LABORATORIO DE BIOESTADISTICA
II SEMESTRE

NOMBRE: ADRIANA GÓMEZ
CODIGO: 971093

INTRODUCCION

Estas guias de laboratorio han sido elaboradas para los alumnos de pregrado del COLEGIO ODONTOLOGICO COLOMBIANO en área de Salud Pública. Dichas guias tienen como objetivo brindarle la posibilidad de aplicar los conocimientos adquiridos en bioestadística.

El docente Lic. en Física y Matemáticas Carlos Ovalle Pérez diseño este trabajo teniendo como referencia el libro "DENTAL PUBLIC HEALTH BIOSTATISTICS" de los doctores Jane A. Weintraub, Chester W. Douglas, Dennis Gilligs, esperando que este material sea una valiosa oportunidad de aprendizaje y obtenga de él, el máximo beneficio.

B I O E S T A D I S T I C A
EN SALUD BUCODENTAL

CAPITULO 1.

DATOS Y TABLAS

INTRODUCCION

Los profesionales que trabajan en el área de la salud bucodental, como clínicos, científicos, docentes, administradores y directivos, deben comprender y utilizar a diario los principios estadísticos. Lo menos que se requiere es capacidad para leer con sentido crítico la bibliografía científica y profesional. El profesor clínico debe, sin embargo, dominar los fundamentos de la lógica estadística y estar en condiciones de aplicar los principios básicos del diseño experimental, la inferencia estadística y la probabilidad.

Este es el primero de una serie de capítulos destinados a enseñarle diversos conceptos básicos y técnicas en bioestadísticas. Aunque cada capítulo de la serie es en sí mismo completo, algunos de los que le siguen pueden requerir el conocimiento de otros previos.

El primero de los tópicos es la presentación de datos descriptivos, que tiene importancia porque los principios estadísticos resumidos en este y en los siguientes dos capítulos cubren muchas de las ideas y métodos utilizados para comprender y comunicar los diferentes tipos de datos.

OBJETIVO

Usted podrá construir una tabla que describa un conjunto determinado de datos.

TERMINOLOGIA BASICA

Los datos existen en muchas formas. Pueden presentarse en forma de símbolos, como por ejemplo las calificaciones de un curso que se registren con una letra del alfabeto, o también en forma numérica. Generalmente, para dar significado a los "datos originales", es necesario ordenarlos en categorías mutuamente excluyentes, de acuerdo con algún tipo de escala. Hay cuatro tipos de escalas diferentes que se pueden utilizar: nominal, ordinal, intervalar y de razón.

En una **escala nominal**, las categorías discretas no guardan una relación cuantitativa entre sí. Son ejemplos de escala nominal el registro de un elector como demócrata, republicano o independiente, o las respuestas si/no a una pregunta.

En una **escala ordinal**, las categorías se pueden ordenar en escala creciente o decreciente. Sin embargo, no se especifica la magnitud de la diferencia entre las categorías. El uso de las letras A, B o C para calificar el nivel obtenido en un curso, es un ejemplo de escala ordinal. A menos que se asignen cantidades numéricas determinadas a cada letra, la diferencia entre una A y una B no es necesariamente la misma que la diferencia entre una B y una C.

En una **escala intervalar** las categorías se ordenan en unidades igualmente espaciadas y es posible medir las diferencias relativas entre cada punto de la escala. No existe un punto cero absoluto. El termómetro centígrado es un ejemplo de una escala intervalar. La diferencia entre 30° y 31°C es la misma que la existente entre 82°C y 83°C. Sin embargo, 100°C no es el doble de caliente de 50°C.

La edad es un ejemplo de **escala de razón**. Existe un punto cero absoluto (el nacimiento) y la magnitud de las diferencias entre edades es la misma. De esta manera, David, que tiene 18 años de edad, duplica la edad de su hermano Miguel, que tiene 9. Un examen en el que el puntaje más bajo posible es cero y el más alto 100, es otro ejemplo de escala de razón.

CONJUNTOS DE DATOS

Habida cuenta de que los datos obtenidos en una investigación son por lo general varios números organizados de alguna manera, nos referiremos a cada uno de ellos como item dentro de los datos y a la totalidad de items como **conjunto de datos.** Por ejemplo, si las calificaciones que un estudiante obtuvo en un curso de anatomía de cabeza y cuello se registraran en una lista, los resultados podrían ser: 85, 92, 94, 89, 95. Cada uno de estos números es un item dentro de los datos y los cinco items constituyen el conjunto de datos.

Problema 1:

Si estudiantes de odontología revisan pacientes en una clínica y llegan a los siguientes diagnósticos: no hay necesidad de tratamiento; es necesario algún tratamiento; se necesita una significativa cantidad e tratamiento; se requiere tratamiento con urgencia.

Qué tipo de escala están utilizando?

Su solución:

Problema 2:

Suponga que usted contó el número de dientes ausentes en seis pacientes odontológicos geriátricos y registró los números siguientes: 8, 12, 16, 10, 14, y 21 dientes ausentes/persona.

Utilizando la terminología que hemos adoptado, cómo se referiría usted a este grupo de seis números y qué tipo de escala representan estos recuentos de dientes ausentes?

Su solución:

Ahora podemos empezar a pensar en la manera de describir datos. Quizás el método más comúnmente utilizado en la mayoría de los trabajos científicos sea la presentación de datos en una tabla como la siguiente.

TABLA 1.1

ESTUDIANTES DE PRIMER AÑO EN FACULTADES DE ODONTOLOGIA
Años Lectivos 1970-71 a 1977-78

Año lectivo	SEXO: Estudiantes varones	SEXO: Estudiantes mujeres	Total de estudiantes de primer año
1970-71	4.4741	94	4.565
1971-72	4.598	147	4.745
1972-73	5.113	224	5.337
1973-74	5.054	391	5.445
1974-75	4.986	631	5.617
1975-76	5.056	707	5.763
1976-77	5.133	802	5.935
1977-78	5.074	880	5.954
TOTAL	39.485	3.876	43.361

Fuente: American Dental Association. Council on Dental Education, Annual Report on Dental Education, 1977-78.

La tabla 1.1, como todas las tablas, comprende varias partes:

1. Título

Es mejor que el título sea breve e indicativo de los datos que se encuentran en cada celda de la tabla. En la Tabla 1.1 estos datos son el número de "Estudiantes de primer año en facultades de Odontología".

2. Encabezamientos

a) El encabezamiento global para la variable representada por las columnas es simplemente "Sexo" en este ejemplo.

b) Los encabezamientos de las columnas casi siempre constituyen categorías mutuamente excluyentes y exhaustivas dentro del encabezamiento global. Obviamente, en este ejemplo existen solo dos categorías para el sexo: varones y mujeres.

c) El encabezamiento de la categoría filas se encuentra a la izquierda del encabezamiento global de las columnas e indica la variable que compone las filas. En la Tabla 1.1 "Año Lectivo" es el encabezamiento al que nos referimos.

d) Los rótulos de las filas se utilizan luego para señalar los componentes del encabezamiento de la categoría.

e) Los totales se usan a menudo tanto para las variables de las filas como para las de las columnas. Los valores totales se denominan a veces "marginales" pues, como usted puede ver, se encuentran en el extremo derecho y en la última fila de la tabla.

f) La fuente de los datos o de la tabla se indica en una nota al pie inmediatamente debajo de la tabla si los datos no son originales de la publicación en que se encuentran.

Suponga que usted es un profesional que trabaja individualmente en una clínica y se le plantea un problema derivado de la necesidad de incorporar más personal. Usted siente que es el momento adecuado para buscar un segundo odontólogo o algún otro auxiliar, pero no está seguro del tipo de odontólogo, higienista, técnico de laboratorio o asistente más adecuado para su clínica. Parece que usted atiende a un montón de niños pero que en definitiva no puede decir a cuántos. Tampoco está usted seguro de la cantidad de prestaciones preventivas y de diagnóstico que brinda ni de los servicios de especialistas a los que podría recurrir. Usted necesitar saber qué tipo de prestaciones recibe en la clínica cada grupo de pacientes para estar en condiciones de determinar cuáles son las áreas en que necesita más ayuda.

Como primer paso decide determinar el tipo de prestaciones que recibe cada grupo etario de pacientes. Usted llega a la conclusión de que lo más acertado es agrupar las edades en <3, 3-6, 7-12. 13-17, 18-24. 25-44, 45-64 y 65 o más años. Contrata a un ayudante para que tabule a todos los pacientes activos que han concurrido a la clínica durante los dos años previos según su edad y tipo de prestación brindada.

Problema 3:

Cómo expondría la información reunida por el ayudante en una tabla que indique la edad del paciente y el tipo de prestación recibida por la población de su clínica?

Su tabla debería ser similar a la tabla 1.4 y llevar como título algo así como "Número de prestaciones por edad del paciente y tipo de prestación". El encabezamiento global de las columnas es "Edad del paciente" y "los encabezamientos de cada columna son los grupos etarios específicos que usted deberá considerar para seleccionar el tipo de personal. El encabezamiento de la categoría filas es "Tipo de prestación" y abarca los principales tipos de prestaciones que constituyen los rótulos de las filas. Los totales son convenientes porque eliminan la necesidad de que el lector realice cálculos aritméticos cuando desea, por ejemplo, conocer de un vistazo el número total de pacientes entre 7 y 12 años que concurren a la clínica o la cantidad de prestaciones brindadas en un rubro específico. De esta manera, el título, los encabezamientos de filas y columnas y los rótulos de cada fila ayudan a clarificar el contenido de la tabla.

Problema 4:

Una vez confeccionada una tabla, la podemos utilizar para examinar el conjunto de datos que incluye. Por ejemplo, en la tabla 1.2 quizás usted quiera conocer el número de prestaciones que la clínica brindó a pacientes menores de 13 y mayores de 65 años. Cuál sería su respuesta?

Tabla 1.2

NUMERO DE PRESTACIONES POR EDAD DEL PACIENTE
Y TIPO DE PRESTACION BRINDADA

Tipo de prestación	Edad del paciente								
	< 3	3-6	7-12	13-17	18-24	25-44	45-65	>65	Total
Preventiva	5	10	24	24	7	42	73	15	200
Diagnóstico	4	13	27	26	9	53	95	53	280
Operatoria	3	25	80	73	21	156	280	163	801
Prot.removible						4	140	108	252
Cirugia			10	5	9	12	30	24	90
Prot.fija					4	118	134	74	330
Endodoncia					2	19	37	10	68
Periodoncia						6	21	3	30
Ortodoncia			7	8					15
Total	12	48	148	136	52	410	810	450	2.066

Su solución:

La lectura de una tabla es bastante sencilla, aunoue su confección resulte a menudo tediosa. Si usted dispone de un conjunto de datos, quizás necesite la ayuda de una computadora. Sin embargo, si los conjuntos de datos son relativamente pequeños, usted puede preparar las tablas a mano. Tal vez el mismo ayudante o gerente comercial que confeccionó la tabla 1.2 haya diseñado gráficos para cada tipo de prestación, controlando las categorías de edad pertinentes mientras revisaba la lista total de historias dentales archivadas en la clínica odontológica (se muestra a continuación un ejemplo del control realizado con respecto a prestaciones quirúrgicas).

Tabla 1.3

PRESTACIONES QUIRURGICAS

		IIII IIII	III	IIII IIII	II IIII IIII	IIII IIII IIII IIII IIII IIII	IIII IIII IIII IIII IIII
< 3	3-6		13-17	18-24	25-44	45-65	> 65

Hay que clasificar sistemáticamente a los pacientes activos haciendo en la columna apropiada una marca por cada uno de ellos. Se anota en bloques de cinco marcas "IIII". Los encabezamientos de cada categoría se escriben al pie de la página y los bloques se disponen de abajo hacia arriba.

En este momento usted ya debe estar preparado para realizar el test final. Por favor, asegúrese de que lo ha contestado correctamente antes de proseguir con el siguiente capítulo.

TEST FINAL 1

En la lista que sigue se encuentran los nombres y fechas de nacimiento de los 55 pacientes más activos de una clínica odontológica. Confeccione una tabla con la edad y el sexo de estos pacientes e indique luego lo que usted aprendió de esta clínica tras el análisis de la tabla. Encontrará una página en blanco para que pueda realizar la tarea asignada. Suponga que la está haciendo el 1º de enero de 1992. Quizás le resulte útil leer primero las dos "ayudas" que le damos a continuación antes de comenzar el test final:

1) Al calcular la edad de cada paciente puede resultar útil colocarla entre paréntesis al lado de la fecha de nacimiento.

2) Cuente el número de varones y mujeres de la lista de pacientes. Luego podrá llevar a cabo un rápido control de la exactitud del trabajo comparando los totales con los correspondientes bloques por edad y sexo.

SUS RESPUESTAS AL TEST FINAL 1

Nombre:

Nombre del paciente	Fecha de nacimiento	Sexo	Nombre del paciente	Fecha de nacimiento	Sexo
Howard Therrel	18/ 4/30	M	V. Howard	22/ 2/24	F
Joel Patterson	29/ 8/83	M	R. Scarborough	1/ 9/80	M
Leslies Maness	12/ 2/47	M	Ethel Shumaker	15/ 1/48	F
W. Fleckenstein	4/ 7/74	M	Pamela Mc Manus	14/ 9/34	F
C. Lee Flanders	15/ 3/53	F	L. Lovingood	3/12/50	M
Susan Snuggs	24/11/30	F	Len Ludwig	12/ 1/62	M
Eugene Ghent	2/ 1/24	M	P. Lussardi	1/10/49	F
Lucile Busby	19/12/32	F	D. Fitz-Henry	20/ 6/84	M
Gladys Nesbit	10/ 4/52	F	L. Bjorklund	24/ 8/35	F
Cecil Putnam	27/ 7/34	M	Wendy Clark	14/ 7/75	F
Jay Swalchick	30/ 5/42	M	Janie Grier	1/ 1/50	F
Anita White	13/ 6/51	F	Ruth Medlin	18/12/64	F
Sally Cornelius	2/ 3/28	F	Joyce Parker	20/ 2/77	F
L. Romano	3/12/61	F	Robin Levine	2/ 5/22	F
Dick Brooks	2/12/76	M	Susan Fisher	19/ 3/74	F
Judy Marcus	9/ 7/77	F	Karen Walter	3/ 6/21	F
Tom Burbank	19/ 1/77	M	M. Monahan	4/ 7/23	F
Debbie Johnson	25/ 8/82	F	B. Bradford	8/11/75	F
Steve Goldman	7/ 5/84	M	Jan Doodley	25/ 7/70	F
Mary Hunter	25/11/60	F	E. O'Brien	22/ 6/84	F
B. Harrison	8/ 8/20	F	Lee Grant	9/ 4/75	M
Julia Jacobs	5/ 9/40	F	Alice Emery	17/ 4/75	F
M. Rappaport	16/ 6/61	F	Mark Fleming	5/10/77	M
Ellen Anderson	19/ 9/19	F	Katie Ford	12/ 5/76	F
Nancy Smith	18/ 9/79	F	E. Friedman	6/ 8/54	F
Bill Greene	11/ 3/82	M	Amy Spring	27/10/17	F
Jill Gallagher	15/ 5/84	F	Linda Philips	10 /6/76	F
Ann Good	4/ 9/55	F			

CAPITULO 2.

TASAS

INTRODUCCION

En el primer capítulo nos hemos ocupado de la confección de tablas para describir datos. Ahora avanzaremos un poco más en materia de datos descriptivos e introduciremos el concepto de tasas, que definen la proporción de individuos con ciertos atributos en el total de la población en estudio.

Por ejemplo, las compañías de seguros están interesadas en la tasa de utilización para cada tipo de prestación odontológica. Suponga usted que sabe que 1.000 personas de una comunidad visitaron al odontólogo durante determinado año. El número 1.000 le indica a usted algo sobre esos individuos, pero no lo dice si la comunidad está utilizando o no los servicios odontológicos. Si la comunidad tiene un tamaño 2.000, el uso de los servicios odontológicos es igual al promedio nacional: 50 por ciento por año. Sin embargo si la comunidad está conformada por 5.000 individuos, usted quizás se interese en investigar por qué tan poca gente recibe atención odontológica.

Una tasa es en realidad una proporción y define la parte que posee un atributo particular en comparación con el todo. La tasa de utilización odontológica puede definirse como la proporción de la población que ha hecho una visita al odontólogo durante el año anterior. Para la comunidad de 2.000 personas en la que 1.000 visitaron al odontólogo, esa tasa sería 1.000/2.000 = 1/2 o 50%.

Problema 1:

Cuál es la tasa que corresponde a la comunidad de 5.000?

Su solución:

OBJETIVO

Definir:

1) ~~Después de estudiar este capítulo, usted podrá~~ definir cuatro tasas: incidencia, prevalencia, sensibilidad y especificidad.

a -
b -
c -
d -

2) Usted también podrá calcular estas tasas si dispone de datos apropiados.

Problema 2:

Ahora, sin fijarse en las definiciones que hemos dado en la página anterior, qué es una tasa?

Su solución:

Durante el año lectivo 1980, en el Colegio Secundario Walt Whitman, que tiene una matrícula de 2.500 estudiantes, 20 sufrieron lesiones "accidentales" en los dientes. Estas lesiones fueron en todos los casos fracturas dentarias que se produjeron durante partidos de fútbol o hockey.

Problema 3:

Cuál fue la tasa de lesiones dentarias en el colegio durante ese año?

Su solución:

INCIDENCIA Y PREVALENCIA

Hay dos tasas importantes que se emplean mucho en epidemiología. Se las denomina tasas de incidencia y de prevalencia y generalmente están referidas a la aparición de enfermedad en una población.

$$\text{Tasa de incidencia} = \frac{\text{Número de nuevos casos de enfermedad que se producen en un lapso determinado}}{\text{Número total de quienes forman la población durante el mismo lapso}}$$

El lapso considerado es a menudo un año, pero también puede ser un día, una semana o varios meses si se trata de una enfermedad infecciosa de fácil diseminación. En el curso de un año la población puede cambiar como consecuencia del traslado de gente a la comunidad o fuera de ella. Por lo tanto, se toma como denominador al número promedio de gente en la población durante el año. La **incidencia** indica cuantos nuevos casos de una enfermedad se está produciendo en relación con la población total, pero nada dice con respecto al número total de casos en un momento determinado. Así llegamos a la definición de **prevalencia**.

$$\text{Tasa de prevalencia: } \frac{\text{Número de casos de enfermedad en un momento determinado}}{\text{Número total de personas integrantes de la población en ese momento}}$$

Lea ahora los enunciados A y B y resuelva el Problema 4.

Enunciado A: durante el mes de octubre de 1981 un equipo de odontólogos examinó a todos los adultos de una reserva indígena del Estado de Nueva York para determinar la proporción de adultos con defectos adamantinos.

Enunciado B: una odontóloga de un servicio hospitalario de Boston contó el número de pacientes con lesiones recién desarrolladas de tipo queratinoso en la mucosa bucal, que concurrieron a su dispensario en noviembre de 1981. La odontóloga también contó el número total de pacientes que atendió allí durante el mismo mes.

Problema 4:

¿Qué tasa procuraron establecer los odontólogos en el Enunciado A?

¿Qué tasa tabuló la odontóloga en el Enunciado B?

Su solución:

SENSIBILIDAD Y ESPECIFICIDAD

Otras dos tasas resultan de mucha utilidad: la de sensibilidad y la de especificidad. Estas tasas de pueden emplear para evaluar una prueba diagnóstica.

Suponga que se utiliza una prueba para clasificar a los pacientes como positivos (presencia de enfermedad) y negativos (ausencia de enfermedad). Es poco probable que una prueba diagnóstica funcione correctamente en todas las ocasiones, por lo que habrá algunos pacientes clasificados como positivos que no tengan la enfermedad (positivos falsos) y otros pacientes incluidos entre los negativos que si la tengan (negativos falsos).

Ahora bien: suponga usted que se analiza la población y que es posible determinar por otros medios (el "patrón de oro") si los individuos tienen realmente la enfermedad. En este caso la población se podrá clasificar en la siguiente tabla:

Tabla 2.1

TABLA PARA EVALUAR UNA PRUEBA DIAGNOSTICA

RESULTADO DE LA PRUEBA	**ESTADO PATOLOGICO**	
	Con enfermedad	**Sin enfermedad**
Positivo	Positivos verdaderos	Positivos falsos
Negativo	Negativos falsos	Negativos verdaderos

La **sensibilidad** de la prueba se define así:

Positivos verdaderos

$$\text{Sensibilidad} = \frac{\qquad}{\text{Positivos verdaderos + negativos falsos}} \times 100$$

(pacientes sin la enfermedad)

o sea, **el porcentaje de personas con la enfermedad** correctamente clasificadas como enfermas.

La **especificidad** se define así:

$$\text{Especificidad} = \frac{\text{Negativos verdaderos}}{\text{Negativos verdaderos + positivos falsos}} \times 100$$

(pacientes sin la enfermedad)

o sea, **el porcentaje de personas sin la enfermedad** correctamente clasificadas como no enfermas.

Figura 2.1

Una forma de distinguir las tasas de sensibilidad de las de especificidad es asociar la sensibilidad con la selección de quienes tienen la enfermedad y la especificidad con la de quienes no la tienen.

En la Figura 2.1 se puede apreciar la relación indirecta que existe entre estas dos tasas.

Hay dos curvas: la que está por encima del eje horizontal de las x representa la distribución de los valores "x" para los pacientes con la enfermedad, y la que está por debajo de ese eje representa dicha distribución para los pacientes sin la enfermedad. Estos valores "x" son el resultado de una prueba diagnóstica, como la presión sanguínea diastólica o el recuento de glóbulos blancos, para la cual es dable encontrar un rango de valores. Como usted puede ver en el diagrama, hay una cierta superposición entre las dos curvas. Para varios valores "x" de esta prueba diagnóstica, hay algunos que tienen la enfermedad y otros que no la tienen.

Si se selecciona un determinado punto de corte, como el que indica la línea vertical designada como "x_1", se considerará que todos loa pacientes con cierto valor en la prueba que estén a la derecha tienen la enfermedad y que los ubicados a la izquierda están libres de ella. Sin embargo, quienes están representados en la parte sombreada de la curva de "pacientes con enfermedad" que están a la izquierda de la línea x_1 serán negativos falsos. En realidad tienen la enfermedad, aunque no lo indique (no pronostique) el resultado de la prueba. Los representados en la parte sombreada de la curva "pacientes sin enfermedad" que están a la derecha del punto de corte x_1 serán positivos falsos.

De conformidad con los criterios establecidos por los resultados de la prueba, se los ha diagnosticado como enfermos, cuando en realidad no lo son. Si se modifican el punto de corte y la ubicación de la línea vertical, cambian la sensibilidad y la especificidad de la prueba.

Figura 2.2

En la figura 2.2 se ha elegido un valor menor de "x" como factor determinante para incluir -o no- a alguien entre los que tienen la enfermedad (por ejemplo, alta presión sanguínea). La línea rotulada x_2 determina este nuevo punto de corte.

Como puede usted observar, el área sombreada debajo de la curva "pacientes con enfermedad" es más pequeña: este cambio entraña menos negativos falsos y más positivos verdaderos y, por tanto, una mayor sensibilidad de la prueba. Empero, aumenta al mismo tiempo el área sombreada que se asocia con la curva de "pacientes sin enfermedad", lo que determina que, por haber más positivos falsos y menos negativos verdaderos, se produzca una disminución de la especificidad de esta prueba diagnóstica.

El trueque entre sensibilidad y especificidad depende a menudo de la comparación de los riesgos con los beneficios derivados de no **tratar** a alguien que tiene la enfermedad u de **tratar** a algún otro que no la tiene. Se han desarrollado técnicas que responden a la denominación genérica de análisis de decisiones, que se puede aplicar cuando hay que resolver si corresponde tratar un caso dudoso.

Las dos tasas, sensibilidad y especificidad, se expresan como porcentajes. En efecto, todo porcentaje en una tasa, porque representa la proporción en que se encuentra determinado atributo, luego de multiplicar por 100. A menudo, las tasas se expresan también como partes de 1.000 ó 100.000. Las tasas de mortalidad se calculan en función del número de muertes que ocurren por año en una población de 1.000 individuos. En cambio, es convencional que las tasas de morbilidad se expresen casi siempre en relación con 100.000 personas.

REVISION

1) $\text{Tasa} = \dfrac{\text{Número de individuos con una característica}}{\text{Número total de personas de la población}}$

2) $\text{Tasa de incidencia} = \dfrac{\text{Número de nuevos casos de la enfermedad que se producen en determinado lapso}}{\text{Número total de personas de la población en el mismo lapso}}$

3) $\text{Tasa de prevalencia} = \dfrac{\text{Número de casos de la enfermedad en un momento determinado}}{\text{Número total de personas de la población en ese momento}}$

4) Tabla para evaluar una prueba diagnóstica:

RESULTADO DE LA PRUEBA	ESTADO PATOLOGICO	
	Con enfermedad	**Sin enfermedad**
Positivo	Positivos verdaderos	Positivos falsos
Negativo	Negativos falsos	Negativos verdaderos

5) Tasa de sensibilidad (quienes tienen la enfermedad)

$$\frac{\text{Positivos verdaderos}}{\text{Positivos verdaderos + negativos falsos}} \times 100$$

6) Tasa de especificidad (quienes no tienen la enfermedad)

$$\frac{\text{Negativos verdaderos}}{\text{Negativos verdaderos + positivos falsos}} \times 100$$

TEST FINAL 2

En septiembre de 1983, 200 estudiantes de primer año de medicina y odontología se instalaron en un amplio edificio residencial ubicado frente al complejo médico de la Universidad. Como parte del curso de introducción a las ciencias de la salud, se pidió a todos los estudiantes que se sometieran a un examen bucodental en la facultad de Odontología. Las historias odontológicas obtenidas pusieron de manifiesto que había dos casos de estudiantes con gingivitis ulceronecrotizante aguda (GUNA). Transcurrió el otoño y los estudiantes estuvieron sumamente ocupados con todos sus cursos de ciencias fundamentales. Durante la semana iniciada el 18 de diciembre, inmediatamente antes de la primera serie de exámenes finales, concurrieron a la guardia del servicio universitario 25 estudiantes que se quejaban de encías inflamadas y dolorosas. El médico de turno los revisó, diagnosticó GUNA en 10 casos y los derivó a todos a los consultorios de la Facultad de Odontología. Los periodoncistas de esta facultad, utilizando pruebas diagnósticas y criterios más específicos, llegaron a la conclusión de que solo siete de los 10 estudiantes con diagnóstico de GUNA tenían en realidad esta enfermedad, al igual que dos de los otros 15 a los que se había clasificado como negativos. Los restantes estudiantes padecían otras formas de enfermedad periodontal con diferente fisiopatología. Todos recibieron el tratamiento que individualmente se les indicó.

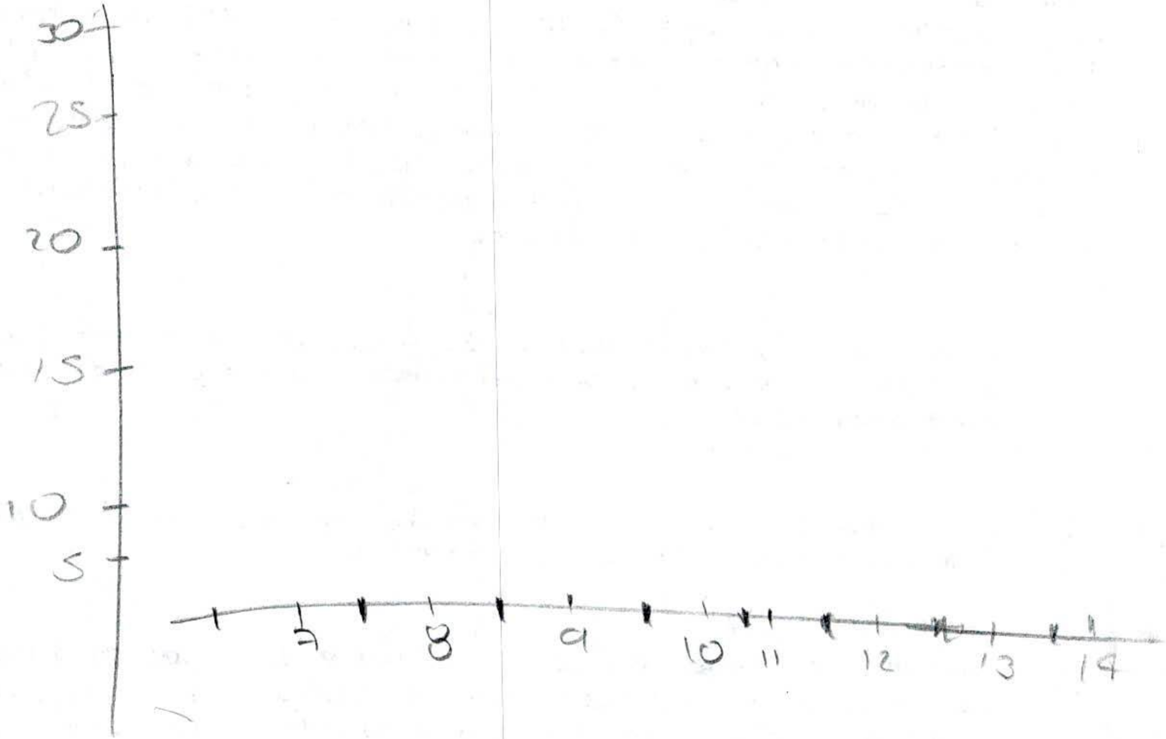
1) Cuál era la prevalencia de GUNA cuando se desarrolló el curso de introducción para estudiantes de medicina y odontología?

2) Cuál era la tasa de incidencia de GUNA durante la semana iniciada el 18 de diciembre?

3) Indique las tasas de sensibilidad y de especificidad del examen diagnóstico que se llevó a cabo en la guardia médica de la Universidad y utilice el diagnóstico definitivo de los periodoncistas como patrón de oro para separar a los pacientes con y sin esta forma aguda de enfermedad periodontal.

Histograma

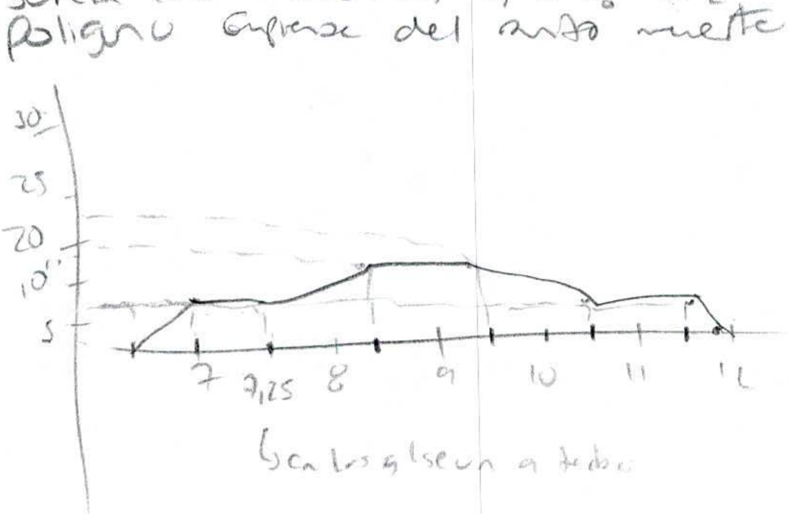
Vamos a tomar el porcentaje.



El histograma.

Poligono

Se trabaja con los puntos medios de los intervalos
se suman los intervalos y se ÷ en 2
El poligono empieza del punto muerto



Sea los q [illegible] a todos

HOJA DE RESPUESTAS

LABORATORIO N. 1

PROBLEMA No 1

Es escala ordinal porque se ordena en forma creciente y se presentan parametros que no los diferencia totalmente pero sin embargo se puede definir su evolución

PROBLEMA No 2

Nos referimos a cada número como un item dentro de los datos y la totalidad de items como conjunto de datos.
y se representa la escala intervalar

PROBLEMA No 3

PROBLEMA No 4

PROBLEMA No 5

PROBLEMA No 6

Tpo prest / Edad	< 3	3-6	7-12	>65	Total
Preventiva	5	10	24	15	54
Diagnostico	4	13	27	56	97
Operatorio	3	25	80	163	271
Prot. Removible	0	0	0	108	108
Cirujia	0	0	10	24	34
Prot. fija	0	0	0	74	74
Endodoncia	0	0	10	0	10
Periodoncia	0	0	0	3	3
Ortodoncia	0	0	7	0	7.
Total	12	48	148	430	658

PROBLEMA No 7

PROBLEMA No 8

PROBLEMA No 9

PROBLEMA No 10

DESARROLLO

CARLOS EDGAR OVALLE PEREZ
VICERRECTOR
Gimnasio J. Sta. Catalina

TEST FINAL No____

Edades de Pacientes Hasta
(1-01-1992)

Edad	Hombre	Mujer	Total
8	2	2	4
9	1	0	1
10	1	1	2
12	1	0	1
13	0	1	1
15	2	2	4
13	1	2	3
17	2	2	4
18	1	1	2
22	0	1	1
28	0	1	1
30	1	0	1
31	0	2	2
32	0	1	1
37	0	1	1
38	0	1	1
39	0	1	1
40	0	1	1
41	0	1	1
42	1	1	2
43	0	1	1
44	0	1	1
45	1	0	1
50	1	0	1
52	0	1	1
57	0	1	1
58	1	1	2
60	1	0	1
62	1	1	2
64	0	1	1
68	1	1	2
69	0	1	1

70	0	1	1
71	0	1	1
72	0	1	1
73	0	1	1
75	0	1	1
Total	18	37	55

CAPITULO 3.

TASAS AJUSTADAS

INTRODUCCION

Algunos de los temas odontológicos que actualmente despiertan más controversias son las tendencias que se observan en enfermedades periodontales, caries dentarias y cáncer bucal.

Para introducirse en el campo de la investigación epidemiológica odontológica, tanto por experiencia personal como por la lectura de la bibliografía correspondiente, resulta necesario comparar en diferentes poblaciones tasas de enfermedad con respecto a algún suceso o característica. Las poblaciones por comparar pueden ser comunidades diferentes, grupos con tratamientos distintos o dos grupos: uno que haya estado expuesto a una enfermedad determinada y otro que no lo haya estado.

Si las dos poblaciones que se comparan están constituidas de modo similar con respecto a factores (como edad, sexo o ingresos) que se asocian con la característica en estudio, no hay ningún problema en comparar sencillamente las tasas **brutas** tal como aparecen; sin embargo, si las muestras o poblaciones no están constituidas de modo similar, la comparación directa de las tasas brutas pueden inducir a error.

En este capítulo se pretende enseñarle: **1) las condiciones** necesarias para reconocer las situaciones que pueden ser engañosas y 2) cómo manejar estadísticamente esas situaciones mediante un procedimiento que elimine los efectos de los factores adicionales (como la edad) en la comparación que interese realizar. Este procedimiento se llama **ajuste de tasas**.

Usted se familiarizará aquí con los métodos directos e indirectos de ajuste de tasas. Empezaremos señalando cuándo tendrá que ajustar las tasas.

OBJETIVO

1) Usted podrá enunciar las cuatro condiciones necesarias para ajustar las tasas y utilizar dichas condiciones para evaluar si el ajuste es apropiado.

2) Usted podrá emplear tanto el método directo como el indirecto para ajustar las tasas.

AJUSTE DE TASAS

Examine brevemente la Figura 3.1. de la siguiente página, en la que se compara la experiencia de mortalidad dentaria de dos ciudades durante 1984.

Trate de formarse una idea general sobre lo que expresa esta ilustración y después continue leyendo.

Figura 3.1

COMPARACION DEL NUMERO DE DIENTES AUSENTES EN DOS CIUDADES

Si usted fuera un epidemiólogo interesado en la relación existente entre la concentración de fluoruro en el agua de consumo y la tasa de mortalidad dentaria (número de dientes ausentes por persona), quizás se dedicaría a estudiar estas dos ciudades, pues podría comparar la mortalidad dentaria de la población que vive en un área con un nivel óptimo de fluoración con otra que tiene carencias en lo que a fluoruro se refiere.

Las tasas de mortalidad dentaria para estas dos ciudades son:

Ciudad A: 953.000 dientes ausentes/138.000 personas = 6.9 dientes ausentes/persona.

Ciudad B: 462.000 dientes ausentes/79.500 personas = 5.8 dientes ausentes/persona.

Tal vez usted se sorprenda y quede desconcentrado al comprobar que la Ciudad A, con óptima fluoración, tiene una tasa más elevada de dientes ausentes por persona.

Problema 1:

En qué ciudad hubiera usted supuesto que encontraría la tasa más elevada, sabiendo que los niveles óptimos de fluoruro ayudan a prevenir las caries que finalmente pueden provocar mortalidad dentaria?

Su solución:

La pregunta lógica que sigue es: "quiénes viven en estad dos ciudades?" Un somero conocimiento de sus poblaciones pueden inducirlo a modificar su interpretación. Fíjese en la Figura 3.2 en la estructura etaria de ambas ciudades.

Figura 3.2

COMPARACION DE NEEDLEPOINT KNOLL CON TERTBOOKTOWN EN FUNCION DE LA EDAD

Ocurre que la Ciudad A, ubicada en un clima templado, es un centro en el que se agrupa un alto porcentaje de jubilados. Personas de edad madura tienden a establecerse allí. En cambio, la Ciudad B es de urbanización reciente y ha crecido alrededor de una importante Universidad. Ha atraído a numerosas personas jóvenes con muchos niños.

Qué factor asociado a las tasas de mortalidad dentaria lo induciría a ajustar su interpretación inicial de la tasa bruta?

Si usted contestó la **edad** está en lo correcto, porque es probable que la diferencia de las tasas brutas se pueda explicar, al menor parcialmente, por el simple hecho de que la Ciudad A alberga a una población más vieja que la Ciudad B. Por consiguiente, lo lógico es que esperemos encontrar relativamente más dientes ausentes en la Ciudad A por la sencilla razón de que viven en ella más personas con alto riesgo de perder sus dientes, en especial por enfermedad periodontal, que se manifiesta más a menudo a medida que aumenta la edad.

La presencia en esta situación de una variable como la **edad** es una de las condiciones necesarias para el cómputo de tasas ajustadas. Decimos que una variable de este tipo es

un factor de confusión pues confunde o enturbia la comparación que queríamos efectuar. Inicialmente pretendíamos comparar las tasas de mortalidad dentaria en dos ciudades con distinto nivel de fluoruro, pero a menos que tengamos también en cuenta los efectos de la edad, los resultados que obtengamos pueden ser engañosos; por lo tanto, la edad es un factor de confusión que hace que sea difícil atribuir la diferencia observada en la tasa bruta de mortalidad dentaria tan solo a una diferencia en la presencia de fluoruro.

Problema 2:

La diferencia que hemos advertido en las tasas brutas se puede explicar, al menos en parte, por la diferencia de ________ de las poblaciones y no enteramente por las diferencias en ________ entre la Ciudad A y la Ciudad B.

Su solución:

Problema 3:

La variable **edad** del ejemplo anterior dificulta la comparación que interesa realizar (mortalidad dentaria en dos ciudades con diferentes niveles de fluoruro). Por consiguiente, es un factor de ________ en la comparación de las tasas brutas de mortalidad dentaria de las Ciudades A y B.

Su solución:

Problema 4:

Si las dos tasas brutas hubieran sido exactamente las mismas, en cierto modo habrían constituido una prueba de que el factor edad no produce ningún efecto en la mortalidad dentaria general.

Su solución:

Verdadero ________ Falso _______

Problema 5:

La existencia de un ______________________________ como la edad es condición fundamental para que sea necesario un ajuste de la tasa.

Su solución:

Consideremos ahora todos los componentes de la situación planteada. Usted está ante un problema que requiere un **ajuste de la tasa** si se dan las cuatro condiciones siguientes:

1) Usted está interesado en la **comparación** de dos o más poblaciones o muestras.

2) El suceso o característica que tiene interés (en este caso la pérdida de dientes) se define, a los efectos del análisis, como una **tasa** (por ejemplo, tasa de mortalidad dentaria) o **proporción**.

3) Su comparación supone tasas brutas o **globales**.

4) Existe un **factor de confusión** que a su juicio puede afectar la comparación.

Figura 3.3

REQUISITOS DEL AJUSTE DE TASAS

Problema 6:

Examine la Figura 3.3 para verificar si en este caso están satisfechas en su totalidad las condiciones siguientes:

a) Quiere usted hacer una comparación? Si es así, qué va a comparar?

b) Se define el suceso que interesa como una tasa o proporción? En caso afirmativo, qué es lo que interesa?

c) Queremos comparar tasas globales?

d) Existe un factor de confusión? En caso afirmativo, cuál es?

Su solución:

Problema 7:

Si en nuestro ejemplo usted estuviera **solamente** interesado en comparar las tasas de mortalidad dentaria de 35 a 40 años, tendría que efectuar un ajuste de la tasa? De dos razones para justificar su respuesta.

Su solución:

Global vs. específica:

Con respecto a la pregunta anterior, usted tendría que saber que algunos epidemiológicos han sostenido que, en realidad, jamás se tiene primordial interés en una tasa global. Es un ataque bastante agresivo que merece cierto análisis! Hagamos una breve digresión para considerar este punto de vista.

Una digresión:

Un epidemiólogo apellidado Woodlsey expresó en 1959 una opinión que comparten varios otros investigadores: "las tasas específicas son esenciales porque únicamente analizando tasas específicas se puede llevar a cabo un estudio seguro y pormenorizado de una variación entre clases de población". En el caso anterior, tendríamos que analizar la tasa de mortalidad dentaria en cada grupo etario y compara la tasa obtenida en cada una de las dos ciudades.

Problema 8:

Algunos epidemiológicos sostienen que nunca debería ajustarse las tasas, pues las únicas significativas son las tasas ....

Su solución:

Problema 9:

Por supuesto, aceptaremos que las tasas específicas son muy importantes para que un análisis sea exacto y pormenorizado. Sin embargo, no estamos de acuerdo en que el uso de tasas _______ no sea nunca apropiado.

Su solución:

Una tasa global puede ser muy útil como **resumen práctico** de la información relativa a una esquema completo de tasas específicas. Si el análisis se oriente hacia un factor que interesa, es más cómodo comparar dos poblaciones utilizando una **única tasa global** en vez de una **serie de muchas tasas específicas**. Sin embargo, las interpretaciones se pueden tornar difíciles cuando la cantidad de tasas específicas es grande.

Usted también debe estar en condiciones de advertir en qué situaciones una tasa global será de cuestionable valor. Hay dos situaciones que usted puede reconocer con claridad:

1. **Si la comparación que interesa se restringe a un subproducto cuya definición es todavía amplia**, puede quedar una variable de confusión (por ejemplo, la edad) y usted tendrá que ajustar sus tasas. Además, el centro de interés será una tasa para el subgrupo y no una tasa global para todo el grupo.

2. **Si hay una incongruencia notoria de las tasas específicas con respecto a una variable de confusión**, (por ejemplo, si las tasas específicas son notoriamente más altas para poblaciones de determinadas edades pero notoriamente inferiores en otras edades), ninguna tasa global para cada población arrojará luz sobre las diferencias referentes a edades específicas. El uso de una tasa global entrañaría una tendencia a ocultar esas diferencias en esta situación.

Problema 10:

Suponga que usted quiere comparar las tasas de mortalidad de dos poblaciones:

	Población A	**Población B**
Jóvenes	12,2/1.000	10,3/1.000
Viejos	9,5/1.000	15,9/1.000

Sería apropiada en este caso una tasa global?

Su solución:

Problema 11:

Cuál es la ventaja primordial de utilizar una tasa global?

Su solución:

Problema 12:

Puede dar usted dos razones para no utilizar tasas globales?

Su solución:

METODOS PARA EL AJUSTE DE TASAS

Volviendo a la investigación epidemiológica dental llevada a cabo en Needlepoint Knoll y Textbooltown, hemos determinado que las tasas de mortalidad dentaria global son de interés para nosotros, pero también que los resultados están sesgados como consecuencia de un factor de confusión, la edad.

Quisiéramos saber cuáles serían las tasas globales de mortalidad dentaria en estas dos ciudades si la estructura etaria de la población fuera la misma en ambas. Es algo que podemos determinar matemáticamente ajustando las tasas que hemos observado para conformar una población conocida.

Tabla 3.2

	Needlepoint Knoll		TTextbooktown		Estándar	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Edad	Población	Tasa	Población	Tasa	Población	Tasa
< 35	23.000	1.0	40.000	3.0	30.000	2.0
35-65	45.000	2.0	25.000	5.0	33.000	5.2
> 65	70.000	12.0	14.500	15.0	12.000	15.0
Total	138.000	6.9	79.500	5.8	75.000	5.5

Si vamos a emplear el método "directo" de ajuste, necesitaremos la distribución etaria específica de la población estándar, que aparece en la columna 6 de la Tabla 3.2 y las tasas de mortalidad dentaria observada en grupos etarios específicos de las dos ciudades, que se indican en la columnas 3 y 5. Si queremos utilizar el método "indirecto" de ajuste, necesitaremos las tasas de mortalidad dentaria de cada grupo etario específico de la población estándar, que se encuentra en la columna 7, la distribución etaria especifica observada en la población de las ciudades, que se registra en las columnas 2 y 4, y las tasas globales de mortalidad dentaria: 6,9 y 5,8. (En general, no se dispone de toda la información que se presenta en la Tabla 3.2).

METODO DIRECTO

Para obtener por el método directo la tasa global ajustada de mortalidad dentaria en Needlepoint Knoll, cada tasa específica observada: 1,0, 2,0 y 12,0, se multiplica por el tamaño de la población estándar que corresponde al grupo etario específico. La suma de estos tres productos se divide luego por el total de la población estándar.

[(1,0) (30.000) + (2,0) (33.000) + (12,0) (12.000)] / 75.000 = 3,2 dientes ausentes por persona.

Problema 13:

Cuál es la tasa global de mortalidad dentaria en Texbooktown, Texas, si se utiliza el método directo?

Su solución:

Como se puede ver, cuando controlamos el factor de confusión, la edad, las tasas globales ajustadas indican lo contrario de lo que señalaban las tasas globales brutas. Si la edad no fuera un factor más, Needlepoint Knoll -con su contenido óptimo de fluoruro- tendría una tasa inferior de mortalidad dentaria.

METODO INDIRECTO

En el método indirecto, el concepto es el mismo, pero varía la influencia que se asigna a cada grupo. En vez de hacer el ajuste en función del **tamaño de la población** estándar correspondiente a un grupo etario específico, se utilizan las **tasas** estándares de los grupos etarios específicos para establecer la tasa de dientes ausentes que esperaríamos encontrar en nuestras poblaciones si se aplican las tasas estándares. Se procede a multiplicar lo números que se encuentran en las columnas 2 y 4 por la tasa correspondiente de la columna 7. La tasa **observada** de dientes ausentes se divide por la rasa **esperada** que se calcula. La relación resultante se llama relación estándar de mortalidad dentaria. Si el suceso que interesara no fuese la mortalidad dentaria, sino otro distinto, su nombre cambiaría en consecuencia.

$$\text{Relación estándar de mortalidad dentaria} = \frac{\text{Tasa observada de dientes ausentes}}{\text{Tasa esperada de dientes ausentes}}$$

Para obtener la tasa global ajustada, esta relación estándar de mortalidad se multiplica por la mortalidad global de la población estándar. En Needlepoint Knoll:

Tasa esperada de dientes ausentes =

[(23.000)(2,0) + (45.000)(5,2) + (70.000)(15,0)]/138.000 =

1.330.000/138.000 = 9.64 dientes ausentes por persona

Tasa observada de dientes ausentes = 6,9

Relación estándar de mortalidad dentaria 6,9/9,64 = 0,72

Tasa ajustada (0.72)(5,5) = 3,96

Problema 14:

Cuál es la tasa global ajustada de mortalidad dentaria en Textbooltown, Texas, si se utiliza el método indirecto?

Su solución:

Aunque los números que se obtienen con los métodos directo e indirecto difieren, las conclusiones son las mismas. Los resultados reales obtenidos también variarán según el estándar que se seleccione. Los ajustes que se hacen más frecuentemente se vinculan con la edad, pero se pueden relacionar con cualquier otro factor de confusión (como el sexo, el nivel de ingresos, el nivel de educación, etc.) si se dispone de la información necesaria. La decisión de utilizar el método directo o el indirecto generalmente se toma sobre la base del tipo de datos disponibles con respecto a los estratos específicos.

Si las tasas ajustadas y las brutas son iguales, entonces el factor por el cual se realizó el ajuste no produce ningún efecto ni esta viciando los resultados.

REVISION

1) Tasa global: una estadística resumida para un factor de interés en la población total.

2) Tasa específica: la proporción que tiene un factor que interesa en un estrato seleccionado de la población.

3) Tasa bruta: una tasa sin ajustar por ningún factor de confusión.

4) Tasa ajustada: una tasa ajustada por un factor de confusión.

5) Tasa ajustada directa: una tasa que se ajustó aplicando las tasas específicas de un estrato, obtenidas de los datos, a una población estándar.

6) Tasa ajustada indirecta: una tasa que se ajustó aplicando las tasas específicas de un estrato de una población estándar a cada estrato del conjunto de datos.

7) Tanto las tasas globales como las específicas pueden ser brutas o ajustadas.

TEST FINAL 3

Suponga usted que es director de odontología de un instituto regional y que tiene un presupuesto limitado para programas odontológicos. Desea averiguar qué partes de la región tienen más necesidad de atención odontológica. Se lleva a cabo un reconocimiento en escuelas seleccionadas de la región para establecer cuál es la salud bucal de los niños de diferentes comunidades. En algunas grandes ciudades las escuelas están situadas en vecindarios con diferente nivel socioeconómico (NSE). Por sus conocimientos de epidemiología odontológica, usted sabe que los factores que determinan la clasificación del NSE, como los ingresos y la educación, guardan relación inversa con la tasa DCPO (número de dientes cariados, perdidos y obturados por persona) de una población. Los datos correspondientes a los niños de dos ciudades son los siguientes:

CIUDAD A			**CIUDAD B**		
NSE	**Población**	**Tasa DCPO**	**NSE**	**Población**	**Tasa DCPO**
Alto	5.000	1,5	Alto	20.000	1,7
Medio	30.000	1,8	Medio	40.000	2,1
Bajo	25.000	5,2	Bajo	10.000	7,3
Total	**60.000**	**3,2**	**Total**	**70.000**	**2,7**

Veamos si puede evaluar si es apropiado o no efectuar un ajuste de las tasas pertinentes.

1) Cuáles son las cuatro condiciones para el ajuste de las tasas?

2) Cuál es el factor de confusión en este ejemplo?

3) Es apropiada la utilización de una tasa global?

4) Cuántas de las cuatro condiciones necesarias para el ajuste de las tasas están satisfechas?

5) Suponga usted que la tasa DCPO para el NSE bajo en la Ciudad B fuera 4,1 en vez de 7,3, utilizaría el ajuste de la tasa?

6) Si usted entiende que el ajuste de tasas está justificado para comparar las tasas DCPO entre las ciudades A y B, considere que la población de la Ciudad C es estándar.

CIUDAD C

NSE	Población	Tasa DCPO
Alto	10.000	1,6
Medio	35.000	2,0
Bajo	15.000	6,0
Total	**60.000**	**2,9**

Calcule las tasas ajustadas utilizando el método directo.

Calcule las tasas ajustadas utilizando el método indirecto.

CAPITULO 4.

DISTRIBUCIONES DE FRECUENCIAS

INTRODUCCION

Es habitual que el lector que eche un vistazo o examine "al vuelo" un conjunto de datos primarios como los presentados más abajo en la Tabla 4.1, no pueda tener más que una impresión muy vaga. Esta afirmación es especialmente válida con respecto a un conjunto que incluya gran cantidad de datos. Es fundamental recurrir a algún tipo de simplificación visual para que su comprensión sea más clara; en este capítulo consideraremos diversas técnicas de presentación de datos. Estas técnicas comprenden la tabla de frecuencias, el histograma (o gráfico de barras), el polígono de frecuencias y la distribución de frecuencias acumuladas.

Tabla 4.1

EDAD EN EL MOMENTO DE ERUPCION DEL PRIMER PREMOLAR SUPERIOR (en años)					
10.5	10.7	9.5	10.5	11.8	9.7
12.0	10.3	13.5	12.3	10.6	11.2
10.7	11.5	11.1	10.6	9.3	9.8
10.4	7.5	10.2	8.7	10.9	12.9
11.7	10.3	10.6	10.5	11.9	9.9
13.9	10.6	10.0	10.8	10.6	11.0
7.3	8.0	8.5	12.5		

OBJETIVOS

A partir de un conjunto adecuado de datos primarios, usted podrá construir las siguientes tablas y gráficos:

1) tabla de frecuencias

2) histograma

3) polígono de frecuencias .

4) distribución de frecuencias acumuladas

TABLA DE FRECUENCIAS

Suponga usted que una Asociación Odontológica haya realizado entre los odontólogos del país una encuesta para establecer el número de horas que trabajan por semana. La siguiente podría ser una muestra de los resultados:

5	10	33	22	21
35	38	40	55	41
45	31	45	40	33
32	35	20	24	34
37	25	50	30	40
10	27	22	50	28
36	36	31	35	35
32	18	35	40	38
15	27	31	35	48
40	38	48	37	20

Lo primero que usted querrá hacer antes de analizar los datos es reordenarlos conforme a una secuencia lógica. A continuación se presentan los datos dispuestos en forma creciente. Esta nueva disposición se conoce con el nombre de **ordenación** u **ordenamiento**.

5	24	32	40
10	25	33	41
10	27	33	41
15	27	34	45
18	28	35	45
20	30	35	48
20	30	35	48
21	31	35	50
22	31	35	50
22	32	35	55

Una manera de resumir estas mediciones consiste en utilizar una tabla de frecuencias como la siguiente:

Tabla 4.2

TABLA DE FRECUENCIAS DEL NUMERO DE HORAS DE TRABAJO POR SEMANA CON ARREGLO A LOS INFORMES DE LOS ODONTOLOGOS

Horas de trabajo por semana	Frecuencia
1-10	3
11-20	4
21-30	10
31-40	25
41-50	8
51-60	1
Total	**50**

Se seleccionan intervalos de clase apropiados para las horas de trabajo (1-10, 11-20, 21-30, 31-40, 41-50, y 51-60) y se cuenta el número de mediciones comprendido en cada uno de ellos. En este caso los intervalos de clase son de igual amplitud, lo que a menudo es deseable aunque no resulte necesario.

Una decisión importante que se debe tomar al confeccionar tablas de frecuencias es la relacionada con la amplitud de cada intervalo de clase, que es la que va a determinar el número de intervalos y, por consiguiente, con cuánto detalle se presentarán los datos.

Para la mayor parte de los objetivos es conveniente disponer de unos 5 a 10 intervalos de clase. Con menos de cinco se pierde mucha información y con más de 10 pueden darse tantos detalles que nos hagan perder la posibilidad de adquirir una impresión general. Así mismo, el número total de las mediciones determina en cierto modo el número de intervalos. No es adecuado tener varios intervalos con tan solo una o dos mediciones ni pocos intervalos con muchas mediciones. En numerosos casos los intervalos se seleccionan con arreglo a la precedencia Si otros investigadores han notificado sus comprobaciones con determinados intervalos, quizás convenga utilizar estos mismos para facilitar la comparación de los datos.

Problema 1:

Construya una tabla de frecuencias con los datos de la tabla 4.1, que muestra las edades de 40 niños en el momento de la erupción de sus primeros premolares superiores.

Su solución:

DISTRIBUCION DE FRECUENCIAS RELATIVAS

La distribución de frecuencias relativas se obtiene dividiendo las frecuencias reales por el número total de observaciones y multiplicándolas después por 100 para transformar el valor en un porcentaje. Advierta que la suma de las frecuencias relativas debe ser aproximadamente el 100% (permita un margen para los errores de redondeo). Así se pueden controlar los cálculos aritméticos. LA distribución de las frecuencias relativas correspondientes a la tabla de frecuencias indicada en la Tabla 4.2, "Número de horas de trabajo por semana con arreglo a los informes de los odontólogos", se presenta a continuación en la Tabla 4.3.

Tabla 4.3

DISTRIBUCION DE FRECUENCIAS RELATIVAS DEL NUMERO DE HORAS DE TRABAJO POR SEMANA CON ARREGLO A LOS INFORMES DE LOS ODONTOLOGOS

Horas de trabajo por semana	Frecuencia	Frecuencia Relativa (%)
1-10	3	6
11-20	4	8
21-30	10	20
31-40	24	48
41-50	8	16
51-60	1	2
Total	**50**	**100**

Problema 2:

Determine la distribución de frecuencias relativas correspondientes a la tabla de frecuencias con las edades de los niños en el momento de la erupción de sus primeros premolares superiores, que usted diseñó para resolver el problema 1.

Su solución:

HISTOGRAMA

El histograma o gráfico de barras se puede confeccionar a partir de una distribución de frecuencias relativas, dibujando una barra para cada intervalo de clase que corresponda ala frecuencia relativa de las mediciones hechas dentro de ese intervalo. Se puede hacer el siguiente histograma de las frecuencias relativas del número de horas de trabajo que se indican el a Tabla 4.3.

Figura 4.1

NUMERO DE HORAS DE TRABAJO POR SEMANA CON ARREGLO A LOS INFORMES DE LOS ODONTOLOGOS

En el gráfico de barras, las barras o columnas son de igual ancho y hay espacios entre las categorías que las barras representan. Las barras pueden ser verticales u horizontales. Los gráficos de barras son excelentes para dar una idea comparativa de conjuntos de datos. En la Figura 4.2 se muestra un gráfico de barras.

Figura 4.2

TASAS DE SOBREVIDA DE CINCO AÑOS EN EL CANCER BUCAL

Aunque a primera vista los histogramas y los gráficos de barras pueden parecer muy similares, los primeros se deben confeccionar con mucho más cuidado. No hay espacio entre las celdas o barras, y la **superficie** incluida en la barra (que se determina multiplicando su altura por su ancho), representa la distribución de frecuencias de los datos. Esto resultará más claro después de estudiar el Capítulo 8. Cuando los intervalos son de igual amplitud, es fácil confeccionar histogramas. Sin embargo, es necesario practicar ajustes si los intervalos no son iguales (por ejemplo, si representan diferentes cantidades de días u horas) para que las superficies delineadas mantengan sus proporciones.

POLIGONO DE FRECUENCIAS

A menudo se utiliza el **polígono de frecuencias** en reemplazo del histograma. En vez de confeccionar una barra para cada intervalo de clase, **se marca un punto a la misma altura del punto medio del intervalo de clase**. Estos puntos **se unen con líneas rectas** para formar un polígono de frecuencias como el que se muestra en la Figura 4.3 Se confecciona muchos gráficos utilizando este principio de representación visual de los datos.

Figura 4.3

NUMERO DE HORAS DE TRABAJO POR SEMANA CON ARREGLOS A LOS INFORMES DE LOS ODONTOLOGOS

Problema 3:

Utilizando la distribución de frecuencias relativas de las edades de los niños en el momento de erupción del primer premolar superior, que usted determinó en el último problema, construya ahora un histograma y un polígono de frecuencias.

Su solución:

DISTRIBUCION DE FRECUENCIAS ACUMULADAS

Por último, **la distribución de frecuencias acumuladas se puede confeccionar calculando el porcentaje de mediciones**

por debajo del límite superior de cada intervalo de clase. Volvamos entonces a nuestro ejemplo del número de horas trabajadas por los odontólogos; la distribución acumulada se puede obtener fácilmente utilizando la distribución de frecuencias relativas de la Tabla 4.3. En la siguiente Tabla 4.4 vemos que el 14% de los odontólogos manifestaron que trabajaban como máximo 20 horas por semana. Es habitual mostrar las frecuencias relativas y acumuladas en una misma tabla.

Tabla 4.4

DISTRIBUCION DE FRECUENCIAS RELATIVAS Y ACUMULADAS DEL NUMERO DE HORAS DE TRABAJO POR SEMANA CON ARREGLOS A LOS INFORMES DE LOS ODONTOLOGOS

Hora de Trabajo por semana	Frecuencia	Frecuencia Relativa (%)		Frecuencia acumulada (%)
1-10	3	6		6
11-20	4	8	+	14
21-30	6	12	+	26
31-40	28	56	+	82
41-50	8	16	+	98
51-60	1	2	+	100
Total	**50**	**100**		

Nota: En algunos casos puede ocurrir que el total no sume exactamente 100% por los errores producidos como consecuencia del redondeo hacia el porcentaje más cercano.

REVISION

1) Recuerde que la tabla de frecuencias es la especificación del número de mediciones que entran en

cada intervalo de clase. Usted tendrá que definir estos intervalos. Para la mayoría de las tablas de frecuencia es conveniente que haya de 5 a 10 intervalos.

2) La distribución de frecuencias relativas es la especificación del número de mediciones que entran en cada intervalo de clase, dividida por el número total de observaciones y multiplicada después por 100 para convertir el valor en porcentaje.

3) El histograma se confecciona a partir de la distribución de frecuencias relativas, dibujando una barra para cada intercalo de clase cuya superficie corresponsa a la frecuencia relativa de las mediciones logradas en ese intervalo.

4) El polígono de frecuencias es un gráfico que se representa colocando un punto en el punto medio del intervalo de clase correspondiente a la frecuencia relativa de las mediciones hechas en ese intervalo. Luego se unen los puntos con líneas rectas.

5) La tabla de distribución de frecuencias acumuladas se confecciona calculando el porcentaje de mediciones por debajo del límite superior de cada intercalo de clase.

TEST FINAL 4

1) Primero, examine los datos confusos y desordenados de la Tabla 4.8. Son las respuestas que dieron 25 personas que hacían fila en un supermercado cuando se les preguntó: Cuándo hizo su última visita al odontólogo?

2) Segundo, examine los mismos datos después de que estén ordenados (Tabla 4.9).

3) Confeccione una tabla de frecuencias rotulándola correctamente: "lapso transcurrido desde la última visita al odontólogo".

4) Luego, calcule y agregue la tabla de distribuciones de frecuencias relativas y acumuladas.

5) Confeccione un gráfico de barras a base de la información de la Tabla 4.9.

Tabla 4.8

LAPSO TRANSCURRIDO DESDE LA ULTIMA VISITA AL ODONTOLOGO datos primarios sin ordenar		
9 meses	4 meses	11 meses
1 año	2 años y 5 meses	1 año y 2 meses
9 meses	5 años	nunca lo visitó
6 años	8 meses	8 meses
1 año y 9 meses	10 meses	8 meses
1 mes	1 año y 6 meses	1 año y 5 meses
5 meses	2 meses	1 año y 8 meses
2 años	6 meses	11 meses
3 años y 1 mes		

Tabla 4.9

LAPSO TRANSCURRIDO DESDE LA ULTIMA VISITA AL ODONTOLOGO datos primarios ordenados		
1 mes	9 meses	1 año y 8 meses
2 meses	10 meses	1 año y 9 meses
4 meses	11 meses	2 años
5 meses	11 meses	2 años y 5 meses
6 meses	1 año	3 años y 5 meses
8 meses	1 año y 2 meses	3 años y un mes
8 meses	1 año y 5 meses	5 años
8 meses	1 año y 6 meses	6 años
9 meses		nunca lo visitó

CAPITULO 5.

MEDIDAS DE TENDENCIA CENTRAL Y DISPERSION

INTRODUCCION

Es importante que los estudiantes de odontología y futuros higienistas dentales puedan comprender el significado de estadísticas descriptivas sencillas, que se utilizan frecuentemente en publicaciones profesionales, disertaciones y clases. Es particular en los turnos de atención en las clínicas y en el consultorio, usted tendrá que tener algunos conocimientos de estadística descriptiva para tomar decisiones con fundamento en materia de medicación con fármacos, métodos de tratamiento e instrumental quirúrgico. Quizás también tenga que calcular medias y desviaciones típicas de muestras de datos en algunos de los cursos a los que concurra o durante las investigaciones que deba emprender como parte de su formación como dentista.

Los prestadores de atención odontológica hacen mediciones de los atributos de los pacientes y establecen comparaciones mentales con una norma (media) para ayudarse en el diagnóstico. Algunas veces consideran cuanta dispersión o variación hay en los datos. La desviación estándar es una medida de dispersión, y al igual que la media, una estadística descriptiva.

En este capítulo se pretende explicar **dos tipos de estadísticas descriptivas:** las que son promedios o **medidas de tendencia central**, como la media, y las que son **medidas de dispersión**, como la desviación estándar.

OBJETIVO

1) Usted podrá calcular la **media**, la **mediana**, la **moda**, la **varianza**, el **rango** y la **desviación estándar** de una determinada muestra de datos.

2) Dados los siguientes enunciados:

   a) La mitad de los odontólogos ejercen más de 38 horas por semana.

   b) La preparación de una cavidad de clase II en un premolar demanda en promedio 17 minutos, y

   c) Los odontólogos que trabajan con tres consultorios o sillones dentales, dos asistentes y una higienista a media jornada, son más que los que tienen otros equipos y distinto personal.

Usted podrá denominar lo que estos enunciados indican y lo que no indican en relación con la población problema en estudio.

TENDENCIA CENTRAL

Las medidas de tendencia central tratan de proporcionar una estadística de la muestra que describa las características de todo un conjunto de datos. Estas estadísticas que representan **promedios**, son tres: la media, la moda y la mediana.

La **media** muestral de un conjunto de datos en el **promedio aritmético**. Para calcular la media, sume los valores de cada observación y divídalos por el número de observaciones.

Problema 1:

Qué sinónimo le parece a usted apropiado para el término media?

Su solución:

Ahora bien, suponga que en el curso de una semana un endodoncista haya examinado a cinco pacientes que necesitaban tratamiento en sus incisivos laterales superiores. Habida cuenta de que una restauración endondóntica debe ajustarse con precisión a la longitud del conducto radicular, el conocimiento de la longitud media de los dientes solo constituye un punto de partida para este procedimiento. En última instancia, habrá que conocer la longitud de cada diente en tratamiento. Por lo tanto, el endodoncista midió la longitud de cada diente conforme a la radiografía y comprobó que las medidas eran 19, 20, 21, 18 y 22mm. La longitud promedio es entonces (19 + 20 + 21 + 18 + 22) / 5 = 20mm. Sustituyendo estos números pro símbolos podemos desarrollar una fórmula para calcular la media de una muestra. Suponga usted que los símbolos x_1, x_2, ... etc., hasta x_n son representativos de lo datos de un conjunto en el cual n es el número total de datos. La media de la muestra, representada por el símbolo X (léase "equis barra"), se calcula por medio de la fórmula:

$$x = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

El símbolo x se utiliza habitualmente en estadística para representar la media de una muestra.

Tal vez usted familiarizando la notación "Σ" (que se lee "sigma"), en la que Σ indica **la suma** de todas las x: por ejemplo, $x_1 + x_2, + \ldots + x_n$, Σ sencillamente una abreviatura conveniente. Utilizando esta notación, la expresión anterior se puede escribir así:

$$x = \frac{\Sigma x}{n}$$

Problema 2:

Suponga usted que midió las longitudes de siete incisivos centrales superiores que necesitaban tratamiento endodóntico. Las longitudes fueron 21, 23, 29, 20, 18, 22 y 21. Cuál es la longitud media de esta muestra de incisivos?

Su solución:

La media es fácil de calcular. Es un promedio, o más técnicamente, una medida de **tendencia central** o de **ubicación central** en un conjunto de datos. Si se nos dijera que el peso medio de una muestra de siete hombres es de 115 kilogramos, tendríamos la impresión de que nos encontramos frente a un grupo de hombres realmente voluminosos.

La **mediana** es otra media de tendencia central. La mediana de una muestra es el **valor central de un conjunto de datos**. Si los datos se disponen en orden creciente, la mediana los divide en dos conjuntos más pequeños pero de igual tamaño. La mitad de los valores están por encima y la otra mitad por debajo de la mediana.

Problema 3:

Trate ahora de precisar el significado del término "mediana" sin releer las líneas anteriores y expóngalo a continuación.

Su solución:

Para calcular la mediana de una muestra, disponga los datos en orden creciente y selecciones luego el que esté en el medio. Algunas veces se plantea una dificultad porque, con un número par de datos en el conjunto, quedan dos en el medio.

En este caso tome usted la media de estos dos valores o puntajes. Con un número impar de datos no hay ningún problema, pues solo puede quedar uno en el medio.

Problema 4:

Trate de calcular algunas medianas, Cuál es la mediana de la longitud de los 7 incisivos centrales superiores del último ejercicio, que median 21, 23, 29, 20, 18, 22 y 21 mm de largo?

Su solución:

Problema 5:

Suponga ahora que hubo otro paciente cuyo incisivo central superior tenía 25mm de largo y que este diente se incluyó para formar un grupo de ocho. Cuál es la mediana de longitud de los ocho dientes?

Su solución:

La **moda** es una tercera medida de tendencia central que se emplea a menudo para describir un conjunto de datos. La moda es **el valor (los valores) que aparece(n) con mayor frecuencia en un conjunto de datos.** No siempre la moda es una sola; en efecto, **es posible que haya varias modas en un conjunto de datos.** Por ejemplo, suponga que a 10 grupos que prestan atención odontològica se les haya preguntado cuántas higienistas emplean y que las respuestas hubieran sido 0, 1, 1, 1, 2, 2, 3, 3, 3, 4. Existen dos modas en ese conjunto de datos: una con puntaje 1 y otra con puntaje 3.

Problema 6:

Suponga que se hubiera formulado la misma pregunta a otros 10 grupos que brindan atención odontológica y que las respuestas hubiesen sido 1, 5, 3, 6, 2, 1, 2, 5, 4, 3, higienistas dentales. Cuál(es) es(son) la(s) moda(s) en este conjunto de datos?

Su solución:

Tal vez usted se pregunte cuáles son las ventajas relativas de utilizar la media, la mediana y la moda como medidas de tendencia central o ubicación central de un conjunto de datos. En la mayoría de las circunstancias se prefiere la media. Sin embargo, tiene la desventaja de estar influida por los valores extremos del conjunto de datos. Por ejemplo, si uno de los incisivos laterales del primer ejemplo hubiera sufrido una malformaciòn y como midiese 10 mm de longitud, la media de la muestra habría sido 3 mm menor. Un dato con un valor extremo tiene menos influencia en la mediana, pues no se utilizan cada uno de los datos para obtener el puntaje medio. Por eso, un solo valor alto no eleva la mediana. En general, la moda es una deficiente medida de tendencia central, pues puede ubicarse en un valor extremo que no proporcione ningún indicio acerca de la ubicación del conjunto de datos. En cambio, una moda puede ofrecer la imagen de lo más frecuente o "popular" en una distribución. Para los vendedores de instrumental quirúrgico y de medicamentos a menudo resulta útil utilizar estadísticas modales de los consultorios odontológicos y de los patrones de tratamiento.

Problema 7:

Suponga que usted ha entrevistado a los pacientes que encontró en las salas de espera de varios consultorios odontológicos de cierta comunidad y les preguntó cuánto tiempo tuvieron que esperar para conseguir hora para un tratamiento que no era de urgencia. Las respuestas pudieron ser 12, 3, 5, 2, 21, 8, 4, 7, 14, 3, 23 y 34 días. Calcule entonces la media, la mediana y la moda de este conjunto de datos.

Qué medida de tendencia central utilizará usted para definir la tendencia o ubicación central de este conjunto de datos?

Su solución:

DISPERSION

Cada una de las tres medidas de tendencia central puede brindar información importante sobre el conjunto de datos. Sin embargo, cualquier resumen que adopte la forma de tendencia central es por sí solo de valor limitado, pues nada nos dice acerca de la variabilidad de los datos. El siguiente ejemplo es ilustrativo.

Después de su graduación en la universidad y mientras trata de decidir la ubicación de su futura clínica odontológica, quizás encuentre interesante consultar una tabla estadística que le dé información sobre la cantidad promedio de habitantes por odontólogo en diferentes comunidades. Si usted ya estaba resuelto a ejercer en su ciudad natal, que tiene 100.000 habitantes y cuenta con ocho odontólogos, pero observa que la cantidad de habitantes por odontólogo es 1250, una cifra que está muy por debajo del promedio de la provincia, quizás experimente dudas sobre la conveniencia de establecerse en ella. Sin embargo, después puede usted enterarse de que esta media se ve muy afectada por tres odontólogos prácticamente retirados que ejercen solo dos medios días por semana y que tienen unos 250 pacientes cada uno. Cuando pregunta cuál es el rango de pacientes por odontólogo, usted se entera de que oscila entre 250 y 2.200. Con esta información usted ya tiene una impresión distinta de la actividad que desarrollan las clínicas odontológicas en la comunidad que le interesa.

El rango de un conjunto de datos es la diferencia entre el valor más alto y el más bajo. El rango es enorme para los valores que forman el precedente conjunto de datos y esta información nos induce a interpretar con cautela la media que originalmente conocimos. Por los general, la información sobre la dispersión es tan importante como la referente a la tendencia central de un conjunto de datos. El rango es una medida de dispersión.

Problema 8:

Se midió la presión arterial de nueve hombres obesos y se verificaron estos valores: 181, 175, 160, 182, 148, 180, 173, 175 y 192 mmHg. Cuál es el rango de la presión arterial de estos hombres?

Su solución:

La segunda medida de dispersión es la **varianza.** Cuanto mayor sea la varianza, tanto más dispersos estarán los datos en relación con el valor medio. Una varianza cero indica que no hay dispersión; en otras palabras, que todos los puntajes tienen el mismo valor. La varianza de una muestra se define como **la suma de los cuadrados de las desviaciones con respecto a la media de la muestra dividida por el número total de observaciones menos uno.** Una desviación referida a la media de la muestra es sencillamente la diferencia entre un dato y la media del conjunto de datos. Utilizando la notación matemática, la varianza puede expresarse así:

$$s^2 = \frac{(x_1 - x_2)^2 + (x_2 - x)^2 + \ldots + (x_n - x)^2}{n - 1}$$

donde s^2 representa la varianza muestral del conjunto de datos, las x, (x_1, x_2, ..., x_n) los valores de los datos, x la media de los valores y n el número total de las observaciones.

Utilizando la notación Σ, la misma fórmula puede escribirse así:

$$s^2 = \frac{\Sigma\ (x - x)^2}{n - 1}$$

Nota: el divisor es $n - 1$, no n. La razón de que sea así es teórica y no nos interesa en este momento. (Si desea temer más información al respecto, consulte la obra de P. Armitage, citada en la página 57).

Problema 9:

Ahora, sin releer la sección anterior, defina con palabras qué es la varianza.

Su solución:

Problema 10:

Ahora usted tendría que calcular una varianza. Se midió la concentración de vapor de mercurio en la atmósfera en cinco sitios distintos de una clínica dental, después de haberse producido un derrame accidental. Los resultados fueron los siguientes: 0,01, 0,01, 0,03, 0,04, 0,06 mgHg/m³. Cuál es la varianza de la muestra para estos cinco niveles de vapor de mercurio?

Su solución:

Hay otra forma de escribir la fórmula de la varianza, que a menudo resulta más fácil de aplicar, especialmente si se utiliza una calculadora. La fórmula alternativa es:

$$s^2 = \frac{x_1^2 + x_2^2 + \ldots + x_n^2 - 1/n\ (x_1 + x_2 + \ldots + x_n)^2}{n - 1}$$

o sea:

(suma de los cuadraros de los valores) − n/1 (suma de los valores)²

$$s^2 = \frac{\qquad\qquad}{n - 1}$$

Utilizando la notación Σ:

$$s^2 = \frac{\Sigma x^2 - \frac{1}{n} (\Sigma x)^2}{n - 1}$$

Recuerde Σx^2 es la suma de los cuadrados de los valores de cada dato y $(\Sigma x)^2$ es la suma de esos valores elevada al cuadrado.

Problema 11:

Repita ahora los cálculos hechos para obtener la varianza de la muestra de los niveles de vapor de mercurio, utilizando la nueva fórmula.

Su solución:

DESVIACION ESTANDAR

Habrá advertido usted que, si bien la media se mide con las mismas unidades utilizadas para los datos, la varianza se determina por el cuadrado de esas unidades. Para suponer este inconveniente, en general se emplea como medida de la dispersión la raíz cuadrada de la varianza en vez de la varianza propiamente dicha. A esta cantidad se la conoce como **desviación estándar**.

s = desviación estándar de la muestra

= $\sqrt{\text{varianza de la muestra}}$

El símbolo matemático más común de la desviación estándar de un conjunto de datos es s, pero frecuentemente se denota sd.

Problema 12:

Cuál es la desviación estándar aproximada del nivel de vapor de mercurio de la muestra anterior?

Su solución:

Una última acotación: se debe tener un poco de cuidado al utilizar el rango como medida de dispersión, habida cuenta de que se ve muy afectado por los valores extremos.

Ya debe estar preparado para el test final. Si hay partes de este capítulo de las que no se siente seguro, leales de nuevo. Si considera que ha comprendido la presentación, trate de resolver el test final. Buena suerte ... pero, antes de comenzar, estudie la revisión.

REVISION

1) Media de la muestra = $\bar{x}$ = Σ x/n

2) Mediana = valor del dato del medio

3) Moda(s) = valor del dato o de los datos que aparece(n) con mayor frecuencia

4) Rango = (valor del dato mayor) − (valor del dato menor)

5) Varianza de la muestra =

$$s^2 = \frac{\Sigma\ (x - x)^2}{n - 1} = \frac{\Sigma x^2 - 1/n\ (\Sigma x)^2}{n - 1}$$

6) Desviación estándar de la muestra = √ varianza de la muestra = s

Problema 13:

Antes de encarar el test final, trate de determinar las estadísticas a que se refieren los enunciados incluidos como OBJETIVO en la primera página de este capítulo.

Su solución:

TEST FINAL 5

Un ortodontista registró la dimensión vertical bucal de 10 nuevos pacientes que necesitaban prótesis completas. Las mediciones fueron: 7, 11, 10, 8, 5, 7, 9, 7, 6, 10 mm.

3 1) Calcule la(s) moda(s), la mediana (4) y la media (5) de esta muestra de mediciones de dimensión vertical.

2) Determine el rango (6) de las mediciones de dimensión vertical en estos diez pacientes.

3) Calcule la varianza (7) muestral y la desviación (8) estándar de las mediciones de dimensión vertical en esta muestra de diez pacientes.

HOJA DE RESPUESTAS
LABORATORIO N.____

PROBLEMA No 1

PROBLEMA No 2

PROBLEMA No 3

PROBLEMA No 4

PROBLEMA No 5

PROBLEMA No 6

PROBLEMA No 7

PROBLEMA No 8

PROBLEMA No 9

PROBLEMA No 10

CAPITULO 6.

PROBABILIDAD

INTRODUCCION

"Juego mi vida a un golpe de dados,
y quiero confiarme a la suerte!"
Shakespeare, Ricardo III

El concepto de probabilidad desempeña un papel importante en el tratamiento de pacientes y en la investigación científica, pues son pocas las decisiones sobre pronóstico de un tratamiento o sobre resultados de un experimento que se pueden predecir con certeza absoluta. Un paciente puede desear saber cuáles son sus posibilidades de verse afectado por un cáncer bucal. Quizás otro paciente con cáncer bucal desee saber cuáles son sus posibilidades de sobrevida en los próximos cinco años. Sin embargo, es difícil pensar en probabilidades en función del paciente individual que nos preocupa en forma directa. Es necesario, por consiguiente, generalizar la situación de un paciente dado y verlo como una de las muchas observaciones hechas en el pasado con pacientes semejantes en circunstancias similares. Empero, en primer lugar hay que comprender las reglas matemáticas básicas que gobiernan el campo de la probabilidad.

OBJETIVO

1. Usted podrá definir **probabilidad** y calcular valores simples de probabilidad.

2. Usted podrá utilizar la **regla de la suma de probabilidades** para determinar la probabilidad de que se produzca uno de dos hechos.

3. Usted podrá usar la **regla de la multiplicación de probabilidades** para establecer la probabilidad de que se produzcan dos hechos.

PROBABILIDAD

Definición. Probabilidad es una medida matemática de la posibilidad de que se produzca un acontecimiento o un hecho en una serie de ensayos repetidos en condiciones similares y con determinada frecuencia.

En 1977, hubo 12.835 aspirantes a ingresar a facultades de odontología que tenían un total de 5.945 vacantes disponibles. Fueron aceptados el 46% de los aspirantes. El resultado obtenido por cada aspirante se puede considerar como un ensayo o hecho aislado, que en este caso se repitió 12.835 veces en condiciones similares. La probabilidad de que se produzca el hecho A, siendo A "un aspirante en condiciones de ingresar en la carrera de odontología en 1977" es 0,46. Esto se escribe con la siguiente notación :

$$P\ (A) = 0{,}46$$

y se lee "la probabilidad de A es 0,46"

Las probabilidades también se pueden expresar en forma de decimales, fracciones o porcentajes. Si se las expresa en forma de decimales o fracciones, el rango de valores posibles va de 0 a 1. Una probabilidad 0 indica que es imposible que se produzca un hecho. Una probabilidad 1 indica en cambio la certeza de que se va a producir. Si se expresa en forma de porcentaje, los valores oscilan entre 0,00 y 100,00 por ciento.

Al calcular probabilidades, hay que determinar todos los resultados posibles que puede asumir un hecho. La suma de las probabilidades asignada a cada uno de los resultados mutuamente excluyentes posibles debe ser igual a 1. En el ejemplo anterior existen dos resultados posibles: o se produce el hecho A o no se produce. Se los llama hechos complementarios. Como la suma de estas dos probabilidades debe ser igual a 1, la probabilidad de un no A, que se escribe P (A) o P (Á), es la siguiente:

$$P\ (A) = 1 - P\ (A)$$

Sustituyendo P (A) por 0,46 podemos determinar que P (A) = 1 - 0,46 = 0,54. Estos se denominan **hechos mutuamente excluyentes**, pues se producirá uno u otro. No pueden producirse ambos como resultado del mismo ensayo. Por cada vez que se arroje una moneda el resultados será cara o cruz (siempre que no sea una moneda adulterada).

Por cada vez que se arroje un dado el resultado será = 1, 2, 3, 4, 5 o 6. Estos son hechos mutuamente excluyentes. Si existe un número "n" de resultados igualmente posibles en que se puede producir el hecho, la probabilidad de que se produzca cada uno de ellos es 1/n. Para cada vez que se arroje el dado existen seis posibles resultados, todos igualmente probables. La probabilidad de obtener un "3" es 1/6 y es igual a P (obtener un 4) = P (obtener un 5), etc.

$$P(1) + P(2) + P(3) + P(4) + P(5) + P(6) = 1$$

Volvamos a nuestros aspirantes a ingresar en facultades de odontología que están esperando nerviosamente la carta o llamada telefónica decisiva. Si los resultados se consideran éxitos o fracasos y hay N_e éxitos y N_f fracasos, el número total de resultados es $N_e + N_f$. En este caso N_e = 5.954 es el número de éxitos y el número total de éxitos y fracasos, 12.835. La probabilidad de obtener el hecho A (éxito) se puede calcular :

$$P(A) = N_e/(N_e + N_f) = 5.954/12.835 = 0.46$$

(Nota : suponemos aquí que todos los aspirantes están igualmente capacitados y tienen la misma posibilidad de ser elegidos. Los que ganan en esta lotería profesional saben por cierto que esto no es verdad).

Problema 1:

En una facultad de odontología, 125 alumnos del último año de estudios rindieron el examen para obtener la licencia habilitante para ejercer la profesión. Pasaron el examen 110 estudiantes. Cuál era la probabilidad de pasar el examen? Cuál era la probabilidad de no pasar el examen?

Su solución:

No en todos los casos cada uno de los resultados tiene la misma posibilidad de producirse. En una población dada :

P (Oclusión Clase I) = 0,70

P (Oclusión Clase II) = 0,25

P (Oclusión Clase III) = 0,05

Son las tres formas posibles de oclusión, mutuamente excluyentes. La suma de sus probabilidades es 1.

REGLA DE LA SUMA DE PROBABILIDADES

La probabilidad de que se produzca cualquiera de dos o más hechos mutuamente excluyentes es la suma de las probabilidades de cada uno de ellos.

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

$$P\ (A\ o\ B\ o\ C) = P\ (A) + P\ (B) + P\ (C)$$

donde A, B y C son mutuamente excluyentes.

La probabilidad de que una persona seleccionada al azar entre la población mencionada más arriba tenga una oclusión de Clase I o de Clase II se expresa de la siguiente forma:

$$P\ (\text{Clase I o Clase II}) = P\ (\text{Clase I}) + P\ (\text{Clase II}) =$$

$$0.70 + 0.25 = 0.95$$

Esto resulta ser lo mismo que la probabilidad de no tener una oclusión de Clase III, lo que se puede expresar así:

$$1 - P\ (\text{Clase III}) = 1 - 0{,}05 = 0{,}95$$

Problema 2:

En el centro de salud de un vecindario con 4.250 pacientes activos, los pagos por las prestaciones odontológicas eran hechos en forma directa por el paciente, por conducto de un seguro de salud estatal o por medio de seguros privados. Si hay 2.210 pacientes que pagan directamente "de sus bolsillos" y 510 pacientes que tienen cobertura con un seguro odontológico privado, cuál es la probabilidad de que un paciente tenga algún tipo de seguro, ya sea estatal o privado?

Su solución:

La situación se hace más complicada cuando los hechos no son mutuamente excluyentes. La superposición entre dos categorías se puede representar en el siguiente diagrama por medio de los dos círculos A y B, que se superponen. Se puede considerar que el área sombreada en el medio, C, pertenece tanto a A como a B.

En un grupo de adultos con sus primeros molares cariados u obturados, puede haber algunos que estén tanto cariados como obturados. Supongamos :

15 dientes están cariados

80 dientes están obturados

5 dientes están cariados y obturados

Tenemos :

P (presencia de caries) =

P (dientes cariados o cariados y obturados) =

0,15 + 0,05 = 0,20

P (presencia de restauración) =

P (dientes obturados o cariados y obturados) =

0,80 + 0,05 = 0,85

Como usted podrá ver, el grupo "cariado y obturado" está contabilizado dos veces y, por consiguiente, la suma de las dos probabilidades no es igual a 1. Así, si queremos considerar las categorías "presencia de caries" y "presencia de restauración" como los dos únicos posibles resultados de este ejemplo, y suponemos que son hechos mutuamente excluyentes, obtendremos una probabilidad mayor que 1 y la suposición habrá sido incorrecta.

En el caso de que dos hechos no sean mutuamente excluyentes, es aplicable la siguiente fórmula:

$$P\ (A \text{ o } B) = P\ (A) + P\ (B) - P\ (A \text{ y } B)$$

Es el caso general de la regla de la suma de probabilidades. Si los dos hechos A y B son mutuamente excluyentes, la probabilidad de que se produzcan A y B es 0 y el último término de la ecuación, P (A y B), se puede suprimir.

De esta manera:

P (presencia de caries o presencia de restauración) =

P (presencia de caries) + P (presencia de restauración) −

P (presencia de caries y restauración) =

$$0{,}20 + 0{,}85 - 0{,}05 = 1{,}00$$

Problema 3:

Otro grupo de adultos con dentición libre de caries se incorpora a la muestra. Los primeros molares en consideración caen ahora dentro de las siguientes categorías:

40 sanos

15 cariados

80 obturados

5 cariados y obturados

Cuál es la probabilidad de que un primer molar seleccionado al azar solo tenga caries, de que tenga caries, de que esté "cariado y obturado", de que esté obturado, de que esté cariado u obturado?

Su solución:

HECHOS INDEPENDIENTES Y LA REGLA DE LA MULTIPLICACION

Si la aparición o no aparición de un hecho no afecta de ninguna forma al otro, se considera que estos dos hechos son independientes. Para determinar la probabilidad de que se produzca la aparición de dos hechos independientes, A y B, solo necesitamos multiplicar las probabilidades de que se produzcan cada uno de los dos hechos individuales. Si A y B son independientes, entonces:

$$P\ (A \text{ y } B) = P\ (A) \cdot P(B)$$

El concepto de hechos independientes es distinto del de hechos mutuamente excluyentes. Dos hechos pueden ser independientes sin ser mutuamente excluyentes. Por ejemplo, si se arrojan simultáneamente dos dados, el resultado de cada dado es independiente; en otras palabras, el lanzamiento de un dado en nada afecta al otro, cualquiera que sea el número que aparezca en cada uno de ellos.

Considere que "A" es el resultado del primer dado, "B" el del segundo y "C"W la suma de los dos resultados. A y B son hechos independientes y los resultados A = 3 y B = 3 no son mutuamente excluyentes, porque sacar "3" con un dado no excluye la posibilidad de sacar también "3" con el otro.

Los hechos A y C no son independientes, porque el resultado C depende tanto de A como de B. Si A = 1, C tiene una probabilidad cero de ser mayor que 7; aquí A = 1 y C = 8 son mutuamente excluyentes. Si B = 6 y C = 11, B y C no son mutuamente excluyentes, porque ambos hechos pueden producirse.

Para dos hechos independientes que no sean mutuamente excluyentes, tenemos:

$$P\ (A\ o\ B) = O(A) - P(B) \cdot P\ (B)\ \ )$$

Esta fórmula es similar a la utilizada en la página 57, pero ahora el último término P (A y B) se puede sustituir por P(A) · P(B).

En el ejemplo anterior de A = 5 y B = 6,

$$P(AoB)\ ?\ (1/6) + (1/6) - (1/6) = 11/36$$

En muchas enfermedades genéticas, como la osteogénesis imperfecta, la probabilidad de que una pareja determinada tenga hijos afectados, es la misma de nacimiento en nacimiento. Esto significa que la probabilidad de que nazca un niño afectado no tiene ningún efecto sobre la probabilidad de que nazcan niños afectados en los siguientes partos. Para una pareja determinada la probabilidad de:

un hijo afectado = 1/4

un hijo normal O 1/4

una hija afectada = O

una hija normal = 1/2

Cuál es la probabilidad de que dos hijos afectados nazca en dos embarazos sucesivos? Como cada nacimiento es un hecho independiente,

P (hijo afectado e hijo afectado) =

P (hijo afectado) • P (hijo afectado) =

1/4 • 1/4 = 1/16

Problema 4:

Para el caso de la pareja mencionada, cuál es la probabilidad de que del primer embarazo nazca un hijo afectado y de que luego aparezca una hija normal? Nuevamente suponga la probabilidad de la presencia de un niño afectado en un nacimiento no influye sobre la probabilidad de que aparezcan hijos afectados en los siguientes partos.

Su solución:

Por desgracia, los hechos en la vida real no son siempre independiente; por eso el cálculo de la probabilidad de su aparición resulta más complejo.

Suponga que usted sabe la probabilidad de que la superficie medial del inciso lateral superior daréis tenga caries es 0.20 y que la probabilidad de que la superficie distal del incisivo central superior derecho tenga caries es 0.15. Si usted utiliza la regla de multiplicación para hechos independientes, tendrá que esperar que la probabilidad de que ambas superficies dentarias tenga caries sea 0.20 x 0.15 = 0.03. Sin embargo, la experiencia clínica permite afirmar que si la superficie mesial de un diente tiene caries es mucho más probable que la superficie distal del diente adyacente también la tenga. La probabilidad de que ambas superficies estén cariadas puede muy bien ser 0.13 en vez de 0.03. Hay que realizar un ajuste en la fórmula para tener en cuenta este tipo de circunstancias.

PROBABILIDADES CONDICIONADAS

Muy frecuentemente es más o menos probable que cierto hecho se produzca si oro hecho ya se ha producido. Por ejemplo, si un individuo ha fumado mucho durante la mayor parte de su vida, es mucho más probable que llegue a tener cáncer bucal que alguno que nunca ha fumado.

Es más probable que un niño pequeño desarrolle caries generalizada si todos sus hermanos mayores que viven con él tienen caries generalizada. Cuando la probabilidad de un hecho resulta modificada o restringida porque se ha producido otro hecho o condición, se llama **probabilidad condicionada:** se escribe P (A/B) que se lee "la probabilidad de A dado que B se ha producido".

La forma general de la regla de la multiplicación que se puede utilizar cuando los hechos no son independientes, sino que uno es dependiente de otro o está condicionado por otro, es:

1) $P\ (A \text{ y } B) = P\ (B/A)\ P(A) = P\ (A/B)\ P\ (B)$

Cuando los hechos A y B son independientes, entonces:

2) $P\ (A/B) = P\ (A)$ y

3) $P\ (B/A) = P\ (B)$

porque la probabilidad de A será la misma, independientemente de que se haya o no producido B.

Con la correspondiente sustitución (3) en la ecuación 1) se obtiene:

$P\ (A y B) = P\ (B) \cdot (A)$

que es equivalente a la fórmula antes utilizada.

Suponga que un viernes a la noche se está tomando un pequeño descanso en el estudio y decide jugar a las cartas con sus amigos. Cuál es la probabilidad de que al retirar un mazo dos naipes en forma sucesiva, sin reemplazar el extraído primero, saque usted un siete y después un ocho? Recuerdo que P (8/7) se lee: Probabilidad de obtener un 8 dado que ya se ha obtenido un 7.

Su solución podría empezar así:

P (7 después 8) = P(7) · P (8/7)

P (/ al retirar la primera carta) = 4/52

Como retiramos las cartas sin reemplazar la ya extraída, si ya sacamos un 7 quedarán 51 naipes de los cuales 4 serán ochos. De esta manera P (8/7) = 4/51 y

$$P\ (7\ y\ 8) = (4/52)\ (4/51) = (1/13)\ (4/51) = 4/663$$

Problema 5:

Cuál es la probabilidad de que usted obtenga primero un 7 y luego otro 7 si retira sucesivamente dos cartas de un mazo de 52 naipes sin reemplazar la primera antes de retirar la segunda?

Su solución:

REVISION

1) **Regla de la suma de probabilidades** o regla "O" en forma genérica:

$$P\ (A\ o\ B) = P(A) + P(B) - P\ (A\ y\ B)$$

Para hechos mutuamente excluyentes:

$$P\ (A\ o\ B) = P(A) + P(B) - [P(A) \cdot P(B)]$$

2) **Regla de la multiplicación de probabilidades** o regla "Y" en forma genérica:

$$P\ (A\ y\ B) = P(B/A)\ P\ (A)$$

Hechos independientes:

$$P\ (A\ y\ B) = P(A) \cdot P(B)$$

$$P\ (B/A) = P\ (B)$$

TEST FINAL 6

Una joven dama, Marcela, se ha mudado recientemente de Carolina del Norte a Boston, una ciudad que no conoce, y tiene cierta dificultad para elegir entre las diferente formas de atención odontológica que se le ofrecen. Le preocupan los costos, la calidad y la comodidad y decide investigar qué le ofrece la clínica de la Facultad de odontología. Cuando se acerca para pedir una cita, se entera de que puede ser tratada por estudiantes novatos, por estudiantes avanzados o por docentes. (Lo que ella en realidad quiere es profilaxis por parte de una higienista dental). Según quien sea el que le brinde el servicio, el tiempo y el costo variarán.

Tabla 6.1

TIEMPO PARA COMPLETAR UNA RESTAURACION DE DOS SUPERFICIES		
Tipo de efector	**No.**	**Horas**
Estudiantes novatos	96	3,3
Estudiantes avanzados	90	1,5
Docentes	14	0,2

a) Si Marcela va a ser atendida por un efector elegido al azar, cuál es la probabilidad de que sea atendida por un docente?

b) Si su hermana Susana decidiera solicitar una cita, cuál es la probabilidad de que sea atendida por un estudiante avanzado?

c) Cuál es la probabilidad de que tanto Marcela como Susana sean atendida por un estudiante novato, suponiendo que sea factible que el mismo alumno trate a los dos?

d) Cuál es la probabilidad de que Marcela o Susana sean atendidas por un estudiante novato?

e) Cuál es la probabilidad de que Marcela y Susana sean atendidas por un estudiante novato, suponiendo que haya una norma en la Facultad que prohibe que un mismo alumno atienda a dos hermanas?

f) En cuál o cuáles casos de las seccione c, d, e son independientes los dos hechos?

CAPITULO 7.

RIESGO EMPIRICO Y LA FORMULA BINOMIAL

INTRODUCCION

Después de haber estudiado el primer capítulo en que nos ocupamos de las probabilidades, usted ya debe estar familiarizado con una forma de definir la probabilidad y con las reglas de la suma y la multiplicación para el cálculo de ciertas probabilidades (por ejemplo, P (A o B) y P (A y B). En este capítulo introduciremos otra definición más general de probabilidad y analizaremos qué se entiende por riesgo empírico. También expondremos otra forma de calcular algunas probabilidades.

OBJETIVO

1) Usted podrá definir la probabilidad de un hecho en función de la frecuencia relativa con que ese hecho se haya producido en el pasado.

2) Usted podrá estimar el riesgo empírico de un hecho a partir de la probabilidad de que ese hecho se produzca conforme a la definición de frecuencia relativa.

3) Usted podrá calcular la probabilidad de que un hecho dado se produzca exactamente r veces en n ensayos independientes idénticos, cuando la probabilidad de que se produzca en uno de esos ensayos sea p.

DEFINICION DE PROBABILIDAD

En el capítulo anterior determinamos la probabilidad de que se produzca un hecho "A" en un experimento que tiene n resultados aislados igualmente posibles y dijimos que es 1/n. Si el hecho se puede producir más de una vez o en más de una forma, la probabilidad de A es la relación entre el número de formas en que A se puede producir y el número total de resultados posibles, o sea A/n. Por ejemplo, si se arrojan un par de dados, hay 36 resultados posibles para el par de números que aparecen en la caras superiores de esos dados. Sin embargo, existen cuatro formas en que se puede producir el hecho A, "suma igual a 5": 1 y 4, 4 y 1, 2 y 3, 3 y 2. Por lo tanto, P (suma de 5) = 4/36 = 0.11. Aunque esta es una definición apropiada, no es lo suficientemente amplia como para cubrir las muchas situaciones en que no existen n resultados aislados igualmente posibles o cuando no tenemos forma de saber cuáles son esos resultados.

Una segunda definición más amplia de probabilidad se basa en el hecho de que si una situación o experimento en particular e repite un número indefinido de veces, la frecuencia relativa de la presencia del resultado específico, A, tiende a converger en un valor constante que denominamos la probabilidad de A.

$$P(A) = a/n$$

donde a = número de veces que se produce A
y n = número total de ensayos

Para cualquier secuencia finita de ensayos correspondientes a un experimento, la frecuencia relativa observada a/n, de A, es una **estimación** de la probabilidad de A. Si el número de ensayos es grande, es razonable que esperemos que la frecuencia relativa observada se acerque al valor de la frecuencia relativa teórica o **probabilidad verdadera** de A. Sin embargo, no olvidemos que en realidad nunca conocemos la probabilidad verdadera de A. Solo podemos estimarla empíricamente, o sea por la observación

Problema 1:

Usted quiere conocer la probabilidad de obtener "carreras" al arrojar una moneda. Para eso lanza la moneda diez veces y comprueba que aparecen cara 4 veces. Sobre la base de este experimento, cuál es la probabilidad de obtener cara al arrojar la moneda una sola vez?

Su solución:

Problema 2:

Usted repite el experimento anterior 4 veces más y obtiene los siguientes resultado:

Tabla 7.1

RESULTADOS			
Ensayo	**Número de lanzamientos**	**Número de caras**	**Frecuencia relativa**
Primero	10	4	0,4
Segundo	10	6	–
Tercero	10	5	–
Cuarto	10	4	–
Quinto	10	5	–
Total	**50**	**24**	**–**

Calcule la frecuencia relativa par cada ensayo y para el total.

Su solución:

RIESGO EMPIRICO

La frecuencia relativa es una forma de encarar la estimación de probabilidades que se emplea habitualmente en medicina clínica porque a veces ocurre que no hay otra forma de determinar las probabilidades deseadas. Es una situación que se presenta a menudo al dar asesoramiento genético y que servirá para introducirlo en el tema de riesgo empírico.

Se conoce el modo de heredar muchas enfermedades genéticas, como la fenilcetonuria (FCU) y la hemofilia A entre otras, y entonces la probabilidad de que una pareja en particular tenga un hijo afectado se puede resolver en términos teóricos. Sin embargo, existen otras enfermedades genética, como la esquizofrenia, el labio fisurado y la diabetes mellitus, en la que el mecanismo genético sigue aún sujeto a controversias. En estos casos, para determinar las probabilidades, debemos recurrir a un enfoque basado en las frecuencias relativas.

Un **riesgo empírico** es una probabilidad que deriva de los datos conjuntos de muchos casos similares, con prescindencia del (de lo) mecanismo(s) genético(s) del problema hereditario en cuestión. Específicamente se encuentra en esta situación la probabilidad de que los hijos de un individual afectado de una enfermedad hereditaria se vean también afectados. Aunque la cifras de riesgo empírico pueden ser guía útiles, hay que tener presente que son solo promedios generales y que no necesariamente estiman el riesgo de una familia en particular.

Por ejemplo, consideremos el "labio fisurado con o sin paladar fisurado", LF (P). Es un defecto con un componente genético definido. Hasta el presente u mecanismo genético exacto continua siendo objeto de debate. Es muy posible, por cierto, que este defecto entrañe una colección heterogénea de causas genéticas. Entre los norteamericanos, la incidencia es de casi 1/1000 en la raza blanca y se aproxima más a 1/2500 en los negros. Si una paciente tiene labio y/o paladar fisurado, la siguiente tabla nos indica algunos riesgos empíricos de que un hijo suyo en particular nazca con uno de esos defectos.

Tabla 7.2

RIESGO QUE UNA MADRE CON LF (P) TENGA UN HIJO CON LF (P)

No. de progenitores de la madre afectados	No. de hermanos de la madre afectados	Riesgo empírico de tener un hijo afectado
0	0	0,001
0	1	0,04
0	2	0,09
1	0	0,04
1	1	0,20

Problema 3:

Cuál es el riego empírico de que una mujer afectada de LF(P), cuyo padre y hermana también están afectados de ese defecto pero cuyo marido y todos sus demás familiares no lo están, tenga un hijo afectado de LF(P)?

Su solución:

FORMULA BINOMIAL

LA fórmula binomial $(p + q)^n$ nos brinda un medio para determinar la probabilidad de observar un número específico de éxitos en n ensayos. Por ejemplo, si la probabilidad de que cierta pareja tenga un niño con FCU (fenilcetonuria) en cualquier embarazo es 1/4 y si esta pareja planea tener cuatro hijos, cuál es la probabilidad de que tengan exactamente:

1) Cuatro hijos con FCU (0 no afectado)?

2) Tres hijos con FCU (1 no afectado)?

3) Dos hijos con FCU (2 no afectados)?

4) Un hijo con FCU (3 no afectados)?

5) Cero hijo con FCU (4 no afectados)?

El desarrollo del binomio puede darnos las respuestas a estas preguntas.

p = probabilidad de aparición en un único ensayo

Si consideramos a FCU como aparición del hecho, entonces p = 1/4

q = probabilidad de que no se produzca la aparición en un solo ensayo =

$$1 - p = 3/4$$

n = número total de ensayos = 4 (cuatro hijos)

El desarrollo del binomio e:

$$(p + q)^n \ 0 \ (q + p)p^4 + 4p^3q^1 + 6p^2q^2 + 4p^1q^3 + q^4$$

$$p^4 = (1/4)^4 = 1/256 =$$

p (exactamente cuatro hijos con FCU) = 0,0039

$$4p^3q^1 = 4(1/4)^3\ (3/4) = 12/256 =$$

p (exactamente tres hijos con FCU) = 0,0469

$$6p^2q^2 = 6(1/4)^2\ (3/4)^2 = 54/256 =$$

p (exactamente dos hijos con FCU) = 0,2109

$$4p^1q^3 = 4(1/4)\ (3/4)^3 = 10/256 =$$

p (exactamente un hijo con FCU) = 0,4219

$$q^4 = (3/4)^4 = 81/256 =$$

p (exactamente cero hijo con FCU) = 0,3164

Tome nota de que la potencia a la que se eleva p indica cuántos son los éxitos (en este caso hijos con FCU) y la potencia a la que se eleva q indica el número de fracaos (en este caso hijos no afectados). Fíjese también en que la suma de las cinco probabilidades es igual a 1,0 y en que para cualquiera de los términos del desarrollo la suma de los exponentes de p y q es igual a n, el número de ensayos.

Problema 4:

Dados los siguientes desarrollos del binomio, responda a las preguntas que siguen:

$$(p + q)^2 = p^2 + 2p^1q^1 + q^2$$

$$(p + q)^3 = p^3 + 3p^2q^1 + 3p^1q^2 + q^3$$

$$(p + q)^4 = p^4 + 4p^3q^1 + 6p^2q^2 + 4p^1q^3 + q^4$$

$$(p + q)^5 = p^5 + 5p^4q^1 + 10p^3q^2 + 10p^2q^3 + 5p^1q^4 + q^5$$

Si una determinada pareja va a tener un total de tres hijos y la probabilidad de que tenga uno con labio y/o paladar fisurado en un alumbramiento en particular es 1/5, cuál es la probabilidad de que ninguno de sus hijos e vea afectado?

Su solución:

Problema 5:

Si la pareja del problema 4 fuera a tener cinco hijos en vez de tres, cuál es la probabilidad de que exactamente cuatro de ellos tenga labio y/o paladar fisurado?

Su solución:

APLICACION DE LAS REGLAS DE MULTIPLICACION Y SUMA

Por qué funciona la fórmula binomial? Cuál es el origen de cada uno de los términos del desarrollo? A decir verdad, no necesitamos calcular las probabilidades con un binomio; podemos utilizar las reglas de multiplicación y la suma que aprendimos en el capítulo anterior. Recuerde la regla de multiplicación:

$$P\ (A \text{ y } B) = P(A) \cdot P(B)$$

donde $P\ (B/A) = P(B)$ si A y B son independientes

Esta regla se puede extender de la siguiente manera:

$$P\ (A \text{ y } B \text{ y } C \text{ y } \ldots N) = P(A) \cdot P(B) \cdot P(C) \,..\, P(N)$$

si A, B, C,... N son independientes

También recuerde la regla de la suma:

$$P(A \text{ o } B) = P(A) + P(B) - P\ (A \text{ y } B)$$

donde P (A y B) = 0 para sucesos mutuamente excluyentes.

Esta regla se puede extender:

$$P\ (A \text{ o } B \text{ o } C \text{ o } \dots N) = P(A) + P(N) + P(C) * \dots + P(N)$$

si A, B, C,... N son mutuamente excluyentes

Como se puede considerar que los distintos nacimientos son **independientes** y que un tipo de descendencia es **mutuamente excluyente** de cualquier otro tipo, son aplicables las extensiones de ambas reglas.

Por ejemplo, si formulásemos la pregunta: "dada una pareja cuya probabilidad de tener en un parto un hijo con agenesia de uno o más dientes es de 1/4, cuál es la probabilidad de que el primero de un total de tres hijos tenga hipodoncia (HD) y los dos últimos no (N)¿

$$P\ (1l.\ HD,\ 2o.\ N \text{ y } 3o.\ N) = P\ (HD) \cdot (N) \cdot P(N)$$

$$(1/4)\ (3/4)\ (3/4) = 9/64$$

Fíjese: si p es la probabilidad de HD en un parto y q es la 1-p, la expresión recién escrita es sencillamente:

$$p^1q^2$$

En el ejemplo anterior la pregunta no era: cuál es la probabilidad de tener un hijo con hipodoncia? sino: "cuál es la probabilidad de que únicamente el primer hijo tenga HD?" Para responder a la pregunta: "cuál es la probabilidad de que exactamente un hijo tenga HD?" debemos considerar las formas en que se pueden tener tres hijos de los cuales exactamente uno esté afectado. Este suceso se puede producir así:

A) 1o. HD y 2o. N y 3o. N = (1/4)(3/4)(3/4) = 9/64

B) 1o. N y 2o. HD y 3o. N = (3/4)(1/4)(3/4) = 9/64

C) 1o. N y 2o. N y 3o. HD = (3/4)(3/4)(1/4) = 9/64

La probabilidad de A, B y C es en cada caso p^1q^2, o sean 9/64. La probabilidad de que nazca exactamente un hijo con HD = P(A o B o C) = P(A) + P(B) ∩P(C) pues A, B, y C son mutuamente excluyentes. Esto es lo mismo que $3p^1q^2$, que es el término del desarrollo binomial $(p + q)^3$ que hubiéramos utilizado para contestar la pregunta.

De esta manera, en un término del desarrollo, la parte exponencial es la probabilidad de que se produzca una ordenación en particular y el coeficiente es el número de diferentes ordenaciones que dan la misma distribución general de éxitos y fracasos.

TERMINO GENERAL DE LA FORMULA BINOMIAL

Con la fórmula siguiente podemos calcular cualquier término de cualquier binomio con mucha más facilidad y rapidez que desarrollando por completo el binomio.

La probabilidad de que se produzcan **exactamente** r éxitos en n ensayos está dada por

$$\frac{n!}{(n-r)!\ r!}\ p^r\ q^{(n-r)}$$

cuando en cada ensayo la probabilidad de un éxito es p y la probabilidad de fracaso (1 - p) es q. La notación n! se lee "factorial de n" y es

$$n \cdot (n-1) \cdot (n-2) \cdot (n-3) \ldots 1$$

Por ejemplo $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$ Se define 0! como igual a 1. Además, debe usted recordar que cualquier número elevado a la potencia cero es igual a 1 (ej., $X^0 = 1$) y que en cualquier número elevado a la potencia 1 es igual a si mismo (ej., $X^1 = X$). Recuerde entonces

$$0! = 1$$

$$X^0 = 1$$

$$X^1 = X$$

Ejemplo: si un hombre con dentinogénesis imperfecta (afección que es el resultado de un gene autosómico dominante) se casa con una mujer normal y tiene cinco hijos, cuál es la probabilidad de que exactamente cero de ellos tenga esa afección? En esta familia, la probabilidad de que un niño en particular reciba el gene de la dentinogénesis imperfecta del padre y llegue a desarrollar la enfermedad, es 1/2.

Respuesta:

$p = p$ (éxito en un ensayo, desarrollo de la afección) $= 1/2$

$q = 1 \cdot p = 1/2 = p$ (ser normal)

$n =$ número de ensayos $= 5$

$r =$ número de éxitos (número hijos con la enfermedad) $= 0$

$$\frac{n!}{(n-r)!\ r!} p^r q^{(n-r)} = \frac{5!}{(5-0)!\ 0!} p^0 q^{(5-0)}$$

$$\frac{5 \times 4 \times 3 \times 2 \times 1}{(5 \times 4 \times 3 \times 2 \times 1)\ (1)} \frac{1}{2}^{0} \frac{1}{2}^{5} = 1 \cdot 1 \cdot \frac{1}{2}^{5} = \frac{1}{32}$$

Problema 6:

En la familia anterior, cuál es la probabilidad de que exactamente un hijo de los cinco tenga la afección?

Su solución:

Problema 7:

Considere el problema de FCU que analizamos antes: recuerde que la probabilidad de FCU en un parto era 1/4 y que determinamos la probabilidad de que exactamente un hijo del total de cuatro tuviera FCU, era 108/256. Para obtener esta respuesta teníamos $n = 4$, $r = 1$, $p = p$ (FCU) $= 1/4$, $q = 3/4$. Conteste ahora la misma pregunta (en esta familia de cuatro hijos, cuál es la probabilidad de que exactamente uno tenga FCU?), pero esta vez defina como éxito a un hijo normal.

Su solución:

$n =$

$r =$

$p = p$ (niño normal) $= 3/4$

$q =$

Nota: no importa cómo defina usted el éxito, siempre que obtenga el número exacto de éxitos para contestar la pregunta (por ejemplo, para responder usted necesitará un éxito entre los cuatro hijos si éxito es FCU y tres éxitos entre los cuatro hijos si éxito es niño normal).

Problema 8:

Para una pareja en particular, la probabilidad de que tenga un hijo con amelogénesis imperfecta es 1/4. Esta pareja desea tener tres hijos. Cuál es la probabilidad de que nazca por lo menos un hijo con esta afección del esmalte?

Su solución:

TEST FINAL 7

Después de evaluar una radiografía panorámica, el odontólogo de Ricardo le informó que tenía cuatro terceros molares retenidos y le sugirió que concertara una cita con un cirujano bucal para que se los extrajese. Ricardo, un consumidor informado, se queda preocupado por los problemas que entrañan radiografías y cirugías odontológicas innecesarias. Decide analizar la situación más a fondo antes de consentir en que lo sometan a un procedimiento irreversible.

Suponga que la probabilidad de que uno de estos molares provoque dolor o infección en el futuro es 1/4 y que la probabilidad de que permanezca no afectado es 3/4. También suponga que lo que pase con uno de sus molares es independiente de lo que pase con los otros.

1) Cuál es la probabilidad de que ninguno de los molares de juicio de Ricardo le cause problemas en el futuro?

2) de que lo haga por lo menos uno?

3) de que lo haga exactamente uno?

4) de que lo haga como máximo uno?

5) de que lo hagan los cuatro?

6) de que lo hagan exactamente dos?

DESARROLLO

TEST FINAL No____

CAPITULO 8.

MUESTREO DE UNA POBLACION

INTRODUCCION

Hasta aquí hemos visto cómo se puede describir datos y expuesto el concepto de probabilidad. Avanzaremos ahora un poco más y aprenderemos los métodos para relacionar las mediciones obtenidas de una muestra con la totalidad de una población.

OBJETIVO

Al terminar su trabajo en este capítulo, usted podrá:

1) Obtener una **muestra aleatoria simple** a partir de una población dada, con la ayuda de una tabla de **números aleatorios**.

2) Reconocer que una determinada muestra es **representativa** o **sesgada** y dar las razones para considerarla de ese modo.

3) Reconocer si una variable en particular es **discreta** o **continua**.

MUESTRA DE UNA POBLACION

Al examinar a un nuevo paciente, usted le toma la presión arterial sistólica. Esta información es de escaso valor, a menos que usted pueda relacionar con la distribución de la presión arterial sistólica en una población que se asemeje a su paciente. Utilizando la distribución de la población, usted podrá decidir si la presión arterial sistólica de un determinado paciente está o no fuera de lo normal. Sin embargo, no es posible obtener lecturas de presión arterial sistólica de una población completa y, aunque lo fuera, sería excesivamente costoso. En lugar de ello es común seleccionar un subconjunto de la población y realizar las mediciones en este subconjunto. En lenguaje estadístico, este subconjunto se denomina **muestra**; es importante que la muestra sea representativa de la población de la cual fue obtenida. Cumplido este requisito, es posible hacer enunciados (o inferencias) sobre la población completa a partir de las mediciones obtenidas en la muestra.

Si una muestra no es representativa de la población que interesa, es una muestra **sesgada**. Por ejemplo, n mediciones de prevalencia de caries, los niños en edad escolar que viven en una comunidad con aguas fluoradas serían una muestra sesgada de la totalidad de los niños.

Problema 1:

Un administrador sanitario desea determinar el tipo de prestaciones odontológicas brindadas en un centro de salud de cierto vecindario. Dispone de un empleado el lunes, miércoles y viernes por la tarde, que durante un mes registra todas las prestaciones brindadas en esos momentos.

Es adecuada esta muestra para determinar los tipos de prestaciones que se ofrecen?

Su solución:

MUESTRA ALEATORIA SIMPLE

Al seleccionar una muestra es deseable utilizar algún procedimiento aleatorio (de azar) para establecer qué unidades se incluyen en la muestra. **Una muestra aleatoria simple es aquella en que cada unidad de la población tiene una posibilidad igual e independiente de ser seleccionada.**

Por ejemplo, se puede seleccionar una muestra aleatoria simple de pacientes de una práctica grupal, confeccionando primero una lista de todos los pacientes y numerándolos en orden. Si hubiera 9500 pacientes, los números irían de 1 a 9500. Se puede utilizar entonces una tabla de números aleatorios que se pueden encontrar en muchos textos de estadística, como ayuda para seleccionar la muestra. Si se requiere de una muestra de tamaño 500 (alrededor del 5% del total), se elige al azar un lugar para comenzar a emplear la tabla y se toman conjuntos de 4 dígitos aleatorios para determinar la muestra. Una tabla de números aleatorios generalmente se presenta de la manera siguiente:

83179	98001	36473	87611
92210	06554	32879	19398

o sea, como una serie de renglones de dígitos con varios grupos de cinco dígitos cada uno. Los renglones se pueden dividir en cinco grupos de cuatro dígitos, de la manera siguiente:

8317/9	980/01	36/473	8/7611
9221/0	065/54	32/879	1/9398

Cada conjunto de 4 dígitos determina números en el rango 0.9999. La muestra de 500 se determina leyendo suficiente cantidad de conjuntos consecutivos de cuatro dígitos. Los números superiores a 9.500 no se tienen encuentra. Tampoco se considera el cero (0000) ni, si se repite, cualquier número de cuatro dígitos, pues en caso contrario se podría incluir dos veces en la muestra de un mismo paciente. A partir de los números aleatorios especificados, se incluirán en la muestra los pacientes con los números 8317, 136, 4738, 7611, 9221, 65, 5432, 8791, 9398. Advierta que en este caso no se toma en consideración el segundo conjunto de cuatro dígitos (9980), pues representa un número superior a 9.500.

Si en la población en estudio hubiera más de 10.000 pacientes, digamos 15.000, habría que utilizar grupos consecutivos de cinco dígitos para determinar la muestra. En este caso, de dejarían a un lado los números aleatorios superiores a 15.000.

Nota: al utilizar una tabla de números aleatorios, es costumbre elegir al azar un punto de partida cada vez que se seleccione una muestra. La forma más elemental de determinar este comienzo aleatorio es cerrar los ojos y poner el dedo "al azar" sobre un punto cualquiera de la tabla.

Problema 2:

Suponga que usted tiene interés en las actitudes de pacientes odontológicos registrados con un sistema recién diseñado. Suponga que hasta la fecha se han registrado 848 pacientes y que, para entrevistarlos, se debe seleccionar una muestra aleatoria simple de tamaño 40. Utilice los siguientes números aleatorios para determinar una muestra aleatoria simple de los pacientes por entrevistar. Suponga que el primer dígito corresponde a un comienzo aleatorio.

21648	78245	50328	94107
73186	40021	89707	34583
21054	54581	64310	98980
38219	32728	26434	54329
87373	21019	08372	61486
21324	83649	27384	97101
54872	93612	49581	02001

Su solución:

OTROS TIPOS DE MUESTRAS

Al seleccionar una muestra a base de probabilidades, en la práctica no se utilizan muy a menudo muestras aleatorias simples. Una razón es que una muestra aleatoria simple puede implicar la selección de una muestra no representativa, pues es una muestra aleatoria simple se pueden producir todas las combinaciones posibles, incluidas las no representativas.

La técnica que más comúnmente se utilizan son equivalentes a muestras aleatorias simples restringidas, en las que las restricciones se seleccionan para reducir al mínimo la posibilidad de obtener una muestra no representativa. La muestra más común de este tipo es una **muestra aleatoria estratificada** en la cual, por ejemplo, se adoptan medidas para asegurar que se seleccionen unidades de cada subgrupo de edad, sexo, raza o nivel social. Existen también métodos de muestreo no probabilísticos, como el muestreo por alicuotas, en el cual se seleccionan, por ejemplo, las diez primeras mujeres, los diez primeros operarios industriales, etc., sin considerar el conjunto total que puedan representar.

DISTRIBUCION DE LA POBLACION

Suponga ahora que estamos tomando una muestra de una población muy grande (por ejemplo, personas que viven en los Estados Unidos de América). Si se aumenta el tamaño de la muestra aleatoria de una población grande y se toman intervalos de clase cada vez más pequeños para construir un histograma de una variable en particular, el histograma se irá convirtiendo paulatinamente en una curva continua.

La línea continua representa la **distribución de la población.** Por ejemplo, un histograma que describa la presión arterial sistólica de una muestra de hombres jóvenes y sanos podría verse así:

Figura 8.1

A medida que aumenta el tamaño de la muestra, disponemos de datos suficientes para utilizar intervalos más pequeños, y entonces el histograma podría verse así:

Figura 8.2

Cuando el tamaño de la muestra es muy grande, el histograma se aproxima a una curva continua.

Figura 8.3

Si X es la viable que se ha medido (en nuestro ejemplo X es la presión arterial sistólica), la curva continua describe la distribución poblacional de la variable X. Es importante comprender que las distribuciones de la poblaciones serán curvas continuas solamente en el caso de variables que se puedan medir, como la presión arterial, la altura, el peso o el nivel de colesterol. En terminología matemática se las llama **variables continuas**, pues se pueden medir con el grado de exactitud que se quiera. La única restricción en materia de exactitud está impuesta por el equipamiento utilizado y por la capacidad del técnico para manejarlo en forma apropiada.

Las variables continuas difieren de las **variables discretas** en que estas últimas únicamente pueden tener un conjunto restringido de valores. Por ejemplo, el número de asistentes empleados en una clínica es una variable discreta, pues las únicas posibilidades admisibles son 0, 1, 2, 3, ... , hasta algún máximo razonable, digamos 10. Las variables discretas se describen en tablas y sus distribuciones poblacionales nunca pueden ser curvas continuas. Los polígonos de frecuencia de estas variables son simple quebrados.

Cabe señalar que las variables continuas se pueden convertir en variables discretas si usted, el analista de los datos, lo desea. En efecto , así lo hace el histograma de una variable continua. La presión arterial sistólica es una variable (medición) continua, pero se puede tratar como discreta seleccionando intervalos de clase y asignando las mediciones a los intervalos apropiados. No es posible proceder a la inversa: las variables originalmente discretas no se pueden convertir en variables continuas.

DISTRIBUCION NORMAL

Hay una distribución de las poblaciones que se observa con mucha frecuencia en las ciencias biológicas. Varias distribuciones de poblaciones corrientes se aproximan mucho a esta distribución, que se llama **distribución normal**. También se la conoce como **distribución Gaussiana**, en honor del famoso matemático J. F. C. Gauss. La distribución normal tiene forma de campana y es simétrica con respecto a su centro. En la figura 8.4 se muestra una típica distribución normal.

Figura 8.4

Un aspecto interesante de la distribución normal es que depende tan solo de los parámetros la media μ (mu) y la desviación estándar σ (sigma) de la población (véase capítulo 5). La desviación estándar es la raíz cuadrada de la varianza. En los trabajos estadísticos, los parámetros de la población se representa generalmente con letras griegas y deben distinguirse de la media y la varianza que se calcularía a partir de la muestra. En el Capítulo 5, la media se representó como x y la desviación estándar con la s, utilizando letras del alfabeto moderno. No son ellas parámetros de la población, sino cantidades que surgen de un conjunto de datos que representan solo una muestra de la población. En el Capítulo 10, usted podrá ver que las cantidades (como x y s) derivadas de datos de una muestra se utilizan para estimar parámetros de la población (como μ y σ).

Volviendo ahora a la distribución normal, la media μ establece la ubicación de la "joroba" de la curva y la desviación estándar σ determina cuánta dispersión o cuánta concentración existe en la distribución. La Figura 8.5 muestra distribuciones normales con las mismas medias, pero con diferentes desviaciones estándar.

Figura 8.5

Cuanto más aplanada es la curva, tanto mayor es la desviación estándar, es decir, tanto mayor es la variabilidad en el conjunto de datos.

La Figura 8.6 muestra tres distribuciones normales con la misma desviación estándar (σ), pero con diferentes medias μ_1, μ_2, μ_3.

Figura 8.6

En el próximo capítulo usted verá cómo se calculan probabilidades a partir de la distribución normal, utilizando tablas estadísticas.

REVISION

1) Una **muestra** es un subconjunto de la población que se está investigando.

2) Una **muestra representativa** es un subconjunto de la población que refleja sus características generales.

3) Una **muestra sesgada** es un subconjunto de la población que no es representativa de la población en su conjunto.

4) Para obtener una **muestra aleatoria simple** primero se decide el tamaño de la muestra. Luego se confecciona una lista en la que se asigna a cada miembro de la población un número consecutivo, comenzando por el 1. Luego, utilizando una tabla de números aleatorios, se seleccionan números aleatorios consecutivos para determinar cuáles se van a incluir en la muestra; no se toman en cuenta los números que se repiten, los mayores que el tamaño de la población ni el cero.

Recuerdo que, en la práctica, las muestras aleatorias simples no se utilizan con frecuencia, pues en este tipo de muestra, en la que pueden aparecer todas las combinaciones posibles, incluidas las no representativas, es posible seleccionar una muestra que no sea representativa. La técnica que más comúnmente se pone en práctica consiste en utilizar la muestra aleatoria estratificada, en la que se eligen los estratos para disminuir la posibilidad de no tener representatividad.

5) **Variables continuas** son la variables que se pueden medir con el grado de exactitud que se quiera. Las únicas restricciones son las derivada de la precisión del equipamiento empleado y de la habilidad del técnico para utilizar los instrumentos de medición. Son ejemplos de estas variables la presión arterial, la altura, el peso y el nivel de colesterol. En el caso de estas variables, la distribución de la población se puede describir con curvas continuas.

6) **Variables discretas** son las que solo pueden tener un conjunto restringido de valores: por ejemplo, el número de dientes presentes o ausentes. Una variable discreta está limitada en cuanto al número de categorías por algún máximo razonable, por ejemplo, 0, 1, 2, ... hasta 32 dientes. En el caso de estas variables, las distribuciones de las poblaciones no son curvas continuas; los polígonos de frecuencia que representan variables continuas son siempre quebrados.

7) La **distribución normal** es una distribución de población que aparece con mucha frecuencia en las ciencias biológicas. Adopta una forma de campana, simétrica con respecto a su centro, y depende únicamente de dos parámetros:

a) la media de la población, (mu), simbolizada por μ.

b) la desviación estándar de la población, (sigma), simbolizada por σ

TEST FINAL 8

Una higienista dental, que ejerce en un hospital de la Administración para los Veteranos de los Estados Unidos en América, tiene interés en desarrollar un programa preventivo y tratamiento no quirúrgico precoz destinado a todos los potenciales pacientes periodontales del centro médico. Considera que necesita alguna información preliminar acerca de los pacientes con riesgo de contraer periodontitis. Le pide a un amigo que enseña bioestadística en una facultad cercana que la ayude a obtener esa información preliminar. El profesor evalua la situación y luego encomienda a un estudiante ya graduado el desarrollo de las tareas pertinentes. El estudiante selecciona una muestra aleatoria simple de 20 pacientes de una lista de 300 que tiene 55 a 80 años de edad y que estaban siendo atendidos en el Hospital. La lista se había confeccionado pocos meses antes para llevar a cabo un estudio de la relación entre estrés y enfermedad periodontal.

Suponga usted que es el estudiante ya graduado que practica la investigación. En las próximas páginas encontrará la lista de los 300 hombres. En la página que sigue a estas dos hay una reproducción de una tabla de números aleatorios en la que está delimitada por parte de esos números.

1) Tome una muestra aleatoria simple de 20 pacientes de la población de 300, utilizando la tabla de números aleatorios. Suponga que el primer dígito de la parte

delimitada corresponde a un comienzo aleatorio en la tabla. Confeccione la lista de números de 20 pacientes seleccionados para formar la muestra.

2) Recuerde que la higienista estaba interesada en todas los potenciales pacientes periodontales. Es la muestra representativa de todos los potenciales pacientes periodontales?

3) Cuando un estudio se refiere a variables como, por ejemplo, el número de personal auxiliar en los servicios hospitalarios privados en Massachusetts, E. U. A:, qué tipo de variable es la que se considera?

4) Al estudiar presión arteria, altura y peso, qué tipo de variable se está considerando? Existe algún tipo de restricción en cuanto a la exactitud de las mediciones?

LISTA DE 300 PACIENTES DE 55 A 80 AÑOS DE EDAD EN UN HOSPITAL DE LA ADMINISTRACION PARA LOS VETERANOS, E.U.A.

1. James Adams
2. Elwin Adderton
3. William Agee
4. Larry Aiken
5. Frank Albertson
6. Carl Albritton
7. Floydd Aldridge
8. James Alford
9. Bob Allen
10. Kevin Allen
11. Junior Allgoodd
12. Bob Alston
13. William Alston
14 Willie Alston
15. Jose Alvarez
16. Bill Amos
17. Hardol Amos
18. Chesster Anderson
19. Eugene Babb
20. Norton Bacon
21. Chestar Baffa
22. Josseph Bailes
23. Ross Bailey
24. Thomas Bain
25. Donald Bain
26. Phil Baisey
27. Fred Baker
28. Timothy Baker
29. Leroy Baker
30. Len Baker
31. Carl Baldwin
32. Dale Baldwin
33. Billy Balkcum
34. Eller Ball
35. Henry Bame
36. Lance Bame
37. Darrell Cable
38. Barnet Cagle
39. Bill Cain
40. Cale Cain
41. James Calhoun
42. Walter Calloway
43. Ted Calomiris
44. Carmen Cambareri
45. Bob Cameron
46. Wilbert Camp
47. Haskell Campbell
48. Floyd Campbell
49. George Campbell
50. Bobby Canada
51. Elmer Canada
52. Milton Cannon
53. Jimmy Canoy
54. William Dabbs
55. Francis Dailey
56. Robert Dailey
57. Veirgis Damron
58. Richard Daniels
59. Mel Daniels
60. Joe Daughtridge
61. San Eaddy
62. Harry Earle
63. Virgil Early
34. Merle Earp
65. Larry East
66. Egbert Easter
67. Clearence Eaton
68. Kyle Eccles
69. Edwared Echerd
70. Lewis Eckert
71. John Edge
72. Baxter Fain
73. Clifford Fain
74. Niles Faricloth
75. Williw Faircloth
76. Lou Fallenstein
77. Azlee Falls
78. Andy Farlow
79. Arnold Farlow
80. Howard Farlow
81. Clarence Farmer
82. Daniel Geenster
83. Brady Felst
84. Lister Gabriel
85. David Gadd
86. Eddy Gaddy
87. John Gaddy
88. Ross Gaddy
89. Richmond Gage
90. Riley Gahagan
91. Conrad Gailey
92. Harold Gaines
93. Simon Gaines
94. Leonard Gallimore
95. Stokes Gallimore
96. Linnerl Haas
97. Bob Kackler
98. Ealt Hackler
99. Neil Hadley
100. Mickey Hager
101. GLen Hagie
102. Val Haley
103. Daniel Hair
104. Ben Hair
105. Francis Hairston
106. Oscar Haithcock
107. Royce Haithcock
108. Roy Iddings
109 Gary Iddins
110. Charles Idol
111. Edd Idol
112. Osca Ijames
113. Dennis Ijames
114. marvin Iked
115. Lauren Iked
116. Eugene Ikner
117. Davis Ilderton
118. Leroy Ingle
119 Pat Ingle
120. Horace Jackson
121. Albert Jackson
122 Dan Jackson
123. David Jackson
124. Billy Jacobs
125. Cal Jacobs
126. Robert Jacobs
127. Earton James
128. Baxter James
129. William James
130. Claude Jarrell
131. Jim Jarrell
132. Julian Kabat
133. Richard Kalte
134. Gary Kanney
135. Himler Kanter

136. Matty Kasias
137. Michael Kats
138. Mitchell Keahey
139 Amos Kearns
140. Bond Kearns
141. Keywood Keaton
142. Edwin Keenan
143. Clarence Keever
144. James Lacey
145. Heath Lacey
146. Donald Lackey
147. Lester Lackey
148. David Lacy
149. Zeke Lacy
150. Albert Larkins
151. Paul Larkins
152. Charles Lambert
153. Billy Lambert
154. Robert Lamb
155. Zack Lamb
156. John Lampe
157. Jared Land
158. David Maas
159. Heck Maas
160. Dewey Mabe
161. Enoch Mabe
162. Garry Mabry
163. Vern Mabry
164. Louis MacAluso
165. Clay Mack
166. Alan MacKeraghan
167. Jen MacKeranghan
168. Tyree Nabors
169. Bert Nalley
170. John Nalley
171. Bart Nance
172. jack Nance
173. Robert Nance
174. Roger Nance
175. Geroge Nash
176. Paul Nash
177. Hal Newman
178. Jin Newman
179. Nick Newsome
180. Charles Oakely
181. Peter Oakley
182. Patrick O'Brien
183. Niles Ockershauser
184. Jerry Odell
185. Mark Odell
186. James DOdum
187. Gilbert Odum
188. Helbert Oglesby
189. Adam O'Ham
190. George O'Ham
191. Mack O'Kelly
192. Mel Pace
193. Joseph Packer
194. Edgar Pargett
195. Fate Padgett
196. Ernest Page
197. Vance Page
198. Bil Palmer
199. Gene Pardue
200. Donad Parham
201. Jay Paris
202. Frank Paris
203. Andrew Parker
204. Frank Queen
205. Jack Queen
206. Lonnie Quesenberry
207. Alan Quesenberry
208. Cleveland Quick
209. Cal Quick
210. Elijah Quicker
211. Max Quicker
212. Stancul Quigley
213. Martin Quigley
214. Hall Quinn
215. Burton Quinn
216. Mortin Rabban
217. Ich Rabon
218. Ronnie Rabon
219. Rob Raby
220. James Raby
221. Karl Ragan
222. Helbert Ragland
223. Lon Ragsfale
224. Thomas Rahsdale
225. Phil Rains
226. Floyd Rains
227. Macy Rakocy
228. Grady Saferigth
229. Leon Sarif
230. Bennie Saintsing
231. Bernei Sale
232. Nelson Sale
233. Stokes Salley
234. Terry Salley
235. Bill Saltes
236. Allie Salway
237. Mose Samet
238. Nestor Samet
239. Edward Sampson
240. Izzy Sampson
241. Bobby Sampsson
242. Edwin Samn
243. Jacob Sams
244. Andy Samuel
245. Robert Samuel
246. Joseph Sanborn
247. James Sanborn
248. Jack Sanford
249. Robert Sandman
250. Greg Saunders
251. Frazier Saunders
252. Fouglas Sawyer
253. Kenneth Sawyer
254. Chuck Scarboro
255. Willie Sacarbono
256. Alden Scarce
257. Leonard Scarce
258. Allard Schnell
259. Gerald Schoen
260. Carroll Stirewalt
261. Joe Stockton
262. Clement Stokes
263. james Stokes
264. Tom Stone
265. Robert Stone
266. Igor Stout
267. James Stout
268. Milt Stratton
269. Rolland Ddtratton
270. Jerry Streeter
271. Charles Stribling
272. Ed Strickland
273. Clarence Strickland
274. Bertie Stround
275. James Tabor
276. Silas Tackett
277. Miles Talbert
278. Francis Talton
279. Harry Tanner
280. Mittie Tate
281. Toger Tate
282. Arnie Taylor
283. Buddy Taylor
284. Gary Taylor
285. Carlyle Teague
286. Royce Teague
287. Clyde Teal
288. Ton Teal
289. James Tedder
290. Desmond Teeter
291. Joe Temple
292. Elton Terrell
293. Tom Terrel
294. Phillip Terry
295. Randall Tharp
296. Rhodes Tharp
297. Branks Thayer
298. Ed Thomas
299. Frank Thomas
300. Paul Thomas

TABLA DE 3.000 DIGITOS ALEATORIOS

59659	72852	06129	94308	05749	19153	82460	00478	75805	32009	23875	68677
82411	90292	30706	97605	48608	54015	51890	17301	99782	19015	57317	68013
13843	93995	51883	58522	61155	71399	58635	85023	51650	76718	21919	73513
10435	78326	64618	93437	07475	30624	83191	56187	35325	94395	73433	94676
53134	66704	32257	59501	47700	00009	06281	98418	76171	96940	09059	69417
72003	34132	73190	69210	97567	74020	66713	12669	93589	94662	43405	10066
28514	51352	08530	95539	05602	63012	18957	88944	94452	93122	64781	06514
12917	97462	25726	94744	35448	13042	50943	79709	67889	29609	62313	79279
83373	39245	92243	97913	17124	85041	58635	45813	53811	37392	24970	79927
07267	86897	51499	43273	18892	82216	80236	57612	14001	62325	38735	84612
76588	26923	31054	60218	17345	39602	59162	57832	84523	41042	11296	30632
49240	63281	04614	33392	07001	84306	16429	53471	25818	63352	46924	04871
35752	04189	97171	14157	74892	95415	33440	30640	51485	68458	10328	33676
64619	77118	41617	20554	18276	44883	99415	16997	53821	94097	75026	81127
88690	61236	33374	60728	14229	29563	34529	88056	68232	68392	76796	56893
79461	84305	43579	72883	38477	67290	94245	99781	10942	80435	00921	14067
80902	36897	95640	36470	62245	03058	08547	50958	83273	19123	09752	08407
01151	21104	97527	58900	05404	60173	80832	51058	85639	39027	46466	51809
37968	64462	56075	08893	37611	99308	04329	93183	30336	21644	58715	38198
23305	26085	92374	32990	26363	91711	69860	05770	29751	97144	07013	86938
59247	16366	96189	97470	30214	44644	98059	22574	91620	60000	23344	37444
36500	60822	90378	97279	23465	08334	02857	55667	47936	56996	89202	83199
70075	54440	64481	84409	41834	60664	47231	82181	43874	30630	26199	45896
86342	24223	44496	59162	48388	74147	98995	81341	34897	34988	36418	02624
84311	88398	64927	44436	58368	20076	99158	96858	50086	65045	38481	07525
54881	63813	61062	90259	29431	43074	28667	69796	93533	22809	99394	69098
14518	41841	98054	66992	22445	17229	43071	14225	88268	00541	71165	14783
50638	67367	02269	41450	41821	68571	23399	93457	89750	47769	21497	89342
69344	86839	20010	07328	29007	80903	54847	30916	17892	82137	53555	20075
47913	28647	19802	76446	97037	70494	69812	09447	32525	42213	09634	72966
26328	57898	18788	24756	85277	52756	63983	56075	33968	14871	02049	60706
50294	16119	83076	90665	84829	49972	46220	16659	46931	92976	65504	06336
83268	99346	59662	48680	42779	03996	94530	98456	01256	74255	91891	12575
07013	87702	50796	70856	97815	32459	12438	76268	47025	97690	79257	36095
46765	50390	54010	78801	07072	98988	33444	99065	94077	10472	23690	92832
08095	47064	95976	87412	70468	75482	08605	29299	35481	73600	64300	33511
60885	24370	73993	25812	84663	70534	56393	53047	35756	69593	79344	90015
27015	06788	34710	77319	00486	64682	87293	41380	21570	14220	87665	74506
00315	92095	71436	53910	71825	35262	29177	02301	62587	61783	55570	55539
53950	65678	20747	31728	09918	25385	77090	01288	94658	74075	88671	49665
99690	59155	81227	47763	47492	04772	94845	40646	56216	21067	42450	19508
66717	11954	01415	45326	51054	41341	34358	85431	82252	87605	36601	06658
26188	30485	81426	23423	23594	09460	41692	05812	82406	15391	36229	17837
90362	41289	37152	07100	23518	50267	32448	62141	01385	36778	27664	59578
82823	69136	09321	04700	63892	27441	67295	65525	52824	23557	96270	69621
42577	06630	63342	07561	50529	70119	79646	51437	92659	13088	04551	00541
23822	36622	50194	18562	02049	10746	05935	51914	76822	08605	11330	73731
32678	80671	56245	21450	03647	04769	21619	45623	61887	83380	85757	41796
23175	01870	06887	58787	91034	07651	98761	15623	11588	16364	82436	89947
10889	22960	89466	62824	01679	49788	03503	79381	92710	16860	46220	73179

HOJA DE RESPUESTAS
LABORATORIO N.____

PROBLEMA No 1

PROBLEMA No 2

PROBLEMA No 3

PROBLEMA No 4

PROBLEMA No 5

PROBLEMA No 6

PROBLEMA No 7

PROBLEMA No 8

PROBLEMA No 9

PROBLEMA No 10

DESARROLLO

TEST FINAL No____

CAPITULO 9.

LA DISTRIBUCION NORMAL

INTRODUCCION

En el último capítulo le hablamos de la **distribución normal**, cuya forma le mostramos de la manera siguiente:

Figura 9.1

Aunque la distribución normal es solo una de las muchas clases de distribución continua de probabilidades, tiene una aplicación muy amplia: quizás es la de más aplicación en el campo de las ciencias biológicas. Por consiguiente, es muy importante que usted comprenda en qué consiste y que pueda calcular las probabilidades que se asocian con ella.

OBJETIVO

El objetivo global de este capítulo es calcular la probabilidad de que un suceso o fenómeno se encuadre dentro de un rango específico de valores. Los valores que se obtienen están regidos por una distribución normal con media (μ) y desviación estándar (σ). Estas probabilidades se calculan utilizando una distribución normal estándar. En particular usted podrá:

1) Calcular probabilidades usando la distribución normal estándar.

2) Convertir un interrogante sobre una distribución normal en el interrogante correspondiente a la distribución normal estándar.

3) Calcular las probabilidades de cualquier distribución normal.

LA DISTRIBUCION NORMAL ESTANDAR

La **distribución normal estándar** es la distribución normal en la que la **media es cero y la desviación estándar es uno**: $\mu = 0$, $\sigma = 1$. Se ha hecho costumbre emplear la letra mayúscula Z para designar la variable asociada con esta distribución. Z es el nombre de la variable y la letra minúscula z representa el valor que asume Z. Z puede tener cualquier valor positivo o negativo, y la probabilidad de que asuma determinados valores en particular está regida por la forma de la curva de la distribución normal estándar.

Figura 9.2

Se han preparado tablas que nos permiten leer las áreas que están por debajo de la curva normal estándar. Se confeccionan diversas tablas con distintos formatos. Al final de este capitulo se incorporan una que se utiliza con frecuencia. Se la llama la distribución normal estándar, y da el área que está a la izquierda de un punto sobre el eje z para gran número de valores de z.

Figura 9.3

Por ejemplo, la probabilidad de que Z sea igual o menor que 0,5, es decir, el área por debajo de la curva normal estándar que está a la izquierda de z = 0., se puede hallar entrando z = 0,5 y leyendo 0,6915 en la tabla.

Figura 9.4

Problema 1:

Pruebe usted, cuál es la probabilidad de que Z tenga un valor que no supere -1,2?

Su solución:

Por favor, tome nota de que para buscar un valor negativo de z en la tabla incluida en el final de este capítulo, se considera el valor de z como si fuera positivo y luego se resta de la unidad el valor de probabilidad encontrado en esa tabla.

Recuerde que una probabilidad no solo da una idea de hasta dónde es posible que un individuo de la población elegido aleatoriamente tenga una propiedad determinada, sino que también indica la proporción (o fracción) de la totalidad de la población que tiene esa propiedad. Por ejemplo, en una población normal estándar, el 40% de sus miembros tienen mediciones de z iguales o menores que -0,253.

Figura 9.5

Observe que tener el 40% de la población normal estándar con valores de z o iguales o menores que −0,253 significa que también el 60% restante tiene valores de z mayores que −0,253.

Problema 2:

A partir de su propio ejemplo, determine la fracción de una población normal estándar con medidas superiores al 2,0.

Su solución:

Ahora ya puede comenzar a entender la forma en que maniobramos para obtener la probabilidad de cualquier intervalo entre valores de z. Utilizamos la tabla para obtener el área total a la izquierda de cualquier valor de z que queramos y restamos las áreas que sean necesarias para obtener el área que buscamos. Por ejemplo. hallemos la probabilidad de que Z esté entre −0,4 y + 1,3. La tabla y el diagrama siguiente nos dan la respuesta.

Figura 9.6

El área total al a izquierda de + 1,43 es 0,9032 y el área total a la izquierda de −0,4 es 0,3446. Por lo tanto, el área entre los dos valores debe ser 0,9032 −0,3446 = 0.5586.

Problema 3:

Usted puede obtener ahora el valor correspondiente a la probabilidad mostrada por el área sombreada de la Figura 9.2: P (0,5 < Z < 1,5).

Su solución:

Ahora usted puede obtener probabilidades a su voluntad en la distribución normal estándar. Solo necesita aprender una pequeña fórmula de conversión que transforma cualquier distribución normal en una Z estándar normalmente distribuida. La teoría matemática ha demostrado que si X tiene la distribución normal con media μ y desviación estándar σ, la fracción $(X - \mu)/\sigma$ tiene la distribución normal estándar; en otras palabras, podemos utilizar la fórmula de conversión:

$$\frac{(X - \mu)}{\sigma} = Z$$

Antonio, un joven paciente, llega a la clínica de un ortodoncista con su madre. La señora está preocupada porque los dientes de su hijo están apiñados y desea saber si necesita aparatos. Antes de formular un diagnóstico, el ortodoncista desea tomar impresiones de los dientes del paciente, examinar los modelos resultantes y analizar varias vistas radiográficas de su dentición y cráneo. Es necesario medir cuidadosamente el ancho de los dientes permanentes y temporarios de Antonio antes de decidir si es apropiado desarrollar un programa de extracciones seriadas.

En un artículo reciente de la publicación ortodóntica del Dr. Y. R. Bender, se ha informado que la distribución normal de las medidas del ancho de los incisivos centrales superiores pueden representarse por una medida de 8,5 mm y una desviación estándar de 0,5 mm. Si llamamos X = ancho del incisivo central, tenemos $\mu = 8{,}5$, $\sigma = 0{,}5$ y la fórmula de conversión es:

$$\frac{X - 8{,}5}{0{,}5} = Z$$

Figura 9.7

Si un incisivo central midiera 9,5 mm, diría usted que es un diente particularmente ancho? Qué fracción de los incisivos centrales alcanzan un ancho de 9,5 mm? EN otras palabras, cuál es P $(X > 9{,}5)$?

Si fuera esta una pregunta con respecto a Z, usted podría consultar directamente la tabla y arribar en un momento a la respuesta buscada. Ahora tenemos que hacer una conversión, sabiendo que la fórmula pertinente nos dice que obtenemos una Z cada vez que restamos μ de X y dividimos por σ. En nuestra pregunta sobre el tamaño del diente, tanto X como 9,5 son valores de X, de manera que haremos su conversión. También recuerde que, en este caso, $\mu = 8{,}5$ y $\sigma = 0.5$.

Así, tenemos lo siguiente:

$$P\ (X > 9{,}5)$$

$$= P\ \frac{(X-8{,}5)}{0{,}5} > \frac{(9{,}5-8{,}5)}{0{,}5}$$

$$= P\quad Z > \frac{(9{,}5\ -\ 8{,}5)}{0{,}5}$$

$$= P\quad Z > \frac{1{,}0}{0{,}5}$$

$$= P\ (Z > 2{,}0)$$

A partir de aquí recurrimos a la tabla y encontramos la respuesta en el lugar correspondiente:

$$P\ (Z >Z\ 2{,}0) = 1 - 0.9773 = 0.0227$$

Es decir, el 2,27% de los incisivos centrales tienen en esta población 9,5 mm o más de ancho. Si se trata de un paciente adulto con necesidad de una prótesis y los fabricantes de dientes artificiales estimaran que este diente es inusualmente grande, usted podría tener dificultades para conseguir el tamaño apropiado para este paciente.

Problema 4:

Usted obtuvo una nota 86 en un test final de BIOESTADISTICA. Se puede considera que los puntajes de su clase están distribuidos en forma normal con una nota media de 82 y una desviación estándar de 8. Dónde se ubica usted en esta clase con respecto a su comportamiento en esta prueba?

Su solución:

Problema 5:

Cuál es la probabilidad de que un estudiante de la clase seleccionado al azar haya obtenido una nota menor de 70?

Su solución:

Es tan solo un poco más trabajoso manejarse con intervalos cerrados en ambos lados. Cuál es la probabilidad de que un estudiante de la clase seleccionado al azar haya mencionado una nota entre 66 y 94?

$$P\ (66 < X < 94)$$

$$= P \quad \frac{(66 - 82)}{8} < \frac{(x - 82)}{8} < \frac{(94 - 82)}{8}$$

$$= P \quad \frac{-16}{8} < Z < \frac{12}{8}$$

$= P\ )-2,0 < Z > 1,5)$
$= 0,9332 - 0,0227 = 0,9105$

Figura 9.8

Problema 6:

En cierta facultad de odontología, el aumento de peso que experimentan los estudiantes entre su primer día de clase y su graduación está normalmente distribuido con una media de 5 libras y una desviación estándar de 2 libras.

Cuál es la probabilidad de que un estudiante seleccionado al azar aumente entre 6 y 11 libras?

Qué proporción de estudiantes pierden peso durante los años que pasan en la facultad de odontología.

Su solución:

DESVIACIONES ESTANDAR

Un último item es útil recordar la proporción de mediciones que se encuadran dentro de una, dos o tres desviaciones estándar de la media en una distribución normal con media μ y desviación estándar σ.

Figura 9.9

En una distribución normal son válidos los siguientes enunciados:

1) el 6826% (= 34,16% + 34,16%) de la población está dentro de una desviación estándar de la media.

2) El 95,46% (O 34,12% + 34,13% + 13,60% + 13,60%) de la población está dentro de dos desviaciones estándar de la media.

3) El 99,74% (= 34,13% + 34,13% + 13,60% + 13,60% + 2,14% + 2,14%) de la población está dentro de tres desviaciones estándar de la media. Por lo común, el enunciado 2 se expresa en forma aproximada:

El 95% de una población normal está dentro de dos desviaciones estándar de la media.

Los enunciados 1, 2 y 3 se pueden visualizar fácilmente si nos remitimos a la tabla de la distribución normal estándar.

Consideremos una distribución normal con media μ y desviación estándar σ. Los valores que pueden asumir los individuos de la población están representados por X.

Enunciado 1:

Queremos que X esté en el rango $\mu - \sigma$, $\mu + \sigma$

$$P \quad (\mu-\sigma) < X < (\mu+\sigma)$$

$$= P \quad \frac{(\mu-\sigma) - \mu}{\sigma} < Z < \frac{(\mu+\sigma) - \mu}{\sigma}$$

$$= P \left(\frac{-\sigma}{\sigma} < Z < \frac{\sigma}{\sigma} \right)$$

$$= P\ (-1 < Z < 1)$$

a) $= 0{,}8413 - 0{,}1587 = 0{,}6826$

o sea que el 68,26% de la población está dentro de una desviación estándar de la media.

Problema 7:

Demuestre que los enunciados 2 y 3 son correctos.

Su solución:

b) $(-2)\ Z = (2)$ | c) $(-3) = Z = (3)$

0.0227 0.9773 | $0.0013 = Z = 0.9987$

0.9546 | 0.9974

TEST FINAL 9

Una encuesta de odontólogos de cierto estado demostró que el número de horas semanales en contacto con pacientes estaba normalmente distribuido con una media de 33,5 horas y una desviación estándar de 2,5 horas.

1) Qué proporción de esta población de odontólogos dedica entre 30 y 35 horas/semana a atender pacientes?

2) Qué proporción de esta población de odontólogos está por lo menos 40 horas/semana en contacto con pacientes?

LA DISTRIBUCION NORMAL ESTANDAR

z	P(Z<z)	z	P(Z<z)	z	P(Z<z)	z	P(Z<z)
0.0000	0.5000	0.8416	0.8000	1.6000	0.9452	2.4000	0.9918
0.1000	0.5398	0.9000	0.8159	1.6450	0.9500	2.5000	0.9938
0.1257	0.5500	1.0000	0.8413	1.7000	0.9554	2.5760	0.9950
0.2000	0.5793	1.0364	0.8500	1.7510	0.9600	2.6000	0.9953
0.2533	0.6000	1.1000	0.8643	1.8000	0.9641	2.7000	0.9965
0.3000	0.6179	1.2000	0.8849	1.8810	0.9700	2.8000	0.9974
0.3853	0.6500	1.2816	0.9000	1.9000	0.9713	2.9000	0.9981
0.4000	0.6554	1.3000	0.9032	1.9600	0.9750	3.0000	0.9987
0.5000	0.6915	1.3410	0.9100	2.0000	0.9773	3.0900	0.9990
0.5244	0.7000	1.4000	0.9192	2.0540	0.9800	3.2000	0.9993
0.6000	0.7257	1.4050	0.9200	2.1000	0.9821	3.2910	0.9995
0.6745	0.7500	1.4760	0.9300	2.2000	0.9861	3.4000	0.9997
0.7000	0.7580	1.5000	0.9332	2.3000	0.9893	3.6000	0.9998
0.8000	0.7881	1.5550	0.9400	2.3260	0.9900	3.7190	0.9999

Para cantidades negativas $-z$, utilice $P(Z<-z) = 1 - P(Z<z)$

HOJA DE RESPUESTAS
LABORATORIO N.____

PROBLEMA No 1

PROBLEMA No 2

PROBLEMA No 3

PROBLEMA No 4

PROBLEMA No 5

PROBLEMA No 6

PROBLEMA No 7

PROBLEMA No 8

PROBLEMA No 9

PROBLEMA No 10

CAPITULO 10.

ESTIMACION

INTRODUCCION

Si quisiéramos saber cuál es el estado global de higiene bucal de cierta población que vive en una remota islita del Pacífico, podríamos en teoría, examinar a la totalidad de la población, determinar su puntaje en un índice como el simplificado de higiene bucal, IHB-S, y calcular la media y otra medida apropiada de tendencia central como las que analizamos en el Capítulo 5. Sin embargo, esto sería totalmente impracticable su estuviésemos estudiando una gran población, como podría ser la de los Estados Unidos de América. En cambio, trataríamos de seleccionar y examinar una muestra que fuese representativa de la población total. Si seleccionásemos más de una muestra, es probable que cada una de ellas fuera ligeramente diferente, con su propia media y su propio desvío estándar. Afortunadamente, podemos soslayar este problema y aún así extraer inferencias acerca de la población total a partir de nuestra única muestra, en virtud del resultado matemático del llamado **teorema del límite central**.

OBJETIVOS

1) Usted podrá enunciar el teorema del límite central.

2) Usted podrá calcular la media y el error estándar de la media en cualquier muestra de observaciones tomadas a una población.

3) Usted podrá calcular e interpretar intervalos de confianza.

TEOREMA DEL LIMITE CENTRAL

Suponga que de una población pudiésemos seleccionar todas las muestras aleatorias posibles de tamaño n y calcular la media x, de cada una. Utilizando las medias calculadas para cada una de estas muestras, las x podríamos calcular una media global. La media de todas las medias de las muestras es equivalente a la media de la población total:

$$X_x = \mu$$

La desviación estándar de todas las x es $\sigma/\sqrt{n}$, donde "n" se refiere al tamaño o número de observaciones de cada muestra. Se la llama **error estándar** de la media de las muestras:

$$SE\ (x) = \sigma_x = \sigma/\sqrt{n}$$

El error estándar de la media de las muestras depende de la variabilidad que se encuentre en la población y del tamaño de la muestra. Sin embargo, la población considerada no necesita tener una distribución normal para que expresemos la distribución de las medias de las muestras en términos como los que ahora hemos estado empleando para describir una población normal. Las observaciones individuales, las x, pueden tener cualquier tipo de distribución de frecuencias. La distribución de las medias de muestras de la población parecerá una distribución normal. A medida que aumente el tamaño de la muestra, n, se asemejará cada vez más a la distribución normal.

El **teorema del límite central** dice que la distribución de las medias, x, de muestras de tamaño n tomadas de una distribución de casi cualquier forma con media μ y desviación estándar σ, se aproxima a una distribución normal con media μ y desviación estándar $\sigma/\sqrt{n}$.

Figura 10.1

Como investigadores, no queremos tomar muchas muestras de una población. Nos gustaría tomar una sola. Cómo nos ayuda el teorema del límite central?

En capítulos anteriores usted aprendió a calcular la varianza muestral, s^2, y la media, x, de una muestra aleatoria simple que contiene n observaciones. La varianza verdadera, σ^2/n, de la distribución aproximadamente normal de las medias de muestras, se puede estimar en función de s^2/n.

Esta varianza muestral de la distribución de las medias de la muestra, que se puede calcular a partir de los datos reunidos en una sola muestra, se simboliza con s_x^2. Por tanto, la varianza aproximada de las medias de las muestras $= s_x^2 = s^2/n$. De modo similar, el error estándar aproximado de la media basado en una muestra y no tomado de la población total es:

$$SE = s_x = \sqrt{(s^2/n)} = s/\sqrt{n}$$

Utilizando los resultados indicados, estamos en condiciones de decir algo con respecto a la distribución de medias de otras muestras de este tamaño, sin tomar realmente muestras adicionales.

Problema 1:

Si se toman 25 muestras de la población de Boston, Massachusetts, E. U. A., cada una con 400 individuos, y se registran los puntajes de higiene bucal de estos individuos, cuál sería la distribución de los puntajes de IHB-S de estas 25 muestras?

Su solución:

Problema 2:

Si la varianza de una de las muestras es 0,36, cuál es el error estándar estimado par ala media sobre la base de esta muestra?

Su solución:

Repasemos: la distribución de las medias de las muestras aleatorias de tamaño n tomadas de casi cualquier tipo de distribución con media μ y desviación estándar σ desconocidas, es aproximadamente normal con media igual a la de la media de la población desconocida y una desviación estándar $\sigma/\sqrt{n}$, que se puede estimar por medio de $s_{\bar{x}} = s/\sqrt{n}$.

AL estudiar las distribuciones normales en el Capítulo 9, usted aprendió que el 95% del área de una distribución normal cae dentro de aproximadamente dos desviaciones estándar a cada lado de la media. Así mismo, el 95% de la distribución de las medias de las muestras estará dentro de dos unidades $\sigma/\sqrt{n}$ a cada lado de μ o dentro del intervalos $\mu \pm 2\sigma/\sqrt{n}$. Lamentablemente, en la mayoría de los casos la varianza verdadera de la población σ^2 no se conoce. Por fortuna, para muestras de tamaño relativamente grande ($n > 30$), podemos utilizar nuestra varianza aproximada s^2 en vez de la varianza desconocida σ^2 y decir que el 95% de las veces la media verdadera, μ estará incluida en el intervalo:

$$x \pm 2s_x = x \pm 2s/\sqrt{n}$$

Esto nos lleva a los intervalos de confianza.

INTERVALOS DE CONFIANZA

A menos que nuestra muestra incluya a la totalidad de la población, nunca sabremos el valor de μ: solo tendremos una estimación de este valor. Nos gustaría fijar el valor de μ dentro de algún rango y que ese rango fuese lo más pequeño posible. Si consideramos que μ es un valor en una escala que se extiende hasta el infinito en ambos sentidos, podemos estar el 100% seguros de que μ está en algún lugar dentro del rango, entre $+\infty$ y $-\infty$. Si estrechamos este intervalo no podemos estar seguros de que el valor verdadero de μ esté dentro o fuera de él. Ya no tenemos un 100% de confianza en nuestros hallazgos.

INTERVALO CON CONFIANZA DEL 95%

Con nuestro conocimiento de una distribución normal, podemos estrechar ese rango son perder mucho de certeza. hay un 95% de posibilidades de que la media verdadera de las muestras esté dentro de 2 s_x a cada lado de x. En promedio, 19 de cada 20 veces la μ desconocida estará incluida en el intervalo. El intervalo o rango que seleccionamos como límites para encerrar el valor de x en una escala con cierto grado de certeza, se llama **intervalo de confianza** para la media. Podemos utilizar los resultados de una muestra tomada al azar para determinar este intervalo. Los límites o extremos del intervalo, x - $2s_x$ y x + $2s_x$ se denominan límites de **confianza**.

Problema 3:

Si el puntaje medio de IHB-S de una muestra de 400 hombres de los Estados Unidos es 1 y la varianza 0.64, cuál es el intervalo con 95% de confianza para μ? Cuáles son los límites de confianza?

Su solución:

De modo similar, un intervalo con 99,74% de confianza se expresaría como ± $3s_x$. Podemos seleccionar cualquier nivel de confianza que deseemos, siempre que se le rotule correctamente. A veces x se expresa junto con el error estándar estimado, s_x para que el lector pueda determinar a su criterio un intervalo de confianza apropiado.

Problema 4:

Si el promedio del índice de tártaro simplificado para una muestra de 625 habitantes de los Estados Unidos cuya edad entre 6 y 74 es 0.35 y el error estándar 0.02, cuáles son los límites con 68.3% de confianza para μ? Y los límites con 99.74% de confianza?

Su solución:

En el capítulo 14 se analizará el intervalo de confianza apropiado para muestras de pequeño tamaño ($n < 30$). EL intervalo con 95% de confianza no se determina por $x \pm t$ s_x. El valor de t (con n-1 "grados de libertad") es ligeramente mayor que 2 y permite que nuestra estimación sea más exacta en los casos de muestras pequeñas. Se basa en la "distribución de t", que es similar a la distribución normal.

REVISION

1) Se puede percibir una población por su media, μ, y su varianza, σ^2 (o desviación estándar σ).

2) Se puede describir una muestra de una población por su media muestral x, y su varianza muestral, s^2 (o desviación estándar muestral, s).

3) El error estándar estimado de la media de la muestra $= SE = S_x = \sqrt{(s^2/n)} = s/\sqrt{n}$.

4) Con arreglo al teorema del límite central, un grupo de media de muestras de tamaño n tienden a distribuirse con forma normal y media igual a la media de la población. Para muestras relativamente grandes ($n > 30$) la varianza verdadera σ^2/n de las medias de las muestras se puede estimar aproximadamente por la varianza de las medias de la muestra $= s^2/n$.

5) El valor de μ se puede estimar utilizando intervalos o límites de confianza. Cuando $n > 30$:

   El intervalo con 95% de confianza es $(x - 2s_x, x + 2s_x)$, que a veces se representa como $x \pm 2s_x$.

   Los límites con 95% de confianza son las cantidades $x - 2s_x$ y $x + 2s_x$.

TEST FINAL 10

Se llevó a cabo una encuesta entre gente que necesitaba que se le extranjera por lo menos un diente, para determinar el número promedio de dientes por extraer y las causas de la extracción. Una muestra de 625 adultos cuya edad oscilaba entre 45 y 64 años registró un número promedio de 6,6 dientes que había que extraer por enfermedad periodontal. La varianza de la muestra fue 196.

Exprese su estimación de la media de la población en función de un intervalo con 95% de confianza.

HOJA DE RESPUESTAS
LABORATORIO N.____

PROBLEMA No 1

PROBLEMA No 2

PROBLEMA No 3

PROBLEMA No 4

PROBLEMA No 5

PROBLEMA No 6

PROBLEMA No 7

PROBLEMA No 8

PROBLEMA No 9

PROBLEMA No 10

LO

INVITO

A

PENSAR

PIENSE

A un herrelo le trajeron 5 trozos de cadena, de tres eslabones cada uno, y le encargaron que los uniera formando una cadena continua.
El herrero comenzó a meditar sobre el número de anillos que tendría necesidad de cortar y forjar de nuevo. Decició que le haría falta abrir y cerrar cuatro anillos.
¿No es posible efectuar el trabajo abriendo y enlazando un número menor de anillos?

PIENSE

El número 30 se puede expresar en función de tres números iguales $5 * 5 + 5 = 30$. ¿Podría usted expresar este mismo número con otras tres cifras iguales?

PIENSE

Dos padres regalaron dinero a sus hijos. Uno de ellos dio a su hijo 150 dólares, el otro entrego a su hijo 100 dólares. Resultó, sin embargo, que ambos hijos juntos aumentaron su capital en 150 dólares sólamente. ¿De qué modo se explica esto?

PIENSE

A un aficionado a los rompecabezas le pregunatron cuantos años tenia. La contestación fué compleja:

Tomad tres veces los años que tendré dentro de tres años, restadle tres veces los años que tenia hace tres años y resultará exactamente los años que tengo ahora.
¿Cuantos años tiene?

PIENSE

cuarto

Se trata de dividir la figura de un curarto creciente de la luna en seis partes, trazando solamente dos (2) lineas rectas. ¿Cómo hacerlo?

PIENSE

Escriba un número de 9 cifras, sin que se repita ninguna de ellas (es decir, que todas las cifras sean diferentes), y que sea divisible por 11.
Escriba el mayor de todos los números que satisfaga estas condiciones.
Escriba el menor de todos ellos.