



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

ELABORACION DE UN TUTOR DE ESTADISTICA
POR COMPUTADORA.

T E S I S

QUE PARA OBTENER EL TITULO DE

A C T U A R I O

P R E S E N T A :

G. Araceli Cruz Martín del Campo

TESIS CON
PALA DE ORIGEN

CIUDAD UNIVERSITARIA

1989



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

INDICE

INTRODUCCION	1
CAPITULO I: CONCEPTOS BASICOS DE ENSEÑANZA	3
1. Didáctica	3
1.1 Fundamentos: aprendizaje, motivación enseñanza	3
1.2 Método Didáctico	5
1.2.1 Definición de objetivos	5
1.2.2 Planeación	6
1.2.3 Realización	7
1.2.4 Evaluación	7
2. Instrucción Programada	9
2.1 Tesis principales de Skinner	9
2.2 Principios de Instrucción Programada...	10
2.2.1 Respuesta activa	10
2.2.2 Mínimo de errores	11
2.2.3 Conocimiento de resultados	12
2.3 Para hacer un buen programa	12

3. Enseñanza Auxiliada por Computadora (E.A.C.)	14
3.1 Ventajas de la E.A.C.	14
3.2 Desventajas de la E.A.C.	15
4. Secuencia cronológica de la Enseñanza Auxiliada por Computadora	17
 CAPITULO II: TUTOR DE ESTADISTICA	 19
1. Generalidades	19
2. Lección 1: Definición y división de la Estadística	23
3. Lección 2: ¿Qué son los datos?. Escalas de medición	45
3.1 Escala nominal	49
3.2 Escala ordinal	51
3.3 Escala de intervalos	52
3.4 Escala de razón	53
3.5 Escala absoluta	55

4. Lección 3: La organización y el resumen de los datos	68
4.1 Datos e información	69
4.2 Tablas de frecuencia e intervalos de clase	70
4.3 Histogramas de frecuencia	83
5. Lección 4: Medidas de tendencia central....	91
5.1 Media aritmética	96
5.2 Moda	102
5.3 Mediana	107
5.4 Media aritmética ponderada	113
6. Lección 5: Medidas de dispersión	115
6.1 Rango	119
6.2 Desviación	120
6.3 Desviación media	122
6.4 Varianza	124
6.5 Desviación estándar	127
6.6 Estimadores Insegados	130
6.7 Puntaje estándar	149

7. Lección 6: Muestreo y distribuciones muestrales	151
7.1 Población y muestra	153
7.2 Conceptos de Probabilidad	156
7.3 Selección del tamaño de muestra	160
7.4 Eventos independientes	174
7.5 Distribución muestral de la media	179
CAPITULO III: PRUEBA PILOTO	183
1. Descripción de la prueba	183
2. Resultados	185
3. Conclusiones	192
4. Recomendaciones	193
4.1 A programadores o autores	193
4.2 A maestros	194
4.3 A alumnos	194
CONCLUSIONES	195
ANEXO A: GLOSARIO DE PILOT	197
BIBLIOGRAFIA	202

INTRODUCCION

INTRODUCCION

Enseñanza Auxiliada por Computadora (E.A.C.) es un término con el que se conoce el proceso de usar una computadora para proveer a una persona de material educativo y ayudarle a que comprenda y asimile esa información.

El autor crea un programa o lección, el cual es ejecutado por el estudiante en la computadora, así, por medio de la interacción con la máquina se enseña alguna materia.

Este proceso se parece a la escritura y lectura de un libro, pues una persona lo escribe una vez, para que después muchas otras puedan utilizarlo.

El presente trabajo hace uso de la Enseñanza Auxiliada por Computadora, con el fin de apoyar a los maestros que imparten alguna materia relacionada con la Estadística.

Consiste en seis lecciones por computadora sobre Estadística Básica en las que se tratan, en forma sencilla y gradual, temas que van desde la definición y división de la Estadística, hasta distribuciones muestrales, pasando por escalas de medición, distribución de frecuencias y medidas de tendencia central y de dispersión.

El lenguaje empleado es simple, amigable y de fácil comprensión, se incluyen representaciones gráficas con cierto movimiento con el fin de que los estudiantes se sientan cómodos y relajados frente a la computadora, favoreciendo así un mejor y más rápido aprendizaje.

Se tomaron en cuenta, al desarrollar el trabajo, algunos principios de Didáctica y de Instrucción Programada, para lograr efectividad en el proceso de enseñanza - aprendizaje.

Este paquete de programas (són 10), puede ser usado en microcomputadoras IBM PC o compatibles de las más pequeñas. Es conveniente que el monitor sea a colores.

En el primer capítulo del trabajo se exponen los conceptos teóricos en que se basan estos programas; en el segundo se encuentra la descripción técnica del tutor incluyendo la de Pilot, el lenguaje de computación utilizado, y la documentación detallada de cada una de las lecciones, en el tercero se describen los resultados de una prueba piloto que se efectuó para evaluar el producto y se hacen algunas recomendaciones a las personas que puedan interesarse en el tema. Al final hay un anexo en el que se da un glosario de instrucciones de Pilot.

CAPITULO I

CONCEPTOS BASICOS DE ENSEÑANZA

CAPITULO I

CONCEPTOS BASICOS DE ENSEÑANZA

1 DIDACTICA

1.1 FUNDAMENTOS:

Para que una sociedad sobreviva es indispensable que transmita a las nuevas generaciones sus conocimientos, lo que se logra por medio del proceso educativo.

En este proceso se observa que, como producto del aprendizaje, se dan cambios en la conducta de las personas. Por ejemplo, adquieren habilidad en el manejo de algún instrumento, o conocimientos e información sobre un tema; forman su capacidad de apreciación de las expresiones artísticas, o llegan a tener una actitud de mayor comprensión y respeto hacia los demás.

El aprendizaje se realiza en la persona a través de sus experiencias, por las diferentes situaciones que se le presentan en la vida diaria, ya sea en forma directa (cosas que le ocurren a ella), o en forma indirecta (experiencias que le son comunicadas) y que influyen en su comportamiento posterior.

La persona está en posibilidad de continuo aprendizaje, en todos lados, en cualquier momento y en todas las circunstancias de su vida.

Para que una persona aprenda se requiere que se enfrente a una situación nueva que exija de ella una respuesta que no conoce de antemano, (situación que debe estar de acuerdo con su preparación, capacidad y madurez); que busque cuáles podrían ser las respuestas adecuadas, seleccionando la que considere mejor y que finalmente, verifique si ha logrado lo que pretendía.

En caso de que la acción haya sido adecuada, se sentirá satisfecha y en otra ocasión tendrá interés por aprender nuevas cosas; en caso de que esta respuesta haya sido inadecuada, puede suceder que ensaye otras acciones, o bien que se sienta fracasada e incompetente y prefiera no volver a intentarlo.

Las personas aprenden efectivamente sólo en la medida que:

- consideran útil e interesante la materia que estudian.
- buscan y seleccionan las respuestas por sí mismas.
- perciben los logros y las deficiencias de su aprendizaje

En ocasiones no se alcanza un aprendizaje adecuado:

- por ignorar cuáles son los objetivos que se buscan.
- por tener deficiente preparación previa.
- por no tener interés en lo que se está estudiando.
- por no encontrar utilidad a lo que se aprende.
- por no poseer la capacidad necesaria para aprender determinada materia.
- por tener problemas de salud, anímicos, etc.

La función del profesor en relación con los alumnos es orientar el aprendizaje guiando, asesorando y coordinando la planeación, realización y evaluación del mismo, tratando que los alumnos aprendan efectivamente a pesar de los obstáculos que se puedan presentar.

La motivación de los alumnos se manifiesta cuando relacionan lo que han estudiado con diversas situaciones de su vida diaria; participan con entusiasmo; consultan, investigan, amplían lo que aprendieron. Cuando se han generado en los alumnos intereses que los mueven a aprender nuevamente. Así es posible hablar de que han aprendido a aprender.

El presente trabajo quiso tener en cuenta todos estos principios de Didáctica, tratando de que los programas que enseñan Estadística por computadora sean ágiles, amigables y atractivos y que motiven a estudiar más sobre esta materia, aprovechando las facilidades que ofrece la tecnología actual.

1.2 METODO DIDACTICO

El método didáctico requiere:

1.2.1 DEFINICION DE OBJETIVOS: Se ha dicho que el aprendizaje es una modificación de algún aspecto de la conducta; la especificación de objetivos determina el comportamiento que el proceso enseñanza - aprendizaje se propone obtener o desarrollar en los alumnos.

Es importante tener objetivos de aprendizaje claros y definidos porque esto facilita la comunicación entre los participantes del proceso de educación. Los objetivos se convierten en los criterios de planeación, realización y evaluación del proceso.

Los cambios que se desea conseguir en la conducta de los alumnos, pueden realizarse en tres grandes grupos:

- área cognoscitiva: comportamientos que se relacionan con los procesos mentales o intelectuales de los alumnos.
- área psicomotriz: habilidades neuromusculares o físicas que requieren diferentes grados de destreza física.
- área afectiva: los valores, actitudes y sentimientos de los alumnos.

Así que al definir el objetivo, debe definirse qué cambio se espera de la persona en lo que se refiere a conocimientos, habilidades y actitudes.

En las lecciones que se presentan en el capítulo II, lo primero que se define son los objetivos en cada uno de estos aspectos.

1.2.2 PLANEACION: Se parte de una situación percibida en un contexto real. Se analiza y se definen los procedimientos o caminos particulares a seguir y los recursos que puedan utilizarse para que los alumnos se enfrenten a esa nueva situación y le den una respuesta satisfactoria, que permita lograr los objetivos propuestos.

Al hacer este trabajo se tuvo en cuenta la problemática de la Facultad de Ciencias, pues es común que algunos maestros se enfrenten con la necesidad de dar un repaso breve a los temas básicos de Estadística, suponiendo que sus alumnos conocen la materia y descubran que esto es falso, pues se requiere un estudio más profundo de estos temas por parte de los estudiantes. Los profesores no pueden

impartir estos conocimientos supuestos ya que les consumiría tanto tiempo que no podrían cubrir su propio temario. Esta es la idea que motivó la realización de este tutor: apoyar a los maestros en la enseñanza de temas básicos de Estadística, por medio de los programas aquí presentados, que permiten que cada alumno aprenda a su propia velocidad, una de las muchas ventajas de la Enseñanza Auxiliada por Computadora.

En la planeación se analizó cuál es el lenguaje que resulta más adecuado y el equipo de cómputo más conveniente. Se concluyó que PILOT es el lenguaje que puede servir más a la Facultad, ya que por ser fácil de manejar y accesible en el precio, es posible que otras personas se interesen en desarrollar este tipo de programas autoinstructivos.

Respecto al equipo, se consideró que las micro-computadoras IBM compatibles son fácilmente accesibles a los alumnos. Dado que ya es común el uso de monitores a color, se aprovechó esto como método pedagógico, usando gráficas de baja resolución, que son las más frecuentemente soportadas.

1.2.3 REALIZACION: Consiste en llevar a cabo las actividades que hayan sido planeadas. Esto es lo que se hizo al desarrollar las seis lecciones que se presentan en el capítulo II.

1.2.4 EVALUACION: Se trata de verificar los resultados obtenidos en relación con los objetivos propuestos, para detectar aciertos y deficiencias y hacer las rectificaciones pertinentes.

La evaluación requiere reunir e interpretar evidencias del cambio de conducta efectuado en los alumnos, como resultado y producto de la acción educativa. Debe tratarse en lo posible que las evidencias que se reúnan sean objetivas, evitando en la apreciación y medición de la conducta, aspectos subjetivos que distorsionen los resultados.

En este proyecto, se llevó a cabo una prueba piloto, con 10 alumnos que se pusieron a estudiar con el tutor. A cada uno de ellos se le pidió llenar una encuesta, cuyos resultados se resumen en el Capítulo III, con el fin de evaluar, y en su caso, mejorar el sistema.

2. INSTRUCCION PROGRAMADA

2.1 Uno de los que trabajaron en los fundamentos de la Enseñanza Auxiliada por Computadora fué B.F. Skinner durante la década de los cincuentas. Sus tesis principales y su argumentación sobre las ventajas de apoyar la labor educativa con el uso de la computadoras, las presenta en el artículo: "Teaching Machines", publicado en Scientific American en noviembre de 1961:

"La máquina en sí no enseña, pone en contacto al estudiante con la persona que elaboró el material. Esto podría parecer producción en masa, pues pone en contacto a un programador (así llama al autor) con un número ilimitado de estudiantes; pero el efecto es, sorprendentemente, similar al de un maestro particular".

"Esto se debe a que:

- "La máquina induce a una actividad sostenida, el estudiante está siempre atento y ocupado.
- "La máquina insiste, al igual que un buen maestro, en que se entienda perfectamente un punto dado.
- "La máquina ayuda a dar la respuesta correcta. Lo hace, en parte gracias a la construcción ordenada del programa, y en parte mediante técnicas de insinuación o sugerencia (una línea punteada indicando el número de letras de la respuesta correcta, un "el" o "la" antes de la respuesta, etc).

- "La máquina, al igual que el preceptor privado, refuerza al estudiante con cada respuesta correcta, aprovechando la comprobación inmediata".

"Con máquinas y programas debidamente planeados, un estudiante lento, con libertad de avanzar a su ritmo normal de trabajo, puede elevarse a niveles de competencia nunca soñados".

"La meta del programador deberá ser la de continuar perfeccionando el programa hasta llegar al punto en que las respuestas del estudiante promedio sean casi siempre correctas."

2.2 Principios de la Instrucción Programada

La tarea del programador, afirma Skinner, consiste en crear materiales de instrucción que logren los siguientes principios:

2.2.1 Respuesta activa.

Una respuesta activa se consigue cuando se exige que el estudiante asimile completamente el material que se le presenta en una unidad de instrucción o cuadro.

Sin embargo, hay diferencia entre exigir una respuesta activa por cada sección de la información, a exigir que el estudiante dé respuesta a cada pregunta por separado, como pretenden otras teorías.

En muchas listas de las características de la instrucción programada encontramos el "principio de los pequeños pasos". La palabra "pequeño" ha tenido efectos desafortunados en el modo en que algunas personas han proyectado la instrucción. Lo que se busca es que se definan unidades que no confundan por el exceso de información, pero que tampoco resulten demasiado elementales. Cuando se fraccionan demasiado las ideas, es fácil perder la visión de conjunto.

2.2.2 Mínimo de errores.

Error es una respuesta no aceptable al programador, que no esperaba o no deseaba que dieran los estudiantes. Alguna vez, el programador puede provocar la respuesta errónea para sacar a la luz un malentendido, pero eso es aceptable y no se considera como un error.

Los programadores intentan eliminar los errores al revisar el programa. Las respuestas equivocadas pueden indicar:

- a) una idea mal expuesta, que no consigue enseñar;
- b) una secuencia en la cual se han suprimido las insinuaciones de una respuesta correcta;
- c) presuponer conocimientos de los que, en realidad, el alumno carece;
- d) análisis deficiente de la materia, lo cual ocasiona confusión que el programador no había previsto.

2.2.3 Conocimiento de los resultados.

Una vez que se ha cometido un error, el estudiante debe enterarse de que se ha equivocado con el fin de evitar que esta situación se repita. Debe además, descubrir cuál es la respuesta correcta. No hay que dársela, sino ayudarle a que la encuentre.

Esto se puede lograr mediante un planteamiento hábil que induzca al estudiante a acertar y a estar seguro de haber acertado.

2.3 Cualidades necesarias para hacer un buen programa:

a) Dominio absoluto de la materia que se va a programar. Para formular los objetivos del programa, establecer en qué secuencia se van a presentar los conocimientos y conceptos, idear ejemplos, relacionar un segmento de la tarea con los demás, para determinar lo adecuado y veraz de los postulados hechos respecto al tema. Si el programador no tiene un conocimiento profundo de lo que está enseñando, debe colaborar estrechamente con alguien que lo posea.

b) Pericia en la comunicación. Se requiere dominio del lenguaje, talento, exactitud y claridad para comunicarse con el estudiante.

c) Respeto en el estudiante cuya conducta se piensa inducir, usando las mejores técnicas disponibles para lograrlo.

d) Habilidad de diagnóstico para observar el proceso de aprendizaje en cada uno de los estudiantes en las sesiones de prueba.

e) Profundidad en el proceso final de la evaluación. Un programa puede resultar poco atractivo para el programador en términos estéticos o personales, este punto carece de importancia, lo esencial es que al aplicarlo con estudiantes, cumpla los objetivos para los que fué hecho.

La clave de la eficacia para el producto final es el procedimiento de prueba. Muchos programadores prefieren hacer la prueba personalmente con unos cuantos estudiantes que trabajan con un programa por vez primera. Hablar con un solo estudiante a la vez puede resultar muchísimo más valioso que recopilar los errores de un grupo numeroso. Si el estudiante no entiende un cuadro en particular, podrá indicar exactamente la palabra o frase que le está causando dificultades. Esta frase se elimina y el cuadro se redacta de otro modo.

3. ENSEÑANZA AUXILIADA POR COMPUTADORA (E.A.C.)

La eficiencia del proceso educativo se alcanza al transmitir más conocimientos y habilidades a más personas en menos tiempo.

Al principio la educación se impartía oralmente, después la imprenta ocasionó cambios radicales, y ahora los avances tecnológicos, específicamente las computadoras, pueden volver a causar un fuerte impacto en los métodos de enseñanza y en el aprendizaje mismo.

Una buena definición de Enseñanza Auxiliada por Computadora es: un sistema interactivo en el que intervienen un estudiante y una computadora, usando un lenguaje natural para comunicarse y que produce cambios en el conocimiento del estudiante.

3.1 Ventajas de la Enseñanza Auxiliada por Computadora

Algunas de las ventajas principales que tiene el uso de esta técnica son:

- a) El estudiante se ve obligado a participar activamente, lo que facilita el aprendizaje
- b) Los estudiantes trabajan individualmente, por lo que tienen la posibilidad de aprender al ritmo que les permite su capacidad.
- c) El profesor puede contar con una evaluación del aprendizaje tan frecuente como lo requiera.
- d) El estudiante puede recibir refuerzos inmediatos y sistemáticos en su proceso de aprendizaje.

e) La computadora es sólo una máquina, por lo que el efecto psicológico que recibe el estudiante al cometer un error, puede ser menos impactante que si la falta se lleva a cabo en presencia del maestro y otros compañeros.

f) Al trabajar con la computadora se obtiene un ambiente más propicio para pensar, y por lo tanto para aprender.

g) La máquina libera al profesor del trabajo repetitivo, brindándole más tiempo para la preparación de sus lecciones o para dar atención personalizada a los alumnos que la requieran.

h) Con lecciones que se pueden usar repetidamente se puede aprovechar la experiencia de los buenos maestros.

i) La Enseñanza Asistida por Computadora permite el manejo de material educativo en lugares donde no se tienen suficientes instructores.

3.2 Desventajas de la E.A.C.

Por supuesto que también tiene desventajas, entre ellas se encuentran las siguientes:

a) Hay poca flexibilidad en los programas, se establece una lección, y el estudiante debe seguirla, aunque ya conozca esa parte del material.

b) A pesar de que es una ventaja el hecho de que no haya presión psicológica en el estudiante, es una desventaja el que tampoco haya un reconocimiento personal de sus logros.

c) El personal docente y administrativo debe ser capacitado y motivado para usar esta nueva forma de enseñar.

d) Lo que actualmente ofrece el mercado para el desarrollo de programas (software) educativos, no permite el desarrollo rápido de los mismos. Se consideran necesarias de 50 a 70 horas de trabajo del autor para que un estudiante tenga media hora de enseñanza en la computadora.

e) Los equipos de computación (hardware) y los programas para elaboración de tutores y los tutores mismos (software) son aún relativamente caros frente a las posibilidades presupuestales de los planteles educativos.

4. SECUENCIA CRONOLOGICA DEL DESARROLLO DE LA E.A.C.

- 1920's - Las bases de la Enseñanza Auxiliada por Computadora se dieron desde los años veinte, cuando el psicólogo Sidney Pressey en 26 inventó la primera máquina de enseñar. Esta máquina permitía que el estudiante seleccionara en un teclado el botón correspondiente a su respuesta, si era la correcta, se pasaba a la siguiente, y en caso contrario, se repetía la pregunta hasta lograr una respuesta satisfactoria.
- 1950's - En los años cincuentas se desarrolló la teoría de la Instrucción Programada, que como ya se dijo, consiste en que el material educativo se dispone de tal forma que, a través de una secuencia de pasos cercanos entre sí, se va guiando al estudiante hacia una comprensión gradual de objetivos cada vez más complejos. El material se presenta en pequeñas unidades escritas de tal forma que se maximizan las posibilidades de éxito.
- 1960,s - A finales de los sesentas y principios de los setentas, se desarrolló uno de los primeros ejemplos de software educativo en la Universidad de Illinois: el sistema PLATO, con costos muy altos, requería una computadora grande y terminales gráficas de alta resolución y sensibles al tacto. Este sistema está en uso en nuestros días.
- 1975 - las microcomputadoras se hicieron más comerciales, empezaron a estar al alcance de más gente por su bajo costo. Esto permitió que algunas escuelas pudieran usar las máquinas como medio educativo y así, la fase experimental de la EAC en Estados Unidos fue una realidad. Como

consecuencia, se extendió ampliamente el desarrollo de software educacional.

- 1977 - Se dió la difusión masiva de las microcomputadoras en Estados Unidos.
- 1982 - El desarrollo de las microcomputadoras en México desde el año de 77 fué de en un 400%, mientras que el crecimiento de las máquinas grandes fué de sólo 285%.
- 1988 - Ha habido una gran aceptación de las computadoras, y el entusiasmo por usarlas en la educación ha crecido mucho. En algunas escuelas, ya se incluyen las micros como un recurso cotidiano para los estudiantes.

Las micros están a precios accesibles y se venden incluso en tiendas de autoservicio.

Ya existen algunas compañías dedicadas a la elaboración profesional de "software" educativo.

- 1992 - Se estima que en este año la razón de estudiantes a microcomputadoras en los Estados Unidos será de 14 a 1. (Byte febrero de 1987), y ya que se ha estimado que la relación de mercado Estados Unidos México es de 1 a 100, podría pensarse que en este año habrá en México una computadora por cada 1400 estudiantes.

CAPITULO II
TUTOR DE ESTADISTICA

GENERALIDADES DEL TUTOR

Requerimientos de equipo mínimos para su uso:

- microcomputadora IBM PC o compatible con 256 Kilobytes de memoria.
- Un "drive" de disco que maneje diskettes de 360 Kb (doble densidad)
- Monitor a color con capacidad para manejar gráficas de baja resolución. (En un monitor monocromático con capacidad gráfica también puede funcionar, aunque con menor claridad, pues en el tutor se usan los colores como apoyo didáctico).

Lenguaje utilizado:

Pilot versión 2.1 es un lenguaje parecido a Basic, pero más potente. Diseñado especialmente para la elaboración de programas educativos, cuenta entre sus características más importantes con:

- un editor de texto simple y fácil de usar
- un editor de gráficas que aunque es un poco limitado, permite crear fácilmente imágenes simples.
- se pueden manejar gráficas con media y alta resolución. El alta resolución permite 640 x 400 "píxeles" y 16 colores.
- un programa para capturar pantallas, lo cual permite tener gráficas tan sofisticadas como se desee y lo permita el "software" utilizado. (No se usó este programa en el presente trabajo, pues se contaba con un software más conocido por la autora, que realizaba la misma función).

- otro programa para comprimir las imágenes que se graban en disco, y al mismo tiempo darles un efecto de movimiento para su presentación en pantalla.

- el lenguaje propiamente dicho, consta de 22 comandos, los cuales unidos a modificadores dan un vocabulario de cerca de 300 expresiones diferentes. Por ejemplo el comando A (Accept = acepta una respuesta), unido al modificador S (Single = solo), esto es, la expresión AS, acepta la respuesta de un estudiante obligando a que ésta sea de un carácter, no es necesario siquiera que se teclee <Return>.

- se pueden encadenar las lecciones, pasando valores de variables de una a otra, lo que permitió en este trabajo, usar el nombre del estudiante que se solicita en la introducción, a lo largo de todas las lecciones.

- se puede guardar registro (en un archivo), de los resultados de un estudiante, para después reportarlos.

- se pueden llamar programas escritos en otro lenguaje.

A continuación se presenta el tutor de manera semejante a como lo ve el estudiante en su pantalla.

**TUTOR DE
ESTADISTICA**

TRABAJO QUE PARA OBTENER

EL TITULO DE ACTUARIO PRESENTA

G. ARACELI CRUZ MARTIN DEL CAMPO

AGOSTO '89

INTRODUCCION

De acuerdo a la hora del reloj interno de la computadora saluda: ¡Buenos días!, ¡Buenas tardes!, o ¡Buenas noches!

¿Cómo te llamas?

Se requiere el nombre, el cual será usado después en las lecciones.

Se despliega un menú de opciones:

CURSO DE ESTADISTICA

por G. Araceli Cruz Martín del Campo

- Lección 1: Definición y división de la Estadística.
- Lección 2: Qué son los datos.
- Lección 3: La organización y el resumen de los datos.
- Lección 4: Medidas de tendencia central.
- Lección 5: Medidas de dispersión.
- Lección 6: Muestreo y distribuciones muestrales.

NOMBRE, selecciona el número de lección (9 para terminar)

Aquí y siempre que la pregunta sea de opción múltiple, se valida en los programas que se haya tecleado alguna de las opciones permitidas.

De acuerdo a la opción elegida, solicita el segundo diskette y/o se encadena con la lección seleccionada.

LECCION 1

DEFINICION Y DIVISION DE LA ESTADISTICA

DEFINICION Y DIVISION DE LA ESTADISTICA

OBJETIVO:

CONOCIMIENTOS: Al terminar la lección el estudiante podrá dar una definición de Estadística, Estadística Descriptiva y Estadística Inferencial

HABILIDADES: Que sepa diferenciar las características que definen específicamente a cada uno de los conceptos aquí expuestos. Que esté familiarizado con la idea de población y muestra.

ACTITUDES: Que se interese en las diferencias de cada concepto.

Bien, ya que vamos a empezar a trabajar juntos, será bueno ponernos de acuerdo en cómo lo haremos:

- LEE siempre TODAS las instrucciones, sólo están las indispensables.
- Puedes contestar con mayúsculas o minúsculas.
- Cuando quieras indicar que tu respuesta está completa, o quieras avanzar en el material, házmelo saber con la tecla <Return>, en el lado derecho de tu teclado. (En alguna máquinas se llama Intro, y en otras tiene una flecha curva dibujada sobre ella). En general, siempre que tengas duda, oprímela
- Si quieres suspender una lección, oprime simultáneamente <Ctrl> y <C> aunque es más recomendable que la termines, todas son pequeñas.

- Si necesitas repasar el texto al que se refiere una pregunta, teclea ? en lugar de la respuesta, y se repetirá esa parte de la lección.

- Cuando te haga una pregunta que requiera respuesta, lo indicaré poniendo las letras en rojo

¿De acuerdo? (con letras rojas)

Si dice algo diferente que SI: lo siento, debes contestar SI para poder continuar.

Muchas gracias, ¡Y aquí vamos...!

CONTINUAMENTE OIMOS

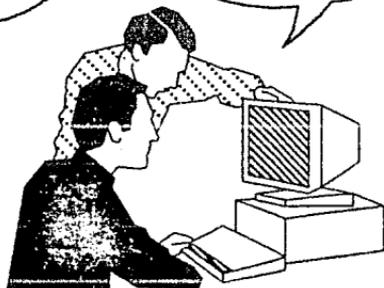
HABLAR DE

LA ESTADISTICA

EN LA EMPRESA

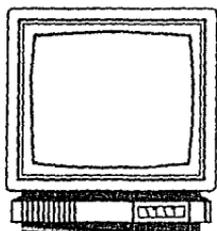
LAS ESTADISTICAS NOS
DICEN QUE LA INFLACION
FUE DEL 7 POR CIENTO

ESTOS DATOS NOS
PERMITEN INFERIR...



EN LA VIDA DIARIA

LA AUDIENCIA DE UN PROGRAMA

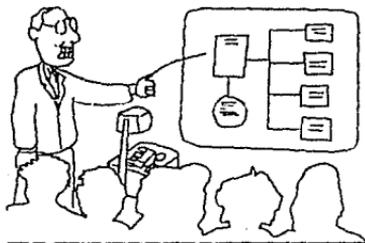


EL NUMERO DE HOMBRES
Y MUJERES QUE HAY
EN EL MUNDO



EN LA ESCUELA

El promedio del grupo A
fué medio punto más
alto que el del grupo B.



La demanda para
inscripción de
nuevos alumnos
se incrementó
este año en 10%.



En junio de 1987
nació el habitante
del mundo número
cinco mil millones

EN LAS NOTICIAS

Las bolsas de todo
el mundo bajaron,
en promedio, 16500
puntos durante el
mes anterior.



PERO...

***¿QUE ES LA
ESTADISTICA?***

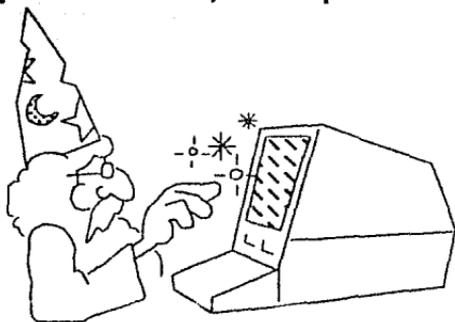
CUADRO 2

De las siguientes definiciones, ¿cuál te parece mejor?

- a) ¿Es una herramienta muy poderosa para predecir valores desconocidos a partir de valores conocidos?
- b) ¿Es la definición de métodos para la obtención y el análisis de datos?
- c) ¿Es una manera especial de organizar y presentar información?
- d) ¿Es un método científico para proceso de información?

Si elige a) (herramienta poderosa para predecir valores desconocidos a partir de valores conocidos), enviar un mensaje:

¡Cuidado!, Es poderosa...



... pero no mágica.

CUADRO 3

Tienes razón, es eso y mucho más. (¿Sabes que?, en realidad todas las respuestas son correctas)

Vamos a tratar de dar una definición más formal y completa de la Estadística, pero antes veamos un ejemplo:

CUADRO 4

Imagina que eres el Director de la Facultad de Ciencias de la UNAM y quieres conocer a los alumnos de la Escuela su edad, sexo, escuela de procedencia, promedio académico, etc.

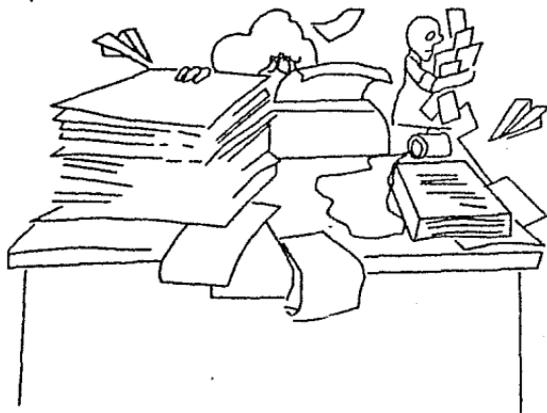
Vas a estudiar a la población a partir de los expedientes que hay en la Sección Escolar. (Claro que también se pueden aplicar los métodos que se mencionan más adelante directamente en el campo, esto se haría, como ya te habrás imaginado, entrevistando a los estudiantes).

CUADRO 5

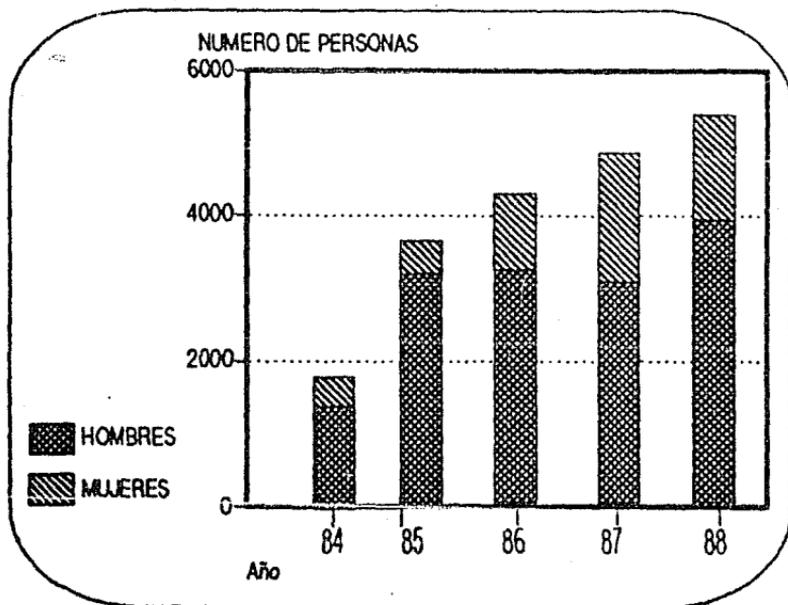
Tienes tres opciones

a) Leer todos los expedientes de los alumnos que están en la Sección Escolar.

Hay 4217 expedientes en 1988.



b) Agrupar los datos que tienes en todos esos expedientes, de manera que los resultados te den una idea aproximada pero confiable de lo que deseas estudiar.



CUADRO 7

c) Leer unos cuantos expedientes y suponer que el resto de los alumnos tiene características similares.

CUADRO 8

¿Cuál crees que sea la forma más adecuada de resolver este problema?

a) Leer todo

b) Agrupar datos y obtener sumas y promedios de ellos para entenderlos mejor

c) Estudiar algunos casos y considerar que representan a toda la Escuela, de manera que, lo que se diga acerca de esta "muestra", sea válido para todos los alumnos.

CUADRO 9

Si responde a) (Leer todo):

¿No crees que va a ser difícil que puedas describir a todas las personas?

¿Qué dirías?

"Juan tiene 19 años, promedio de 8, viene de Prepa 6; Ana tiene 22 años, promedio de 7, viene de CCH..."

... y así ...

No, definitivamente hay un camino más corto y eficaz.

Si responde b) (agrupar y obtener promedios):

¡Bien!, lo que estás haciendo es DESCRIBIR la población por medio de algunas de sus características mas importantes.

Ahora puedes decir:

"El 80% de los alumnos tienen entre 19 y 24 años. La calificación promedio del nivel escolar anterior (Prepa, CCH, Vocacional, etc) es de 7.26.

Hay 1943 alumnos que vienen de preparatorias de la UNAM, 1039 de CCH's, 374 de vocacionales y 861 de preparatorias particulares".

Como ves, estos datos te dan idea de cómo es la población que estás estudiando. Esta es la forma de trabajo de la ESTADISTICA DESCRIPTIVA.

Si responde c) (inferir a partir de una muestra):

Tienes razón, tal vez estudiar a todos los alumnos representa demasiado esfuerzo para los resultados que vas a obtener. No hay personal para hacer este estudio tan completo, y además los alumnos de la Facultad tienen muchas cosas en común, (no sería lo mismo si se tratara de estudiar a las personas que asisten a una función de cine), pues todos son estudiantes, la mayoría jóvenes, están a nivel licenciatura y ya seleccionaron la carrera que quieren estudiar. Es una población bastante homogénea.

La ESTADISTICA INFERENCIAL proporciona técnicas que permiten obtener conclusiones válidas para una población a partir de una muestra.

(En cualquiera de las dos alternativas anteriores b y c, se muestra al estudiante el párrafo que se le habría mostrado en caso de que él hubiera optado por la otra respuesta).

Ahora, con tu ayuda, daremos las definiciones formales.

Selecciona, entre las siguientes características, DOS que definan exclusivamente a:

(Respuestas correctas)

- | | |
|-----------------------------|------|
| (1) Estadística | b, c |
| (2) Estadística Descriptiva | a, e |
| (3) Estadística Inferencial | d, f |

a) Resume mediante el empleo de MEDIDAS SUMARIZADAS la información desordenada de toda una población.

b) Su materia prima son datos numéricos que se comportan regularmente.

c) Se refiere al estudio, desarrollo y empleo de técnicas efectivas de colección, presentación, resumen y elaboración de conclusiones a partir de datos numéricos.

d) Se usa para sacar CONCLUSIONES de un todo A PARTIR DE UNA MUESTRA usando la teoría de probabilidades.

e) Se trabaja sobre el CONJUNTO TOTAL de los elementos, en donde se ha practicado el conteo o medición. Por ejemplo un censo.

f) Verifica que los procesos de extracción y selección del material a estudiar sean tales que permitan una CORRECTA INFERENCIA o deducción.

Evaluación pregunta (1)

Si a,e:

La Estadística es más que eso, también puede trabajar con muestras. Esta definición sólo corresponde a la Estadística Descriptiva.

Si b,c:

Estás en lo cierto, esta es una de las características que definen a la Estadística

Si d,f:

No, esto es sólo una parte de la Estadística, la Inferencial.

Evaluación pregunta (2)

Si a,e:

¡Correcto!, esta es una característica propia de la Estadística Descriptiva

Si b,c:

No, esta no es una característica exclusiva de la Estadística Descriptiva, también la tiene la Inferencial.

Si d,f:

No, la Estadística Descriptiva estudia a toda la población, nunca a una parte de ella, como lo hace la Estadística Inferencial.

Evaluación pregunta (3)

Si a,e:

No, en general la Estadística Inferencial no estudia poblaciones completas, sólo a parte de ellas

Si b,c:

No es característica que defina a la Estadística Inferencial, también la tiene la Descriptiva

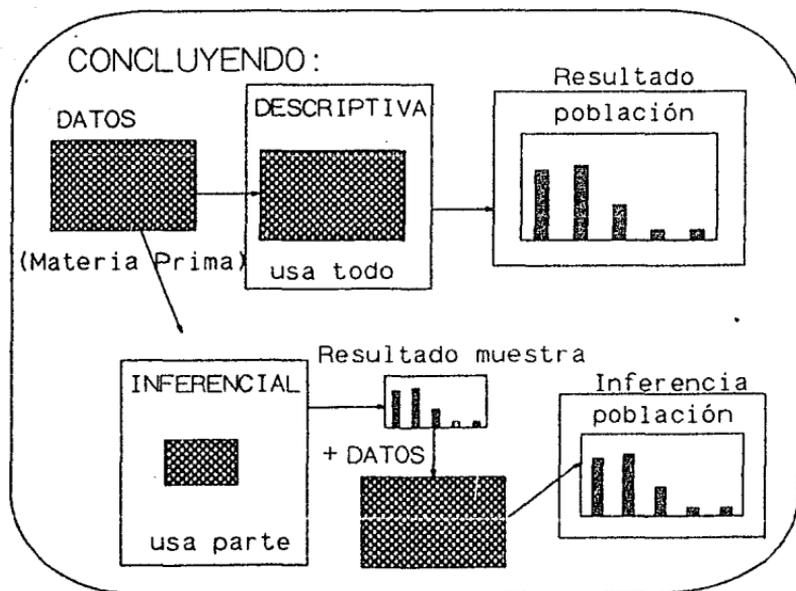
Si d,f:

Tienes razón, es una característica propia de la parte inferencial de la Estadística

Si al estar haciendo este ejercicio, el estudiante teclea ?, se le muestra el siguiente párrafo y se le solicita que conteste otra vez.

La Estadística busca definir técnicas para el manejo adecuado de datos numéricos, y se divide en dos partes principales;

- Descriptiva - considera TODOS los elementos de una población.
- Inferencial - considera sólo UNA PARTE, o muestra de los elementos de la población.



ESTADISTICA

Se refiere al estudio, desarrollo y empleo de técnicas efectivas de colección, presentación, resumen y elaboración de conclusiones a partir de datos numéricos. Su materia prima son datos numéricos que se comportan regularmente.

ESTADISTICA DESCRIPTIVA

Se trabaja sobre el conjunto total de los elementos donde se ha practicado el conteo o medición. Por ejemplo un censo. Resume mediante el empleo de medidas sumariadas la información desordenada de una población.

ESTADISTICA INFERENCIAL

Se usa para sacar conclusiones de un todo a partir de una muestra usando la teoría de probabilidades. Verifica que los procesos de extracción y selección del material a estudiar sean tales que permitan una correcta inferencia o deducción.

Conviene aclarar que esta división es principalmente con fines pedagógicos, porque en la práctica sucede frecuentemente que al estudiar las muestras se les calculen medidas de Estadística Descriptiva, como si fueran una población completa.

CUADRO 14

43.

Hemos hablado aquí de población y muestra de una manera un tanto intuitiva. Demos sus definiciones formales.

POBLACION es un conjunto de elementos que tienen al menos una característica en común. Por ejemplo, los estudiantes de la Facultad de Ciencias, las personas sometidas a cierta intervención quirúrgica, etc.

MUESTRA Es una parte de la población. Por ejemplo, los alumnos del 1er. semestre de la Carrera de Biología; los pacientes que se someten a amigdalectomía en el mes de enero, etc.

Algunas veces la muestra es REPRESENTATIVA de la población y otras no. Es importante seleccionar muestras que sí representen la característica de la población que se está estudiando.

CUADRO 15

Selecciona cuál es la muestra que representa más adecuadamente

El peso en bebés recién nacidos.

- a) Los que nacieron un día determinado.
- b) Los bebés varones.
- c) Los de cierto hospital.

Si a):

Tienes razón, esta muestra es representativa pues se toma al azar los niños que nacen ese día, sin importar su sexo, o su posición social, como en los

44.
otros dos casos. Esto hace que el valor obtenido al pesar a estos niños, sea parecido al de toda la población.

Si b):

No, porque generalmente los hombres pesan un poco más que las mujeres. Tu muestra haría que se supusiera un peso mayor del real. Es mejor considerar a los que nacen en cierto día, pues no importa su sexo, ni su posición social.

Si c):

No, porque si el hospital atiende principalmente a personas de escasos recursos, los niños tenderían a pesar menos; mientras que en el caso contrario, en un hospital de gente con recursos económicos, el peso calculado sería mayor al real. Es mejor considerar a los que nacen en cierto día, pues no importa su sexo, ni su posición social.

Bueno, con esto acabamos la primera lección.

(Se despliegan los resultados obtenidos en las preguntas hechas a lo largo de la lección).

¡Hasta la próxima!

LECCION 2

¿QUE SON LOS DATOS?

ESCALAS DE MEDICION

L E C C I O N 2

QUE SON LOS DATOS

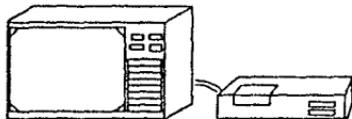
OBJETIVOS:

CONOCIMIENTOS: Al terminar la lección, el alumno podrá definir las cinco escalas de medición: nominal, ordinal, de intervalo, de razón y absoluta. Podrá definir y diferenciar constante, variable y parámetro. Podrá también distinguir la diferencia entre una variable continua y una discreta.

HABILIDADES: El alumno debe poder seleccionar la escala adecuada para medir distintos eventos, o describir las propiedades de un ente.

ACTITUDES: Se combatirá el error de creer que el solo hecho de usar números para representar a un evento, da la facilidad de hacer con ese evento las mismas operaciones que se hacen con los números.

Actualmente recibimos información por muchos medios:



La televisión



Los periódicos



El teléfono



Los censos de población

y muchos mas...

Todos estos medios nos proporcionan datos.

La materia prima de la Estadística son los datos.

A veces son datos obtenidos "de primera mano", pero otras, los científicos deben conformarse con datos de fuentes indirectas.

CUADRO 22

¿Y qué son esos datos que maneja la Estadística?

- a) Es cualquier tipo de información: imágenes, sonidos, palabras, números
- b) O es algo más específico

¿Tu qué opinas?

CUADRO 23

Si responde a) (Cualquier información):

No, no es posible que se usen los métodos estadísticos en una gama tan amplia de objetos. Recuerda que esta herramienta (especialmente la Estadística Inferencial), está basada en la Teoría de las Probabilidades, la cual a su vez está basada en las Matemáticas, que trabaja con números.

La realidad es muy compleja, es por eso que sólo podemos analizar una parte de ella a la vez, una de sus propiedades o atributos. Esta propiedad debe poder "medirse".

La materia prima de la Estadística son sólo valores numéricos, cantidades resultado de una medición.

CUADRO 24

Si b) (Algo más específico):

Tienes razón, no es posible que se usen los métodos estadísticos en una gama tan amplia de objetos como la que se menciona en el inciso a). Recuerda que esta herramienta (sobre todo la Estadística Inferencial), está basada en la Teoría de las Probabilidades, la cual a su vez está basada en las Matemáticas, que trabaja con números.

La materia prima de la Estadística son sólo valores numéricos, cantidades resultado de una medición.

CUADRO 25

No siempre los datos que se quieren estudiar son números. Es por ello que, hábilmente, se usan varias escalas de medición que permiten "convertir" cosas no numéricas a números manipulables con los métodos estadísticos. No siempre son números, pero siempre es posible que se "vuelvan" números.

Así que, en sentido amplio, todas las cosas pueden ser objeto de estudio de la estadística, siempre que se use la escala adecuada para convertirlas en números.

Estos valores numéricos representan alguna propiedad específica de lo que se está observando, y hay que tener mucho cuidado con el manejo que se haga de ellos.

CUADRO 26

Hay cinco diferentes escalas de medición que son, en orden ascendente según su precisión:

- Nominal.
- Ordinal.
- De Intervalo.
- De razón.
- Absoluta.

Las características de una escala, que determinan su nivel de precisión son:

- Orden en los números.
- Distancia entre los números.
- Un origen llamado cero

CUADRO 27

La primera escala que veremos es

NOMINAL: (del latín NOMEN = nombre) al usarla, se "etiqueta" a los nombres con números, (también pueden usarse otros símbolos, como letras, para hacerlo).



BUDISMO = 1



ISLAMISMO = 2



JUDAISMO = 3



CRISTIANISMO = 4



Avión = 1



Coche = 2



Barco = 3



Tren = 4

En ninguno de estos casos, los números quieren decir que el coche (2) es más que el avión (1), ni que el coche (2) es menos que el barco (3) o que el tren (4), es decir, no hay un orden en estos valores.

CUADRO 28

ORDINAL: Sigue el principio de ORDEN de los números reales.

En la recta real:

—1 0 1 2 3 4 5 6 7 ...

1 es mayor que 0 y menor que 2

$$0 < 1 < 2$$

Veamos un ejemplo: se pueden asociar los rangos militares en el Ejército Mexicano a los números reales:

- 17 = Presidente de la República. (Jefe del Ejército).
- 16 = Secretario de la Defensa. (General de Cuatro Estrellas).
- 15 = General de División
- 14 = General de Brigada
- 13 = General Brigadier
- 12 = Coronel
- 11 = Teniente Coronel
- 10 = Mayor
- 9 = Capitán Primero
- 8 = Capitán Segundo
- 7 = Teniente
- 6 = Subteniente

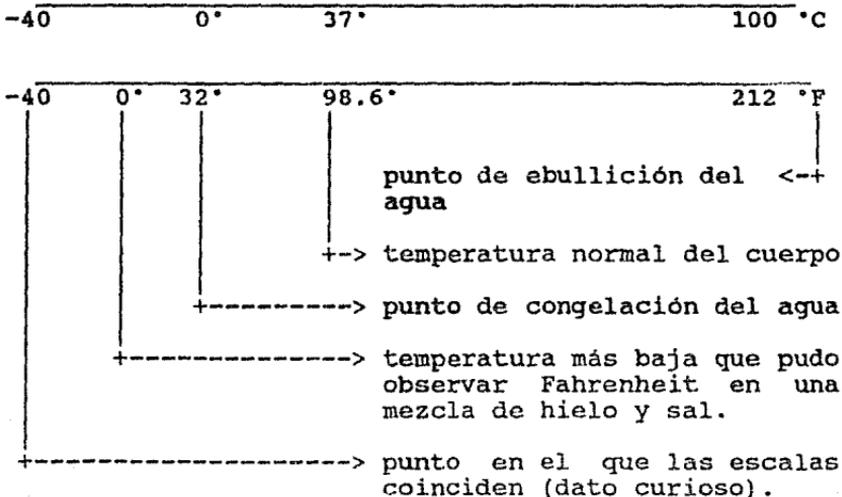
- 5 = Sargento Primero
- 4 = Sargento Segundo
- 3 = Cabo
- 2 = Soldado de Primera
- 1 = Soldado Raso

¡Pero cuidado!. Esto no quiere decir que "Subteniente es el doble de Cabo y la mitad de Coronel" como se podría decir que 6 es el doble de 3 y la mitad de 12.

En esta escala se sigue el principio de orden de los números reales, PERO NO EL DE DISTANCIA.

CUADRO 29

DE INTERVALO: sigue el principio de orden de los números reales (como la ordinal), y además el de distancia, pero no el de origen.



Observa que se guarda proporcionalidad entre las dos escalas. Los INTERVALOS son del mismo tamaño. La distancia del 0°C al 37°C es la misma que del 32°F al 98.6°F.

Veamos esto con números: calculando la proporción que hay de cada uno de estos puntos al punto más alto en la escala, tenemos:

$$^{\circ}\text{C} \quad (100 - 0) / (100 - 37) = 1.59$$

$$^{\circ}\text{F} \quad (212 - 32) / (212 - 98.6) = 1.59$$

Así que la proporcionalidad de las diferencias (intervalos) se preserva al pasar de una escala a otra.

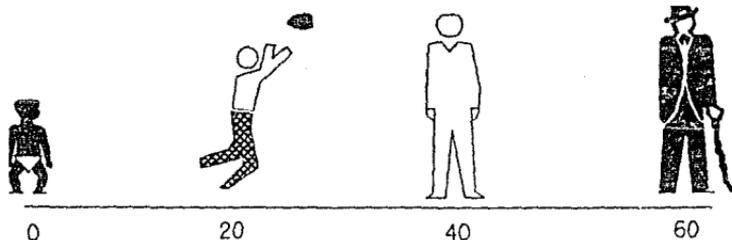
Nota que no importa con qué escala se mida, el punto de ebullición o de congelación del agua siempre es el mismo, aunque el valor numérico cambie.

El cero en estas escalas fué fijado convencionalmente, no indica ausencia de temperatura, no es el origen, se puede cambiar si así se desea.

CUADRO 30

DE RAZON: Sigue los principios de orden, distancia y origen de los números reales.

Por ejemplo, la edad de las personas



Esta escala cumple los principios:

- a) orden: bebé < joven < edad madura
- b) distancia: 25 años \times 2 = 50 años
- c) origen: 0 = nacimiento

Sin embargo, la edad puede medirse en años, o en meses, en días y hasta en horas o minutos si así lo quisieramos, y el resultados sería equivalente.

CUADRO 30-A

ABSOLUTA: Existen fenómenos en los que no se puede cambiar la unidad en que se miden. Veamos un ejemplo: si quisieras contar las personas que hay en un salón de clases, puedes pensar que conviene usar la escala de razón, pues este evento cumple los siguientes principios:

a) hay orden, pues tres personas son menos que cinco personas ($3 < 5$).

b) hay distancia, ya que si en un salón hay 10 personas y en otro cinco, en el primero hay el doble que en el segundo.

c) hay origen, pues cero significa que no hay nadie en el salón.

Pero además de esto, sólo puedes contar personas. No se puede usar otra unidad, como "alumnos", porque dentro del salón también podría haber maestros o inspectores, o visitantes. Tampoco se podría pensar en brazos porque si, desafortunadamente a alguna de las personas le faltara uno, la cuenta ya no sería equivalente.

Esto es lo que caracteriza a la escala absoluta: una unidad de medida única.

CUADRO 31

Resumiendo:

PRINCIPIOS DE LOS NUMEROS REALES

ESCALA	ORDEN	DISTANCIA	ORIGEN	UNIDAD UNICA
Nominal	no	no	no	no
Ordinal	SI	no	no	no
Intervalo	SI	SI	no	no
De Razón	SI	SI	SI	no
Absoluta	SI	SI	SI	SI

CUADRO 32

Selecciona la escala adecuada para medir los siguientes eventos

- Usa: a) escala nominal
 b) escala ordinal
 c) escala de intervalo
 d) escala de razón
 e) escala absoluta

(Tip: trata de usar primero la escala de razón, que es la que más se parece a los números reales, si no se puede, intenta con la de intervalo; si no es posible, la ordinal y, por último la nominal.)

- 1.- Número de hijos en un grupo de familias
- 2.- El tiempo (es decir, ¿qué escala usa la carátula de un reloj de manecillas?).

- 3.- Partido político al que apoya una persona.
- 4.- La velocidad
- 5.- Los médicos usan para medir el dolor las siguientes convenciones Si el paciente reporta poco dolor, es un dolor +; si reporta regular, ++; y si reporta mucho, +++. Así pueden indicar si aumentó de + a ++, o si disminuyó de +++ a ++. ¿A qué escala corresponde esto?

Evaluación:

- 1.- No. de hijos de una familia

Si a) (Nominal):

Es cierto que a cada hijo debe dársele un nombre, pero recuerda que siempre que se usa una escala es con el fin de hacer los datos mas manipulables. Y de seguro es mucho más fácil decir "tengo 4 hijos", que decir "tengo como hijos a Juan, Ana, Laura y Javier.

Si b) (Ordinal):

¿No crees que te estás complicando la vida?. Estás convirtiendo de una escala que sí cumple el principio de distancia de los números reales, a otra que no lo cumple, lo cual hace que los valores numéricos te proporcionen menos información.

Si c) (de intervalo):

Si así fuera, debería cumplir los principios de orden y de distancia de los números reales, pero no así el de origen. Pero tener 0 hijos significa no tener, o sea el origen. Así que no es esta la escala adecuada.

Si d) (De razón):

¡Muy bién! Este ejemplo cumple con los principios de orden, de distancia y de origen de los número reales.

2.- El tiempo (es decir, ¿qué escala usa la carátula de un reloj de manecillas?).

Si a) (Nominal):

Esta variable puede medirse con la escala de intervalo, pues a las 4 00 es más temprano que a las 6 00, es decir 4 es menor que seis, así que esta no es la escala adecuada. (Hay orden y distancia).

Si b) (Ordinal):

Es cierto que el tiempo cumple el principio de orden de los números reales, pero cumple también el de distancia, pues 3 horas es igual a 3 veces una hora, así que no es esta la escala adecuada.

Si c) (De intervalo):

Tienes toda la razón, pues la carátula de un reloj se asemeja a los reales en cuanto a orden y distancia, pero no en cuanto a origen, pues el cero no representa la "no hora", sino que es un punto convencional que representa al mismo tiempo las 0 00 horas y las 12 00 horas.

Si c) (De razón):

No, pues el cero no representa la "no hora", sino que es un punto convencional que representa al mismo tiempo las 0 00 horas y las 12 00 horas. No se puede considerar, por lo tanto, un origen.

3.- Partido político que apoya una persona

Si a) (Nominal):

Tienes razón. Si se le asignara un número a cada partido político, sería únicamente para manejar más fácilmente los datos, no porque uno estuviera por encima de otro.

Si b) (Ordinal):

No se puede usar esta escala, porque, aparte de las preferencias personales de cada quien, no hay un criterio objetivo para decidir cuál partido va antes y cuál después.

Si c) (De intervalo):

No puedo imaginar qué intervalos definirías.

Si d) (De razón):

¿Realmente lo crees?. ¿Cuál es el origen?, ¿cuál la distancia y cuál el orden?

4.- La velocidad

Si a) (Nominal):

Qué nombre le pondrías a la velocidad que hay entre los 60 km/h y los 60.5 km/h. Recuerda que al usar esta escala, se trata de "bautizar" a cada evento con un nombre que sea equivalente a un número, pero los datos que estás manejando ya son números, por lo tanto, esta escala no te sirve.

Si b) (Ordinal):

Esta escala se usa cuando los valores de la variable que se estudia no comparten con los números reales los principios de distancia y de origen. Pero esta variable si los tiene.

Si c) (De Intervalo):

Lo siento, hubiera sido una buena elección si la velocidad 0 no existiera (estar sin movimiento), es decir, si el principio de orden de los números reales no se cumpliera aquí.

Si d) (De razón):

¡Muy bien! Esta es la escala adecuada para medir la velocidad.

5.- Los médicos usan para medir el dolor las siguientes convenciones Si el paciente reporta poco dolor, es un dolor +; si reporta regular, ++; y si reporta mucho, +++. Así pueden indicar si aumentó de + a ++, o si disminuyó de +++ a ++. ¿A qué escala corresponde esto?

Si a) (Nominal):

No es posible dar un nombre a cada grado de dolor, pues son muchísimos. Pero si es posible decir que un dolor es mayor o menor que otro; así que esta no es la escala adecuada.

Si b) (Ordinal):

Tienes razón. Esta es la escala correcta, pues si se cumple el principio de orden de los números reales, pero (a pesar de que como convención se aumenta el número de +'s), no es correcto decir que un dolor ++ es el doble de un dolor +, simplemente se puede afirmar que es mayor.

Si c) (De intervalo):

Estás en un error, un error muy común, pues la convención de usar un cierto número de +'s genera confusión. Recuerda que esta escala se caracteriza por cumplir el principio de distancia de los números reales. Se piensa que un dolor + es la mitad de otro ++, pero esto no es posible. Hay muchísimos grados de dolor, pero no se puede hacer operaciones con ellos como con los números, y no es posible "sacar mitad de un dolor".

Si d) (De razón):

Aunque esta variable cumple el principio de origen de los números reales, (se puede decir que cuando no hay dolor, estamos en el origen), no cumple el de distancia. Hay una confusión al pensar que un dolor + es la mitad de otro ++, pero esto no es posible. Hay muchísimos grados de dolor, y no se pueden hacer operaciones con ellos como con los números, así que no es posible "sacar mitad de un dolor".

CUADRO 33

Veamos ahora qué se mide con estas escalas.

Seguramente has oído los términos CONSTANTE, VARIABLE Y PARAMETRO. Vamos a ponernos de acuerdo en su definición para usarlos adecuadamente a lo largo de estas lecciones.

CONSTANTE: Es un valor fijo, que no cambia. Por ejemplo, la velocidad de la luz (300,000 km/seg)

VARIABLE: Es cualquier valor de un dato. Por ejemplo, el pH de una solución. (pH mide la acidez o la alcalinidad, es la concentración de los iones de Hidrógeno de una solución).

PARAMETRO: Es una variable que permanece constante en un problema dado. Por ejemplo la temperatura a que se realiza un experimento. (Bueno, los científicos estarían felices de que esto fuera posible, ... en realidad es un concepto más teórico que real) Esta es la idea menos clara, porque está en medio de las otras dos.

Veamos un ejemplo que aclare un poco más el concepto de parámetro

Recuerdas la ecuación de la recta?

$$y = mx + b$$

donde:

x, y - son las coordenadas de los puntos que caen sobre una recta. Son las variables, pueden tomar una gran cantidad de valores distintos.

m - es la pendiente de la recta. Es la tangente del ángulo que forma la recta con el eje de las Y 's.

b - es la ordenada al origen. El punto donde la recta corta al eje vertical.

¿Cuál de estos valores hace a una recta diferente de otra?

Evaluación:

Si x o y :

No, repasa un poco más la ecuación de la recta

Si m:

No sólo ese valor, también es importante saber dónde corta al eje Y. Pues no es lo mismo una recta que corta sobre el punto (0,3) que otra que corta en el punto (0,-100)

Si b:

No solamente ese valor, pues aunque se está fijando un punto, sobre él pueden pasar muchísimas rectas diferentes con sólo cambiar el ángulo de inclinación.

Lo que distingue una a recta de otra son m y b, que son los parámetros de la recta, pues una vez que se fijan para una recta dada, permanecen constantes.

Otro ejemplo más cotidiano de lo que es un parámetro, es el salario mínimo. Cuando se calculan los impuestos que deben pagar los empleados de un Departamento, el salario mínimo vigente es el que permanece constante para todas las personas.

CUADRO 34

Hablemos ahora de los tipos de variables que existen:

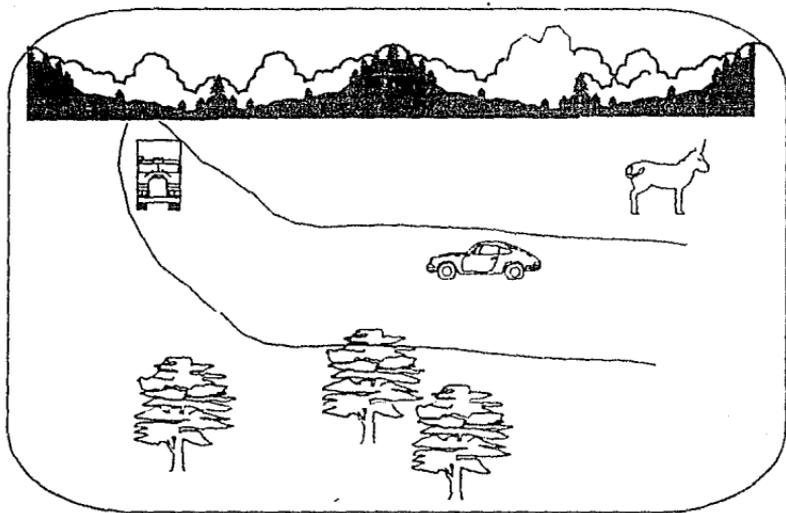
Las variables pueden ser:

- Continuas
- Discretas.

Son CONTINUAS: si teóricamente pueden tomar cualquier valor de entre dos valores posibles de una escala de medición. Por ejemplo edad, distancia, concentración.

Su gráfica es continua.

La distancia es una variable continua.

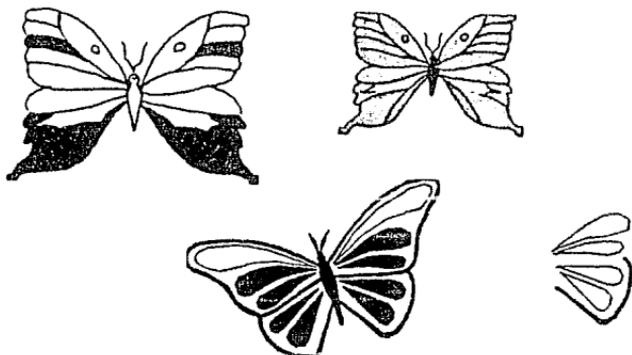


Cuando queremos ir de un lugar a otro, tenemos que pasar por todos los puntos de enmedio, aunque no nos guste. Aún si tomamos otro camino, en este atajo debemos recorrer todos los puntos, en forma continua.

CUADRO 35

Son **DISCRETAS**: cuando sólo pueden tomar valores escalonados o discontinuos. Hay saltos o interrupciones en los valores que pueden tomar.

El número de ejemplares
en una muestra es discreto.



No tiene sentido hablar de media mariposa.

CUADRO 36

A veces sucede que una variable es continua, pero la "regla" con la que la medimos no tiene la precisión que requerimos, y entonces nos vemos obligados a considerar a la variable como discreta.

Por ejemplo: Supongamos que se quiere medir la estatura de las personas, no se acostumbra decir que alguien mide 1.65743543 metros porque no hay reglas con divisiones tan precisas. Peor aún, puede ser que el observador no tenga regla y sólo pueda manejar una escala nominal (propia exclusivamente de las variables discretas), en la que los valores sean alto, mediano y bajo.

Hay que distinguir estos casos de aquellos en que la variable es intrínsecamente discreta, por ejemplo el sexo.

CUADRO 37

Asocia ahora qué escala debe usarse con qué tipo de variable

- | | |
|-----------------|----------------------|
| 1) Nominal | a) Variable discreta |
| 2) Ordinal | b) Variable continua |
| 3) De intervalo | |
| 4) de razón. | |

Forma las parejas con un número y una letra que representen la escala y la variable que puede relacionarse adecuadamente. Indica todas las posibles parejas (sin espacios ni comas, por ejemplo: 1a)

Son seis las respuestas correctas.

Evaluación:

Si 1a) Nominal - discreta.

¡Bien! Por ejemplo los posibles estadios de una mariposa.

Si 1b) Nominal - continua.

No, no puedes enumerar todos los puntos de una

recta, (que es la representación gráfica de todos los valores de una gráfica continua), por lo tanto no puedes usar esta escala si la variable es continua.

Si 2a) Ordinal - discreta.

¡Bien! Por ejemplo los grados académicos.

Si 2b) Ordinal - continua.

No, esta combinación no es posible, aunque tal vez pensaste en algo así como el dolor, que es continuo. Por deficiencias en los procedimientos de evaluación, tenemos que considerar al dolor como discreto.

Si 3a) De intervalo - discreta.

¡Bien!, por ejemplo el año de nacimiento, pues el cero (ya no vive nadie que haya nacido el año 0 de nuestra era) no representa el origen, pero si se conserva el principio de distancia.

Si 3b) De intervalo - continua.

¡Bien! Por ejemplo el tiempo.

Si 4a) De razón - discreta.

De acuerdo. Por ejemplo el número de enfermos que responden a un fármaco.

Si 4b) De razón - continua.

¡Claro!, por ejemplo la presión arterial.

(Se despliegan los resultados obtenidos por el alumno a lo largo de la lección).

¡Hasta la vista!

LECCION 3

LA ORGANIZACION Y EL RESUMEN DE LOS DATOS

La organización y el resumen de los datos

OBJETIVOS

CONOCIMIENTOS: Al terminar la lección el alumno podrá enumerar las ventajas de organizar y resumir los datos y podrá definir los conceptos de frecuencia, frecuencia acumulada, frecuencia relativa y frecuencia relativa acumulada.

HABILIDADES: Podrá elaborar una tabla, ordenando y agrupando datos para presentarlos con mayor claridad; será capaz de reconocer y entender un histograma de distribución de frecuencias, y podrá dibujar un polígono de frecuencias.

ACTITUDES: El alumno deberá querer agrupar organizadamente los datos para obtener información más completa de ellos.

NOTA IMPORTANTE

Si en cualquier momento necesitas usar una calculadora, puedes llamarla tecleando <Alt> C. Debes oprimir estas dos teclas al mismo tiempo. Si quieres ayuda al usar la calculadora, teclea H (Help). Si quieres traer un resultado obtenido con la calculadora oprime <Alt> F.

Para calificarte adecuadamente, por favor en todas las respuestas que no sean enteras, usa sólo 2 decimales, redondeando si es necesario. (0-4 al dígito anterior y 5-9 al superior). Si no lo haces así, se te puede considerar mal una respuesta correcta.

Vamos a practicar llama a la calculadora y calcula $12.1 * 5.33$. (multiplicación). Trae el resultado a la lección con <Alt> f. O retira la calculadora con <Esc> y teclea tú el resultado.

Si responde 64.493:

Bién, pero olvidaste redondear a dos dígitos decimales significativos. Dame nuevamente el resultado.

Si el resultado es 64.49:

¡Perfecto!, comprendiste la idea. Si quieres, practica un poco más. Recuerda que los signos de operación son más +; menos -; por * y entre /. (Si ya no quieres practicar da <Return>).

Si la respuesta es algo diferente:

No, lee con atención como se usa, y vuelve a contestar.

CUADRO 41

Hay una gran diferencia entre tener datos revueltos, sin orden ni concierto, y organizarlos adecuadamente para que nos proporcionen información.

Por ejemplo:

DATOS

Luis López	Doctor
30 años	Premio Nobel
experimento	Medicina
1834	gatos
Juan Pérez	cáncer

INFORMACION

El Doctor Juan Pérez, de sólo 30 años de edad, recibió el Premio Nobel de Medicina al haber desarrollado experimentos que permitieron avanzar en la lucha contra el cáncer.

CUADRO 42

Quando se hacen observaciones de una variable cualquiera, casi siempre los valores observados quedan desordenados. Veamos un ejemplo:

Se aplicó un test a 38 personas para medir su coeficiente intelectual (IQ), y los resultados obtenidos fueron:

154	100	110	131	109	111	128	116
133	105	93	130	127	145	130	116
116	150	104	142	124	151	154	97
128	118	103	125	121	118	154	
95	97	108	113	115	122	108	

Estos datos así presentados, no dan una imagen clara de lo que significan. Se necesita un rato de observación para averiguar información tan elemental como cuál es el dato mayor y cuál el menor.

¿Cuál es el menor?

Si responde 93, ¡Bien!

Si no, se le da la respuesta correcta (todas las preguntas de este tipo, se manejan así).

¿Y cuál el mayor?

71.

Si responde 154:

!Ok!

Si no elige bien ninguno de los valores:

No acertaste ninguna, ¿que no estás poniendo atención?

CUADRO 43

De cualquier manera, si tuvieramos la tabla ordenada sería muy fácil responder estas preguntas e incluso otras más complicadas.

Entonces ordenémosla. Tú veme dando los números del menor al mayor, y yo los acomodo

(Se presenta la tabla con los valores de los coeficientes, pero ahora con guiones en el lugar donde estaba el 93).

154	100	110	131	109	111	128	116
133	105	--	130	127	145	130	116
116	150	104	142	124	151	154	97
128	118	103	125	121	118	154	
95	97	108	113	115	122	108	

Ya habíamos dicho que el primero es 93, (nota que en la tabla ya se tachó).

¿Cuál es el segundo?

Si la respuesta es 95:

¡Bién! Lo tachó (y lo va marcando en la tabla)

¿Cuál es el tercero?

72.

Si responde 97:

¡Correcto! Lo tacho (dos veces, porque se repite en la tabla)

Bueno, como ordenar es una función más propia de una máquina (yo) que de un ser pensante (tú), yo termino.

La tabla ya ordenada queda:

93	-->	95	-->	97	-->	97	100	103	104	105
108	108	109	110	111	113	115	116			
116	116	118	118	121	122	124	125			
127	128	128	130	130	131	133	142			
145	150	151	154	154	154					

Como verás, ahora es muy fácil obtener el valor mayor y el menor.

Más aún, los datos así ordenados nos permiten:

- Ver qué valores no se presentaron, por ejemplo ni 96 ni 99.
- Ver cuáles se repiten y cuántas veces. Por ejemplo el valor 116 se repite tres veces.

CUADRO 44

Observa cuanta información adicional se logra al ordenar los datos. Si calculamos la frecuencia de cada valor, tendremos todavía más riqueza de estos datos. Ayúdame un poco a obtener esta frecuencia:

¿Cuántas veces se presentó cada dato? (Va solicitando los valores respectivos y evaluando las respuestas)

93: 1	95: 1	97: 2	100: 1
103: 1	104: 1	105: 1	108: 2
109: 1	110: 1	111: 1	113: 1
115: 1	116: 3	118: 2	121: 1
122: 1	124: 1	125: 1	127: 1
128: 2	130: 2	131: 1	133: 1
142: 1	145: 1	150: 1	151: 1
154: 3			

Con esta tabla ya podemos decir la frecuencia con que ocurre cada valor de la variable, de "un vistazo".

CUADRO 45

Sin embargo, la tabla anterior no nos permite ver si hay o no tendencia hacia algún valor. Es decir, no sabemos si la población que estamos observando tiene en conjunto, un IQ alto, medio o bajo.

Este tipo de preguntas se resuelve agrupando los datos. Definiendo intervalos para los valores observados y calculando la frecuencia de los valores observados que caen en cada intervalo.

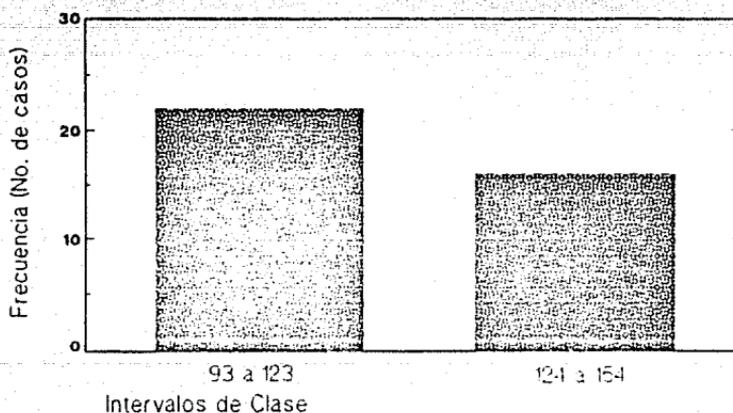
Seguramente querrás saber cómo se definen estos intervalos. Para ello debemos responder dos preguntas:

- 1a. ¿cuántos intervalos son convenientes?
- 2a. ¿de qué tamaño debe ser cada uno?

¿Cuántos intervalos?:

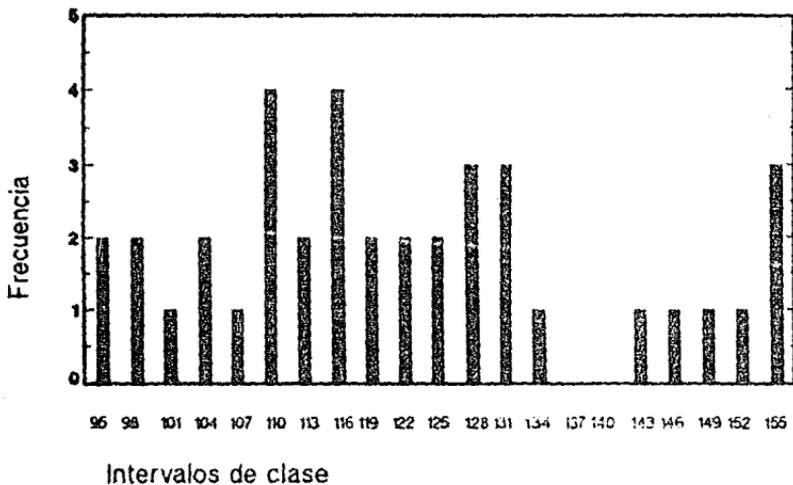
La mejor guía para responder esto es el conocimiento que se tenga de los datos. Pero también hay algunas ideas que nos pueden guiar

No es conveniente usar tan pocos intervalos que se pierda información relevante.



La única información que proporciona esta gráfica (en esta lección, un poco más adelante, aprenderás cómo se dibuja), es que hay más personas con IQ menor a 124 que mayor o igual a ese número.

Pero tampoco tantos
que no se logre una síntesis.



Es difícil concluir algo de una gráfica como esta. Nos da información similar a la tabla de frecuencias que recién hicimos, no tiene caso trabajar para obtener algo que ya sabemos.

CUADRO 48

Hay una fórmula, dada por Sturges para calcular el número de los intervalos:

$$k = 1 + 3.322 (\log n)$$

donde: k - no. de intervalos
n - no. de observaciones

La respuesta obtenida por esta Regla de Sturges no debe considerarse como definitiva, sino sólo como guía.

En el ejemplo que venimos estudiando, tenemos:

$$\begin{aligned} n &= 38 \\ k &= 1 + 3.322 (\log 38) \\ &= 1 + 3.322 (1.5798) \\ &= 6.25 \end{aligned}$$

Usemos entonces seis INTERVALOS DE CLASE (Se llaman de clase, porque definen una clase, una categoría para los datos). Podríamos usar 5, 7 u 8 si así lo consideráramos conveniente, por alguna razón.

CUADRO 49

La segunda pregunta es: ¿de qué tamaño deben definirse los intervalos?

Es recomendable que, siempre que sea posible, los intervalos tengan la misma amplitud.

La forma de calcular esta amplitud está dada por los siguientes pasos:

1o. ¿Cuántos valores posibles tenemos?

$$(\text{valor mayor} - \text{valor menor}) + 1$$

En nuestro caso:

$$(154 - 93) + 1 = 62 \text{ posibles casos}$$

2o. Si queremos 6 intervalos, cada uno debe tener

$$62 / 6 = 10.33 \text{ de amplitud}$$

Hagamos entonces cada intervalo de 11 puntos.

3o. Para calcular los extremos de cada intervalo

$$\text{límite superior} = \text{límite inferior} + (\text{amplitud} - 1)$$

4o. Entonces nuestros intervalos quedan definidos como sigue:

I	I	I	I	I	I	I					
93	103	104	114	115	125	126	136	137	147	148	158

El límite del último intervalo (158) es superior a nuestro valor observado más grande (154). Esto es necesario porque de otra forma el último intervalo sería de una amplitud menor a los demás.

Calcula la frecuencia de cada intervalo de clase
(Cuántos datos caen en cada uno de los intervalos)

1er. Intervalo 93 a 103	:	6
2o. Intervalo 104 a 114	:	8
3er. Intervalo 115 a 125	:	10
4o. Intervalo 126 a 136	:	7
5o. Intervalo 137 a 147	:	2
6o. Intervalo 148 a 158	:	5

Ahora veamos los resultados que habíamos calculado en forma de tabla

Intervalo	Frecuencia
93 - 103	6
104 - 114	8
115 - 125	10
126 - 136	7
137 - 147	2
148 - 158	5

	38

Esta tabla que hicimos se llama DISTRIBUCION DE FRECUENCIAS porque muestra la forma en que se distribuyen los valores de la variable observada.

CUADRO 51

(Preguntas de repaso)

¿Cómo se llaman los intervalos que venimos trabajando?

intervalos de clase

¿Cómo se llama la tabla que hiciste en el cuadro anterior?

distribución de frecuencias

Si no responde bien a ninguna de las dos preguntas:

Son importantes los conceptos pero también los nombres, ¡pon más atención!.

CUADRO 52

A veces puede ser útil conocer, más que el número, la proporción de valores que caen en un intervalo de clase. Es decir, considerar que el total de observaciones (en nuestro ejemplo 38), es el 100%, y calcular qué porcentaje representa el número de observaciones en cada intervalo.

Esta es la FRECUENCIA RELATIVA, relativa al total de los casos.

Vamos a calcular esta frecuencia relativa a los datos del ejemplo. Dejaremos a la vista nuestra tabla de distribución de frecuencias, para calcular las frecuencias relativas.

Intervalo	Frecuencia
93 - 103	6
104 - 114	8
115 - 125	10
126 - 136	7
137 - 147	2
148 - 158	5

Si 38 es el 100%, ¿qué porcentaje representa el 6 del primer intervalo? Recuerda que puedes usar la calculadora (con <Alt> C). Sólo da 2 dígitos decimales en la respuesta, redondeando si es necesario.

Evaluación:

Si la respuesta es 15.79:

¡Muy bien!

Si es 15.789 ó 15.78:

No olvides que debes redondear a dos decimales. Ya no te lo voy a recordar más adelante, es la última vez que la tomo como buena. ¿OK?

Si responde otra cosa, o solicita ayuda (tecleando el signo ?):

Repasemos la regla de tres:

si 38 es el 100%
6 qué porcentaje será

Se resuelve multiplicando 6 por 100 y dividiendo entre 38. Vuelve a contestar

¿Qué porcentaje representa el 8 del segundo intervalo?

Si responde 21.05:

¡Bien!

Si cualquier otra cosa:

No, esa no es la respuesta correcta. Vuelve a contestar

Como siempre, yo continuo el trabajo. La tabla queda:

Intervalo	Frecuencia	Frecuencia relativa
93 - 103	6	15.79%
104 - 114	8	21.05%
115 - 125	10	26.32%
126 - 136	7	18.42%
137 - 147	2	5.26%
148 - 158	5	13.16%
TOTALES	38	100.00%

La frecuencia relativa es muy útil para comparar la distribución de los valores de una serie con otras series similares de tamaño diferente.

Si quisieras saber, por ejemplo, cuántos valores caen del 104 al 125 (Porque observaste que en esos intervalos está el mayor número de los casos), podrías usar la frecuencia acumulada, es decir, la suma de las frecuencias de los intervalos que te interesan. En este caso $8 + 10 = 18$ para la frecuencia absoluta (que generalmente se llama sólo frecuencia), y $21.05 + 26.32 = 47.37$ para la frecuencia relativa.

Pueden acumularse las frecuencias y frecuencias relativas para facilitar la obtención de información acerca de la frecuencia dentro de dos o más intervalos de clase continuos.

CUADRO 54

Ahora vamos a terminar, con tu ayuda la tabla completa:

Yo te indico ("flashando") qué resultado requiero, y tu me lo vas dando

Intervalo	Frecuencia	Frecuencia acumulada	Frecuencia relativa	Frecuencia relativa acumulada
93 - 103	6	6	15.79%	15.79%
104 - 114	8	—	21.05%	—%
115 - 125	10	—	26.32%	—%
126 - 136	7	—	18.42%	—%
137 - 147	2	—	5.26%	—%
148 - 158	5	—	13.16%	—%
TOTALES	38	—	100.00%	—%

(Va solicitando y evaluando cada uno de los datos que faltan en esta tabla hasta llegar a la frecuencia relativa acumulada del tercer intervalo y entonces:)

Para no variar, te evito la tarea repetitiva, y la hago yo ¿OK?

Intervalo	Frecuencia	Frecuencia acumulada	Frecuencia relativa	Frecuencia relativa acumulada
93 - 103	6	6	15.79%	15.79%
104 - 114	8	14	21.05%	36.84%
115 - 125	10	24	26.32%	63.16%
126 - 136	7	31	18.42%	81.58%
137 - 147	2	33	5.26%	86.84%
148 - 158	5	38	13.16%	100.00%
TOTALES	38	38	100.00%	100.00%

Ahora si, ya con la tabla completa, si quieres conocer los valores que caen entre el 104 y el 125, puedes restar 6 (del intervalo 93-103) al 24 (del intervalo 115-125), y obtienes el mismo 18 que ya habíamos calculado antes.

CUADRO 55

Estos resultados pueden verse más fácilmente en forma gráfica. A la gráfica que veremos aquí se le llama HISTOGRAMA.

Para construir el histograma se considera:

En el eje horizontal - los valores de la variable observada, considerando los intervalos definidos.

En el eje vertical - Su frecuencia (absoluta o ^{84.} relativa, simple o acumulada, la que quieras graficar).

Por encima de cada intervalo de clase, se levanta sobre el eje vertical, una barra cuya altura representa la frecuencia correspondiente al intervalo.

CUADRO 56

Dado que el IQ es una variable continua, las barras del histograma deben quedar unidas, y para ello se considera que los límites verdaderos de cada intervalo son:

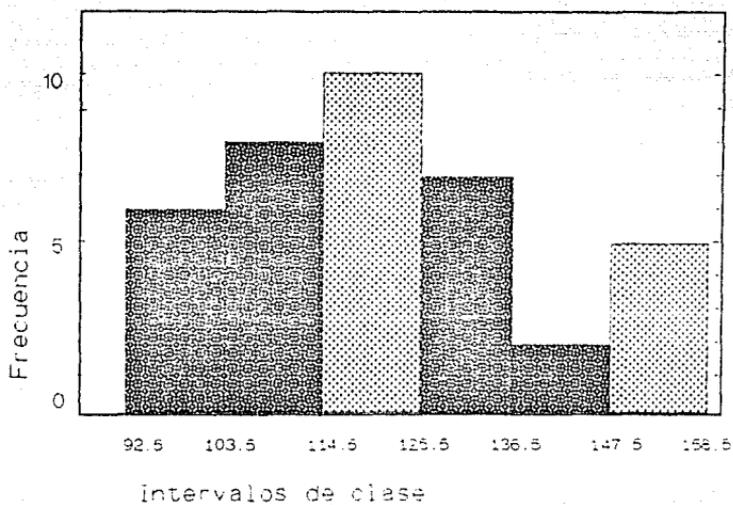
$$\text{límite inferior verdadero} = \text{límite inferior} - 0.5$$

$$\text{límite superior verdadero} = \text{límite superior} + 0.5$$

Entonces los límites verdaderos de nuestros intervalos son:

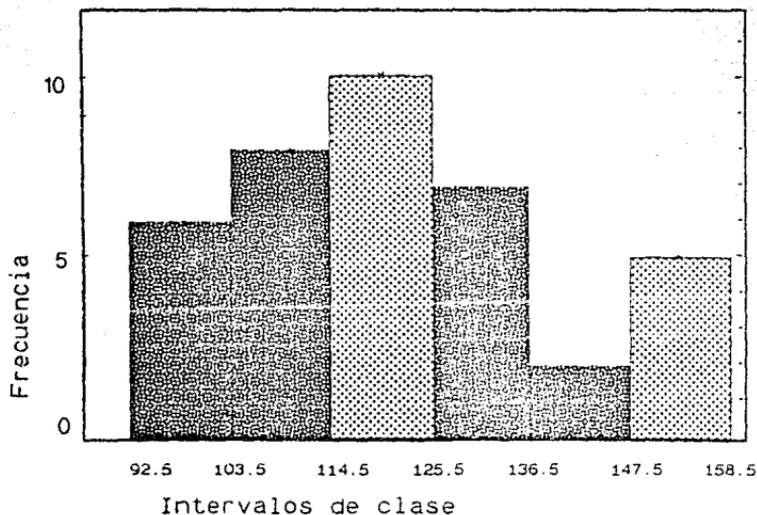
92.5 - 103.5
103.5 - 114.5
114.5 - 125.5
125.5 - 136.5
136.5 - 147.5
147.5 - 158.5

Y el histograma de frecuencias quedará:



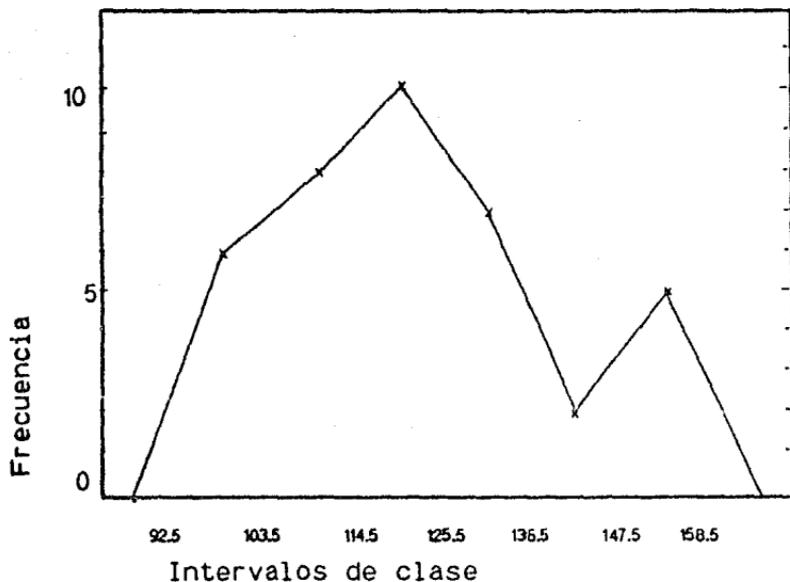
Hay otra forma gráfica para una distribución de frecuencias. Se llama POLIGONO DE FRECUENCIAS, y se obtiene de la siguiente manera:

- 1.- Se marca el punto medio de cada intervalo de clase en la parte superior de la barra asociada. Estos puntos se llaman "marcas de clase".



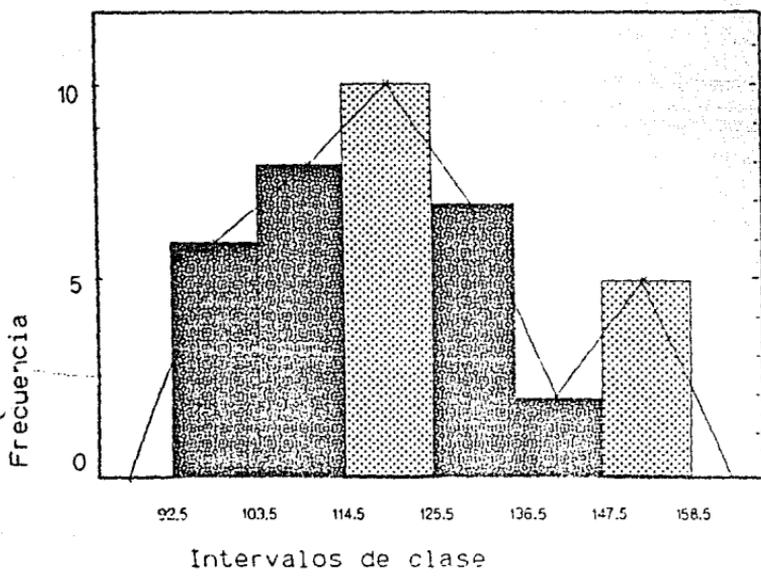
CUADRO 58

2.- Se unen los puntos así obtenidos, llegando en el eje horizontal hasta los puntos que serían el punto medio si hubiera un intervalo adicional a cada extremo del histograma.



CUADRO 59

Un dato interesante es que siempre, el área total bajo el polígono de frecuencias es igual al área bajo el histograma.



CUADRO 60

En esta gráfica se puede ver que:

1o. Hay menos casos observados en los intervalos del principio y del final, y más en el los intervalos centrales, con excepción del último intervalo, en el que se ve un incremento.

2o. La mayor parte de nuestras observaciones caen en el intervalo 115-125, donde también cae el valor medio de los datos observados, que está dado por: $(156 - 93)/2 = 124.5$.

3o. Hay primero una tendencia a crecer, se llega al valor máximo, y después hay decrecimiento. Nuevamente, el último intervalo no se comporta como el resto.

Estas observaciones nos permiten concluir que la población que estamos estudiando, tiene un Coeficiente Intelectual medio (alrededor de 125), con algunos casos de personas con un IQ bastante más alto.

CUADRO 66

Es necesario enfatizar la necesidad de relativizar en Estadística.

Algunos de los errores más comunes en las manipulaciones de datos al aplicar la Estadística, es precisamente no relativizar, que es tanto como no dar un marco de referencia.

Un buen ejemplo lo encontramos en los informes del Gobierno. Por ejemplo, en la propaganda que dice "en 1987 se construyeron 20,000 aulas en México, el doble que en 1977", se puede estar omitiendo información muy importante porque probablemente en 1977, 10,000 aulas representaron relativamente más que 20,000 en 1987. Para que este dato fuera valioso, habría que relativizarlo al número de niños en edad escolar en cada una de estas fechas.

(Se muestran los resultados obtenidos por el alumno durante la lección).

Bueno, con esto terminamos las tercera lección, esperamos haya sido de tu agrado, y seguir contando con tu agradable compañía.

LECCION 4

MEDIDAS DE TENDENCIA CENTRAL

LECCION 4

Medidas de Tendencia Central

OBJETIVOS:

CONOCIMIENTOS: Al terminar la lección, el alumno podrá definir las medidas de tendencia central: moda, mediana, media aritmética y media aritmética ponderada. Podrá enumerar las principales características de cada medida, y decir qué medida de tendencia central ayuda a explicar cada variable, cuando se está usando tal o cual escala para representarla.

HABILIDADES: El alumno podrá saber cuando debe usarse cierta medida de tendencia central para describir una población.

ACTITUDES: El alumno debe querer conocer una población por medio de sus parámetros.

NOTA IMPORTANTE:

¿Ya sabes usar la calculadora?

Si responde si:

Entonces ya no hace falta estudiarla, pasamos directamente a la lección.

Si responde no:

Si en cualquier momento necesitas usar una calculadora, puedes obtenerla tecleando <Alt> C. Debes oprimir estas dos teclas al mismo tiempo. Si quieres ayuda al usar la calculadora, teclea H (Help). Si quieres traer un resultado obtenido con la calculadora oprime <Alt> F.

Para calificarte adecuadamente, por favor en todas las respuestas que no sean enteras, usa sólo 2 decimales, redondeando si es necesario. (0-4 al dígito anterior y 5-9 al superior). Si no lo haces así, se te puede considerar mal una respuesta correcta.

Vamos a practicar: llama a la calculadora y calcula $12.1 * 5.33$. (multiplicación) Trae el resultado con <Alt> f

Si responde 64.493:

¡Bien, pero olvidaste redondear a dos dígitos decimales significativos. Dame nuevamente el resultado.

Si responde 64.49:

¡Perfecto!, comprendiste la idea. Si quieres, practica un poco más. Recuerda que los signos de operación son: más +; menos -; por * y entre /. Para dejar de usarla, da <Esc> y luego <Return>.

CUADRO 81

Si quieres describir una población, tienes dos opciones:

a) Enunciar exhaustivamente los elementos que la componen. (Recuerda los expedientes de la facultad de Ciencias que mencionamos en la lección 1).

b) Hallar unas cuantas medidas que la describan, es decir, hallar valores que sean típicos, representativos, para ese conjunto de datos.

¿Recuerdas que en la lección 1 concluimos que es mejor la segunda opción?. En esta lección y la siguiente hablaremos de esas medidas descriptivas, que se dividen en dos grupos:

1.- Medidas de Tendencia Central.

2.- Medidas de Dispersión.

Estas medidas se aplican tanto a poblaciones como a muestras. Aunque ahora nos referiremos principalmente a poblaciones, a lo largo de esta lección también veremos algo relativo a muestras.

CUADRO 82

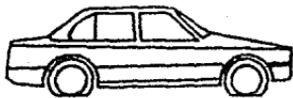
Medidas de Tendencia Central:

Los investigadores en muchos campos usan el término "promedio" en respuesta a preguntas tales como:



¿Cuántos cigarrillos fuman los jóvenes?

¿Cuál es la demanda de ingreso a la Universidad?



¿Cuántos accidentes automovilísticos son causados por el abuso del alcohol?

Estas preguntas se pueden responder con las medidas de tendencia central, que son diferentes "estilos" de obtener promedios.

CUADRO 83

Las medidas de tendencia central son, entre otras:

- Media aritmética.
- Moda.
- Mediana.
- Media aritmética ponderada.

Las tres primeras son las más importantes y las más utilizadas.

CUADRO 84

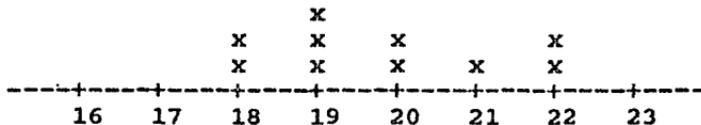
MEDIA ARITMETICA

Es la medida más importante y también la más conocida. Es en la que la mayoría de las personas piensan cuando se habla del "promedio". Se calcula sumando todos los datos, y dividiendo entre el número de datos.

Por ejemplo, supongamos que se quiere conocer la edad promedio de un grupo de 10 personas, de quienes se tienen los siguientes datos:

Carlos: 20, Elena: 19, Francisco: 19, Rosa: 22,
 Juan: 18, Enrique: 20, Ana: 21, Ernesto: 22,
 María: 19 e Irma: 18

Esto se puede representar gráficamente:



Para calcular la media,

- que se designa por μ si es media de la población
- o por \bar{X} con una línea horizontal arriba (le llamaremos \bar{X} barra porque no puedo dibujarla), si es la media muestral,

se proceda como sigue:

$$\begin{aligned}\mu &= (20+19+19+22+18+20+21+22+19+18) / 10 \\ &= 198 / 10 \\ &= 19.8 \text{ años promedio.}\end{aligned}$$

CUADRO 85

Para generalizar estos conceptos, usémos la siguiente notación.

- . X - (con mayúscula), variable en estudio, llamada variable aleatoria.
- . x - (con minúscula) valores específicos de la variable aleatoria. Para distinguir una observación de otra, se usa un subíndice. En general, el valor de una variable será x_i , y el último valor de una población con N elementos será x_N .
- . N - es el número de elementos de la población.
- . μ - la letra griega mu que representa la media de la población.
- . \bar{X} - representa la media de la muestra.

Veamos algunas preguntas para reafirmar estos conceptos:

En nuestro ejemplo, ¿Cómo se llama la variable en estudio?, ¿Qué representa X?

La edad

¿Cuánto vale N? (¿Cuál es el número de elementos de la población?)

vale 10

Considerando el orden en que se presentan, ¿cuánto vale x_6 ?

la edad de Enrique: 20 años

¿Cuánto vale x_2 ?

la edad de Elena: 19 años

¿Cuánto vale x_N ?

x_N es el último elemento, en este ejemplo x_{10} y representa la edad de Irma: 18 años.

Si responde bien a todo: ¡Todas buenas! Muy bien.

CUADRO 86

La fórmula general para calcular la media de una población con N elementos es:

$$\frac{\sum_{i=1}^N x_i}{N}$$

El símbolo: $\sum_{i=1}^N$

se lee, suma de los valores desde que i vale 1 hasta N. El símbolo Σ es la letra griega sigma, inicial de la palabra "suma".

La media aritmética se usa con variables CONTINUAS o DISCRETAS, siempre que se usen escalas DE INTERVALO o DE RAZON. No se puede sacar un promedio de datos manejados con escala ordinal, (¿te acuerdas del ejemplo de los grados de dolor?); y mucho menos con escala nominal.

Calcula la media de las siguientes poblaciones: Recuerda que la calculadora se llama con <Alt> C.

a) La duración en días de la convalecencia de personas a quienes se les practicó apendicectomía (extracción del apéndice).

15,19,20,26,27,28,32,35

¿Cuál es el valor de la media?

25.25

b) El costo de dicha operación (en miles de pesos).

1400,1420,1450,1450,1475,1480,1480,2800

¿Cuál es el valor de la media?

1619.38

c) El número de personas que fueron operadas durante cada mes de 1987.

2,2,4,4,4,5,7,7,8,9,9,10

¿Cuál es el valor de la media? (Recuerda redondear)

5.92

CUADRO 87

En estos ejemplo, pudiste observar las propiedades de la media aritmética:

1) Es única, en cada ejemplo sólo había un posible resultado correcto.

2) Es fácil de calcular y de comprender.

3) Se ve afectada por cada uno de los valores de la población. Como en el segundo ejercicio, aunque casi a todas las personas les costó alrededor de \$1'450,000 pesos la operación, la media es bastante mayor (\$1'619,380 pesos), debido a lo que pagó la última persona. (¡Seguramente se operó en el extranjero!)

CUADRO 88

La media puede ser considerada como el punto alrededor del cual, en cualquier distribución de frecuencia, se alternan los valores por encima y por debajo. Esto puede verse fácilmente calculando las desviaciones, es decir, cuánto se aleja cada dato de la media ($x - \mu$)

Veamos un ejemplo en el cual ya habíamos calculado la media:

a) La duración en días de la convalecencia de personas a quienes se les practicó apendicectomía (extracción del apéndice).

$$20, 35, 15, 19, 26, 32, 28, 27 = 25.25$$

Si calculamos la desviación para cada uno de los datos, y obtenemos la suma de los valores positivos y la suma de los valores negativos, y las sumamos, la diferencia es cero.

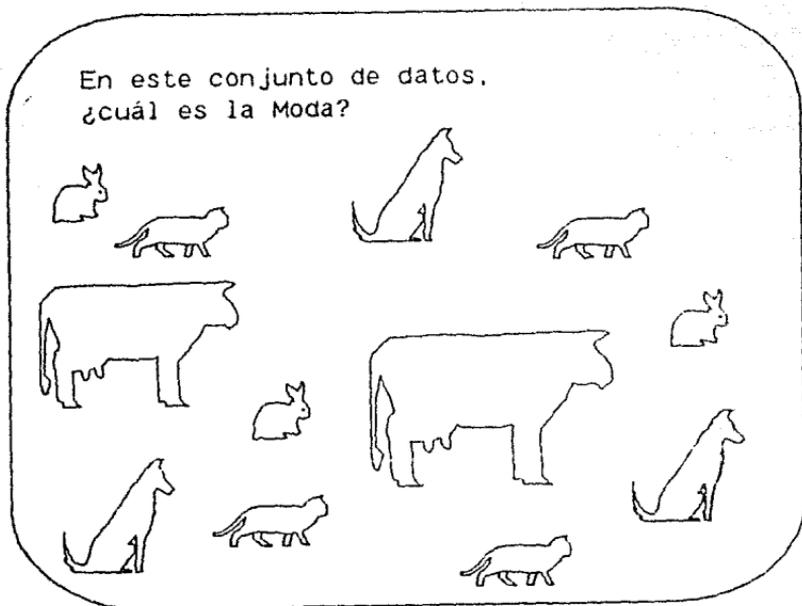
20:	$20 - 25.25 =$	-5.25	Suma negativos
35:	$35 - 25.25 =$	9.75	-21.75
15:	$15 - 25.25 =$	-10.25	
19:	$19 - 25.25 =$	-6.25	Suma positivos
26:	$26 - 25.25 =$	0.75	21.75
32:	$32 - 25.25 =$	6.75	
28:	$28 - 25.25 =$	2.75	Diferencia
27:	$27 - 25.25 =$	1.75	0

En cualquier ejemplo sucede lo mismo. La media puede considerarse así como el centro de gravedad de una distribución de frecuencias.

MODA (Mo):

Es el valor que "está de moda", el que más se repite.

La forma de encontrarla es obteniendo una tabla de frecuencias, y viendo qué dato tiene la frecuencia mayor, esa es la moda.



La tabla de frecuencias de las variables es:

Conejo = 3

Gato = 4

Perro = 3

Vaca = 2

Entonces, ¿qué animal es la moda?

La moda es gato porque es el que más veces se repite (4 veces).

Para convertir esto a números, podemos usar la escala nominal:

Conejo = 1

Gato = 2

Perro = 3

Vaca = 4

Y entonces tendríamos la serie 1,1,1,2,2,2,2,3,3,3,4,4 con moda = 2

CUADRO 90

La moda es una medida útil cuando la variable es DISCRETA, y se está usando la escala NOMINAL, es decir, es buena medida de tendencia central para datos cualitativos que no pueden ser operados numéricamente.

Por ejemplo, los pacientes que se atendieron en una clínica de salud mental durante un determinado año pudieron haber recibido uno de los siguientes diagnósticos: retraso mental, síndrome cerebral orgánico, psicosis, neurosis y alteración de la personalidad. El diagnóstico que se haya presentado con más frecuencia en el grupo de pacientes, es el que se llamará diagnóstico modal.

Esta medida también es útil para describir la tendencia central de una población donde se presentan algunos valores extremos, (como en el ejemplo del costo de la operación, que alteran mucho el promedio)

CUADRO 91

Calcula la moda de las siguientes poblaciones. (Si hay mas de una, sepáralas con comas).

a) La duración en días de la convalecencia de personas a quienes se les practicó apendicectomía (extracción del apéndice).

15,19,20,26,27,28,32,35

¿Cuál es el valor de la moda?

No hay moda en esta población, ningún valor se repite más que otros.

b) El costo de dicha operación (en miles de pesos).

1400,1420,1450,1450,1475,1480,1480,2800

¿Cuál es el valor de la moda?

Hay dos modas: 1450 y 1480.

c) El número de personas que fueron operadas durante cada mes de 1987.

2,2,4,4,4,5,7,7,8,9,9,10

¿Cuál es el valor de la moda?

4

CUADRO 92

En estos ejemplos, puedes observar las propiedades de la moda:

1) La moda no es única, algunas distribuciones tienen dos o más modas.

Algunas distribuciones tienen más de una moda.



Unimodal



Bimodal



Multimodal

2) Es fácil de calcular, se obtiene por observación.

3) No se ve afectada por todos los valores de la población, sino sólo por aquel que se presenta más frecuentemente. Como la moda en la ropa femenina, no importa si una persona se viste anticuada, la moda permanece igual.

CUADRO 93

MEDIANA (Md):

Es el "punto de en medio" de una distribución. Corta a la población en dos partes iguales. De cada lado de ella queda el 50% de los datos.

La palabra mediana da la idea de estatura, y resulta claro que si se habla de estaturas, los datos debe "estar formados", esto es, ordenados. Para ello se requiere que se esté usando una escala de razón, de intervalo, o al menos ordinal. (No se puede calcular la mediana a datos como los animales que vimos antes). La variable bajo estudio puede ser discreta o continua.

La mediana puede usarse cuando se tiene una población grande de datos ordenados, pues con esta medida nos damos una idea de la tendencia central sin hacer cálculos.

CUADRO 94

Para calcularla:

1.- Si tenemos un número impar de casos, la mediana será el valor que esté en medio de la formación.

Por ejemplo: En la serie 3,7,9,12,17,21,34,42,50

La mediana es 17, pues hay cuatro elementos antes que ese número y cuatro después.

2.- Cuando el número de observaciones es par, no se tiene una sola observación en medio, sino dos. En este caso, se ha establecido la convención de que la mediana sea el promedio aritmético de esas dos observaciones.

Por ejemplo: La serie 1,5,12,15,23,26,39,40

$$\begin{aligned} \text{Md} &= (15 + 23) / 2 \\ &= 38 / 2 \\ &= 19 \end{aligned}$$

CUADRO 95

En términos generales, si la serie de datos está ordenada, la posición de la mediana está dada por la fórmula:

$$\text{Pos. Md} = (N + 1) / 2$$

donde N - Número de elementos de la distribución.

1.- Si N es impar, el valor $(N+1)/2$ nos da el lugar que está en medio de la formación.

En el ejemplo que vimos antes: 3,7,9,12,17,21,34,42,50

$$\begin{aligned} N &= 9 \\ \text{Pos. Md} &= (9+1)/2 = 5 \\ \text{Md} &= 17 \text{ (ocupa el 5o. lugar)} \end{aligned}$$

2.- Si N es par, $(N+1)/2$ no es entero, es decir, esta fórmula nos dará un lugar que no existe, por ejemplo, el lugar 4.5. Esto significa que la mediana será el promedio de los valores que ocupen los lugares 4o. y 5o. en la serie.

En el ejemplo anterior: 1,5,12,15,23,26,39,40

$$N=8$$

$$\text{Pos. Md} = (8+1)/2 = 4.5$$

$$\text{Md} = (15+23)/2 = 38/2 = 19$$

CUADRO 96

Calcula la mediana de las siguientes poblaciones:

a) La duración en días de la convalecencia de personas a quienes se les practicó apendicectomía (extracción del apéndice).

15,19,20,26,27,28,32,35

¿Cuál es el valor de la mediana?

26.5

b) El costo de dicha operación (en miles de pesos).

1400,1420,1450,1450,1475,1480,1480,2800

¿Cuál es el valor de la mediana?

1462.5

c) El número de personas que fueron operadas durante cada mes de 1987.

2,2,4,4,4,5,7,7,8,9,9,10

¿Cuál es el valor de la mediana?

6

CUADRO 97

En estos ejemplo, pudiste observar las propiedades de la mediana:

- 1) La mediana es única, sólo existe una mediana para un determinado conjunto de datos.
- 2) Es fácil de calcular.
- 3) No se ve afectada por valores extremos, es decir, si los valores de en medio de una serie, son pequeños, la mediana será pequeña; o al contrario, si los valores centrales son grandes, la mediana será grande, independientemente de que los valores en los extremos sean pequeños o grandes.

En la siguiente gráfica, se tiene un resumen de estas tres medidas:

La Media es única, igual que la mediana.
Pero no así la moda.

YO ESTOY EN EL CENTRO,
PERO CONSIDERO SIEMPRE
A LOS DE IZQUIERDA Y A
LOS DE DERECHA.

\bar{x}

Md

YO ME FORMO CON LOS DE
EN MEDIO. NO SOY NI
GRANDE, NI PEQUEÑA.

NOSOTROS, COMO
BUENAS CHICAS MODERNAS,
SEGUIMOS LO QUE MAS
SE LLEVA.

Mo

Mo

Mo

CUADRO 99

Veamos algunos ejemplos que ilustren la utilidad de estas medidas:

Al analizar la situación económica de un país, podemos pensar en considerar representativo el ingreso per capita, pero este número se ve afectado por valores extremos (por ejemplo el sueldo del Presidente), es entonces mucho más significativo el salario mínimo, que es la MODA.

Un ejemplo del uso de la MEDIA es calcular cuál es la edad promedio en que se casan los jóvenes en la República Mexicana. Este dato no es fácil de obtener para toda la población del país, pero si es factible usar las técnicas de la Estadística Inferencial (que permiten obtener conclusiones válidas sobre una población grande a partir de una muestra), obtener una muestra representativa de la población que se quiere estudiar, calcular la media muestral, aplicar los resultados a la población, y obtener así conclusiones verdaderamente interesantes.

La MEDIANA se puede usar para saber si el sueldo que gana una persona en una empresa es alto o bajo. Al considerar el sueldo mediano, se tiene el valor por encima del cual gana la mitad de la empresa y por debajo del cual gana la otra mitad. Aunque nunca es bueno compararse, la mediana da una posición más exacta para determinar si ese sueldo en particular es alto o bajo con respecto a los demás.

CUADRO 100

MEDIA ARITMETICA PONDERADA

Si alguna vez quieres calcular la media de dos o más poblaciones, y ya conoces la media y el tamaño de cada una de ellas, ésta es la medida indicada.

En vez de sumar todos los datos otra vez, puedes encontrar la media de la población combinada a partir de la información que ya tienes.

La notación usada, en forma general, será:

k - número de distintas poblaciones.

$\mu_1, \mu_2, \dots, \mu_k$ - las medias de cada población.

n_1, n_2, \dots, n_k - el número de elementos de cada población.

N - número total de elementos ($n_1+n_2+\dots+n_k$)

μ - media de la población combinada.

La media de esta población combinada resultará de la suma ponderada de las medias de las poblaciones originales.

$$\mu = (n_1/N)*\mu_1 + (n_2/N)*\mu_2 + \dots + (n_k/N)*\mu_k$$

Por ejemplo:

Población 1

$$X_i = 2, 7, 9$$

$$\mu_1 = (2+7+9)/3 = 18/3 = 6$$

$$n_1 = 3$$

Población 2

$$X_i = 4, 18$$

$$\mu_2 = (4+18)/2 = 22/2 = 11$$

$$n_2 = 2$$

En la población que conjuga estas dos, $N=5$

$$\begin{aligned} \text{Media Aritmética Ponderada} &= (3/5)*6 + (2/5)*11 \\ &= 3.6 + 4.4 \\ &= 8 \end{aligned}$$

$$\begin{aligned} \text{Que es igual a } (2 + 7 + 9 + 4 + 18) / 5 \\ &= 40 / 5 \\ &= 8 \end{aligned}$$

En este ejemplo son pocos datos, y el esfuerzo con un procedimiento y con otro es equivalente, pero imagina que cada una de las poblaciones tuviera 100 datos, y entonces verás la ventaja de usar esta medida de tendencia central.

(Se muestran los resultados de la lección).

Si tuvo correctas todas las respuestas:

¡Qué bárbaro, no tuviste ninguna mala! ¡Qué bien!

Gracias por tu atención. Nos seguimos viendo...

LECCION 5
MEDIDAS DE DISPERSION

L E C C I O N 5

Medidas de Dispersión

OBJETIVOS:

CONOCIMIENTOS: Al terminar la lección el estudiante podrá definir las medidas de dispersión: rango, desviación, desviación media, varianza, desviación estándar y puntaje estándar. Podrá seleccionar cuál es la medida más conveniente para indicar la dispersión de una serie de datos y aprenderá a calcular las medidas de dispersión para datos agrupados.

HABILIDADES: El estudiante deberá ser capaz de calcular cada una de las medidas de dispersión y de obtener conclusiones a partir de los datos calculados.

ACTITUDES: Se interesará en medir la dispersión en series de datos para poder entender mejor la información.

NOTA IMPORTANTE

¡Hola!, ¿Ya sabes usar la calculadora?

Si responde si:

OK, continuamos...

Si no sabe usarla:

NOTA IMPORTANTE:

Si en cualquier momento necesitas usar una calculadora, puedes obtenerla tecleando <Alt> C. Debes oprimir estas dos teclas al mismo tiempo. Si quieres ayuda al usar la calculadora, teclea H (Help). Si quieres traer un resultado obtenido con la calculadora oprime <Alt> F.

Para calificarte adecuadamente, por favor en todas las respuestas que no sean enteras, usa sólo 2 decimales, redondeando si es necesario. (0-4 al dígito anterior y 5-9 al superior). Si no lo haces así, se te puede considerar mal una respuesta correcta.

Vamos a practicar: llama a la calculadora y calcula $12.1 * 5.33$. (multiplicación) Trae el resultado con <Alt> f

Si responde: 64.493

Bién, pero olvidaste redondear a dos dígitos decimales significativos. Dame nuevamente el resultado.

Si la respuesta es: 64.49

¡Perfecto!, comprendiste la idea. Si quieres, practica un poco más. Recuerda que los signos de operación son: más +; menos -; por * y entre /. Para dejar de usarla, da <Esc> y luego <Return>.

Cualquier otra respuesta:

No, lee con atención como se usa, y vuelve a contestar.

CUADRO 111

Las medidas de tendencia central, que vimos en la lección anterior dan una idea de los datos que "representan", pero no los definen totalmente.

Veamos un ejemplo: tres personas tuvieron los siguientes ingresos (en miles de pesos) en cada uno de los meses que se indican.

	Enero	Febrero	Marzo	Abril	Mayo
Pedro	500	500	500	500	500
Juan	500	-1000	0	500	2500
Luisa	100	200	500	500	1200

Nota: El sueldo de Juan es negativo en febrero, esto se debe a que pidió un préstamo de un millón de pesos, y lo pagó con el sueldo que le corresponde durante ese mes y el siguiente.

CUADRO 112

¿Qué se puede concluir de estos valores?

Veamos primero qué obtenemos si calculamos la media, mediana y moda de estas series.

¿Cuál es la MEDIA de la primera serie (Pedro)

El resultado es 500

¿Cuál es la MEDIANA de la segunda serie (Juan)

El resultado es 500

¿Cuál es la MODA de la tercera serie (Luisa)

El resultado es 500

La tabla completa queda:

SERIE	MEDIA	MEDIANA	MODA
Pedro	500	500	500
Juan	500	500	500
Luisa	500	500	500

!Todos los resultados son iguales!... parece que las medidas de tendencia central no nos dan información completa sobre el comportamiento de las series, pues en promedio (tres tipos diferentes de promedio), parecería que los sueldos de estas personas fueron equivalentes, y sabemos que esto no es así.

Llegan entonces en nuestra ayuda, las medidas de dispersión. Ellas nos indican el grado de variación dentro de cada serie, es decir qué tan lejos se encuentran los valores unos de otros.

CUADRO 113

Las medidas de dispersión que veremos aquí son:

- Rango.
- Desviación.
- Desviación Media.
- Varianza.
- Desviación Estándar.
- Puntaje estándar ("standard score")

CUADRO 114

RANGO O RECORRIDO

----- - -----

Es la DIFERENCIA QUE EXISTE ENTRE EL VALOR MENOR Y EL MAYOR DEL CONJUNTO DE OBSERVACIONES.

En el ejemplo que estamos estudiando:

Rango

Pedro	$500 - 500 = 0$
Juan	$2500 - (-1000) = 3500$

Calcula tú el de Luisa. Si la serie es:

Luisa	100	200	500	500	1200
-------	-----	-----	-----	-----	------

¿cuánto es el rango? (No se vale indicar las operaciones, dame un solo resultado).

El valor del rango en esta serie es 1100

Esta medida sólo considera los valores extremos de la serie, y sin embargo ya podemos afirmar, por los resultados obtenidos, que Pedro ganó lo mismo todos los meses, Juan tuvo una variación muy grande en el ingreso (debido a su préstamo), y Luisa tuvo ingresos más constantes, aunque con un rango de variación de más del doble del ingreso promedio, que es de 500 mil pesos mensuales.

CUADRO 115

DESVIACION

De esta medida ya habíamos hablado cuando vimos la media aritmética. Es la DIFERENCIA ENTRE CADA OBSERVACION DE LA POBLACION Y SU MEDIA.

Repasemos algunos puntos importantes de esta medida:

1o. Su fórmula es $x_i - \mu$

2o. El valor de la suma de todas las desviaciones de una serie es: (selecciona una opción)

- a) siempre igual
- b) siempre cero
- c) siempre diferente

Si responde a:

Tienes razón, pero además de igual, siempre es cero.

Si la respuesta es b:

Esta es la respuesta exacta, pues aunque el inciso a) es correcto, esta respuesta es más precisa.

Si c:

La respuesta correcta es la del inciso b)

Esto se debe a que la media es el centro de masa o centro de gravedad de la distribución.

La suma de las desviaciones es cero siempre. Compruébalo con las series que venimos estudiando.

	----- Xi - μ -----					Suma
Pedro	0	0	0	0	0	0
Juan	-1500	-500	0	0	2000	0
Luisa	-400	-300	0	0	700	0

La desviación es una medida útil, nos permite ver cuánto se aleja cada valor de la media. Y esto nos da más idea de la dispersión de los datos.

Sin embargo, no podemos obtener un sólo valor para cada serie, pues al sumar las desviaciones, las positivas se cancelan con las negativas, y el resultado final es siempre cero.

CUADRO 116

Pensemos en otras posibilidades para obtener un sólo número que por él mismo nos pueda indicar el grado de dispersión de la serie.

Una alternativa es usar el promedio del valor absoluto de las desviaciones. Recuerda que el VALOR ABSOLUTO es una función que afecta el signo de los números y se define como:

$$\begin{aligned} | A | &= A \text{ si } A \geq 0 \\ &= -A \text{ si } A < 0 \end{aligned}$$

(Observa que, en realidad, lo único que hace esta función es convertir todos los valores a números positivos).

La fórmula sería algo como:

$$\Sigma | (x_i - \mu) | / N$$

Como no hay nada nuevo bajo el sol, esto ya lo pensaron los Estadísticos y le llamaron:

DESVIACION MEDIA

que se define como la MEDIA ARITMETICA DEL VALOR ABSOLUTO DE LAS DESVIACIONES.

Calculemos este valor en nuestra series: (Yo hago las dos primeras y tú la última ¿OK?).

Recuerda que $\mu = 500$.

	----Valor absoluto de la desv----					Desv. Media
Pedro	0	0	0	0	0	0
Juan	0	1500	500	0	2000	$4000/5 = 800$

Si la última serie es:

Luisa 100 200 500 500 1200

¿cuánto vale el 1er. valor absoluto que estamos calculando? $|x_1 - \mu|$

Este valor está dado por $|100 - 500| = |-400| = 400$

¿cuánto vale el segundo valor? $|x_2 - \mu|$

Este valor está dado por $|200 - 500| = |-300| = 300$

¿cuánto vale el tercer valor?

Este valor está dado por $|500 - 500| = |0| = 0$

¿cuánto vale el cuarto valor?

Este valor está dado por $|500 - 500| = |0| = 0$

¿y el quinto valor?

Este valor está dado por $|1200 - 500| = |700| = 700$

Entonces, si los resultados que calculaste son 400, 300, 0, 0 y 700 ¿cuál es el valor de la desviación media para esta serie?

Este valor está dado por $(400 + 300 + 0 + 0 + 700) / 5 = 280$

CUADRO 117

¿Qué te parece si, como otra alternativa para eliminar los signos negativos, elevamos al cuadrado cada desviación?.

!Ya también nos ganaron!, ¡Con estos Estadísticos no se puede! Esta es la manera en que se obtiene la medida que nombraron como:

VARIANZA

que es la MEDIA ARITMETICA DE SUMA DE LAS DESVIACIONES ELEVADAS AL CUADRADO.

Su fórmula para poblaciones es:

$$\sigma^2 = \Sigma(x_i - \mu)^2 / N$$

donde:

- . σ^2 - letra griega sigma (minúscula), elevada al cuadrado, es la varianza.
- . μ - letra griega mu, es la media de la población.
- . x_i - cada una de las observaciones.
- . N - número de elementos de la población.

CUADRO 118

Si usamos esta fórmula para calcular la varianza de los sueldos de nuestros amigos, tenemos:

$$\text{-----}Y = (X_i - \mu)^2 \text{-----} \text{Var} = \Sigma Y/5$$

Pedro	0	0	0	0	0	0
Juan	2,250,000	250,000	0	0	4,000,000	1,300,000

A ti te toca, como siempre, la tercera serie:

Si los valores son:

Luisa	100	200	500	500	1200
-------	-----	-----	-----	-----	------

y ya calculaste $(x_i - \mu)$ como:

-400	-300	0	0	700
------	------	---	---	-----

¿cuánto vale $(x_1 - \mu)^2$? (Escribe sólo números, no uses comas)

Este valor está dado por $(-400)^2 = 160000$

¿cuánto vale $(x_2 - \mu)^2$?

Este valor está dado por $(-300)^2 = 90000$

¿cuánto vale $(x_3 - \mu)^2$?

Este valor está dado por $0^2 = 0$

¿cuánto vale $(X_4 - \mu)^2$?

Este valor está dado por $0^2 = 0$

¿Y $(X_5 - \mu)^2$?

Este valor está dado por $700^2 = 490000$

Entonces, si ya calculamos los valores 160000, 90000, 0, 0 y 490000 ¿cuál es el valor de la varianza para esta serie? (si quieres cambiar de lugar la calculadora teclea <Ctrl> y la flechita en la dirección deseada simultáneamente.)

Este valor está dado por $(160000 + 90000 + 0 + 0 + 490000) / 5 = 740000 / 5 = 148000$

Los valores resultan:

	Varianza
Pedro	0
Juan	1,300,000
Luisa	148,000

Esto significa que el sueldo de Pedro no tuvo dispersión, todos los meses fué igual; el ingreso de Juan está muy disperso, y el de Luisa no tanto.

Observa que, sin embargo, los resultados son demasiado grandes. Al elevar al cuadrado, los valores crecen demasiado.

CUADRO 119

¿Y si calculamos la raíz cuadrada de la varianza?... ¿qué crees? ... los Estadísticos ya lo hicieron así, y llegaron a la definición de LA MAS IMPORTANTE de las medidas de dispersión, la:

DESVIACION ESTANDAR

que se define como la RAIZ CUADRADA DE LA VARIANZA.

Ellos encontraron que es la más adecuada porque:

- no es un número tan grande como la varianza. Al sacar la raíz cuadrada, se disminuye mucho la magnitud de los números que se usan.
- tiene propiedades lineales que facilitan la manipulación algebraica o matemática.
- se mide en las mismas unidades que la población.

Tú ya estás en posibilidades de definir la fórmula para calcular la desviación estándar. ¿Cuál es esa fórmula, si

- σ - es la letra griega sigma (minúscula), que representa a la desviación estándar de una población.
- x_i - es cada uno de los elementos de la población.
- μ - es la media poblacional.
- N - es el número de elementos de la población?

Selecciona la fórmula correcta para calcular la desviación estándar [LA RAIZ COMPRENDE TODO LO QUE ESTA ENTRE PARENTESIS CUADRADOS]

$$a) \sigma = \sqrt{[\Sigma(\mu - x_i)^2 / N]}$$

$$b) \sigma = \sqrt{[\Sigma(x_i - \mu)^2 / N]}$$

$$c) \sigma = \sqrt{[\Sigma(x_i - \mu)^2] / N}$$

Si a:

Tienes razón, es lo mismo $\Sigma(\mu - x_i)^2$ que $\Sigma(x_i - \mu)^2$

Si la respuesta es b:

Esta fórmula es correcta, aunque nota que el inciso a) también lo es.

Si responde c:

La raíz cuadrada debe comprender a todo el cociente, no sólo el dividendo.

CUADRO 120

Calculemos la desviación estándar de los sueldos de nuestros amigos. Como la calculadora que está en la máquina no tiene la función de raíz cuadrada, yo te daré el resultado.

	Varianza	Desviación Estándar
Pedro	0	0
Juan	1300000	1140.2
Luisa	148000	384.7

Observa que es más cómodo y fácil de interpretar y trabajar con los valores que se obtienen al calcular la desviación estándar.

Tiene sentido decir que Luisa tuvo un sueldo promedio de \$500,000 pesos mensuales con una variación de \$384,700 pesos durante los cinco meses que estamos considerando ¿no te parece?.

Sería mas complicado usar la varianza, pues tendríamos que hacer afirmaciones como "Juan ganó en promedio quinientos mil pesos mensuales, con una variación de mil trecientos millones de pesos cuadrados", o algo así de extraño.

CUADRO 121

Todas estas medidas las hemos calculado para una población; son medidas descriptivas. Pero la función principal de la Estadística es hacer afirmaciones confiables donde hay incertidumbre, es decir obtener valores de los parámetros de una población, a partir de datos de una muestra que la represente adecuadamente.

Por razones de claridad, las personas que trabajan con esto, convinieron en usar una notación diferente para la población y la muestra.

Así tenemos:

	POBLACION	MUESTRA
MEDIA	μ (mu)	X barra
No. ELEMENTOS	N	n
VARIANZA	σ^2 (sigma)	s ²
DESVIACION ESTANDAR	σ (sigma)	s

Observarás que las letras griegas se usan para representar a los parámetros de la población, y las latinas para representar a las estadísticas de la muestra.

CUADRO 122

Las medidas obtenidas de una muestra, llamadas estadísticas, ESTIMAN los valores de los parámetros de la población. Así, la media muestral \bar{X} , es un ESTIMADOR de μ .

Similarmente:

la varianza muestral estima la varianza poblacional σ^2

la desviación estándar muestral s estima la desviación estándar poblacional σ .

Para que un estimador sea bueno, debe ser INSESGADO.

Este concepto se comprenderá mejor con un ejemplo: supongamos que tengo una población de la cual

- conozco sus parámetros (por ejemplo, la media poblacional)
- obtengo todas las posibles muestras de un tamaño dado
- calculo la estadística que corresponde al parámetro de la población para todas las muestras (por ejemplo la media muestral)
- promedio los valores obtenidos en cada muestra

el valor que resulta es el del parámetro (en el ejemplo, la media poblacional).

CUADRO 123

Veámoslo con valores numéricos:

Consideremos la población:

1,2,3,4,5,6,7

su media poblacional es:

$$\mu = (1+2+3+4+5+6+7)/7 = 28/7 = 4$$

Tomemos muestras de tamaño 5.

Tal vez recuerdes, o si no lo sabías, la fórmula para calcular todas las posibles combinaciones de N elementos tomados de n en m (en nuestro caso $N = 7$ y $n = 5$), es:

$$N \text{ en } m = \frac{N!}{(N-m)! * m!}$$

donde $N!$ (factorial de N ó N factorial), se define como el producto de N por todos los números menores a él, hasta llegar a la unidad. En fórmula:

$$N! = N * (N-1) * (N-2) * \dots * 2 * 1$$

Entonces el número total de muestras diferentes es:

$$\begin{aligned} &= 7! / (7-5)! * (5!) \\ &= 7! / 2! * 5! \\ &= 5040 / 2 * 120 \\ &= 5040 / 240 \\ &= 21 \end{aligned}$$

Ahora con tu ayuda, calcularemos la media de las 21 muestras diferentes, de la población (recuerda que es 1,2,3,4,5,6,7 y que su media es 4, es decir, $\mu = 4$), y obtendremos su media muestral:

Veamos las 7 primeras. Yo indico algunas, y muestro flashando las que a ti te tocan... ¿sale?

						Media muestral
Muestra 1	1	2	3	4	5	3.0
Muestra 2	1	2	3	4	6	
Muestra 3	1	2	3	4		<u>3.4</u>
Muestra 4	1	2	3	5	6	
Muestra 5	1	2	3	5	7	<u>3.6</u>
Muestra 6	1	2	3		7	3.8
Muestra 7	1	2	4	5	6	<u> </u>

(se van pidiendo y evaluando las respuestas, los valores que faltan en la tabla anterior)

Bueno, siguiendo el mismo mecanismo, la tabla ya completa queda:

						Media muestral
Muestra 1	1	2	3	4	5	3.0
Muestra 2	1	2	3	4	6	3.2
Muestra 3	1	2	3	4	7	3.4
Muestra 4	1	2	3	5	6	3.4
Muestra 5	1	2	3	5	7	3.6
Muestra 6	1	2	3	6	7	3.8
Muestra 7	1	2	4	5	6	3.6
Muestra 8	1	2	4	5	7	3.8
Muestra 9	1	2	4	6	7	4.0
Muestra 10	1	2	5	6	7	4.2
Muestra 11	1	3	4	5	6	3.8
Muestra 12	1	3	4	5	7	4.0
Muestra 13	1	3	4	6	7	4.2
Muestra 14	1	3	5	6	7	4.4
Muestra 15	1	4	5	6	7	4.6
Muestra 16	2	3	4	5	6	4.0
Muestra 17	2	3	4	5	7	4.2
Muestra 18	2	3	4	6	7	4.4
Muestra 19	2	3	5	6	7	4.6
Muestra 20	2	4	5	6	7	4.8
Muestra 21	3	4	5	6	7	5.0

Calcula el valor de la media de los promedios muestrales. ¿Cuánto te da?

El resultado correcto es 4

¿ya te diste cuenta que este resultado es igual al promedio de la población?

Parece magia ¿verdad?, pero no es.

Esto sucede para cualesquiera valores de la población. Puedes comprobarlo si tienes la paciencia, o una hoja de cálculo electrónica en la máquina que estás usando.

Con esto concluimos que la MEDIA MUESTRAL es un ESTIMADOR INSESGADO de la media poblacional.

Este es un resultado de mucha importancia, ¡apréndetelo bien!

CUADRO 124

Demos un repaso de conceptos que vimos anteriormente:

Relaciona el símbolo con la medida que representa, formando las parejas que correspondan, por ejemplo 1a.

- | | |
|-------------------------------------|--------------------|
| 1) Media poblacional | a) σ^2 |
| 2) No. de elementos de la muestra | b) \bar{X} barra |
| 3) Varianza muestral | c) n |
| 4) Media muestral | d) s^2 |
| 5) Varianza poblacional | e) μ |
| 6) No. de elementos de la población | f) N |

Selecciona una pareja:

Las respuestas correctas son: 1e, 2c, 3d, 4b, 5a y 6f

CUADRO 125

Vamos a hacer un ejercicio con las tres series siguientes:

SERIE X Observac.	Y	Z	SERIE X Observac.	Y	Z		
1	0	1	0	13	9	6	5
2	0	1	0	14	6	4	3
3	0	1	0	15	4	3	3
4	0	1	0	16	2	3	2
5	1	1	0	17	2	2	2
6	1	2	1	18	1	2	1
7	2	2	4	19	1	1	1
8	2	3	4	20	0	1	1
9	4	3	7	21	0	1	1
10	6	4	9	22	0	1	1
11	9	6	8	23	0	1	1
12	10	10	6				
		# de elementos		60	60	60	

Estas son tablas de distribución de frecuencia, que significan por ejemplo, que el dato 9 se observó 4 veces en la serie X, 3 en la serie Y y 7 en la serie Z.

Para obtener la media de las observaciones 7 a 9 de la serie X, que tienen los números siguientes 7: 2, 8: 2 y 9: 4, se calcula:

$$\begin{aligned}
 X &= (7+7+8+8+9+9+9)/8 \\
 &= 66/8 = 8.25
 \end{aligned}$$

Que en forma simplificada sería:

$$\begin{aligned} X &= ((7*2)+(8*2)+(9*4))/8 \\ &= 66/8 = 8.25 \end{aligned}$$

Generalizando esto, la media para datos agrupados es:

$$\mu = \Sigma (fx) / N$$

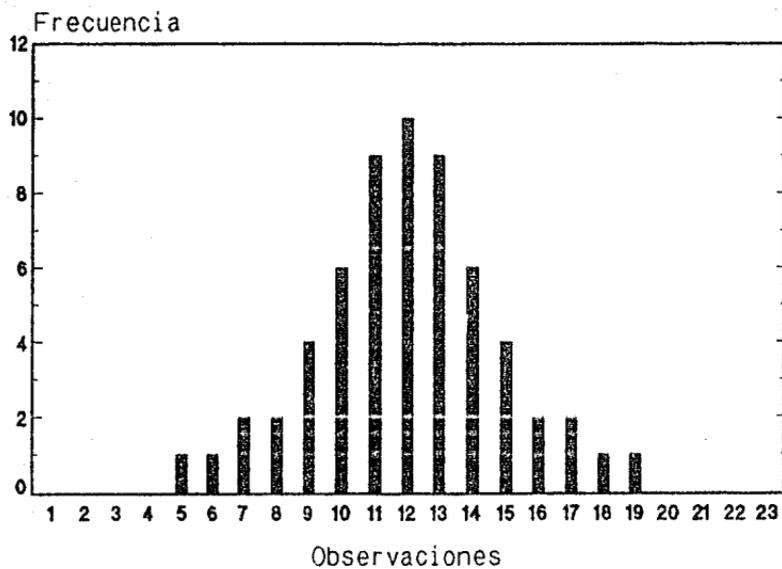
donde f - frecuencia observada de cada valor de x

Nota que el símbolo usado en la fórmula es μ y no X barra, esto se debe a que estamos trabajando con una población y no con una muestra.

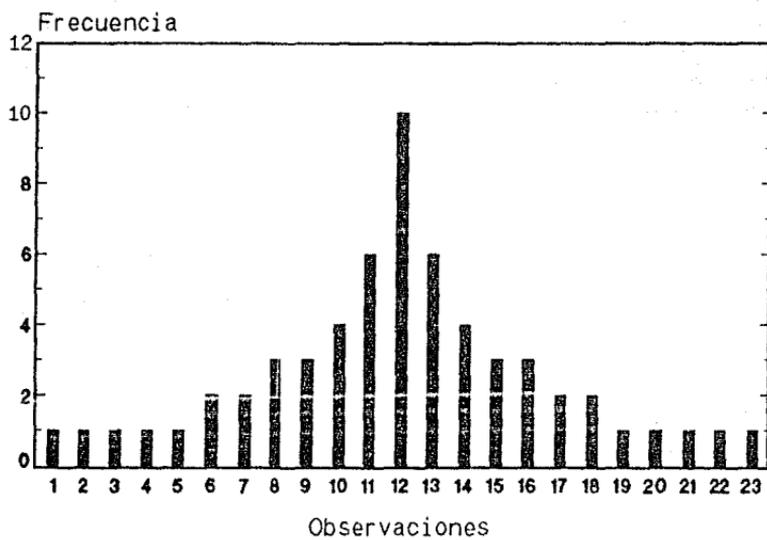
CUADRO 126

Ahora observa las gráficas y piensa qué puedes concluir de ellas, considerando que las tres series tienen el mismo tamaño: 60 y la misma media: 12.

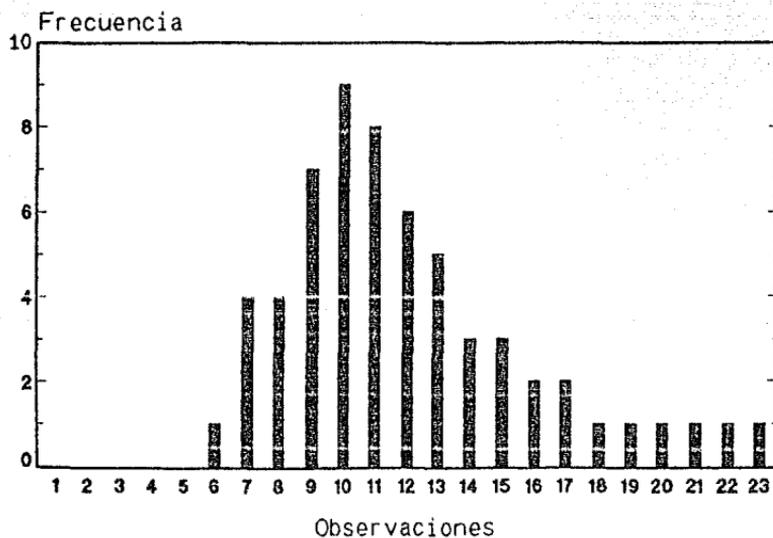
Distribución Serie X
Media = 12



Distribución Serie Y
Media = 12



Distribución Serie Z
Media = 12



¿Hay alguna con mayor dispersión?

(SI NECESITAS VOLVER A VER LAS GRAFICAS PARA CONTESTAR ESTA PREGUNTA Y LA SIGUIENTE, TECLEA <VER GRAFICAS> PARA VOLVER A MOSTRARTELAS)

¿Tu qué opinas?

Si responde SI:

Estoy de acuerdo contigo, no se ve que las barras esten distribuidas con la misma dispersión en cada una de las series.

Cualquier otra respuesta:

Yo creo que si hay mayor dispersión en una de las series.

¿Cuál es?

Si responde Y:

Asi parece, vamos a analizar esto más detenidamente.

Cualquier otra respuesta:

No estoy de acuerdo contigo, de cualquier manera...

CUADRO 127

Veámoslo con números, para ello, calculemos las medidas que aquí se vieron:

Como los datos están agrupados, la fórmula que usaremos para calcular la desviación estándar es:

$$\sigma = \Sigma [f(x-\mu)^2] / N$$

Es muy parecida a la que siempre hemos manejado, sólo que cada dato se multiplica por su frecuencia.

Las series con las que estamos trabajando son:

-----SERIE X-----							
x	f	f*x	f((x-μ) ²)	x	f	f*x	f((x-μ) ²)
1	0	0	0	13	9	117	9
2	0	0	0	14	6	84	24
3	0	0	0	15	4	60	36
4	0	0	0	16	2	32	32
5	1	5	49	17	2	34	50
6	1	6	36	18	1	18	36
7	2	14	50	19	1	19	49
8	2	16	32	20	0	0	
9	4	36	36	21	0	0	0
10	6	60	24	22	0	0	0
11	9	99	9	23	0	0	0
12	10	120	0				

# Elementos	60
Media	12.00
Varianza	
Desv. Estándar	<u>2.80</u>

¿Cuánto vale la varianza?

$$\begin{aligned} &\text{Vale } (49+36+50+32+36+24+9+9+24+36+32+50+36+49)/60 \\ &= 472/60 = 7.87 \end{aligned}$$

-----SERIE Y-----							
x	f	f*x	$f((x-\mu)^2)$	x	f	f*x	$f((x-\mu)^2)$
1	1	1	121	13	6	78	6
2	1	2	100	14	4	56	16
3	1	3	81	15	3	45	27
4	1	4	64	16	3	48	48
5	1	5	49	17	2	34	50
6	2	12	72	18	2	36	72
7	2	14	50	19	1	19	49
8	3	24	48	20	1	20	64
9	3	27	27	21	1	21	81
10	4	40	16	22	1	22	100
11	6	66	6	23	1	23	121
12	10	120	0				

# Elementos	60
Media	12.00
Varianza	
Desv. Estándar	<u>4.60</u>

¿Cuánto vale la varianza?

$$\text{Vale } (121+100+81+64+49+72+50+48+27+16+6) * 2 / 60$$

$$= 634 * 2 / 60 = 1268 / 60 = 21.13$$

-----SERIE Z-----							
x	f	f*x	f((x- μ) ²)	x	f	f*x	f((x- μ) ²)
1	0	0	0	13	5	0	605
2	0	0	0	14	3	0	300
3	0	0	0	15	3	0	243
4	0	0	0	16	2	0	128
5	0	0	0	17	2	0	98
6	1	0	36	18	1	0	36
7	4	0	100	19	1	0	25
8	4	0	64	20	1	0	16
9	7	0	63	21	1	0	9
10	9	0	36	22	1	0	4
11	8	0	8	23	1	0	1
12	6	0	0				

# Elementos	60
Media	12.00
Varianza	14.73
Desv. Estándar	3.84

Las varianzas de las series resultaron:

Serie X:	7.87
Serie Y:	21.13
Serie Z:	14.73

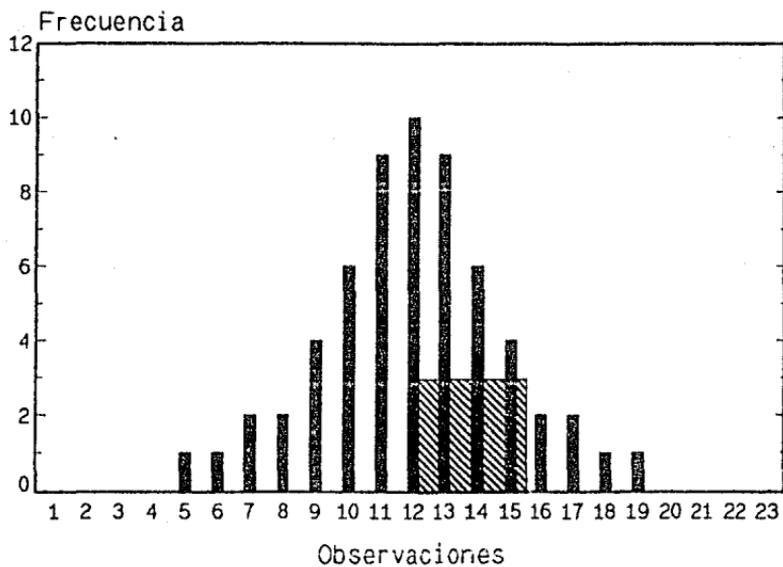
Así que la serie con mayor dispersión es la serie Y, como habíamos pensado al ver sus gráficas.

CUADRO 128

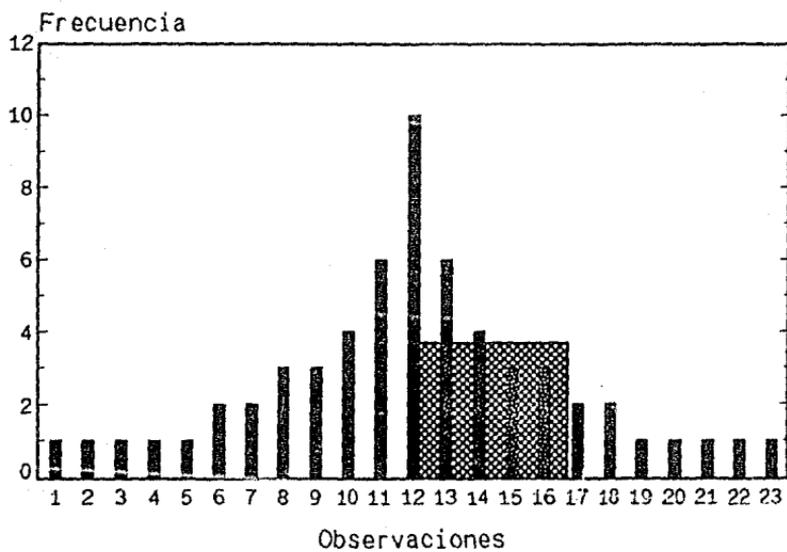
La varianza expresada gráficamente representa un cuadrado que es más pequeño, mientras menos dispersa está la serie, y crece junto con la dispersión:

En las gráficas que siguen, se dibujó un cuadrado cuyo lado es del tamaño de la desviación estándar (recuerda que es la raíz cuadrada de la varianza); y se colocó en el origen.

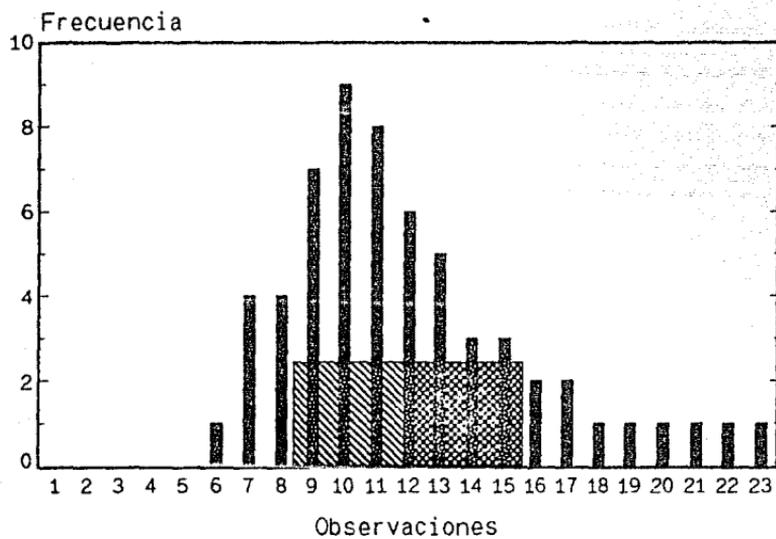
Distribución Serie X
Media = 12



Distribución Serie Y
Media = 12



Distribución Serie Z
Media = 12



CUADRO 129

Veamos ahora unas aplicaciones prácticas de todo esto:

Se ha observado que se puede esperar que del 90% al 100% de las observaciones de una población, caigan en el intervalo

$$\mu \pm 3\sigma$$

Como ya hemos hecho en otras ocasiones, veámoslo con números:

Calculemos el intervalo $\mu - 3\sigma$, $\mu + 3\sigma$ para los sueldos de Pedro, Juan y Luisa.

Los datos originales son:

						--- σ ---
Pedro	500	500	500	500	500	0
Juan	500	-1000	0	500	2500	1140.2
Luisa	100	200	500	500	1200	384.7

Los intervalos, y la cantidad de datos en cada uno de ellos son:

	<u>$\mu - 3\sigma$</u>	<u>$\mu + 3\sigma$</u>	<u>datos en el intervalo</u>
Pedro	500	500	100%
Juan	-2921	3921	100%
Luisa	-654	1654	100%

Coo verás esto es bastante útil, pues si tienes una población de la que conoces solamente la media y la desviación estándar, puedes determinar un intervalo el cual esté al menos el 90% de ella.

CUADRO 130

Hay otra medida muy útil, basada en las que hemos visto en esta lección, que permite comparar los parámetros de dos poblaciones.

Supongamos que a un maestro se le aplican dos encuestas (test psicológicos). Con la primera se pretende medir la habilidad que tiene para transmitir sus conocimientos, y con la segunda el liderazgo que ejerce sobre los alumnos.

Los resultados que obtuvo el maestro fueron:

	No. de respuestas correctas
1a. prueba (pedagogo)	36
2a. prueba (lider)	65

Estos exámenes ya se han aplicado a muchas personas con distintas ocupaciones y se ha visto que para ellas los resultados son:

	Media	Desv. estándar
1a. prueba	20 correctas	10
2a. prueba	45 correctas	15

La pregunta es: ¿en qué salió calificado más alto este maestro? ¿Es mejor pedagogo o mejor lider?

Parece que mejor lider ¿no?

¿Tú qué dices? ¿Pedagogo o lider?

Vamos a guardar esta respuesta, para compararla después con la que nos de esta medida que vamos a estudiar.

CUADRO 131

¿Qué hacemos para comparar estos resultados?

Primero veamos cuánto se aleja el valor obtenido por nuestro "profe" de la media de la población.

1a. prueba	36-20 = 16 respuestas a la prueba 1
2a. prueba	65-45 = 20 respuestas a la prueba 2

Pero estos números no son comparables porque las unidades con que se mide el 16 son "respuestas a la prueba 1", mientras que las unidades del 20 son "respuestas a la prueba 2"

Si dividimos entre la desviación estándar de la población, que se mide en las mismas unidades, resolveremos el problema.

Esto es:

$$\begin{aligned} z &= (16 \text{ resp. a prueba 1}) / (10 \text{ resp. a prueba 1}) \\ &= 16/10 \\ &= 1.6 \text{ (sin unidades)} \end{aligned}$$

$$\begin{aligned} z &= (10 \text{ resp. a prueba 2}) / (15 \text{ resp. a prueba 2}) \\ &= 20 / 15 \\ &= 1.3 \text{ (sin unidades)} \end{aligned}$$

Ahora ya podemos concluir que es mejor pedagogo que líder.

Tu habías dicho que (lo que haya respondido), pero tal vez sólo por intuición, ahora ya tenemos un método confiable para hacer afirmaciones de este estilo con la certeza de que son ciertas.

CUADRO 132

Esta medida se llama puntaje estándar ("standard score"), y su fórmula general es:

$$z = (x - \mu) / \sigma$$

donde:

z - puntaje estándar

x - una observación particular de una población.

μ - media de la población

σ - desviación estándar de la población.

Esta medida ajusta las diferencias tanto en variabilidad como en tendencia central de una observación.

CUADRO 133

Un estudiante obtuvo 8.4 en el examen final de Matemáticas, en el cual la media del grupo fue de 7.6, y la desviación estándar 1. En su examen final de Física el promedio de grupo fue 8.2 y la desviación estándar 1.6. El obtuvo 9 de calificación.

¿En qué materia salió mejor?

Si dice que MATEMATICAS:

Tienes razón, porque al calcular z , se obtiene:

$$\text{Matemáticas } z = (8.4 - 7.6) / 1 = 0.8$$

$$\text{Física } z = (9 - 8.2) / 1.6 = 0.8 / 1.6 = 0.5$$

Por lo tanto, tuvo una mejor actuación en Matemáticas.

Cualquier otra respuesta:

No, porque los resultados son:

$$\text{Matemáticas } z = (8.4 - 7.6) / 1 = 0.8$$

$$\text{Física } z = (9 - 8.2) / 1.6 = 0.8 / 1.6 = 0.5$$

Por lo tanto, tuvo una mejor actuación en Matemáticas.

(Se muestran los resultados obtenidos a lo largo de la lección.)

Así terminamos la quinta lección...

LECCION 6

MUESTREO Y DISTRIBUCIONES MUESTRALES

LECCION 6

Muestreo y distribuciones muestrales

OBJETIVOS:

CONOCIMIENTOS: Al terminar la lección, el estudiante podrá definir los siguientes conceptos: espacio muestral, evento, variable aleatoria, distribución muestral y valor esperado. Podrá decir por qué es malo un experimento que conste de una sola observación.

HABILIDADES: Podrá calcular la probabilidad de que ocurra un evento, y utilizará fórmulas para el cálculo de la media y la varianza de una distribución muestral.

ACTITUDES: Aprenderá que la Estadística Inferencial debe aplicarse adecuadamente para obtener resultados satisfactorios, pues de lo contrario, las inferencias pueden resultar poco valederas para la población muestreada.

NOTA IMPORTANTE:

¿Ya has usado la calculadora?

Si responde sí:

Bien, entonces continuamos con la lección...

Si responde no:

Si en cualquier momento necesitas usar una calculadora, puedes llamarla tecleando <Alt> C. Debes oprimir estas dos teclas al mismo tiempo. Si quieres ayuda al usar la calculadora, teclea H (Help). Si quieres traer un resultado obtenido con la calculadora oprime <Alt> F.

Para calificarte adecuadamente, por favor en todas las respuestas que no sean enteras, usa sólo 2 decimales, redondeando si es necesario. (0-4 al dígito anterior y 5-9 al superior). Si no lo haces así, se te puede considerar mal una respuesta correcta.

Vamos a practicar: llama a la calculadora y calcula $12.1 * 5.33$. (multiplicación). Trae el resultado a la lección con <Alt> f. O retira la calculadora con <Esc> y teclea tú el resultado.

Si contesta 64.493:

Bién, pero olvidaste redondear a dos dígitos decimales significativos. Dame nuevamente el resultado.

Si la respuesta es 64.49:

¡Perfecto!, comprendiste la idea. Si quieres, practica un poco más. Recuerda que los signos de operación son: más +; menos -; por * y entre /. (Si ya no quieres practicar da <Return>).

Cualquier otra respuesta:

No, lee con atención como se usa, y vuelve a contestar.

CUADRO 141

¿Recuerdas que en una lección anterior vimos la diferencia entre población y muestra?:

Completa el siguiente texto:

La población se refiere a _____ el conjunto de datos, mientras que la muestra se refiere sólo a _____ del conjunto de datos.

¿Qué respondes a la primera parte?

Si responde TODO:

¡Muy bien!

Si responde otra cosa:

Población = TODO el conjunto de datos.

¿Cuál es la respuesta a la segunda parte?

Si responde PARTE:

¡Muy bien!

Si responde diferente:

Muestra = PARTE del conjunto de datos.

Sucede muy frecuentemente que al estudiar una población no se tiene disponible la información de todos sus elementos, sino sólo de una parte de ellos, y se usan estos para "inferir" sobre el comportamiento de la población.

CUADRO 142

Veamos unos ejemplos:

1) Un productor quiere saber cuánta aceptación tiene un producto alimenticio que quiere introducir en el mercado. No le es posible conocer la opinión de todos los compradores potenciales, por lo que entrevista sólo a algunos clientes preguntándoles su opinión sobre el producto.

2) Se está investigando qué tan efectivo es un nuevo fármaco al ser administrado en enfermos que padecen diabetes. Se selecciona un grupo de pacientes de prueba y se analizan los efectos que produce dicho medicamento en ellos.

3) En una línea de producción, se prueban al azar (esto significa que cualquiera puede ser escogido) algunos artículos terminados, para aceptar o rechazar todo el lote de artículos producidos.

CUADRO 143

Puedes observar que en estos ejemplos se está trabajando con muestras, al tomar una muestra, se hacen experimentos. En cada uno de nuestros ejemplos un experimento sería:

1) Cada opinión de un comprador.

2) La reacción de un paciente a una dosis de medicamento.

3) Las pruebas a un artículo que sale de la línea de producción.

¿Qué puedes observar como característica común en estos experimentos

- a) Es elemental (uno a la vez)
- b) Puede repetirse
- c) Depende de la persona que toma la muestra.

Si responde a:

No siempre sucede así. Un experimento puede consistir en tirar 2 dados juntos, o en sacar 5 cartas de una baraja. O en nuestro primer ejemplo, la encuesta para la entrevista puede estar formada por varias preguntas.

Si responde b:

Tienes razón, esta es la principal característica de un experimento.

Si la respuesta es c:

Esto no siempre se cumple. Cuando lo que se observa es, por ejemplo, cuántos centímetros cúbicos de lluvia cayeron en una región en un mes dado; o, como en nuestro segundo ejemplo, la efectividad de un fármaco al ser administrado a un enfermo, es claro que el resultado de estos experimentos no depende de la persona que está tomando la muestra.

CUADRO 144

La idea de que un experimento pueda repetirse muchas veces, es crítica en el desarrollo y comprensión de la estadística inferencial pues una gran parte de las conclusiones que se obtienen se hacen a partir de planteamientos del tipo: ¿qué pasaría si este experimento se repitiera indefinidamente?

Los métodos estadísticos nos enseñan a obtener conclusiones sobre una población a partir de experimentos repetibles.

Aquí los conceptos de probabilidad juegan un papel muy importante, pues como ya lo hemos mencionado antes, la Probabilidad es la base de la Estadística Inferencial. Demos pues, un repaso a la teoría de Probabilidad.

Un ejemplo clásico de un evento probabilístico es el experimento que consiste en lanzar una moneda al aire. Se ocurren entonces preguntas como:

¿Si elijo águila, cuánta probabilidad tengo de atinarle? (Hablando con propiedad, ¿Cuál es la probabilidad del evento águila?)

No se necesita pensar mucho para concluir que si la moneda está bien balanceada (si es "honestá"), más o menos la mitad de las veces cae sol y la otra mitad águila. Así que podemos decir que la probabilidad de que caiga la cara elegida es ____

¿Cuánto?

Si responde un medio:

Claro, ¿verdad que es algo lógico?. Veamos ahora una definición formal.

Si responde 50%:

Tu respuesta no es muy exacta, pues la probabilidad puede tomar valores entre 0 y 1 únicamente.

Si da cualquier otra respuestas:

Dudas ¿eh?, bueno vamos a ver cómo se asigna probabilidad a un evento de una manera más formal para que se aclare por qué es $1/2$ este valor.

CUADRO 145

La probabilidad de un evento se define como la FRECUENCIA RELATIVA DE ESTE EVENTO, si el experimento subyacente se repite INDEFINIDAMENTE.

¡Se me pasó la mano de formalidad!

Veámoslo con el ejemplo de la moneda. Se simuló el experimento de lanzar una moneda al aire muchas veces (con un programa de computadora que genera al azar números 0 y 1, y relacionando: si el número generado es 0 significa que cayó aguilá y si es 1, entonces cayó sol). Los resultados obtenidos con este programa fueron:

Número de experimentos:

	100	200	300	400	500	600	700
águila(0)	48	106	147	216	243	291	346
sol (1)	52	94	153	184	257	309	354

Frecuencia Relativa (en porcentaje):

águila	48.00	53.00	49.00	54.00	48.60	48.50	49.43
sol	52.00	47.00	51.00	46.00	51.40	51.50	50.57

Que graficamente se ve:

PROBABILIDAD DE UN "VOLADO"

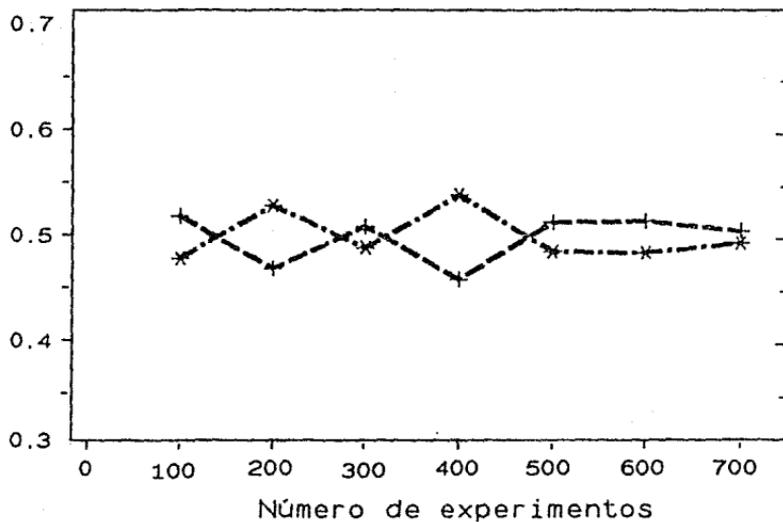
AGUILA

*—

SOL

-+—

Probabilidad



¿Observaste como se acerca cada vez más a $1/2$?, así que si se efectuara el experimento un número infinito de veces, la frecuencia relativa, que es la probabilidad, sería de exactamente $1/2$. (Si quieres volver a observar las gráficas teclea VER GRAFICA)

CUADRO 146

Otras definiciones de Probabilidad:

ESPACIO MUESTRAL: El conjunto de todas los posibles resultados de un experimento. Por ejemplo, al lanzar al aire una moneda hay dos posibles resultados: aguilá o sol; al tirar un dado hay seis posibles resultados: 1,2,3,4,5 y 6.

EVENTO: Es un elemento del espacio muestral. V.gr.: que caiga aguilá.

VARIABLE ALEATORIA: Es una regla para asignar valores numéricos a los eventos de un espacio muestral. En el ejemplo, definimos que aguilá = 0 y sol = 1. Comúnmente se da ese nombre tanto a la regla de asignación como a la variable que toma el valor asignado.

CUADRO 147

Como regla práctica, para calcular la probabilidad de un evento A, designada por $P(A)$, (considerando que cada uno de los eventos del espacio muestral tienen la misma oportunidad de presentarse), se usa la siguiente razón:

$$P(A) = \text{casos favorables} / \text{casos posibles}$$

La probabilidad de que al lanzar una moneda caiga aguila, es $1/2$ pues se tiene un caso favorable (aguila), de dos posibles (aguila y sol).

La probabilidad se designa con la letra "p" y siempre toma un valor entre 0 y 1.

CUADRO 149

Ya con estos conceptos de Probabilidad en mente, ¿quieres ver cómo un experimento malo para estimar un parámetro de la población, puede convertirse en uno bueno mediante una mejor selección de la muestra?.

Para tener mejor perspectiva, conozcamos primero los valores de los parámetros de la población, y después veremos cómo seleccionar la muestra que nos proporcione buenos estimadores.

La tienda Electrónica S.A. maneja una línea de radios de mesa. Vende modelos AM o FM, con reloj y sin reloj.

El margen de utilidad es:

\$6,000 en cada radio AM con o sin reloj
 \$12,000 en cada radio FM sin reloj
 \$18,000 en cada radio FM con reloj

El día de ayer se vendieron 12 radios, dejando las siguientes ganancias.

	Ganancia unitaria	Ganancia total
4 AM sin reloj	\$6,000	\$24,000
3 AM con reloj	\$6,000	\$18,000
4 FM sin reloj	\$12,000	\$48,000
1 FM con reloj	\$18,000	\$18,000

		\$108,000

CUADRO 150

Calculemos la media y la varianza de esta población:

Para no hacer muchas operaciones, y ya que los datos son iguales, hagamos los cálculos sin ceros, (que la unidad de medida sea miles de pesos), además usemos la fórmula para datos agrupados. Recuerda que para ello se requiere primero la tabla de distribución de frecuencias.

x	f
6	7 (3+4)
12	4
18	1

Recordaras que las fórmulas son:

$$\text{Media: } \mu = \Sigma (f)(xi) / N$$

$$\text{Varianza: } \sigma^2 = \Sigma [f(xi-\mu)^2] / N$$

¿Cuánto vale la media?

Si responde 9:

¡Bien!

Si responde algo diferente:

La forma de calcularla es:

$$\begin{aligned} \mu &= [(4*6)+(3*6)+(4*12)+(1*18)] / 12 \\ &= 108 / 12 = 9 \end{aligned}$$

La ganancia promedio fue de \$ 9 mil pesos por radio

¿Cuánto vale la varianza?

Si la respuesta es 15:

¡OK!

Si no da el resultado correcto:

La forma de calcularla es:

$$\begin{aligned}\sigma^2 &= (7*(6-9)^2 + 4*(12-9)^2 + 1*(18-9)^2) / 12 \\ &= (7*9) + (4*9) + (1*81) / 12 = (63+36+81) / 12 \\ &= 180 / 12 = 15\end{aligned}$$

Por lo tanto, la desviación estándar es

$$\sigma = \sqrt{15} = 3.87 \text{ mil pesos o } 3 \text{ mil } 870 \text{ pesos.}$$

CUADRO 151

Ahora ya conocemos los valores de nuestra población

Número de elementos: 12 (es finita, su fin es el elemento 12)

Media: 9 mil pesos

Desviación estándar: 3.87 mil pesos o \$3,870 pesos

En lo que viene a continuación vamos a utilizar mucho estos tres parámetros de la población en estudio. ¡Procura recordarlos!

CUADRO 152

Resulta que en Electrónica S.A., al empleado que registra las ventas, Juan Pérez Oso, no le gusta trabajar. Por ejemplo, hoy en la mañana, cuando su jefe el gerente de ventas, al ver las doce notas por la venta de radios, le preguntó cuánto se había ganado por estas ventas, el muy flojo tomó una de las notas (que correspondía a un radio AM, con el que se ganan \$ 6,000), y respondió, generalizando la ganancia obtenida al vender ese radio a todas las ventas: "seis mil pesos por cada uno"

Lo que hizo este empleado es un experimento (mal hecho), vamos a analizarlo desde un punto de vista estadístico.

CUADRO 153

Pero antes repasemos algunos conceptos, completa las siguientes frases:

Las doce notas de venta representan el espacio -----

¿Cuál es la respuesta?

Si responde espacio muestral, se le dice que es correcto, si no lo hace así se despliega la respuesta.

Cada una de las notas es un -----

¿Cuál es la respuesta?

Si responde que cada nota es un evento del espacio muestral. se le indica que está bien, en caso contrario se da la respuesta correcta.

Ahora si, desde el punto de vista estadístico, el valor de la variable aleatoria (utilidad) está dado, de acuerdo al experimento ideado por el empleado flojo, por la siguiente regla: extrae una nota, si corresponde a un radio AM, la variable vale 6; si es FM sin reloj vale 12; y si es FM con reloj vale 18.

Para analizar más a fondo este experimento, construyamos una tabla con los posibles valores de la variable aleatoria, asignando la probabilidad que corresponde a cada evento o nota:

Te indico el valor solicitado como siempre ¿OK?

#	No.	Nota	Modelo	Probabilidad	Valor variable aleatoria
-----	-----	-----	-----	-----	-----
1)	2		AM s/reloj	1/12	6
2)	4		AM s/reloj	-----	--
3)	5		AM s/reloj	1/12	6
4)	9		AM s/reloj	1/12	6
5)	6		AM c/reloj	1/12	6
6)	8		AM c/reloj	1/12	6
7)	11		AM c/reloj	1/12	6
8)	1		FM s/reloj	-----	--
9)	3		FM s/reloj	1/12	12
10)	7		FM s/reloj	1/12	12
11)	10		FM s/reloj	1/12	12
12)	12		FM c/reloj	-----	--

CUADRO 154

Esta tabla puede re-escribirse considerando sólo los diferentes valores de la variable aleatoria, y la probabilidad asociada con cada uno de ellos, y entonces resulta:

Variable Aleatoria (x)	Probabilidad P(x)
6	$7/12 = 0.5833$
12	$4/12 = 0.3333$
18	$1/12 = 0.0833$

	1.0 0.9999

Estos valores representan lo que pasaría si el experimento del empleado flojo se repitiera una y otra vez (considerando teóricamente que el experimento se pudiera repetir indefinidamente). La probabilidad $7/12$ indica que el 58.3% de las veces, el empleado reportará una ganancia de \$6,000 pesos por cada venta; el 33.3% de las veces reportará una ganancia de doce mil pesos por cada uno, y el 8.33% reportará una utilidad de 18 mil pesos. (Recuerda que la ganancia media de la población es de \$9,000 pesos).

A esto se le llama DISTRIBUCION DE PROBABILIDAD DE LA VARIABLE ALEATORIA. Nota el parecido con una tabla de distribución de frecuencias, sólo que aquí, en lugar de anotar el número de veces que se repite una observación, se consigna el número de veces que se PUEDE repetir un evento.

CUADRO 155

La distribución de la variable aleatoria se parece mucho a la distribución de la población. En realidad, la única diferencia es que el peso que se da a los diferentes valores de la variable (en nuestro caso: utilidad), son probabilidades en lugar de frecuencias, es decir, frecuencias relativas de las repeticiones de un experimento en vez de observaciones de la población.

Esto es:

La distribución de una población mide: FRECUENCIAS

La distribución de una variable aleatoria mide: PROBABILIDAD.

CUADRO 156

Calculemos la media y la varianza de la distribución de nuestra variable aleatoria.

Las fórmulas son: (otra vez aprecia la analogía con la distribución de una población).

$$\text{Media:} \quad \mu = \Sigma [x * P(x)]$$

$$\text{Varianza:} \quad \sigma^2 = \Sigma [(x-\mu)^2 * P(x)]$$

(Observa que la frecuencia relativa f/N es equivalente a la probabilidad de x , $P(x)$)

Con los valores que tenemos (en miles de pesos), completa la tabla:

x	P(x)	x * P(x)
6	7/12	-----
12	4/12	
18	1/12	

Da el valor de $x * P(x)$

Si responde 42/12, ¡bién!, si no se le muestra la respuesta correcta

Yo continuo con esta parte

x	P(x)	x*P(x)
6	7/12	42/12
12	4/12	48/12
18	1/12	18/12

		108/12

Y tú completa el resto:

x	P(x)	x*P(x)	$(x-\mu)^2$	$(x-\mu)^2 * P(x)$
6	7/12	42/12	9	63/12
12	4/12	48/12	-	36/12
18	1/12	18/12	81	
		-----		-----
		108/12		

¿Cuál es el valor de $(x-\mu)^2$ si $x = 12$?

Si responde la respuesta correcta que es $(12-9)^2 = 9$, ¡bién!, si no se le muestra.

¿Cuánto vale $(x-\mu)^2 * P(x)$ si $x = 18$?

La respuesta es $81 * 1/12 = 81/12$. Si la da ¡OK!, si no es así, se le despliega este valor.

Entonces la tabla completa queda:

x	P(x)	x*P(x)	(x-μ) ²	(x-μ) ² *P(x)
6	7/12	42/12	9	63/12
12	4/12	48/12	9	36/12
18	1/12	18/12	81	81/12
		108/12		180/12

Calculemos ahora la media y la varianza. Recuerda que

$$\text{Media: } \mu = \Sigma [x * P(x)]$$

$$\text{Varianza: } \sigma^2 = \Sigma [(x-\mu)^2 * P(x)]$$

¿Cuánto vale la media?

La respuesta correcta es $\mu = 108 / 12 = 9$, si da algún otro resultado, se le indica el correcto.

¿Cuánto vale la varianza?

Si responde 15 está bien, si no se le hace ver que lo correcto es $\sigma^2 = 180 / 12 = 15$

¡Son los mismos valores que en la población! Es agradable comprobar que la distribución de la población y la de la variable aleatoria son similares.

A la media de la distribución de la variable aleatoria se le llama VALOR ESPERADO. Dado que la probabilidad es una frecuencia relativa a "largo plazo", podemos interpretar la media de una variable aleatoria como EL VALOR PROMEDIO PARA LA VARIABLE SI EL NUMERO DE EXPERIMENTOS SE INCREMENTA SIN LIMITE.

En nuestro ejemplo el valor esperado es 9. Aunque nunca ocurre, es lo que se espera en el sentido de promedio, es decir, es el promedio esperado.

CUADRO 157

Regresando al caso de Electrónica S.A., como nosotros sí conocemos el valor de los parámetros de la población, ($\mu = 9$ y $\sigma = 3.87$ mil pesos), sabemos que su experimento no es conveniente, pues la mayor parte de las veces (el 58.33%) va a reportar una utilidad de 6000 pesos, aunque eso no sea cierto.

El error fundamental es que realiza un solo experimento. En general, este tamaño de muestra nunca es aceptable; vamos nosotros a tomar una muestra de tamaño 2.

(Dado el tamaño de la población, Juan Pérez bien podría haber hecho todas las operaciones que hicimos al principio y podría tener información completa de la población, pero creo que ya te dije que es flojo, así que nos sentiremos contentos con que trabaje con una muestra de dos elementos).

CUADRO 158

Una de las propiedades más deseables de una muestra es que represente en miniatura la población de donde fué tomada. Un buen método para que esto suceda es tomar la muestra al azar, es decir, que cada elemento de la población tenga la misma oportunidad de ser seleccionado para formar parte de la muestra. ¿Qué te parece el siguiente método para seleccionar una muestra de 2 notas?

1. - Tenemos 12 posibles resultados. Necesitamos un instrumento que nos proporcione ese número de opciones. Podemos pensar en tirar AL MISMO TIEMPO un dado (6 posibles resultados) y una moneda (2 posibles resultados), ambos "honestos"

2.- Numeremos las notas por el reverso, del 1 al 12.

3.- Establezcamos la siguiente relación:

Si sale	Se incluye en la muestra la nota no.	Probabilidad
-----	-----	-----
águila y 1	1	1/12
águila y 2	2	1/12
águila y 3	3	1/12
águila y 4	4	1/12
águila y 5	5	1/12
águila y 6	6	1/12
sol y 1	7	1/12
sol y 2	8	1/12
sol y 3	9	1/12
sol y 4	10	1/12
sol y 5	11	1/12
sol y 6	12	1/12

CUADRO 159

Para que realmente cada nota tenga la misma probabilidad de formar parte de la muestra, es necesario "muestrear con reemplazo", es decir, tirar el dado y la moneda una vez, seleccionar la nota según la regla definida en la tabla anterior, y luego repetir el experimento exactamente igual; esto significa que puede llegar a salir dos veces seleccionada la MISMA nota.

Al operar de esta forma se obtienen experimentos independientes entre sí, que significa que el resultado obtenido en uno no afecta en nada el resultado del otro.

¿Cuál es el espacio muestral?. Recuerda que son todos los posibles eventos.

Ya que el espacio muestral es muy amplio, y enumerarlo sería muy pesado, únicamente, responde la siguiente pregunta:

¿Cuántas son todas las posibles parejas que se pueden formar con los números de las notas?. (Considera que puede repetirse la misma nota)

Si responde 144:

¡Muy bien!

Si la respuesta es 66:

Son más, por todas las que se repiten. Si consideras las combinaciones de 12 tomadas de 10 en 10, sería lo mismo la muestra 1-2 que la 2-1

Las enumeramos a continuación:

1-1	1-2	1-3	1-4	1-5	1-6	1-7	1-8	1-9	1-10
								1-11	1-12
2-1	2-2	2-3	2-4	2-5	2-6	2-7	2-8	2-9	2-10
								2-11	2-12
3-1	3-2	3-3	3-4	3-5	3-6	3-7	3-8	3-9	3-10
								3-11	3-12
4-1	4-2	4-3	4-4	4-5	4-6	4-7	4-8	4-9	4-10
								4-11	4-12
5-1	5-2	5-3	5-4	5-5	5-6	5-7	5-8	5-9	5-10
								5-11	5-12
6-1	6-2	6-3	6-4	6-5	6-6	6-7	6-8	6-9	6-10
								6-11	6-12
7-1	7-2	7-3	7-4	7-5	7-6	7-7	7-8	7-9	7-10
								7-11	7-12
8-1	8-2	8-3	8-4	8-5	8-6	8-7	8-8	8-9	8-10
								8-11	8-12
9-1	9-2	9-3	9-4	9-5	9-6	9-7	9-8	9-9	9-10
								9-11	9-12
10-1	10-2	10-3	10-4	10-5	10-6	10-7	10-8	10-9	10-10
								10-11	10-12
11-1	11-2	11-3	11-4	11-5	11-6	11-7	11-8	11-9	11-10
								11-11	11-12
12-1	12-2	12-3	12-4	12-5	12-6	12-7	12-8	12-9	12-10
								12-11	12-12

CUADRO 160

Una vez que obtenemos al azar dos notas, lo que sigue es calcular la media aritmética de estos dos experimentos. Esta es la MEDIA MUESTRAL que se representa por \bar{X} barra y se calcula como ya sabemos:

$$\bar{X} \text{ barra} = \Sigma x_i / N$$

En la siguiente tabla mostraremos todos los posibles valores de la media muestral para un tamaño de muestra 2. Ayúdanos a completarla por favor ¿sí?. (Por razones de espacio abreviaré \bar{X} barra como X_b).

Si el primer valor de la variable aleatoria (v.a.) es I Y el segundo valor de la variable aleatoria es:

I	6	12	18
---	---	----	----

	I			
6	I	Xb = 6	Xb = 9	Xb = 12
12	I	Xb = -	Xb = -	Xb = -
18	I	Xb = -	Xb = -	Xb = -
	I			

¿Cuánto vale X barra en el primer caso?

Si responde 9, está bien. Si no se le indica que la respuesta correcta es $(12+6)/2 = 9$

¿Cuánto vale X barra si los valores de la variable aleatoria son 18 y 12?

Funciona de la misma forma que en la pregunta anterior, y la respuesta que se considera correcta es $(12+18)/2 = 15$

La tabla ya terminada queda:

Si el primer valor de la variable aleatoria (v.a.) es I Y el segundo valor es:

I	6	12	18
---	---	----	----

	I			
6	I	Xb = 6	Xb = 9	Xb = 12
12	I	Xb = 9	Xb = 12	Xb = 15
18	I	Xb = 12	Xb = 15	Xb = 18

CUADRO 161

Aquí cabe otro paréntesis de probabilidad.

Dos eventos son independientes si el hecho de que se dé uno, no causa ningún efecto sobre la probabilidad del otro. En otras palabras, la probabilidad de un evento es la misma, ocurra o no el otro evento.

Cuando esto sucede se cumplen dos reglas, las cuales definiremos usando la siguiente notación:

Sean A y B dos eventos independientes

por ejemplo A: el día está soleado

B: llueve

$P(A)$ - probabilidad de que ocurra el evento A.

$P(B)$ - probabilidad de que ocurra el evento B.

CUADRO 160

Las reglas son:

1a. Regla:

$$P(A \cup B) = P(A) + P(B)$$

En palabras, en eventos independientes, la probabilidad de la unión es igual a la suma de las probabilidades.

Por ejemplo, de un pueblito de México se conoce que:

$P(A) = 1/2$ - probabilidad de que el día esté soleado

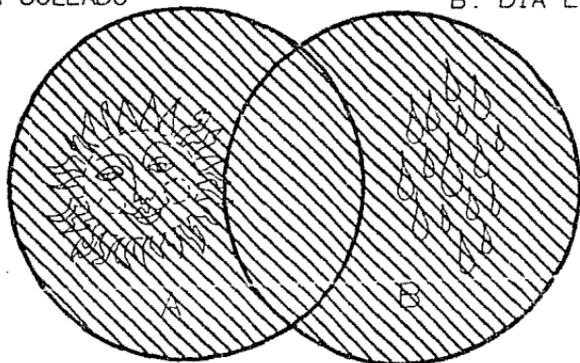
y $P(B) = 1/4$ - probabilidad de que llueva (son datos inventados)

Entonces: $P(A \cup B) = 1/2 + 1/4 = 3/4 = 6/8$
Esto es se puede esperar que, de cada ocho días, seis de ellos haya sol, o llueva, o ambas cosas en ese lugar.

UNION $A \cup B$

A: DIA SOLEADO

B: DIA LLUVIOSO

 $A \cup B$ - QUE EL DIA ESTE SOLEADO O LLUVIOSO

AL PERTENECER A UNO DE LOS CONJUNTOS (O EVENTOS)
SE PERTENECE A LA UNION

CUADRO 162

2a. Regla:

$$P(A \cap B) = P(A) * P(B)$$

Que en palabras es, la probabilidad de la intersección es igual al producto de las probabilidades, siempre y cuando los eventos sean independientes.

Con los datos del ejemplo,

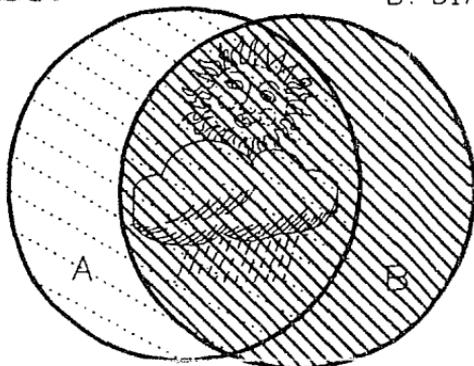
$$P(A \cap B) = 1/2 * 1/4 = 1/8$$

Entonces, en el pueblo del que estamos platicando, se puede esperar que un día de cada ocho llueva con sol (y haya arcoiris).

INTERSECCION $A \cap B$

A: DIA SOLEADO

B: DIA LLUVIOSO

 $A \cap B$ - QUE EL DIA ESTE SOLEADO Y LLUVIOSO

SE REQUIERE ESTAR EN AMBOS CONJUNTOS PARA
PERTENECER A LA INTERSECCION.

CUADRO 164

Calculemos entonces, la probabilidad que tiene cada uno de los valores obtenidos en la tabla que veníamos manejando, la tabla es:

Si el primer valor de la variable aleatoria (v.a.) es	I	Y el segundo valor es		
	I	6	12	18
6	I	Xb = 6	Xb = 9	Xb = 12
12	I	Xb = 9	Xb = 12	Xb = 15
18	I	Xb = 12	Xb = 15	Xb = 18

Como el segundo resultado no depende de lo que se haya obtenido en el primero, estos es, como nuestros eventos son independientes, se cumple que $P(A \cap B) = P(A) * P(B)$

Entonces calculemos la probabilidad de cada evento conjunto, considerando las probabilidades de los eventos por separado:

Hagamos como en otras ocasiones, te muestro la tabla y te indico qué valor estoy solicitando por medio de "flasheo"

Si el primer valor de la variable aleatoria (v.a.) es	I	Y el segundo valor es		
	I	6	12	18
	[P(x)]	[7/12]	[4/12]	[1/12]
6	[7/12]	-----	28/144	7/144
12	[4/12]	28/144	-----	4/144
18	[1/12]	7/144	4/144	-----

¿Cuánto?

Si responde 38.89, tienes toda la razón, en caso contrario se le señala esta respuesta.

Y que disminuyó la probabilidad de obtener un valor por debajo del real (6 mil pesos) de 58.33 cuando la muestra era de tamaño 1 a 34.02% ahora que es de tamaño 2.

CUADRO 166

La media de \bar{X} se calcula igual que si fuera cualquier otra variable aleatoria, se multiplican los valores de la variable aleatoria por sus probabilidades y se obtiene la suma:

$$\begin{aligned} \mu &= \sum X \text{ barra} * P(X \text{ barra}) \\ &= [(6)(49/144) + (9)(56/144) + (12)(30/144) + \\ &\quad + (15)(8/144) + (18)(1/144)] \\ &= [294/144 + 504/144 + 360/144 + 120/144 + 18/144] \\ &= 1296/144 \\ &= 9 \end{aligned}$$

Observa que resultado más interesante:

$$\mu = \mu$$

\bar{X}

LA MEDIA O VALOR ESPERADO DE LA DISTRIBUCION MUESTRAL DE LA MEDIA ES IGUAL A LA MEDIA DE LA POBLACION MUESTREADA.

Esto se cumple sin importar el tamaño de la muestra. ¿Recuerdas que ya habíamos hablado de esto en la lección anterior, al tratar el tema de los estimadores insesgados?

CUADRO 167

Es lógico preguntar ahora qué sucede con la desviación estándar de la distribución muestral de la media.

En la siguiente tabla podemos ver el comportamiento de estos valores para muestras de tamaño 2,3,4 y 5. Si quieres puedes calcular todos estos valores para comprobar que son los correctos.

Tamaño de la muestra: n	Media X barra	Varianza s'	Desv. Estándar s
2	9.0	7.50	2.74
3	9.0	5.00	2.24
4	9.0	3.75	1.94
5	9.0	3.00	1.73

Es fácil observar que a medida que el tamaño de la muestra aumenta, la desviación estándar de la media muestral disminuye.

Después de observar con cuidado, nuestros amigos estadísticos, que ya conoces como son de "sesudos", hallaron la relación:

$$\sigma_{\text{X barra}} = \sigma / \sqrt{n}$$

La desviación estándar de la media muestral se llama también ERROR ESTANDAR DE LA MEDIA MUESTRAL, o más brevemente ERROR ESTANDAR DE LA MEDIA. La razón del término "error" es que la dispersión de la distribución muestral de la media refleja el error que puede encontrarse al usar la media muestral en lugar de la media poblacional.

(Se muestran los resultados del alumno).

Y con esto terminamos el curso, esperando que haya cubierto tus expectativas y te haya sido de utilidad.

CAPITULO III
PRUEBA PILOTO

DESCRIPCION DE LA PRUEBA PILOTO

Con el fin de conocer las virtudes y defectos del trabajo que aquí se presenta, se decidió aplicar una prueba del programa en el campo, solicitando a algunas personas que estudiaran las lecciones del Tutor frente a una computadora, y llenaran la encuesta que sigue:

Nombre _____

Edad ___ Sexo (M) (F) ¿Dónde estudiaste Prepa? _____

¿Qué grado escolar cursas actualmente o terminaste?

Califica de 1 a 10 lo siguiente:

- 1.- Grado de conocimiento que tenías de la Estadística ANTES de estudiar el Tutor.

NULO 1 2 3 4 5 6 7 8 9 10 TOTAL

- 2.- Grado de conocimiento que tienes de la Estadística DESPUES de estudiar el Tutor.

NULO 1 2 3 4 5 6 7 8 9 10 TOTAL

- 3.- Gusto por esta materia ANTES.

NULO 1 2 3 4 5 6 7 8 9 10 TOTAL

- 4.- Gusto por esta materia DESPUES.

NULO 1 2 3 4 5 6 7 8 9 10 TOTAL

- 5.- Necesidad de la materia.

NULA 1 2 3 4 5 6 7 8 9 10 TOTAL

Llena el siguiente cuadro:

LEC- CION	TIEMPO PARA ESTUDIARLA	R E S P U E S T A S		INTENTOS
		CORRECTAS	ERRONEAS	
1				
2				
3				
4				
5				
6				

Se aplicaron 10 encuestas. La mayor parte de las personas encuestadas son estudiantes del primer semestre de la carrera de Biología en la Facultad de Ciencias de la UNAM, que es la población a la que va dirigido el Tutor. También se hizo la encuesta a profesionistas con mayor preparación para que la prueba fuera más completa y se tuvieran críticas más estrictas.

RESULTADOS DE LA PRUEBA

Los resultados promedio de esta encuesta fueron:

- 1.- Grado de conocimiento de la Estadística ANTES de estudiar el Tutor.

X (3.6)

NULO	1	2	3	4	5	6	7	8	9	10	TOTAL
------	---	---	---	---	---	---	---	---	---	----	-------

- 2.- Grado de conocimiento de la Estadística DESPUES de estudiar el Tutor.

X (6.6)

NULO	1	2	3	4	5	6	7	8	9	10	TOTAL
------	---	---	---	---	---	---	---	---	---	----	-------

- 3.- Gusto por esta materia ANTES.

X (3.8)

NULO	1	2	3	4	5	6	7	8	9	10	TOTAL
------	---	---	---	---	---	---	---	---	---	----	-------

- 4.- Gusto por esta materia DESPUES.

X (8.2)

NULO	1	2	3	4	5	6	7	8	9	10	TOTAL
------	---	---	---	---	---	---	---	---	---	----	-------

- 5.- Necesidad de la materia.

X (9.8)

NULA	1	2	3	4	5	6	7	8	9	10	TOTAL
------	---	---	---	---	---	---	---	---	---	----	-------

Sobre las lecciones:

LEC- CION	TIEMPO PARA ESTUDIARLA	R E S P U E S T A S		INTENTOS
		CORRECTAS	ERRONEAS	
1	11.9	6.5	3.3	9.5
2	17.9	9.9	3.7	13.4
3	29.9	22.5	5.2	27.8
4	38.8	10.8	4.2	15.0
5	61.2	28.0	5.4	33.8
6	38.4	17.7	7.0	24.7

Estos resultados en números porcentuales, y el número de preguntas que tiene cada lección puede verse en el cuadro de la página siguiente:

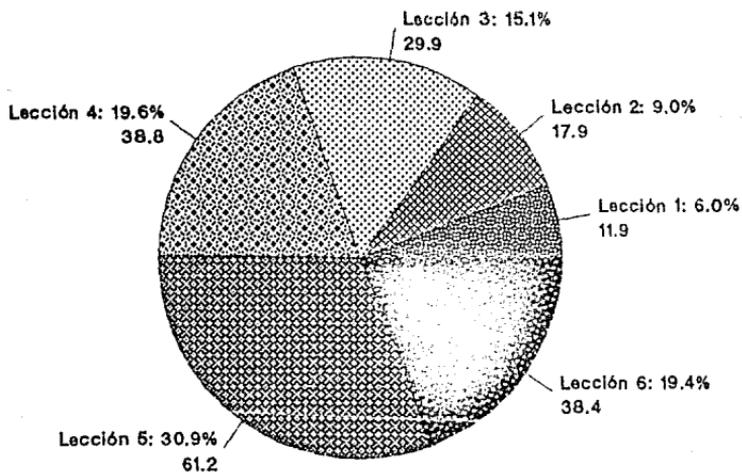
LEC- CION	R E S P U E S T A S		INTENTOS	NO. DE PREGUNTAS
	CORRECTAS	ERRONEAS		
1	68.42	34.74	100.0	9
2	73.88	27.61	100.0	12
3	80.94	18.71	100.0	25
4	72.0	28.0	100.0	15
5	82.84	15.98	100.0	34
6	71.66	28.34	100.0	26

La razón de que haya más intentos que preguntas en la lección, es que el programa contabiliza como intento cada vez que el estudiante interactúa con la máquina, ya sea oprimiendo <Return> o <?> que significa solicitud de ayuda.

En las láminas que están a continuación pueden apreciarse estos resultados gráficamente:

Distribución del tiempo empleado

Total: 198 minutos = 3 horas 18 min.



Resultados reportados de las lecciones

No. de encuestas realizadas: 10

Respuestas

correctas



Respuestas

erróneas

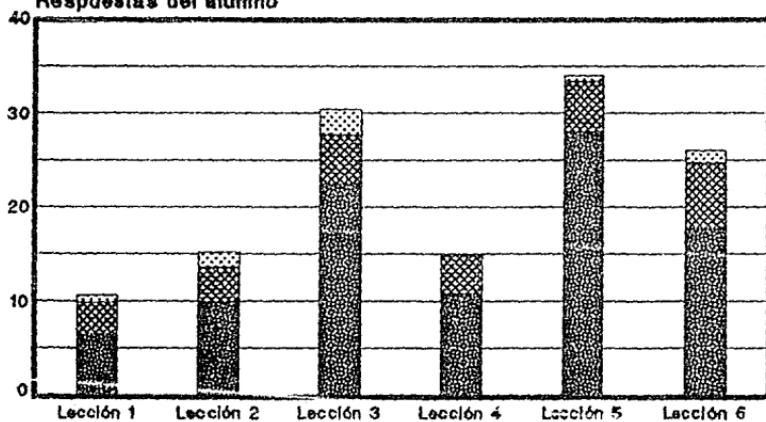


Otros

intentos



Respuestas del alumno



Resultados reportados de las lecciones

No. de encuestas realizadas: 10

Respuestas

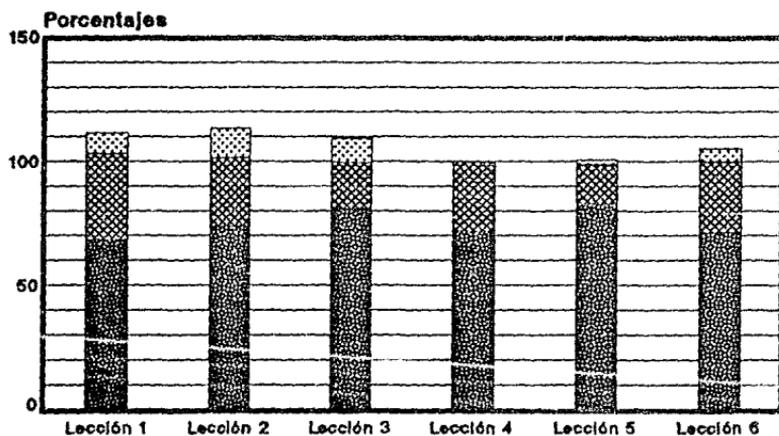
correctas

Respuestas

erróneas

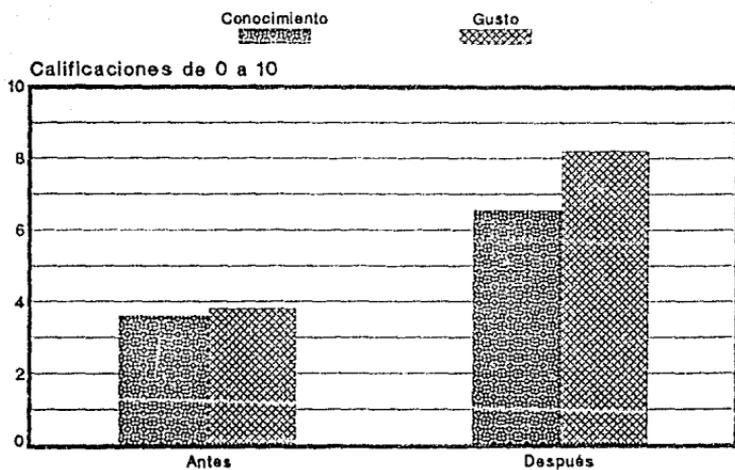
Otros

Intentos



Opinión del Tutor de Estadística

No. de encuestas realizadas: 10



CONCLUSIONES

Por los resultados presentados en el punto anterior, puede concluirse que:

Sobre los estudiantes:

- existe la constante de que consideran necesario aprender Estadística.
- fueron motivados suficientemente para terminar el Tutor, ya que dedicaron más de tres horas a estudiarlo.
- puede afirmarse que el programa cubrió sus expectativas, pues lo terminaron.

Sobre los logros tangibles del Tutor:

- los estudiantes consideran haber incrementado su conocimiento sobre la materia en un 30%.
- las respuestas correctas representan un 76.8% del total de las preguntas (son 121), por lo que puede afirmarse que los estudiantes alcanzaron un conocimiento satisfactorio de estos temas.
- un hecho muy notable de este programa es que aumentó el gusto de los estudiantes por la materia en un 44%. Este es un punto que merece atención especial por la influencia tan decisiva que se puede tener en la calidad de vida estudiantil, (y por lo tanto profesional) de las personas. Es decir, si se logra, por medio de los tutores, hacer que los estudiantes tengan gusto por el aprendizaje, la enseñanza puede volverse una tarea más sencilla y agradable.

RECOMENDACIONES

A Programadores o autores:

El lenguaje empleado para hacer este programa, Pilot, es un lenguaje muy poderoso que puede dar resultados satisfactorios con sólo una parte de los comandos que lo componen, pero usándolo en toda su capacidad, se pueden lograr programas más eficientes y de rápido mantenimiento. La recomendación consiste en que se explote al máximo las capacidades del lenguaje que se emplee en la elaboración de los tutores.

Este Tutor ofrece la facilidad de que en cada pregunta, si el estudiante necesita un repaso de la teoría, se teclea una clave que se interpreta como petición de ayuda y el programa regresa a la parte que convenga. Originalmente no se tenía esta facilidad; se incluyó por solicitud de los estudiantes y por ser hecho a posteriori resultó muy complicado. Se recomienda planear estas rutinas antes de programar y hacerlas comunes.

En general, es conveniente hacer común lo más que sea posible, por ejemplo la cuenta de aciertos y errores, la definición de colores, el reporte de resultados, etc.

Tres recomendaciones interesantes que ya no se incluyeron en este Tutor son:

- Cuando se hace alguna pregunta, se pueden usar recuadros con la información que se quiere que el estudiante memorice.

- Sugerir al estudiante que tome nota de los datos más importantes, para reforzar el aprendizaje.

- Hacer un examen final de evaluación.

A maestros:

Una recomendación que podría salir sobrando, pero que es importante, es que consideren a los programas de Enseñanza Auxiliada por Computadora como sus aliados, que exploten sus beneficios y aprovechen lo que de bueno tienen. El maestro es irremplazable y el fin de estos programas es ser un apoyo didáctico en la instrucción.

En el uso concreto de este Tutor, se recomienda seguirlo primero personalmente para conocerlo a fondo, y después darlo a los alumnos.

Se agradecerá mucho que se hiciera llegar a la autora cualquier sugerencia de mejora de estos programas.

A Alumnos

Para aprender se requiere principalmente humildad; dicen los japoneses que hay que vaciarse de uno mismo para poder adquirir sabiduría. Si la actitud al estudiar es receptiva se asimilará mucho más que si se toman posturas soberbias o críticas.

CONCLUSIONES

CONCLUSIONES

De la experiencia de haber hecho un tutor, se puede concluir que:

La Enseñanza Auxiliada por Computadora ofrece, junto con otros adelantos tecnológicos, una oportunidad excelente para mejorar los métodos de enseñanza y aprendizaje. Permite una gran creatividad, ya que los recursos de la máquina, como gráficas, colores y respuesta inmediata, entre otros, dan más alternativas que los métodos convencionales.

Las ayudas gráficas permiten expresar más claramente una idea y dejarla más grabada en la mente del estudiante.

La interacción que la computadora tiene con el estudiante hace que, en caso de cometer un error, la retroalimentación sea inmediata, por lo que la falla se conoce en el momento que ocurre.

El trabajo requerido para desarrollar un programa como el que aquí se presenta es mucho. Este Tutor es fruto de más de 500 horas de trabajo. Aunque no todo fué trabajo directo sobre el programa, pues se hizo necesario aprender el lenguaje Pilot y la técnica misma de Enseñanza Auxiliada por Computadora, además de que al aplicarlo con algunos alumnos, se recibieron opiniones que ocasionaron cambios significativos en los programas.

Pero aunque es trabajo pesado, vale la pena hacerlo porque al final se tiene un apoyo valioso en la instrucción, que permanece cuando ya se olvidó el esfuerzo de haberlo hecho.

Este trabajo puede servir como punto de partida para elaborar un programa más completo, que cubra los temas de Estadística considerados importantes en la formación de los alumnos de la Facultad de Ciencias. Podría formarse un grupo con las personas interesadas en desarrollar este curso y así, con menos esfuerzo, elaborar un Tutor más completo.

ANEXO A
GLOSARIO DE PILOT

GLOSARIO DE COMANDOS DE PILOT:

- A: ACCEPT Solicita información en la pantalla. La respuesta se almacena en las variables que se le indiquen, o si se omiten en un "buffer" o memoria auxiliar llamada %B de 80 caracteres
- AS: ACCEPT SINGLE Sólo recibe un carácter como respuesta.
- AJ: ACCEPT (AHEAD) Para solicitar la respuesta en la misma línea que se hizo la pregunta.
- C: COMPUTE Calcula una expresión que puede ser numérica o alfanumérica.
- D: DIMENSION Para definir vectores.
- E: END Termina un programa. También una subrutina en cuyo caso puede indicarse a qué etiqueta regresará.
- F: FILE Permite acceso a los archivos.
- FI: INPUT Lee un archivo y lo deja en una variable dimensionada previamente al tamaño adecuado.
- FD: Escribe en un archivo
- FX: Abre o cierra un archivo

- J:** **JUMP** Brinca a una etiqueta o una instrucción.
- J:** @A Regresa al último comando ACCEPT (A) ejecutado.
- J:** @P Brinca al siguiente control de programa (P).
- J:** @M Brinca al siguiente comando MATCH (M).
- K:** **KEEP** Guarda la expresión que se le indique en un archivo K:REC, en el cual se lleva registro de los resultados de los alumnos.
- L:** **LINK** Se enlaza con otro programa de PILOT, puede brincar a una etiqueta dentro del programa.
- M:** **MATCH** Compara la respuesta del estudiante contra lo que se le indique en esta instrucción y da un resultado booleano: si o no.
- MJ:** **MATCH OR JUMP** Si no se cumple la comparación, brinca al siguiente comando MATCH.
- P:** **PROGRAM CONTROL** Permite definir opciones para un programa.
- Estas son:
- E:** **(ESCAPE)** Permite que el estudiante use el comando ESCAPE.

F:	(FUNCTION KEY)	Permite el uso de la tecla <ESC>.
G:	(GOTO)	Permite el uso del comando GOTO.
L:	(LOWER)	Convierte a minúsculas la respuesta del estudiante.
U:	(UPPER)	Convierte la respuesta a mayúsculas.
S:	(SPACES)	Remueve los espacios en la respuesta.
R:	REMARKS	Para incluir comentarios.
S:	SOUND	Emite un sonido (se indica el tono de agudo a grave, y la duración)
T:	TYPE	Muestra el texto en pantalla o el contenido de las variables que se le indiquen.

Dentro de esta instrucción se pueden usar:

#n	Cambia el color del texto
\$n	Cambia el color del fondo
#nombre	Muestra el contenido de la variable numérica llamada "nombre"
\$nombre	Muestra el contenido de la variable alfanumérica llamada "nombre"

TH:	HANG	No avanza una línea ("line feed").
TX:	CLEAR	Pone el cursor al inicio de la ventana.
TS:	SCREEN	Define la pantalla: ventanas, colores, etc.
	Dn DELAY	Se espera n 60'avos de segundo
	Fn FOREGROUND	Color del texto.
	Bn BACKGROUND	Color del fondo.
	En ERASE BACKG.	Cambia el fondo al color n.
	Gx,y GOTO	Mueve el cursor a la posición x,y
	V VIEWPORT	Para definir una ventana
	*n(..)	Repite n veces lo que está entre paréntesis.
U:	USE	Llama una subrutina dentro del mismo programa.
V:	VARIABLE	Permite llamar subrutinas ejecutables (que esten en lenguaje de máquina) almacenadas en alguna variable.
W:	WAIT	El tiempo (en décimas de segundo) que el programa hace una pausa. Si se oprime una tecla cualquiera, se suspende la espera.

X: EXECUTE Ejecuta la expresión que se le indique, interpretándola como un comando de PILOT.

Caracteres que se pueden usar con la instrucción MATCH:

!	CONCATENACION	Para varias opciones ("or")
&	Y	Para unir preguntas ("and")
*		Representa cualquier caracter
%		Representa un espacio o fin, se usa cuando se requiere que el MATCH sea exacto.

Condicionales:

(Se dan a continuación ejemplos con la instrucción TYPE, pero esto funciona con todas las instrucciones que resulten válidas)

TY: Muestra el texto si se cumplió la condición en la última instrucción MATCH.

TN: Ejecuta la instrucción si la condición no se cumplió.

TE: Ejecuta la instrucción si se da la condición de error.

Tn: n representa el número de veces que el estudiante ha pasado por el último ACCEPT. Se ejecuta la instrucción cuando se ha contestado el número de veces que se indica.

T(var=0) Ejecuta la instrucción si la expresión entre paréntesis es verdadera.

BIBLIOGRAFIA

Graciela Pérez Rivera, Ma. Esther Aguirre Lora y
Martín Arredondo Galván.
Manual de Didáctica. Curso Introductorio.
Centro de Didáctica UNAM.
México, 1972.

Naiman Aleline
A Hard Look al Educational Software. In the rush
to equip schools with computers, we risk ignoring
some important issues.
Byte Febrero 1987.
Mc. Graw Hill.

Cañedo Dorantes Luis
Investigación Clínica
Nueva Editorial Interamericana S.A. de C.V.
México D.F., 1987.

Sokal Robert & Rohlf James
Biometry. The principles and practice of statistics
in biological research.
W.H. Freeman and Co.
San Francisco U.S.A., 1969.

Summers, Peters and Armstrong
Basic Statistics in Business and Economics.
Wadsworth.
U.S.A., 1973
2nd. edition 1977.

Hoel Paul G.
Elementary Statistics
Wiley International.
Canada, 1960.
3th. edition, 1971.

Snedecor George W. & Cochran William G.
Métodos Estadísticos.
Compañía Editorial Continental S.A.
México, 1987.

Daniel W. Wayne.
Bioestadística. Base para el análisis de las ciencias
de la salud.
Editorial Limusa.
México, 1987.

Huff Darrell
How to Lie with Statistics.
Penguin Books.
England 1954.
7th edition 1979.

Soto Franco Mario Antonio y Gómez Pérez Rocío.
Enseñanza Asistida por Computadora. Tesis para
obtener el título de Ingeniero en Cibernética y
Ciencias de la Computación.
Universidad La Salle.
México, 1985

Ortega Andrade Eduardo y Ortega Cuellar Gabriela
Panorámica de la Enseñanza Auxiliada por Computadora
y Propuesta del Diseño para un Sistema Tutorial.
Tesis para obtener el título de Licenciado en
Informática.
Universidad Anahuac.
México, 1986.

Meyer Markle Susan
Instrucción Programada. Análisis de cuadros buenos y
malos.
Limusa Wiley
México, 1971

Choroever Stephan L.
Cautions on Computers in Education. Effects on the
student-teacher relationship.
Byte, junio 1984
Mc. Graw Hill