de análisis estadístico de Pearson:

Para la verificación de que ambas variables cuantitativas continuas se distribuían o no de manera semejante a la de la curva normal se realizó el cálculo del sesgo y de la curtosis a través del método de momentos y se encontraron los siguientes resultados que muestran que NO existía la semejanza requerida

Variable cuantitativa continua	Sesgo	Curtosis
Concentración de glucosa	1.27	3.32
Concentración de colesterol	0.26	1.87

Para la verificación de que las varianzas de ambas variables eran semejantes (homoscedásticas)

se calculó el valor $F_{max} = \frac{\text{Varianza mayor}}{\text{Varianza menor}} = \frac{s^2_{mayor}}{s^2_{menor}} = \frac{(27.98)^2}{(25.92)^2} = 1.17$; tal valor resultó menor

que el valor crítico de 3.44 de la tabla de la distribución F para un nivel de significancia de 0.05, con 8 grados de libertad tanto para el numerador como para el denominador (ver tablas de valores críticos de F).

En vista de lo anterior, no se pudo rechazar la hipótesis de que las varianzas eran diferentes; por tanto se aceptó que las varianzas de ambas variables eran homogéneas, situación que constituye otro de los requisitos de la prueba que se está revisando.

Tercer paso, Planteamiento de hipótesis estadísticas:

Hipótesis estadística nula $H_0: r_{\text{spearman}} = 0$

Hipótesis estadística alterna $H_a: r_{\text{spearman}} \neq 0$

Cuarto paso, cálculo del coeficiente de correlación:

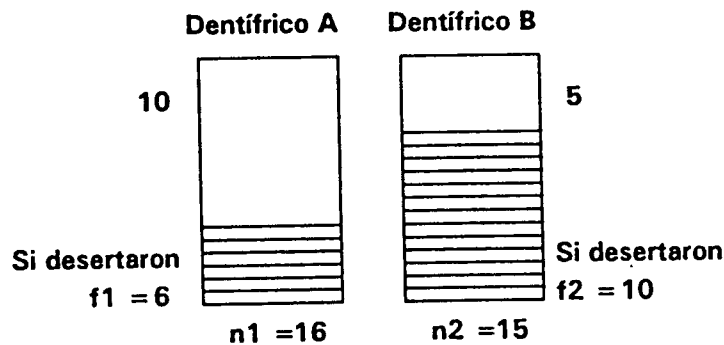
Para el cálculo se empleó la fórmula:

$$r_s = 1 - \frac{6 \sum d^2}{n(n-1)(n+1)}$$

Ejemplo y Resumen:**Comparación de dos grupos a través de sus proporciones mediante la distribución z****Relato:**

En un estudio sobre el uso de los dentífricos en los niños, interesaba saber si el sabor era un factor de influencia para el abandono del hábito del lavado dental. Durante tres meses se administró un dentífrico A de agradable y suave sabor a un grupo de 16 niños y un dentífrico B de picante sabor a menta a otro grupo de 15 niños.

Luego de tres meses se observó que habían abandonado el hábito de lavado dental seis niños del grupo A y diez niños del grupo B.

1o. Diseño y variables:

Variable independiente: Tipo de dentífrico (A / B): (cualitativa nominal con dos modalidades)

Variable dependiente: Deserción (si / no): (cualitativa nominal con dos modalidades)

2o. Medidas de resumen:**Proporciones:**

$$p_1 = \frac{f_1}{n_1} \quad p_1 = \frac{6}{16} = 0.375$$

$$p_2 = \frac{f_2}{n_2} \quad p_2 = \frac{10}{15} = 0.667$$

$$q_1 = 1 - p_1 \quad q_1 = 1 - 0.375 = 0.625$$

$$q_2 = 1 - p_2 \quad q_2 = 1 - 0.667 = 0.333$$

3o. Verificación de las condiciones

$$\begin{array}{cccc}
 n_1 p_1 \geq 5 & n_1 q_1 \geq 5 & n_2 p_2 \geq 5 & n_2 q_2 \geq 5 \\
 (16)(0.375) = 6 & (16)(0.625) = 10 & (15)(0.667) = 10 & (15)(0.333) = 5
 \end{array}$$

4o. Planteamiento de hipótesis estadísticas:

$$H_o: p_1 = p_2$$

$$H_a: p_1 \neq p_2$$

4o. Realización de prueba de análisis estadístico

$$z_{\text{observado}} = \frac{p_1 - p_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{donde: } \hat{p} = \frac{f_1 + f_2}{n_1 + n_2} \quad \text{y} \quad \hat{q} = 1 - \hat{p}$$

$$\hat{p} = \frac{6+10}{16+15} = 0.52 \quad \hat{q} = 1 - 0.52 = 0.48$$

$$z_{\text{observado}} = \frac{0.375 - 0.667}{\sqrt{(0.52)(0.48)\left(\frac{1}{16} + \frac{1}{15}\right)}} = \frac{-0.292}{\sqrt{0.250(0.063+0.067)}} = \frac{-0.292}{\sqrt{0.032}} = \frac{-0.292}{0.180} = -1.63$$

5o. Comparación de $z_{\text{observado}}$ con $z_{\text{crítico}}$ para verificar si se llega a una de las dos regiones de rechazo de la hipótesis estadística nula:

En la curva normal, al definir dos regiones extremas y simétricas que, en conjunto, valgan 0.05 se encuentra que los valores z valen -1.96 por el lado izquierdo y +1.96 por el lado derecho.

En vista de que el valor z calculado no rebasa al valor crítico, entonces no se llega a ninguna de la dos regiones de rechazo de la hipótesis nula y, por ende, no se puede decir que ambas proporciones sean diferentes con un nivel de significancia de 0.05

Problema: **Comparación de dos grupos
a través de proporciones**

Relato 1:

Se diseñó un experimento para investigar el papel del hábito tabáquico como agente carcinogénico en los tumores pulmonares. En el experimento se usaron 74 ratones, de los cuales 38 constituyeron el grupo experimental y 36 el grupo control. Los ratones del grupo control se conectaron en una jaula conectada a una máquina consumidora de cigarrillos. La máquina fue calibrada para consumir 180 cigarrillos por día, durante seis días a la semana, a lo largo de 6 meses.

Al final del experimento, los animales de los dos grupos fueron sacrificados. Se encontraron 21 casos de cáncer pulmonar en el grupo experimental y 9 en el grupo control.

El investigador concluyó: *Hubo muy ligera preponderancia de tumores en el grupo experimental sobre el grupo control que no fue significativa en el análisis estadístico... Los resultados de esta investigación muestran que el papel del cigarrillo es mínimo o no tiene efecto alguno en la incidencia de tumores pulmonares en los ratones.*

Problemas: Realice la prueba de análisis estadístico apropiada para verificar las anteriores conclusiones. ¿Cuáles son sus propias conclusiones?

Relato 2:

Un indicador de la capacidad de atracción y retención de pacientes en las unidades de salud es la proporción de consultas subsecuentes con relación al total. En vista de que en dos unidades médicas se había apreciado que la proporción de consultas subsecuentes era muy baja, se decidió que en una de ellas se aplicaría un estricto programa de mejoría de la atención que incluía modificaciones en los procedimientos de recepción, consulta, archivo y visitas domiciliarias.

Al cabo de un año se encontró lo siguiente:

Unidad Médica	SIN programa	CON programa
Consultas de primera vez	7, 453	2, 155
Consultas Subsecuentes	9, 322	13, 226

Problemas: Identifique si existieron diferencias significativas entre las proporciones de consultas subsecuentes.

Relato 3:

120 personas asistieron a una comida de bodas. El listado siguiente muestra los datos de cada uno de los invitados en relación con la ingestión o no de cada uno de los tres alimentos que se ofrecieron en el festejo y en relación con la presentación de diarrea. En cada caso la letra S significa Si y la letra N significa NO

Núm. de caso	Comió SOPA	Comió ARROZ	Comió MOLE	Tuvo DIARREA	Núm. de caso	Comió SOPA	Comió ARROZ	Comió MOLE	Tuvo DIARREA
1	S	N	N	S	31	N	S	N	N
2	N	S	S	N	32	S	N	S	S
3	S	N	N	N	33	S	N	N	S
4	S	N	S	S	34	S	N	S	S
5	S	N	N	S	35	S	S	N	S
6	N	S	S	N	36	S	S	N	S
7	S	S	N	S	37	N	N	N	N
8	S	S	N	S	38	S	N	N	N
9	S	N	S	N	39	S	N	S	N
10	S	S	S	S	40	S	S	N	S
11	S	S	N	S	41	N	N	S	N
12	S	S	N	S	42	N	S	N	N
13	N	S	N	N	43	S	S	N	S
14	S	S	N	S	44	N	N	N	N
15	N	N	N	S	45	N	N	N	N
16	N	S	S	N	46	S	N	S	N
17	S	S	N	S	47	N	S	N	S
18	S	S	N	S	48	S	N	S	S
19	S	N	S	S	49	S	S	N	N
20	S	N	S	S	50	N	N	N	N
21	S	N	N	S	51	N	N	N	N
22	S	N	N	S	52	N	S	N	N
23	S	N	S	S	53	N	N	S	N
24	S	N	S	S	54	S	S	N	S
25	S	S	S	S	55	N	N	N	N
26	N	N	N	S	56	N	N	S	N
27	S	N	S	S	57	S	S	S	N
28	N	S	N	S	58	N	N	N	N
29	S	S	N	S	59	N	N	N	N
30	S	S	S	S	60	S	S	S	N

Problemas:

1. Considere como dos grupos a quienes ingirieron y no ingirieron sopa y verifique si existen diferencias significativas entre las proporciones de enfermos de diarrea en ambos grupos.
2. Considere como dos grupos a quienes ingirieron y no ingirieron arroz y verifique si existen diferencias significativas entre las proporciones de enfermos de diarrea en ambos grupos.
3. Considere como dos grupos a quienes ingirieron y no ingirieron mole y verifique si existen diferencias significativas entre las proporciones de enfermos de diarrea en ambos grupos.

PROBLEMAS EJEMPLARES:

Comparación de dos grupos a través de sus proporciones mediante la distribución z

Problema 1

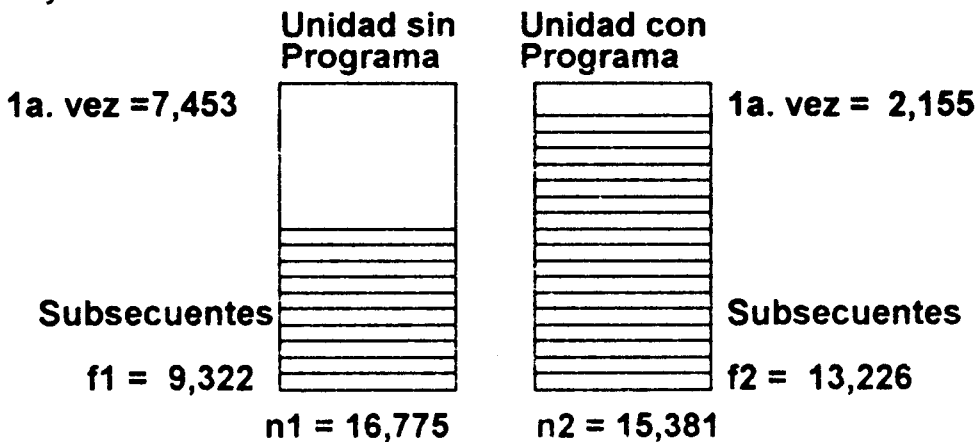
Un indicador de la capacidad de atracción y retención de pacientes en las unidades de salud es la proporción de consultas subsecuentes con relación al total.

En vista de que en dos unidades médicas se había apreciado que la proporción de consultas subsecuentes era muy baja, se decidió que en una de ellas se aplicaría un estricto programa de mejoría de la atención que incluía modificaciones en los procedimientos de recepción, consulta, archivo y visitas domiciliarias.

Al cabo de un año se encontró lo siguiente:

	Unidad Médica SIN programa	Unidad Médica CON programa
Consultas de primera vez	7,453	2,155
Consultas Subsecuentes	9,322	13,226

1o. Diseño y variables:



Variable independiente: Tipo de Unidad (Sin / Con): (cualitativa nominal con dos modalidades)
 Variable dependiente: Tipo de consulta (1a. / Subs): (cualitativa nominal con dos modalidades)

2o. Medidas de resumen (proporciones):

$$p_1 = \frac{f_1}{n_1} \quad p_1 = \frac{9,322}{16,775} = 0.556 \quad p_2 = \frac{f_2}{n_2} \quad p_2 = \frac{13,226}{15,381} = 0.860$$

$$q_1 = 1 - p_1 \quad q_1 = 1 - 0.556 = 0.444 \quad q_2 = 1 - p_2 \quad q_2 = 1 - 0.860 = 0.140$$

3o. Verificación de las cuatro siguientes condiciones

$$n_1 p_1 \geq 5 \quad (16,775)(0.556) = 9,327 \quad n_1 q_1 \geq 5 \quad (16,775)(0.444) = 7,448$$

$$n_2 p_2 \geq 5 \quad (15,381)(0.860) = 13,228 \quad n_2 q_2 \geq 5 \quad (15,381)(0.140) = 2,153$$

4o. Planteamiento de hipótesis estadísticas:

$$H_0: p_1 = p_2$$

$$H_a: p_1 \neq p_2$$

5o. Realización de prueba de análisis estadístico

$$z_{\text{observado}} = \frac{p_1 - p_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{donde: } \hat{p} = \frac{f_1 + f_2}{n_1 + n_2} \quad \text{y} \quad \hat{q} = 1 - \hat{p}$$

$$\hat{p} = \frac{9,322 + 13,226}{16,775 + 15,381} = 0.70 \quad \hat{q} = 1 - 0.70 = 0.30$$

$$z_{\text{observado}} = \frac{0.556 - 0.860}{\sqrt{(0.70)(0.30)\left(\frac{1}{16,775} + \frac{1}{15,381}\right)}} = \frac{-0.304}{\sqrt{0.21(0.00006 + 0.00007)}} = \frac{-0.304}{\sqrt{0.00000007}} = \frac{-0.292}{0.00009} = -3,358.32$$

6o. Comparación de $z_{\text{observado}}$ con $z_{\text{crítico}}$ para verificar si se llega a una de las dos regiones de rechazo de la hipótesis estadística nula:

En la curva normal, al definir dos regiones extremas y simétricas que, en conjunto, valgan 0.05 se encuentra que los valores z valen -1.96 por el lado izquierdo y +1.96 por el lado derecho.

En vista de que el valor z calculado rebasa muy ampliamente, por el lado izquierdo de la curva, al valor crítico de -1.96, entonces se llega a la región izquierda de rechazo de la hipótesis nula y, por ende, se puede decir que ambas proporciones son diferentes con un nivel de significancia menor que 0.05

En términos del problema, se puede decir que *existe diferencia estadísticamente significativa entre las proporciones de consultas subsecuentes entre las dos unidades médicas.*

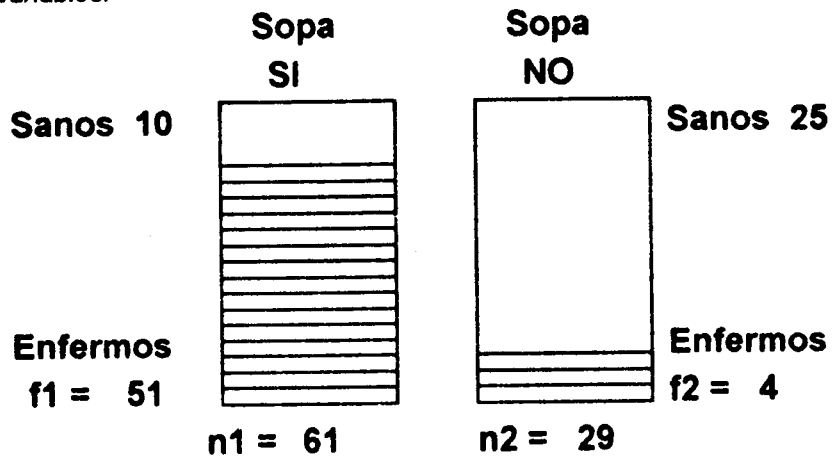
Problema 2

90 personas asistieron a una comida de bodas. El listado siguiente muestra los datos de cada uno de los invitados en relación con la ingestión o no de sopa y en relación con la presentación de diarrea. En cada caso la letra S significa SI y la letra N significa NO

Ver listado en hoja anexa >>>>>

Considerando como dos grupos a quienes ingirieron y a quienes no ingirieron sopa, verifique si existen diferencias significativas entre las proporciones de enfermos de diarrea en ambos grupos.

1o. Diseño y variables:



Variable independiente: Ingestión de sopa (Si / No):

(cualitativa nominal con dos modalidades)

Variable dependiente: Diarrea (Si / No):

(cualitativa nominal con dos modalidades)

2o. Medidas de resumen (proporciones):

$$p_1 = \frac{f_1}{n_1}$$

$$p_1 = \frac{51}{61} = 0.84$$

$$p_2 = \frac{f_2}{n_2}$$

$$p_2 = \frac{4}{29} = 0.14$$

$$q_1 = 1 - p_1$$

$$q_1 = 1 - 0.84 = 0.16$$

$$q_2 = 1 - p_2$$

$$q_2 = 1 - 0.14 = 0.86$$

3o. Verificación de las cuatro siguientes condiciones

$$n_1 p_1 \geq 5 \quad (61)(0.84) = 51$$

$$n_1 q_1 \geq 5 \quad (61)(0.16) = 10$$

$$n_2 p_2 \geq 5 \quad (29)(0.14) = 4$$

$$n_2 q_2 \geq 5 \quad (29)(0.86) = 25$$

Observación: En sentido estricto NO podría realizarse la prueba de análisis estadístico, pues una de la cuatro condiciones no está cubierta; sin embargo, con propósitos solamente ilustrativos se continuará el procedimiento no sin antes insistir en que las conclusiones podrían ser inválidas.

4o. Planteamiento de hipótesis estadísticas:

$$H_0: p_1 = p_2$$

$$H_a: p_1 \neq p_2$$

5o. Realización de prueba de análisis estadístico

$$z_{\text{observado}} = \frac{p_1 - p_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{donde: } \hat{p} = \frac{f_1 + f_2}{n_1 + n_2} \quad \text{y} \quad \hat{q} = 1 - \hat{p}$$

$$\hat{p} = \frac{51 + 4}{61 + 29} = 0.61 \quad \hat{q} = 1 - 0.61 = 0.39$$

$$z_{\text{observado}} = \frac{0.84 - 0.14}{\sqrt{(0.61)(0.39)\left(\frac{1}{61} + \frac{1}{29}\right)}} = \frac{0.70}{\sqrt{0.24(0.016 + 0.035)}} = \frac{0.70}{\sqrt{0.012}} = \frac{0.70}{0.11} = 6.36$$

6o. Comparación de $z_{\text{observado}}$ con $z_{\text{crítico}}$ para verificar si se llega a una de las dos regiones de rechazo de la hipótesis estadística nula:

En la curva normal, al definir dos regiones extremas y simétricas que, en conjunto, valgan 0.05 se encuentra que los valores z valen -1.96 por el lado izquierdo y +1.96 por el lado derecho.

En vista de que el valor z calculado rebasa muy ampliamente, por el lado derecho de la curva, al valor crítico de -1.96, entonces se llega a la región derecha de rechazo de la hipótesis nula y, por ende, se puede decir que ambas proporciones son diferentes con un nivel de significancia menor que 0.05

En términos del problema, se puede decir que, en caso de que se cumplan las condiciones para efectuar la prueba, existe diferencia estadísticamente significativa entre las proporciones de enfermos de diarrea entre quienes si comieron sopa y los que no la comieron.

Núm. de caso	Comió Sopa	Tuvo Diarrea
1	S	S
2	N	N
3	S	N
4	S	S
5	S	S
6	N	N
7	S	S
8	S	S
9	S	N
10	N	S
11	S	S
12	S	S
13	N	N
14	S	S
15	S	S
16	N	N
17	S	S
18	S	S
19	S	S
20	S	S
21	S	S
22	S	S
23	S	S
24	S	S
25	S	S
26	N	S
27	S	S
28	N	N
29	S	S
30	S	S

Núm. de caso	Comió Sopa	Tuvo Diarrea
31	N	N
32	S	S
33	S	S
34	S	S
35	S	S
36	S	S
37	N	N
38	S	N
39	S	N
40	S	S
41	N	N
42	N	N
43	S	S
44	N	N
45	N	N
46	S	N
47	N	N
48	S	S
49	S	N
50	N	N
51	N	N
52	N	N
53	N	N
54	S	S
55	N	N
56	N	N
57	S	N
58	S	S
59	N	N
60	S	N

Núm. de caso	Comió Sopa	Tuvo Diarrea
61	S	S
62	N	S
63	S	N
64	S	S
65	S	S
66	N	N
67	S	S
68	S	S
69	S	N
70	N	N
71	S	S
72	S	S
73	N	N
74	S	S
75	S	S
76	N	N
77	S	S
78	S	S
79	S	S
80	S	S
81	S	S
82	S	S
83	S	S
84	S	S
85	N	N
86	N	S
87	S	S
88	N	N
89	S	S
90	S	S

Texto:

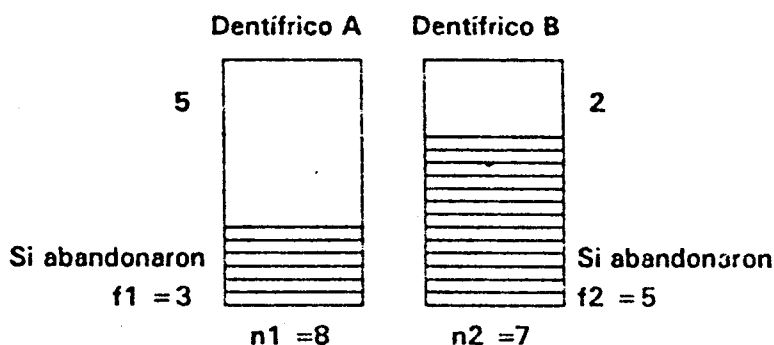
Comparación de dos grupos a través de sus proporciones mediante la Prueba de Fisher

Relato:

En un estudio sobre el uso de los dentífricos en los niños, interesaba saber si el sabor era un factor de influencia para el abandono del hábito del lavado dental. Durante tres meses se administró un dentífrico A de agradable y suave sabor a un grupo de ocho niños y un dentífrico B de picante sabor a menta a otro grupo de siete niños.

Luego de tres meses, se observó que habían abandonado el hábito de lavado dental tres niños del grupo A y cinco niños del grupo B.

Diseño y variables:



Variable independiente: Tipo de dentífrico (A / B): (cualitativa nominal con dos modalidades)

Variable dependiente: Abandono (si / no): (cualitativa nominal con dos modalidades)

Medidas de resumen:

Proporciones:

$$p_1 = \frac{f_1}{n_1} \quad p_1 = \frac{3}{8} = 0.375$$

$$p_2 = \frac{f_2}{n_2} \quad p_2 = \frac{5}{7} = 0.714$$

$$q_1 = 1 - p_1 \quad q_1 = 1 - 0.375 = 0.625$$

$$q_2 = 1 - p_2 \quad q_2 = 1 - 0.714 = 0.286$$

Verificación de las condiciones

$$\begin{array}{cccc}
 n_1 p_1 \geq 5 & n_1 q_1 \geq 5 & n_2 p_2 \geq 5 & n_2 q_2 \geq 5 \\
 (8)(0.375) = 3 & (8)(0.625) = 5 & (7)(0.714) = 4.99 & (7)(0.286) = 2.00
 \end{array}$$

Planteamiento de hipótesis estadísticas:

$$H_0: p_1 = p_2$$

$$H_a: p_1 \neq p_2$$

Realización de prueba de análisis estadístico

En vista de que NO se cubrieron tres de las cuatro condiciones, entonces NO se puede realizar la prueba de comparación de dos proporciones mediante la fórmula:

$$z_{\text{observado}} = \frac{p_1 - p_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{donde:} \quad \hat{p} = \frac{f_1 + f_2}{n_1 + n_2} \quad \text{y} \quad \hat{q} = 1 - \hat{p}$$

$$\hat{p} = \frac{3+5}{8+7} = 0.53 \quad \hat{q} = 1 - 0.53 = 0.47$$

Por lo anterior se usa la prueba de Fisher con el siguiente procedimiento:

Primero:

Ordenamiento de los datos y cálculo de los totales marginales en una tabla de contingencia como la del siguiente modelo:

A	B	(A + B)
C	D	(C + D)
(A + C)	(B + D)	N

	A	B	(A + B)
	5	2	7
	C	D	(C + D)
Si Abandonaron	3	5	8
	(A + C)	(B + D)	N
	8	7	15

Segundo:

En la tabla de valores críticos de la prueba de Fisher, localizar el valor correspondiente al de la celdilla B y a los totales marginales (A + B) y (C + D).

Una vez en el renglón correspondiente, identificar el valor crítico para la celdilla D; si el valor D calculado es menor o igual al de la tabla se puede rechazar la hipótesis estadística nula de igualdad de proporciones al nivel de significancia seleccionado en la tabla.

Ante la eventualidad de que en la tabla no se encuentre un valor B semejante al observado, entonces se puede usar el valor de la celdilla A, con lo cual el valor crítico será comparado con el observado en la celdilla C.

Ante la eventualidad de que en la tabla no exista una combinación de valores (A + B) y (C + D) semejantes a los observados, se deben reacomodar los renglones de los valores originales, tal como se muestra para el ejemplo:

	A	B	(A + B)
	3	5	8
	C	D	(C + D)
Si Abandonaron	5	2	7
	(A + C)	(B + D)	N
	8	7	15

Siguiendo con el ejemplo: al buscar en la tabla, en el renglón correspondiente a (A + B) = 8, (C + D) = 7 y B = 5, se encuentra un valor crítico para D de 0 al nivel de significancia de .05; cómo el valor observado es 2 (que es mayor que el valor crítico), entonces no se puede rechazar la hipótesis estadística nula de igualdad de proporciones; en términos del problema: parece ser que las proporciones de niños que abandonaron el hábito de lavado dental con uno u otro dentífrico son iguales.

Una tercera eventualidad que se puede presentar consiste en que alguno de los dos totales marginales ($A + B$) o ($C + D$) rebase el valor de 15 (que es el máximo que muestra la tabla de valores críticos para la prueba de Fisher), habrá de procederse también al reordenamiento de los datos tal como se muestra en el siguiente caso:

Sexo	Si Votó	No Votó	
Hombres	A	B	(A + B)
	12	4	16
Mujeres	C	D	(C + D)
	3	8	11
	(A + C)	(B + D)	N
	15	12	27

Al no existir un renglón de la tabla que sirva para totales marginales mayores de 15 (como en este caso ocurre con el total ($A + B$) que vale 16), los datos se reordenan de la siguiente forma:

Votó	Hombres	Mujeres	
Si	A	B	(A + B)
	12	3	15
No	C	D	(C + D)
	4	8	12
	(A + C)	(B + D)	N
	16	11	27

y se procede a la búsqueda del valor crítico para D o para C. En este caso, al no haber valor semejante a B se usa el de A que es 12. En la tabla se encuentra que el valor crítico para C es de 5 al nivel de significancia de .05; como en el ejemplo se aprecia que el valor observado de C es 4 y este es inferior al crítico, entonces se puede rechazar la hipótesis estadística nula de igualdad de proporciones.

Extracto de la tabla de
Valores Críticos en la
Prueba de Fisher
en siguiente página >>>

Extracto de la tabla de Valores Críticos en la Prueba de Fisher

*(Rechazar la hipótesis estadística nula si el valor
calculado es menor o igual al de la tabla)*

Totales del margen derecho		Valor Crítico para D (o para C en caso de haber usado a la celdilla A) a los niveles de significancia de .05 y de .01
-------------------------------	--	---

A + B	C + D	B (o A)	.05	.01
3	3	3	0	-
4	4	4	0	-
	3	4	0	-
5	5	5	1	0
	4	4	0	-
		5	1	0
		4	0	-
	3	5	0	-
2	5	0	-	

• • • • •
• • • • •

8	7	8	3	2
		7	2	1
		6	1	0
		5	0	-

• • • • •
• • • • •

15	12	15	8	7
		14	7	5
		13	6	4
		12	5	3
		11	4	2
		10	3	1
		9	2	1
		8	1	0
		7	1	0
		6	0	-
		5	0	-

Table of critical values of D (or C) in the Fisher test

Totals in right margin	B (or A)†	Level of significance			
		.05	.025	.01	.005
$A+B=3$ $C+D=3$	3	0	—	—	—
$A+B=4$ $C+D=4$ $C+D=3$	4	0	0	—	—
	4	0	—	—	—
$A+B=5$ $C+D=5$ $C+D=4$ $C+D=3$ $C+D=2$	5	1	1	0	0
	4	0	0	—	—
	5	1	0	0	—
	4	0	—	—	—
	5	0	0	—	—
	5	0	—	—	—
$A+B=6$ $C+D=6$ $C+D=5$ $C+D=4$ $C+D=3$ $C+D=2$	6	2	1	1	0
	5	1	0	0	—
	4	0	—	—	—
	6	1	0	0	0
	5	0	0	—	—
	4	0	—	—	—
	6	1	0	0	0
	5	0	0	—	—
	6	0	—	—	—
	6	0	—	—	—
$A+B=7$ $C+D=7$ $C+D=6$ $C+D=5$ $C+D=4$ $C+D=3$ $C+D=2$	7	3	2	1	1
	6	1	1	0	0
	5	0	0	—	—
	4	0	—	—	—
	7	2	2	1	1
	6	1	0	0	0
	5	0	0	—	—
	4	0	—	—	—
	7	2	1	0	0
	6	1	0	0	—
	5	0	—	—	—
	7	1	1	0	0
	6	0	0	—	—
	5	0	—	—	—
7	0	0	0	—	
6	0	—	—	—	
7	0	—	—	—	

† When B is entered in the middle column, the significance levels are for D . When A is used in place of B , the significance levels are for C .

Totals in right margin	B (or A)†	Level of significance			
		.05	.025	.01	.005
A+B=9 C+D=6	9 8 7 6 5	3 2 1 0 0	2 1 0 0 1	1 0 0 1 0	1 0 0 1 0
C+D=5	9 8 7 6 5	2 1 0 0 1	1 0 0 1 0	1 0 0 1 0	1 0 0 1 0
C+D=4	9 8 7 6 5	1 0 0 0 1	1 0 0 1 0	1 0 0 1 0	1 0 0 1 0
C+D=3	9 8 7 6 5	0 0 0 0 1	0 0 0 1 0	0 0 0 1 0	0 0 0 1 0
C+D=2	9 8 7 6 5	0 0 0 0 1	0 0 0 1 0	0 0 0 1 0	0 0 0 1 0
A+B=10 C+D=10	10 9 8 7 6 5 4	6 4 3 2 1 0 0	5 3 2 1 0 0 4	3 2 1 0 0 3 2	3 2 1 0 0 3 2
C+D=9	10 9 8 7 6 5	4 3 2 1 0 0 4	3 2 1 0 0 4 2	2 1 0 0 4 2 1	2 1 0 0 4 2 1
C+D=8	10 9 8 7 6 5	3 2 1 0 0 3 2	2 1 0 0 3 2 1	1 0 0 3 2 1 0	1 0 0 3 2 1 0
C+D=7	10 9 8 7 6 5	3 2 1 0 0 3 2	2 1 0 0 3 2 1	1 0 0 3 2 1 0	1 0 0 3 2 1 0

Totals in right margin	B (or A)†	Level of significance			
		.05	.025	.01	.005
A+B=8 C+D=8	8 7 6 5 4	4 2 1 0 0	3 2 1 0 2	2 1 0 2 1	2 0 0 2 1
C+D=7	8 7 6 5 4	3 2 1 0 0	2 1 0 0 2	2 1 0 2 1	1 0 0 2 1
C+D=6	8 7 6 5 4	2 1 0 0 0	2 1 0 0 2	1 0 0 2 1	1 0 0 2 1
C+D=5	8 7 6 5 4	2 1 0 0 0	1 0 0 0 2	1 0 0 2 1	0 0 0 2 1
C+D=4	8 7 6 5 4	1 0 0 0 0	1 0 0 0 2	0 0 0 2 1	0 0 0 2 1
C+D=3	8 7 6 5 4	0 0 0 0 0	0 0 0 0 2	0 0 0 2 1	0 0 0 2 1
C+D=2	8 7 6 5 4	0 0 0 0 0	0 0 0 0 2	0 0 0 2 1	0 0 0 2 1
A+B=9 C+D=9	9 8 7 6 5 4	5 3 2 1 0 0	4 3 1 1 0	3 2 1 0 3	3 1 0 0 3
C+D=8	9 8 7 6 5 4	4 3 2 1 0 0	3 2 1 0 0	3 1 0 0 3	2 1 0 0 3
C+D=7	9 8 7 6 5	3 2 1 0 0	3 2 1 0 0	2 1 0 0 3	2 0 0 0 3

7.13

Totals in right margin	B (or A)†	Level of significance			
		.05	.025	.01	.005
A+B=11 C+D=8	11 10 9 8 7 6 5	4 3 2 1 1 0 0	4 3 2 1 0 0	3 2 1 0 0	3 1 1 0 — — —
C+D=7	11 10 9 8 7	4 3 2 1 1 0 0	3 2 1 1 0 0	2 1 1 0	2 1 0 0
C+D=6	11 10 9 8 7 6	3 2 1 1 0 0 0	2 1 1 0 0	2 1 0 0	1 0 0
C+D=5	11 10 9 8 7 6	3 2 1 1 0 0 0	2 1 1 0 0	1 0 0	1 0 0
C+D=4	11 10 9 8 7	2 1 1 0 0 0	1 0 0	1 0 0	— — —
C+D=3	11 10 9 8	1 0 0 0 0	0 0 0	0	— — —
C+D=2	11 10	0 0	0	0	— —
A+B=12 C+D=12	12 11 10 9 8 7 6 5 4	8 6 5 4 3 2 1 0 0	7 5 4 3 2 1 0 0	6 4 3 2 1 0 0	5 4 2 1 1 0 — — —

Totals in right margin	B (or A)†	Level of significance			
		.05	.025	.01	.005
A+B=10 C+D=6	10 9 8 7 6 5	3 2 1 1 0 0	2 1 1 0	2 1 0	1 0 0
C+D=5	10 9 8 7 6	2 1 1 0 0	2 1 0	1 0	1 0
C+D=4	10 9 8 7	1 1 0 0	1 0	0	0
C+D=3	10 9 8	1 0 0	0	0	0
C+D=2	10 9	0 0	0	0	— —
A+B=11 C+D=11	11 10 9 8 7 6 5	7 5 4 3 2 1 0 0	6 4 3 2 1 0 0	5 3 2 1 0 0	4 3 2 1 0
C+D=10	11 10 9 8 7 6 5	6 4 3 2 1 1 0 0	5 4 3 2 1 0	4 3 2 1 0	4 2 1 0 0
C+D=9	11 10 9 8 7 6 5	5 4 3 2 1 1 0 0	4 3 2 1 1 0	4 2 1 1 0	3 2 1 0 0

Totals in right margin	B (or A)†	Level of significance			
		.05	.025	.01	.005
A+B=12 C+D=6	12 11 10 9 8 7 6 12 11 10 9 8 7	3 2 1 1 0 0 0 2 1 1 0 0 0	3 2 1 0 0 0 2 1 0 0 0 1 0 0 0	2 1 0 0 0 1 0 0 0 1 0 0 0 0	2 1 0 0 0 1 0 0 0 0 0 0 0 0
C+D=5	12 11 10 9 8 7 6 12 11 10 9 8 7	2 1 1 0 0 0 0 2 1 1 0 0 0	2 1 0 0 0 1 0 0 0 1 0 0 0 0	1 0 0 0 1 0 0 1 0 0 0 0 0 0	1 0 0 0 1 0 0 1 0 0 0 0 0 0
C+D=4	12 11 10 9 8 7 6 12 11 10 9 8 7	2 1 1 0 0 0 0 2 1 1 0 0 0	2 1 0 0 0 1 0 0 0 1 0 0 0 0	1 0 0 0 1 0 0 1 0 0 0 0 0 0	1 0 0 0 1 0 0 1 0 0 0 0 0 0
C+D=3	12 11 10 9 8 7 6 12 11 10 9 8 7	2 1 1 0 0 0 0 2 1 1 0 0 0	2 1 0 0 0 1 0 0 0 1 0 0 0 0	1 0 0 0 1 0 0 1 0 0 0 0 0 0	1 0 0 0 1 0 0 1 0 0 0 0 0 0
C+D=2	12 11 10 9 8 7 6 12 11 10 9 8 7	2 1 1 0 0 0 0 2 1 1 0 0 0	2 1 0 0 0 1 0 0 0 1 0 0 0 0	1 0 0 0 1 0 0 1 0 0 0 0 0 0	1 0 0 0 1 0 0 1 0 0 0 0 0 0
A+B=13 C+D=13	13 12 11 10 9 8 7 6 5 4 13 12 11 10 9 8 7 6 5	9 7 6 4 3 2 2 1 0 0 8 6 5 4 3 2 1 1 0 0 8 6 5 4 3 2 1 1 0 0	9 7 6 4 3 2 2 1 0 0 8 6 5 4 3 2 1 1 0 0 8 6 5 4 3 2 1 1 0 0	6 4 3 2 1 0 0 5 4 3 2 1 0 0 5 4 3 2 1 0 0 5 4 3 2 1 0 0 5 4 3 2 1 0 0	6 4 3 2 1 0 0 5 4 3 2 1 0 0 5 4 3 2 1 0 0 5 4 3 2 1 0 0 5 4 3 2 1 0 0
C+D=12	13 12 11 10 9 8 7 6 5 4 13 12 11 10 9 8 7 6 5	13 12 11 10 9 8 7 6 5 4 13 12 11 10 9 8 7 6 5	13 12 11 10 9 8 7 6 5 4 13 12 11 10 9 8 7 6 5	13 12 11 10 9 8 7 6 5 4 13 12 11 10 9 8 7 6 5	13 12 11 10 9 8 7 6 5 4 13 12 11 10 9 8 7 6 5

Totals in right margin	B (or A)†	Level of significance			
		.05	.025	.01	.005
A+B=12 C+D=11	12 11 10 9 8 7 6 5 12 11 10 9 8 7 6 5	7 5 4 3 2 1 1 0 0 6 5 4 3 2 1 0 0 6 5 4 3 2 1 0 0 6 5 4 3 2 1 0 0	6 5 3 2 1 1 0 0 5 4 3 2 1 0 0 5 4 3 2 1 0 0 5 4 3 2 1 0 0 5 4 3 2 1 0 0	5 4 2 2 1 0 0 4 3 2 1 0 0 4 3 2 1 0 0 4 3 2 1 0 0 4 3 2 1 0 0 4 3 2 1 0 0	5 3 2 1 0 0 4 3 2 1 0 0 4 3 2 1 0 0 4 3 2 1 0 0 4 3 2 1 0 0 4 3 2 1 0 0 4 3 2 1 0 0
C+D=10	12 11 10 9 8 7 6 5 12 11 10 9 8 7 6 5	6 5 4 3 2 1 1 0 0 5 4 3 2 1 0 0 5 4 3 2 1 0 0 5 4 3 2 1 0 0 5 4 3 2 1 0 0	6 5 3 2 1 1 0 0 5 4 3 2 1 0 0 5 4 3 2 1 0 0 5 4 3 2 1 0 0 5 4 3 2 1 0 0	5 4 2 2 1 0 0 4 3 2 1 0 0 4 3 2 1 0 0 4 3 2 1 0 0 4 3 2 1 0 0 4 3 2 1 0 0	5 3 2 1 0 0 4 3 2 1 0 0 4 3 2 1 0 0 4 3 2 1 0 0 4 3 2 1 0 0 4 3 2 1 0 0 4 3 2 1 0 0
C+D=9	12 11 10 9 8 7 6 5 12 11 10 9 8 7 6 5	5 4 3 2 1 1 0 0 4 3 2 1 0 0 4 3 2 1 0 0 4 3 2 1 0 0 4 3 2 1 0 0 4 3 2 1 0 0	5 4 3 2 1 1 0 0 4 3 2 1 0 0 4 3 2 1 0 0 4 3 2 1 0 0 4 3 2 1 0 0 4 3 2 1 0 0	4 3 2 1 0 0 3 2 1 0 0 3 2 1 0 0 3 2 1 0 0 3 2 1 0 0 3 2 1 0 0 3 2 1 0 0	4 3 2 1 0 0 3 2 1 0 0 3 2 1 0 0 3 2 1 0 0 3 2 1 0 0 3 2 1 0 0 3 2 1 0 0 3 2 1 0 0
C+D=8	12 11 10 9 8 7 6 5 12 11 10 9 8 7 6 5	4 3 2 2 1 1 0 0 3 2 2 1 0 0 3 2 2 1 0 0 3 2 2 1 0 0 3 2 2 1 0 0 3 2 2 1 0 0	4 3 2 2 1 1 0 0 3 2 2 1 0 0 3 2 2 1 0 0 3 2 2 1 0 0 3 2 2 1 0 0 3 2 2 1 0 0	3 2 1 1 0 0 2 1 1 0 0 2 1 1 0 0 2 1 1 0 0 2 1 1 0 0 2 1 1 0 0 2 1 1 0 0	3 2 1 0 0 2 1 0 0 2 1 0 0 2 1 0 0 2 1 0 0 2 1 0 0 2 1 0 0 2 1 0 0 2 1 0 0 2 1 0 0
C+D=7	12 11 10 9 8 7 6 5 12 11 10 9 8 7 6 5	3 2 2 2 1 1 0 0 2 1 2 1 0 0 2 1 2 1 0 0 2 1 2 1 0 0 2 1 2 1 0 0 2 1 2 1 0 0	3 2 2 2 1 1 0 0 2 1 2 1 0 0 2 1 2 1 0 0 2 1 2 1 0 0 2 1 2 1 0 0 2 1 2 1 0 0	2 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0	2 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0

Totals in right margin	B (or A)†	Level of significance				
		.05	.025	.01	.005	
A+B=13 C+D=6	13	3	3	2	2	2
	12	2	2	1	1	1
	11	2	1	1	0	0
	10	1	1	0	0	0
	9	1	0	0	0	0
	8	0	0	0	0	0
	7	0	0	0	0	0
	13	2	2	1	1	1
	12	2	1	1	0	0
	11	1	1	0	0	0
	10	1	0	0	0	0
	9	0	0	0	0	0
	8	0	0	0	0	0
A+B=14 C+D=14	14	10	9	8	7	7
	13	8	7	6	5	5
	12	6	6	5	4	4
	11	5	4	3	3	3
	10	4	3	2	2	2
	9	3	2	2	1	1
	8	2	2	1	1	1
	7	1	1	0	0	0
	6	1	0	0	0	0
	5	0	0	0	0	0
	4	0	0	0	0	0

Totals in right margin	B (or A)†	Level of significance				
		.05	.025	.01	.005	
A+B=13 C+D=11	13	7	6	5	5	5
	12	6	5	4	4	3
	11	4	4	3	3	2
	10	3	3	2	2	1
	9	3	2	1	1	1
	8	2	1	0	0	0
	7	1	0	0	0	0
	6	0	0	0	0	0
	5	0	0	0	0	0
	13	6	6	5	4	4
	12	5	4	3	3	3
	11	4	3	2	2	2
	10	3	2	1	1	1
9	2	1	1	0	0	
8	1	1	0	0	0	
7	1	0	0	0	0	
6	0	0	0	0	0	
5	0	0	0	0	0	
13	5	5	4	3	3	
12	4	4	3	2	2	
11	3	3	2	1	1	
10	2	2	1	1	0	
9	2	1	1	0	0	
8	1	1	0	0	0	
7	0	0	0	0	0	
6	0	0	0	0	0	
5	0	0	0	0	0	
13	4	4	3	3	3	
12	3	3	2	2	2	
11	2	2	1	1	1	
10	2	1	1	0	0	
9	1	1	0	0	0	
8	1	0	0	0	0	
7	0	0	0	0	0	
6	0	0	0	0	0	
5	0	0	0	0	0	
13	4	4	3	3	3	
12	3	3	2	2	2	
11	2	2	1	1	1	
10	2	1	1	0	0	
9	1	1	0	0	0	
8	1	0	0	0	0	
7	0	0	0	0	0	
6	0	0	0	0	0	
5	0	0	0	0	0	

Totals in right margin	B (or A)†	Level of significance			
		.05	.025	.01	.005
A = B = 14 C + D = 9	14 13 12 11 10 9 8 7 6	6 4 3 3 2 1 1 0 0	5 4 3 2 1 1 0 0	4 3 2 1 1 0 0	4 3 2 1 1 0 0
C + D = 8	14 13 12 11 10 9 8 7 6	5 4 3 2 2 1 0 0 0	4 3 2 2 1 0 0 0	4 2 2 1 1 0 0	3 2 1 1 0 0
C + D = 7	14 13 12 11 10 9 8 7 6	4 3 2 2 1 0 0 0 0	3 2 2 1 1 0 0	3 2 1 1 0 0	2 1 1 0 0
C + D = 6	14 13 12 11 10 9 8 7	3 2 2 1 1 0 0 0	3 2 1 1 0 0 0	2 1 1 0 0	2 1 0 0
C + D = 5	14 13 12 11 10 9 8	2 2 1 1 0 0 0 0	2 1 1 0 0 0	1 1 0 0	1 0 0 0

Totals in right margin	B (or A)†	Level of significance			
		.05	.025	.01	.005
A + B = 14 C + D = 13	14 13 12 11 10 9 8 7 6 5	9 7 6 5 4 3 2 1 1 0 0	8 6 5 4 3 2 1 1 0 0	7 5 4 3 2 1 1 0	6 5 3 2 2 1 0 0
C + D = 12	14 13 12 11 10 9 8 7 6 5	8 6 5 4 3 3 2 2 1 0 0	7 6 4 3 3 2 1 0 0	6 5 4 3 2 1 0 0	6 4 3 2 1 1 0
C + D = 11	14 13 12 11 10 9 8 7 6 5	7 6 5 4 3 3 2 1 0 0 0	6 5 4 3 2 1 1 0 0	6 4 3 2 1 1 0 0	5 4 3 2 1 0 0
C + D = 10	14 13 12 11 10 9 8 7 6 5	6 5 4 3 3 2 1 1 0 0 0	6 4 3 3 2 1 1 0 0	5 4 3 2 1 0 0 0	4 3 2 1 1 0 0

Totals in right margin	B (or A)†	Level of significance				
		.05	.025	.01	.005	
A+B=15 C+D=13	15 14 13 12 11 10 9 8 7 6 5	9 7 6 5 4 3 2 2 1 0 0	8 7 5 4 3 2 2 1 0 0	7 6 4 3 2 2 1 0 0	7 5 4 3 2 2 1 0 0	7 5 4 3 2 2 1 0 0
C+D=12	15 14 13 12 11 10 9 8 7 6 5	8 7 6 5 4 3 2 2 1 0 0	7 6 5 4 3 2 2 1 0 0	7 5 4 3 2 2 1 0 0	7 5 4 3 2 2 1 0 0	6 4 3 2 2 2 1 0 0
C+D=11	15 14 13 12 11 10 9 8 7 6 5	7 6 5 4 3 2 2 1 0 0	7 5 4 3 2 2 1 0 0	7 6 4 3 2 2 1 0 0	7 5 4 3 2 2 1 0 0	5 4 3 2 2 1 1 0 0
C+D=10	15 14 13 12 11 10 9 8 7 6 5	6 5 4 3 3 2 2 1 1 0 0	6 5 4 3 3 2 2 1 1 0 0	6 5 4 3 3 2 2 1 1 0 0	6 5 4 3 3 2 2 1 1 0 0	5 3 2 2 2 1 1 0 0

Totals in right margin	B (or A)†	Level of significance				
		.05	.025	.01	.005	
A+B=14 C+D=4	14 13 12 11 10 9	2 1 1 0 0 0	1 1 0 0 0 0	1 0 0 — — —	1 0 0 — — —	1 0 0 — — —
C+D=3	14 13 12 11 10	1 0 0 0 0	1 0 0 — —	0 0 — —	0 — — —	0 — — —
C+D=2	14 13 12	0 0 0	0 0 —	0 — —	0 — —	— — —
A+B=15 C+D=15	15 14 13 12 11 10 9 8 7 6 5	11 9 7 6 5 4 3 2 1 1 0 0	10 8 6 5 4 3 2 1 1 0 0	9 7 5 4 3 2 1 1 0 0	9 7 5 4 3 2 1 1 0 0	8 6 5 4 3 2 1 1 0 0
C+D=14	15 14 13 12 11 10 9 8 7 6 5	10 8 7 6 5 4 3 2 1 1 0	9 7 6 5 4 3 2 1 1 0	8 6 5 4 3 2 1 1 0	8 6 5 4 3 2 1 1 0	7 6 4 3 2 1 1 0

Problemas: **Diseño:** Comparación de dos grupos independientes

Prueba: Comparación de dos promedios independientes a través de la distribución t de Student

En un estudio multicéntrico interesaba averiguar si las situaciones de stress producían aumento de peso de las glándulas suprarrenales.

Investigadores de dos diferentes institutos sometieron en cada sitio a un grupo experimental de ratas a situaciones de tensión y, luego de sacrificarlas, pesaron sus glándula suprarrenales.

Para comparar, también hicieron lo mismo con grupos control de ratas no sometidas a situaciones de stress.

A continuación se muestran sus resultados:

Instituto A

Ratas del grupo experimental	Ratas del grupo control
3.8	4.2
6.8	4.8
8.0	4.8
3.6	2.3
6.1	6.5
4.5	4.9
3.9	3.6
5.9	2.4
6.0	3.2
5.7	4.9
5.6	
4.5	

Instituto B

Ratas del grupo experimental	Ratas del grupo control
3.2	4.3
7.2	4.8
7.9	4.8
3.6	4.3
3.9	4.5
4.5	4.9
3.9	3.6
5.9	3.7
6.0	3.8
5.7	4.4
5.6	4.3
4.5	4.2
5.2	

- Calcule las medidas de resumen pertinentes para describir estadísticamente a cada una de las cuatro series de valores.
- Una vez que haya verificado si es válido describir a las cuatro series a través de promedios, efectue la o las pruebas de análisis estadístico pertinentes para verificar si las situaciones de stress hacen que aumente el peso de las glándulas suprarrenales en las ratas con los datos de uno, otro o ambos institutos.

Texto: **Diseño:** Comparación de dos grupos independientes

Prueba: Comparación de dos promedios independientes a través de la distribución *t* de Student

Jesús Reynaga Obregón

INTRODUCCION:

Uno de los diseños de investigación más extensamente usado es el que consiste en la comparación de dos grupos independientes en los que se ha medido una variable dependiente de tipo cuantitativo continuo.

En este tipo de diseño existe una variable independiente, o antecedente, de tipo cualitativo nominal con dos modalidades; ambas modalidades son las que definen la existencia de dos grupos independientes.

Cuando se desea efectuar dicha comparación suele anticiparse que la prueba de análisis estadístico pudiera consistir en la comparación del promedio de dicha variable en un grupo con el promedio de la misma variable en el otro grupo para averiguar si existen diferencias estadísticamente significativas.

Obsérvense en la siguiente tabla dos conjuntos de datos independientes que servirán de ejemplo para ilustrar el procedimiento de análisis estadístico.

**Tiempo de vaciamiento gástrico (minutos y fracciones)
en dos grupos de pacientes
sometidos a diferente anticolinérgico cada uno**

Anticolinérgico A	Anticolinérgico B
38.25	42.75
68.75	48.50
80.25	48.25
36.50	23.25
61.25	65.50
45.75	49.75
39.75	36.75
59.50	24.50
60.50	32.25
57.25	49.25
56.50	
45.75	

Promedio	54.17	42.01
Desviación Estándar	13.29	13.03

PRIMER PASO, IDENTIFICACION DEL DISEÑO:

Existe una variable independiente de tipo cualitativo nominal con dos modalidades: *Tipo de anticolinérgico*; el hecho de existir dos modalidades para dicha variable antecedente origina que existan los dos grupos diferentes, ajenos o independientes.

Existe una variable dependiente de tipo cuantitativo continuo: *Tiempo de vaciamiento gástrico*.

SEGUNDO PASO, VERIFICACION DE QUE SE CUMPLEN LAS CONDICIONES DE LA PRUEBA DE ANALISIS ESTADISTICO:**Primera condición, semejanza a la distribución normal:**

En vista de que la prueba que se está ilustrando consiste en averiguar si existen diferencias estadísticamente significativas entre los dos promedios de la variable dependiente, medida en ambos grupos, se debe comprobar si los valores de la variable cuantitativa continua siguen una distribución semejante a la de la curva normal.

A continuación se muestran los valores de sesgo y curtosis, calculados a través del método de momentos, para los tiempos de vaciamiento gástrico en ambos grupos:

Medida de resumen	Anticolinérgico A	Anticolinérgico B
Sesgo	0.32	0.05
Curtosis	2.32	2.32

Si se adoptan los criterios de que un sesgo que se encuentre en el intervalo que va desde -0.5 hasta +0.5 y de que una curtosis que se encuentre en el intervalo que va desde 2 hasta 4 expresan semejanza con la curva normal, entonces puede continuarse con el procedimiento que se describe a continuación.

Segunda condición, homoscedasticidad (homogeneidad de varianzas):

Existen dos fórmulas a utilizar para la comprobación de semejanza o diferencia entre los promedios de uno y otro grupo:

la primera se utiliza cuando se ha verificado que las varianzas de la variable son semejantes en ambos grupos (*ver prueba Fmax de homoscedasticidad*):

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left[\frac{\left(\sum x_1^2 - \frac{(\sum x_1)^2}{n_1} \right) + \left(\sum x_2^2 - \frac{(\sum x_2)^2}{n_2} \right)}{n_1 + n_2 - 2} \right] \left(\frac{n_1 + n_2}{n_1 n_2} \right)}}$$

la segunda se utiliza cuando se ha comprobado que las varianzas de la variable son diferentes en ambos grupos (*ver prueba Fmax de homoscedasticidad*):

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}}$$

Así pues, para verificar la condición de homogeneidad de varianzas u homoscedasticidad se utiliza la prueba Fmax que se ilustra a continuación:

$$\text{para los datos el valor Fmax calculado es } = \frac{\text{varianza mayor}}{\text{varianza menor}} = \frac{(s_1)^2}{(s_2)^2} = \frac{(13.29)^2}{(13.03)^2} = 1.04$$

El valor Fmax crítico de la tabla de la distribución F para un nivel de significancia de 0.05, con $12 - 1 = 11$ grados de libertad para la varianza mayor y con $10 - 1 = 9$ grados de libertad para la varianza menor es aproximadamente de 3.13

En vista de que el valor calculado para Fmax es menor que el valor crítico de la tabla, entonces no se puede rechazar la hipótesis estadística nula de que ambas varianzas son iguales. Lo anterior quiere decir que sí existe homoscedasticidad.

Por lo anterior, en el caso de los tiempos de vaciamiento gástrico, la prueba de análisis estadístico requiere el uso de la fórmula para el caso de varianzas homogéneas:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left[\frac{\left(\sum x_1^2 - \frac{(\sum x_1)^2}{n_1} \right) + \left(\sum x_2^2 - \frac{(\sum x_2)^2}{n_2} \right)}{n_1 + n_2 - 2} \right] \left(\frac{n_1 + n_2}{n_1 n_2} \right)}}$$

TERCER PASO, PLANTEAMIENTO DE HIPOTESIS ESTADISTICAS:

Hipótesis estadística nula: $H_0: \bar{x}_1 = \bar{x}_2$

Hipótesis estadística alterna: $H_0: \bar{x}_1 \neq \bar{x}_2$

CUARTO PASO, CALCULO DE t OBSERVADA:

El cálculo de la fórmula se facilita con el empleo de una tabla auxiliar de trabajo como la siguiente:

Anticolinérgico A		Anticolinérgico B	
x_1	x_1^2	x_2	x_2^2
38.25	1463.06	42.75	1827.56
68.75	4726.56	48.50	2352.25
80.25	6440.06	48.25	2328.06
36.50	1332.25	23.25	540.56
61.25	3751.56	65.50	4290.25
45.75	2093.06	49.75	2475.06
39.75	1580.06	36.75	1350.56
59.50	3540.25	24.50	600.25
60.50	3660.25	32.25	1040.06
57.25	3277.56	49.25	2425.56
56.50	3192.25		
45.75	2093.06		
650.00	37,150.00	420.75	19,230.19

sustituyendo los valores en la fórmula se tiene:

$$t = \frac{54.17 - 42.01}{\sqrt{\left[\frac{\left(\frac{37,150.00 - \frac{(650.00)^2}{12}}{12+10-2} \right) + \left(\frac{19,230.19 - \frac{(420.75)^2}{10}}{12+10-2} \right)}{\left(\frac{12+10}{(12)(10)} \right)} \right]}$$

$$t = \frac{12.16}{\sqrt{\left[\left(\frac{(1,941.17) + (1527.13)}{20} \right) (0.183) \right]}} = \frac{12.16}{\sqrt{31.73}} = 2.16$$

QUINTO PASO. COMPARACION DEL VALOR t OBSERVADO CON UN VALOR CRITICO TABULAR Y EVALUACION DE LAS HIPOTESIS ESTADISTICAS:

El valor t calculado se compara con un valor t crítico en la tabla de la distribución t de Student. Dicho valor crítico se localiza en el cruce de una columna de nivel de significancia (por ejemplo 0.05) con un renglón que corresponda a los grados de libertad de los datos. Los grados de libertad para el caso de datos procedentes de dos grupos homoscedásticos se calculan mediante la fórmula $n_1 + n_2 - 2$ que en este caso equivale a $12 + 10 - 2 = 20$.

En la tabla de la distribución t se encuentra que para 20 grados de libertad y un nivel de significancia de 0.05 el valor crítico que debe rebasarse para poder rechazar la hipótesis estadística nula es de 2.086; en vista de que el valor observado de 2.16 rebasa al valor crítico tabular de 2.086, puede entonces rechazarse la $H_0: \bar{x}_1 = \bar{x}_2$ para el nivel de significancia de 0.05

Cabe señalar aquí que si las varianzas de los dos grupos hubieran resultado no homogéneas, la fórmula para el cálculo de t observada hubiera requerido el empleo de la fórmula:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}}$$

Adicionalmente, el procedimiento para el cálculo de los grados de libertad sería de la siguiente manera: para cada grupo se utilizaría la fórmula $n - 1$ de manera individual, siempre con un mismo nivel de significancia (por ejemplo 0.05), y se tendría que obtener un promedio de los dos valores tabulares de t . Tal promedio se utiliza como el valor t crítico que se contrasta con el valor t observado

SEXTO PASO. CONCLUSION EN TERMINOS DEL PROBLEMA DE INVESTIGACION:

Al menos para los dos grupos estudiados, puede considerarse que el tiempo de vaciamiento gástrico es diferente en función del tipo de anticolinérgico suministrado.

t Distribution

d.f.	Level of significance for two-tailed test			
	0.10	0.05	0.02	0.01
	1	6.314	12.706	31.821
2	2.920	4.303	6.965	9.925
3	2.353	3.182	4.541	5.841
4	2.132	2.776	3.747	4.604
5	2.015	2.571	3.365	4.032
6	1.943	2.447	3.143	3.707
7	1.895	2.365	2.998	3.499
8	1.860	2.306	2.896	3.355
9	1.833	2.262	2.821	3.250
10	1.812	2.228	2.764	3.169
11	1.796	2.201	2.718	3.106
12	1.782	2.179	2.681	3.055
13	1.771	2.160	2.650	3.012
14	1.761	2.145	2.624	2.977
15	1.753	2.131	2.602	2.947
16	1.746	2.120	2.583	2.921
17	1.740	2.110	2.567	2.898
18	1.734	2.101	2.552	2.878
19	1.729	2.093	2.539	2.861
20	1.725	2.086	2.528	2.845
21	1.721	2.080	2.518	2.831
22	1.717	2.074	2.508	2.819
23	1.714	2.069	2.500	2.807
24	1.711	2.064	2.492	2.797
25	1.708	2.060	2.485	2.787
26	1.706	2.056	2.479	2.779
27	1.703	2.052	2.473	2.771
28	1.701	2.048	2.467	2.763
29	1.699	2.045	2.462	2.756
30	1.697	2.042	2.457	2.750
40	1.684	2.021	2.423	2.704
60	1.671	2.000	2.390	2.660
120	1.658	1.980	2.358	2.617
∞	1.645	1.960	2.326	2.576

Texto: **Diseño:** Comparación de dos grupos apareados o correlacionados

Prueba: Comparación de dos promedios apareados a través de la distribución *t* de Student

Jesús Reynaga Obregón

INTRODUCCION:

Un diseño de investigación muy ampliamente usado es el que consiste en la comparación de dos grupos de mediciones efectuadas a un solo conjunto de individuos en los que se ha medido una variable dependiente de tipo cuantitativo continuo.

En este tipo de diseño existe una variable independiente, o antecedente, de tipo cualitativo nominal con dos modalidades; ambas modalidades son las que definen la existencia de los dos grupos de valores que, por pertenecer a un solo conjunto de individuos, se denominan grupos apareados o correlacionados.

Tal como será descrita a continuación, la prueba de análisis estadístico, puede efectuarse siempre y cuando la variable dependiente de tipo cuantitativo continuo tenga una distribución semejante a la de la curva normal.

En esencia, la prueba consiste en la comparación del promedio de dicha variable en un grupo de valores con el promedio de la misma variable en el otro grupo de valores para averiguar si existen diferencias estadísticamente significativas entre dichos promedios que, por pertenecer a un solo conjunto de individuos, se denominan promedios apareados o correlacionados.

Obsérvense en la siguiente tabla dos grupos de datos correlacionados que servirán de ejemplo para ilustrar el procedimiento de análisis estadístico.

Concentraciones de glucosa en en sangre en un grupo de diabéticos
antes y después de un programa de dieta y ejercicio
(mgrs / 100 ml)

Paciente	Antes	Después
ART	152	97
PHG	125	99
FVA	142	72
GML	101	110
MELZ	180	124
Promedio	140.0	100.4
Desviación Estándar	29.6	19.2

PRIMER PASO, IDENTIFICACION DEL DISEÑO:

Existe una variable independiente o antecedente de tipo cualitativo nominal con dos modalidades: *Tiempo en relación con el programa dieta-ejercicio*; el hecho de existir dos modalidades, antes y después, para dicha variable antecedente origina que existan los dos grupos de valores apareados o correlacionados.

Existe una variable dependiente de tipo cuantitativo continuo: *Concentración de glucosa en sangre*.

SEGUNDO PASO, VERIFICACION DE QUE SE CUMPLEN LAS CONDICIONES DE LA PRUEBA DE ANALISIS ESTADISTICO:**Condición de semejanza a la distribución normal:**

En vista de que la prueba que se está ilustrando consiste en averiguar si existen diferencias estadísticamente significativas entre los dos promedios apareados de la variable dependiente, se debe comprobar si los valores de tal variable cuantitativa continua siguen una distribución semejante a la de la curva normal.

A continuación se muestran los valores de sesgo y curtosis, calculados a través del método de momentos, para las concentraciones de glucosa en sangre:

Medida de resumen	Antes	Después
Sesgo	0.03	- 0.35
Curtosis	2.03	2.25

Si se adoptan los criterios de que un sesgo que se encuentre en el intervalo que va desde -0.5 hasta +0.5 y de que una curtosis que se encuentre en el intervalo que va desde 2 hasta 4 expresan semejanza con la curva normal, entonces puede continuarse con el procedimiento que se describe a continuación.

TERCER PASO, PLANTEAMIENTO DE HIPOTESIS ESTADISTICAS:

Hipótesis estadística nula: $H_0: \bar{x}_{\text{antes}} = \bar{x}_{\text{después}}$

Hipótesis estadística alterna: $H_a: \bar{x}_{\text{antes}} \neq \bar{x}_{\text{después}}$

CUARTO PASO, CALCULO DE t OBSERVADA:

El cálculo del valor observado para t se efectúa con la siguiente fórmula:

$$t = \frac{\sum d}{\sqrt{\frac{n \sum d^2 - (\sum d)^2}{n-1}}}$$

El cálculo de la fórmula se facilita con el empleo de una tabla auxiliar de trabajo como la siguiente:

Paciente	Antes	Después	Diferencias d	Diferencias cuadráticas d^2
ART	152	97	55	3,025
PHG	125	99	26	676
FVA	142	72	70	4,900
GML	101	110	- 9	81
MELZ	180	124	56	3,136
Sumatorias			198	11,818

sustituyendo los valores en la fórmula se tiene:

$$t = \frac{\sum d}{\sqrt{\frac{n \sum d^2 - (\sum d)^2}{n-1}}} \quad t = \frac{198}{\sqrt{\frac{[(5)(11,818)] - (198)^2}{5-1}}} = \frac{198}{\sqrt{\frac{19,886}{4}}} = \frac{198}{70.51} = 2.8$$

QUINTO PASO. COMPARACION DEL VALOR t OBSERVADO CON UN VALOR CRITICO TABULAR Y EVALUACION DE LAS HIPOTESIS ESTADISTICAS:

El valor t calculado se compara con un valor t crítico en la tabla de la distribución t de Student (ver tabla en hoja anexa).

Dicho valor crítico se localiza en el cruce de una columna de nivel de significancia (por ejemplo 0.05) con un renglón que corresponda a los grados de libertad de los datos.

Los grados de libertad para el caso de datos procedentes de dos grupos apareados de datos se calculan mediante la fórmula $n - 1$, que en este caso equivale a $5 - 1 = 4$

En la tabla de la distribución t se encuentra que para 4 grados de libertad y un nivel de significancia de 0.05 el valor crítico que debe rebasarse para poder rechazar la hipótesis estadística nula es de 2.776

En vista de que el valor observado de 2.81 rebasa al valor crítico tabular de 2.776, puede entonces rechazarse la $H_0: \bar{x}_{\text{antes}} = \bar{x}_{\text{después}}$ para el nivel de significancia de 0.05

SEXTO PASO. CONCLUSION EN TERMINOS DEL PROBLEMA DE INVESTIGACION:

Al menos para los dos grupos de valores estudiados, puede considerarse que la concentración de glucosa en sangre es diferente antes y después del programa de dieta - ejercicio.

Valores Críticos para la distribución t de Student

g.l.	Nivel de significancia a dos colas			
	0.10	0.05	0.02	0.01
1	6.314	12.706	31.821	63.657
2	2.920	4.303	6.965	9.925
3	2.353	3.182	4.541	5.841
4	2.132	2.776	3.747	4.604
5	2.015	2.571	3.365	4.032
6	1.943	2.447	3.143	3.707
7	1.895	2.365	2.998	3.499
8	1.860	2.306	2.896	3.355
9	1.833	2.262	2.821	3.250
10	1.812	2.228	2.764	3.169
11	1.796	2.201	2.718	3.106
12	1.782	2.179	2.681	3.055
13	1.771	2.160	2.650	3.012
14	1.761	2.145	2.624	2.977
15	1.753	2.131	2.602	2.947
16	1.746	2.120	2.583	2.921
17	1.740	2.110	2.567	2.898
18	1.734	2.101	2.552	2.878
19	1.729	2.093	2.539	2.861
20	1.725	2.086	2.528	2.845
21	1.721	2.080	2.518	2.831
22	1.717	2.074	2.508	2.819
23	1.714	2.069	2.500	2.807
24	1.711	2.064	2.492	2.797
25	1.708	2.060	2.485	2.787
26	1.706	2.056	2.479	2.779
27	1.703	2.052	2.473	2.771
28	1.701	2.048	2.467	2.763
29	1.699	2.045	2.462	2.756
30	1.697	2.042	2.457	2.750
40	1.684	2.021	2.423	2.704
60	1.671	2.000	2.390	2.660
120	1.658	1.980	2.358	2.617
α	1.645	1.960	2.326	2.576

Problemas:**Diseño: Comparación de dos grupos apareados****Prueba: Comparación de dos promedios correlacionados a través de la distribución t de Student**

RELATO

Se sabe que la exploración física de los pacientes ha de ser minuciosa y que debe tener una duración suficiente para permitir el seguimiento preciso del padecimiento. También se sabe que las actividades hospitalarias adoptan un ritmo que se caracteriza por una dinámica acelerada por las mañanas y una mayor tranquilidad por las tardes y noches.

En un hospital se ignoraba si la duración de la exploración física que los médicos efectuaban a sus pacientes estaba influida por el horario en que la efectuaban.

Se suponía que la duración de la exploración física de los pacientes podría tener una duración diferente según el turno en que se realizaba.

Si tal suposición fuera cierta, entonces un mismo médico, explorando a su mismo paciente en diversos turnos, tardaría diferente tiempo en explorarlo.

Al azar fueron observados siete médicos mientras efectuaban la exploración a sus pacientes en los turnos diurno y nocturno. Aunque los observadores sabían que el tiempo es una variable cuantitativa continua no pudieron encontrar un cronómetro y se conformaron con usar un reloj común. A continuación se muestra la duración de la exploración física que los siete médicos efectuaron.

Duración (minutos) de la exploración física efectuada por siete médicos según turno

Médico	Turno diurno	Turno nocturno
1	27	28
2	26	37
3	30	38
4	24	28
5	30	41
6	27	29
7	33	49

Soluciones:**Diseño:** Comparación de dos grupos apareados**Prueba:** Comparación de dos promedios correlacionados
a través de la distribución *t* de Student**RELATO**

Se sabe que la exploración física de los pacientes ha de ser minuciosa y que ...

Duración (minutos) de la exploración física efectuada por
siete médicos según turno

Médico	Turno diurno	Turno nocturno
1	27	28
2	26	37
3	30	38
4	24	28
5	30	41
6	27	29
7	33	49

Tabla Auxiliar de trabajo:

Médico	Turno diurno	Turno nocturno	Diferencias <i>d</i>	Diferencias cuadráticas <i>d</i> ²
1	27	28		
2	26	37		
3	30	38		
4	24	28		
5	30	41		
6	27	29		
7	33	49		

Cálculos:

$$t = \frac{\sum d}{\sqrt{\frac{n \sum d^2 - (\sum d)^2}{n-1}}}$$

$$t = \frac{[(\quad) (\quad)] - (\quad)^2}{-1} = \sqrt{\quad} = \quad = \quad = \quad =$$

M U E S T R E O

Una muestra podemos definirla como una parte del universo o población de referencia. Entre las ventajas que ofrece el estudiar a una muestra en lugar de a toda la población tenemos:

- a) se reducen los costos;
- b) se reduce el tiempo requerido para la investigación;
- c) se reduce el número de personas que se requiere estudiar; y
- d) permite una investigación más detallada de las unidades de observación.

I.- Tipos de muestra

1. Probabilísticas.- En este caso, la probabilidad que tiene cada individuo de ser seleccionado en la muestra, es conocida. Los tipos más comunes de muestras probabilísticas son las siguientes:

1.1 Muestreo aleatorio simple (Random sampling)

En este tipo de muestra, cada individuo tiene la misma probabilidad de ser incluido.

Procedimiento de selección:

- a) Numerar a todos los elementos de la población del 1 al N.
- b) Utilizando tablas de números aleatorios, seleccionar un número entre 1 y N.
- c) Seleccionar un segundo número entre 1 y N. En caso de ser repetido, se descarta y se selecciona otro.
- d) Continuar de esta misma manera hasta seleccionar n distintos números entre 1 y N, " n " corresponde al tamaño de la muestra.

Cuando la población total no es muy numerosa, un método alternativo es la lotería. En este caso, después de numerar a todos los elementos de la población del 1 al N, se hacen papелitos con estos números, se mezclan y se seleccionan " n " papелitos.

1.2 Muestro sistemático (Systematic sampling)

En este caso, los elementos de una lista se seleccionan aplicando un intervalo de amplitud constante, después de un inicio seleccionado al azar.

Procedimiento de selección:

- a) Numerar a todos los elementos de la población del 1 al N.
- b) Se selecciona un intervalo de muestreo (k). Si se desea una muestra de tamaño "n", generalmente $k=N/n$, redondeando el resultado de tal manera que quede una cifra entera.
- c) Al azar se selecciona un número entre 1 y k.
- d) Los individuos que se van a seleccionar serán: el seleccionado al azar entre 1 y k, y cada k elemento de la lista.

Ejemplo:

- a) Población total (N)= 2,000
muestra (n)= 200
- b) $k=N/n= 2,000/200= 10$
- c) Seleccionar al azar un número entre el 1 y el 10, digamos el 6.
- d) De la lista de elementos, seleccionaremos al 6, 16, 26, etc.

1.3 Muestro estratificado (Stratified sampling)

Es el resultado de dividir a la población en grupos o estratos de acuerdo a cierta característica. Los estratos se forman de manera tal, que cada elemento de la población se asigne únicamente en uno de los estratos.

Procedimiento de selección:

- a) Dividir la población en estratos de acuerdo a cierta característica.
- b) Determinar el tamaño de la muestra para cada estrato. El tamaño de la muestra puede ser igual para todos los estratos. Sin embargo, en algunas ocasiones es necesario dar distintos pesos a cada estrato. Por ejemplo, grupos minoritarios pueden ser representados en la muestra por una proporción mayor a la que existe en la población total. Si representan un 10% en la población, en la muestra se seleccionará más del 10% para asegurar un grupo de comparación adecuado.
- c) Seleccionar una muestra probabilística dentro de cada estrato, por ejemplo, una muestra aleatoria simple.

1.4 Muestreo por conglomerados (Cluster sampling)

Es un tipo de muestreo probabilístico en que las unidades de muestreo son una colección o agrupa-

miento de elementos. Por ejemplo, seleccionar a todos los habitantes de una muestra de manzanas, en lugar de una muestra de individuos dispersos por la ciudad.

Procedimiento de selección:

- a) Especificar un agrupamiento o conglomerado adecuado. En la mayoría de los casos, los agrupamientos consisten en elementos de la población que están físicamente juntos y por lo tanto son similares entre sí.
- b) Seleccionar una muestra probabilística de conglomerados.
- c) Recolectar datos en todos los elementos de la población que se encuentran en los conglomerados seleccionados.

1.5 Muestreo por área (Area sampling)

Es un método de muestreo probabilístico en que las unidades de área (países, estados, etc.), son las unidades de muestreo. Un requisito importante es que las fronteras deberán estar bien definidas y deberán de ser permanentes por lo menos por el tiempo que dure el estudio. Por ejemplo, un país se puede dividir en estados, y posteriormente seleccionar una muestra aleatoria simple de estados.

Este tipo de muestreo con frecuencia se da en varias etapas (muestreo de etapas múltiples). En cada una de las etapas se van seleccionando submuestras de unidades de área cada vez más pequeñas.

Ejemplo:

- 1ª etapa: seleccionar estados de un país;
- 2ª etapa: seleccionar ciudades dentro de los estados;
- 3ª etapa: seleccionar colonias dentro de las ciudades.

1.6 Muestreo multifásico (Multistage sampling)

Más que tener relación con la manera de seleccionar la muestra, el muestreo multifásico se refiere a la cantidad de información que se colecta. Consiste en tomar datos básicos de una muestra numerosa e información más detallada en la submuestra.

Es importante resaltar que en el caso de un muestreo en etapas múltiples, la misma cantidad de

información se obtiene en todas las unidades muestreo.

2. **No Probabilísticas.** En este tipo de muestras se desconoce la probabilidad que tiene cada individuo de ser seleccionado en la muestra, y por lo tanto, desconocemos si esta es verdaderamente representativa de la población de referencia. Por desconocerse la representatividad, no es posible generalizar los resultados obtenidos en la muestra, a la población de referencia. En este tipo de muestras es muy difícil evitar introducir sesgos al seleccionar a los individuos.

Los tipos más comunes de muestras no probabilísticas son las siguientes:

2.1 **Selección casual (Haphazard selection)**

En este caso, la muestra está integrada por individuos con los que se encuentra casualmente recolector de datos o entrevistador, por ejemplo en un centro comercial, en la calle, etc.

2.2 **Muestreo a juicio del recolector de datos (Judgment sampling)**

En este método de selección se indica a los entrevistadores que, utilizando su propio criterio seleccionen a los individuos que ellos consideran "representativos" de la población en estudio.

2.3 **Muestreo por expertos (Expert sampling)**

En este caso, alguien experto en el asunto en estudio, selecciona una muestra que a su criterio es "representativa". La selección de la muestra para un estudio de casos, cae dentro de esta categoría.

2.4 **Muestreo con un propósito definido (Purposive sampling)**

La muestra está integrada por ciertos subgrupos de la población en los que interesa probar alguna hipótesis específica. Por ejemplo, seleccionar un grupo de migrantes rurales para probar su adaptabilidad a las condiciones urbanas.

2.5 **Muestreo por cuotas (Quota sampling)**

Método de selección en que se indica a los entrevistadores que tienen que completar un número pre establecido de individuos según cierta(s) características. Ejemplo: por edad, sexo, ocupación, etc. Este tipo de muestras tienen la ventaja que aseguran la representación de los principales estratos de la población.

II. Determinación del tamaño de la muestra.-

No existe una fórmula mágica para determinar el tamaño de la muestra. Son varios los factores que se deben tomar en cuenta:

- a) la prevalencia de la condición o características que se estudia: enfermedades o fenómenos raros que requieren de muestras más numerosas que las de presentación más frecuente;
- b) la variabilidad entre las unidades de observación: poblaciones homogéneas pueden ser representadas por muestras más pequeñas que cuando son heterogéneas;
- c) tipo de estudio: estudios de casos y controles requieren de muestras menores que en estudios de cohortes;
- d) cuando existen pequeñas diferencias o asociaciones débiles, se requieren muestras más numerosas;
- e) el nivel de confiabilidad deseado y la magnitud del error de muestreo que se está dispuesto a aceptar; y
- f) el costo, en relación a los recursos disponibles, etc.

BIBLIOGRAFIA

DOCUMENTOS BASICOS

- 1.- The Clark-Omran System of research design in epidemiology: A simplified Narrative and Illustration.- A. Omran.- University of North Carolina at Chapel Hill, 1977
2. Introduction to Sample Survey Methodology (Class Notes) W. Kalsbeek.- University of North Carolina at Chapel Hill, 1980

OTRAS REFERENCIAS

3. Statistics with applications to the biological and health sciences.- Remington & Schork.- Prentice Hall, 1979 (p 87-111)
4. The sample survey: theory and practice.- D. Warwick & Ch. Lininger.- McGraw-Hill, 1975 (p 69-110)

E J E R C I O S S O B R E M U E S T R E OINGRESOS HOSPITALARIOS EN CINCO SERVICIOS DEL HOSPITAL INFANTILDE MEXICO. ENERO 1980

(DATOS FICTICIOS)

NOMBRE	SEXO	EDAD	SERVICIO
Aguilera Vargas Luis	M	5	U
Álvarez Valdes Josefina	F	8	I
Arroyo García Jaime	M	0	N
Bernal Esquivel José	M	13	H
Camargo Leon Alvaro	M	9	C.P
Cano Méndez Juan	M	14	C.P
Coronado Acosta Rosa	F	1	N
Díaz Pérez Soledad	F	5	H
Díaz Torres M. Guadalupe	F	8	U
Escobedo Corral Manuel	M	11	CP
Galicia Ortega Nicolas	M	8	H
García Juárez Alfonso	M	6	I
González Quezada Aurora	F	2	N
Guzmán Pérez Alejandro	M	8	I
Hernández Aranda Enrique	M	7	U
Herrera ReyesMartha	F	5	H
Ibañez Vega Ramón	M	3	N
Juárez Guerra M. Teresa	F	9	CP
López Macias Antonio	M	14	U
Noriega Zamora Pedro	M	8	H
Palacios Flores Elsa	F	5	I
Paz	M	7	CP

Ramírez Estrada Ignacio	M	4		U
Reyes Calderón Ma. Eugenia	F	2		N
Río García Carlos	M	4		N
Rodríguez Escobedo Adolfo	M	10		U
Rodríguez Mendoza Ignacio	M	5		CP
Rojas Mejía Ma. Eugenia	F	3		N
Romero Gómez Arturo	M	9		H
Ruíz Silva Gloria	F	7		I
Salazar Bonilla Alfonso	M	6		CP
Salgado Salgado Jesus	M	6		U
Sánchez Fernández Ramón	M	8		I
Sepulveda Alvarez Armando	M	9		H
Solano Reyes Jesus	M	9		I
Topete Uribe Sergio	M	11		U
Torres Jiménez Rosalinda	F	15		CP
Valencia López Gustavo	M	12		H
Vázquez López Felipe	M	6		I
Zavala Peña José	M	3		N

CLAVES: SEXO: M : MASCULINO F: FEMENINO

EDAD: EN AÑOS CUMPLIDOS

SERVICIO: U: UROLOGÍA

H: HEMATOLOGÍA

I : INFECTOLOGIA

C.P. CIRUGÍA PEDIATRICA

N: NUTRICION

No. 1: MUESTREO ALEATORIO SIMPLE

De la lista de 40 ingresos hospitalarios, seleccione una muestra aleatoria simple de 15 individuos. Utilice la tabla de números aleatorios que se presenta a continuación:

Tabla G Números aleatorios.

85967	73152	14511	85285	36009	95892	36962	67835	63314	50162
07483	51453	11649	86348	76431	81594	95848	36738	25014	15460
96283	01598	61414	33525	04231	13604	75339	11730	85423	60698
49174	12074	98551	37895	93547	24769	09404	76548	05393	96770
97366	39941	21225	93629	19574	71565	33413	56087	40875	13351
90474	41469	16812	81542	81652	45554	27931	93994	22375	00953
28599	64109	09497	76235	41383	31555	12639	00619	22909	29563
25254	16210	89717	65997	82667	74624	36348	44018	64732	93589
28785	02760	24359	99410	77319	73403	58993	61098	04393	48245
84725	86576	86944	93296	10081	82454	78810	52975	10324	15457

Enliste a los 15 individuos seleccionados en orden alfabético según apellido paterno.

- 1.- _____
- 2.- _____
- 3.- _____
- 4.- _____
- 5.- _____
- 6.- _____
- 7.- _____
- 8.- _____
- 9.- _____
- 10.- _____
- 11.- _____
- 12.- _____
- 13.- _____
- 14.- _____
- 15.- _____

EJERCICIO:**No. 2: muestreo sistemático .-**

De la misma lista de 40 ingresos hospitalarios, selecciona muestra sistemática de 10 individuos. Especifique los valores N, n, k y del número entre 1 y k (seleccionado de acuerdo a tabla de números aleatorios del ejercicio no. 1)

$$N = 40$$

$$n = 10$$

$$k = N/n = \frac{40}{10} = 4$$

número entre 1 y $k = 4$

a sea = (1, 2, 3, 4)

Enliste a los 10 individuos seleccionados, en orden alfabético según apellido paterno.

1.- _____

2.- _____

3.- _____

4.- _____

5.- _____

6.- _____

7.- _____

8.- _____

9.- _____

10.- _____

EJERCICIO

No. 3 : MUESTRO ESTRATIFICADO .-

De la lista de 40 ingresos hospitalarios, seleccione una muestra estratificada de 12 elementos, de acuerdo a edad (integre estratos de igual amgnitud de intervalo) . Seleccione 3 elementos por estratos.

a) Cuales son los rangos de edad de cada uno de los estratos ?

estrato 1 : 1.- _____
 2.- _____
 3.- _____
 4.- _____

5.- _____
 6.- _____
 7.- _____

estrato 2: 1.- _____
 2.- _____
 3.- _____
 4.- _____

5.- _____
 6.- _____
 7.- _____
 8.- _____

9.- _____
 10.- _____
 11.- _____
 12.- _____
 13.- _____
 14.- _____

estrato 3: 1.- _____
 2.- _____
 3.- _____
 4.- _____
 5.- _____
 6.- _____
 7.- _____

3.- _____
 9.- _____
 10.- _____
 11.- _____
 12.- _____
 13.- _____
 14.- _____

estrato 4: 1.- _____
 2.- _____
 3.- _____
 4.- _____
 5.- _____

c) Seleccione una muestra aleatoria simple de 3 elementos por estrato, utilizando la siguiente tabla de números aleatorios:

85967	73152	14511	85285	36009	95892	36962	67835	63314	50162
07483	51453	11649	86348	76431	81594	95848	36738	25014	15460
96283	01898	61414	83525	04231	13604	75339	11730	85423	60698
49174	12074	98551	37895	93547	24769	09404	76548	05393	96770
97366	39941	21225	93629	19574	71565	33413	56087	40875	13351
90474	41469	16812	81542	81652	45554	27931	93994	22375	00953
28599	64109	09497	76235	41383	31555	12639	00619	22909	29563
25254	16210	89717	65997	82667	74624	36348	44018	64732	93589
28785	02760	24359	99410	77319	73408	58993	61098	04393	48245
84725	86576	86944	93296	10081	82454	76810	52975	10324	15457
41059	66456	47679	66810	15941	84602	14493	65515	19251	41642
67434	41045	82830	47617	36932	46728	71183	36345	41404	81110
72766	68816	37643	19959	57550	49620	98480	25640	67257	18671
92079	46784	66125	94932	64451	29275	57669	66658	30818	58353
29187	40350	62533	73603	34075	16451	42885	03448	37390	96328
74220	17612	65522	80607	19184	64164	66962	82310	18163	63495
03786	02407	06098	92917	40434	60602	82175	04470	78754	90775
75085	55558	15520	27038	25471	76107	90832	10819	56797	33751
09161	33015	19155	11715	00551	24909	31894	37774	37953	78837
75707	48992	64998	87050	39333	00767	45637	12538	67439	94914

estrato 1: (Use el 1er grupo de número aleatorios)

- 1.- _____
- 2.- _____
- 3.- _____

estrato 2 : (Use el 2º grupo de números aleatorios)

- 1.- _____
- 2.- _____
- 3.- _____

estrato 3 : (Uso el 3er grupo de número aleatorio)

- 1.- _____
- 2.- _____
- 3.- _____

estrato 4: (Use el 4º grupo de números aleatorios)

- 1.- _____
- 2.- _____
- 3.- _____

Estos son los 12 individuos que forman muestra estratificada por cada.

Ejercicio No. 4 : POR CONGLOMERADOS .-

De la lista de 40 ingresos hospitalarios seleccionados una muestra de 16 elementos , por conglomerados. Tome como conglomerados a los diferentes servicios .

a) Seleccione por muestreo alatorio simple 2 de los 5 servicios. Utilice el 5º grupo de número alatorios de la tabla del ejercicio no. 3. Horizontal o vertical

- 1.- Urología
- 2.- Hematología
- 3.- Infectología
- 4.- Cirugia pediátrica
- 5.- Nutrición

Los servicios seleccionados son:

- 1.- _____
- 2.- _____

