

Sesión 1

Tema: Conceptos básicos de estadística.

I. **Objetivos de la sesión:** conocer los conceptos básicos para iniciar el tema de estadística descriptiva.

II. Tema:

1. Introducción: Permanentemente recibimos información referente al área en que trabajamos y es necesario hacer uso de ella, puesto que será útil para el proyecto en que estamos trabajando.

La información es importante para la toma de decisiones en muchos problemas. Para esto necesitamos un procesamiento adecuado de los datos de, para que nos arroje conclusiones certeras. En caso contrario, si no se aplica un buen procesamiento, es posible que en base a los resultados tomemos una mala decisión.

Información buena	} ⇒	buena decisión
Información suficiente		
Procesamiento correcto		

La estadística es un campo del conocimiento que permite al investigador deducir y evaluar conclusiones acerca de una población a partir de información proporcionada por una muestra.

Específicamente, la estadística trata de teoremas, herramientas, métodos y técnicas que se pueden usar en:

- Recolección, selección y clasificación de datos.
- Interpretación y análisis de datos.
- Deducción y evolución de conclusiones y de su confiabilidad, basada en datos muestrales.

Los métodos de la estadística fueron desarrollados para el análisis de datos muestreados, así como para propósitos de inferencia sobre la población de la que se selecciono la muestra.

La estadística como ciencia, cubre un extenso campo donde poder aplicarla. Se agrupa en 2 grandes áreas: estadística descriptiva y estadística inferencial, que desempeñan funciones distintivas, pero complementarias en el análisis.

Es importante que todo profesional que utilice la estadística como herramienta auxiliar de trabajo, posea un mínimo de conocimientos y habilidades

prácticas en aquellas técnicas que le facilitarán el buen desarrollo de esta actividad.

1.1 Estadística descriptiva.

La estadística descriptiva comprende las técnicas que se emplean para resumir y describir datos numéricos.

Son sencillas desde el punto de vista matemático y su análisis se limita a los datos coleccionados sin inferir en un grupo mayor.

El estudio de los datos se realiza con representaciones gráficas, tablas, medidas de posición y dispersión.

1.2 Estadística inferencial.

El problema crucial de la estadística inferencial es llegar a proposiciones acerca de la población a partir de la observación efectuada en muestras bajo condiciones de incertidumbre. Ésta comprende las técnicas que aplicadas en una muestra sometida a observación, permiten la toman de decisiones sobre una población o proceso estadístico. En otras palabras, es el proceso de hacer predicciones acerca de un todo basado en la información de una muestra.

La inferencia se preocupa de la precisión de los estadígrafos descriptivos ya que estos se vinculan inductivamente con el valor poblacional.

2. Definimos conceptos fundamentales: población, muestra y variable.

2.1 Población.

Es el conjunto de todos los elementos que presentan una característica común determinada, observable y medible. Por ejemplo, si el elemento es una persona, se puede estudiar las características edad, peso, nacionalidad, sexo, etc.

Los elementos que integran una población pueden corresponder a personas, objetos o grupos (por ejemplo, familias, fábricas, empresas, etc).

Las características de la población se resumen en valores llamados parámetros.

2.2 Muestra.

La mayoría de los estudios estadísticos, se realizan no sobre la población, sino sobre un subconjunto o una parte de ella, llamado muestra, partiendo del supuesto de que este subconjunto presenta el mismo comportamiento y

características que la población. En general el tamaño de la muestra es mucho menor al tamaño de la población.

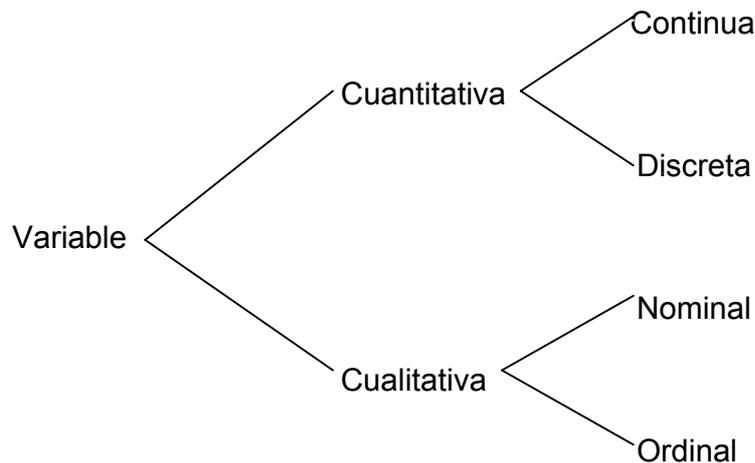
Los valores o índices que se concluyen de una muestra se llaman estadígrafos y estos mediante métodos inferenciales o probabilísticos, se aproximan a los parámetros poblacionales.

2.3 Variable.

Se llama variable a una característica que se observa en una población o muestra, y a la cual se desea estudiar.

La variable puede tomar diferentes valores dependiendo de cada individuo.

Una variable se puede clasificar de la siguiente manera.



a) Variable cuantitativa: es aquella que toma valores numéricos. Dentro de ella, se subdividen en:

Continua: son valores reales. Pueden tomar cualquier valor dentro de un intervalo. Ej. Peso, estatura, sueldos.

Discreta: toma valores enteros. Ej. N° de hijos de una familia, n° de alumnos de un curso.

b) Variable cualitativa: es aquella que describe cualidades. No son numéricas y se subdividen en:

Nominal: son cualidades sin orden. Ej. Estado civil, preferencia por una marca, sexo, lugar de residencia.

Ordinal: son cualidades que representan un orden y jerarquía. Ej. Nivel educacional, días de la semana, calidad de la atención, nivel socioeconómico.

3. Obtención de los datos

Como se ha puesto de manifiesto, gran parte del trabajo de un estadístico profesional se hace con muestras. Estas son necesarias porque las poblaciones son casi siempre demasiado grandes para estudiarlas en su totalidad. Exigiría demasiado tiempo y dinero estudiar la población entera, y tenemos que seleccionar una muestra de la misma, calcular el estadístico de esa muestra y utilizarlo para estimar el parámetro correspondiente de la población.

La obtención de la información se puede realizar por diversos medios. Una forma es a través de una encuesta a un grupo de individuos, donde a cada uno se le hacen las mismas preguntas.

Otra forma es a través de experimentos donde la respuesta a la variable es el resultado del experimento. Puede también recolectarse los datos en forma directa, es decir, la información se extrae de alguna base de datos seleccionando una muestra de ellos.

En cualquiera de estos casos contamos con una selección de información llamada muestra y que se procede a analizar.

Existen diferentes técnicas para realizar el muestreo y que dependerán cada caso, cual usar. Algunas de ellas son:

3.1 Muestreo aleatorios simple: todos los elementos de la población tiene igual posibilidad de ser escogido y se eligen al azar.

3.2 Muestreo sistemático: los elementos se seleccionan a un intervalo uniforme en una lista ordenada. Una preocupación del muestreo sistemático es la existencia de factores cíclicos en el listado que pudieran dar lugar a un error.

3.3 Muestreo estratificado: los elementos de la población son primeramente clasificados en grupos o estratos según una característica importante. Luego, de cada estrato se extrae una muestra aleatoria simple.

3.4 Muestreo por conglomerado: los elementos de la población están subdivididos en grupos y se extraen aleatoriamente algunos de estos grupos completos

III. Actividad previa: leer textos indicados en bibliografía, referentes al tema.

IV. Actividad post sesión: realizar la guía de ejercicios que a continuación se presenta.

Ejercicios.

1. De las siguientes afirmaciones ¿cuál se asemeja más a los conceptos: un parámetro, un dato, una inferencia a partir de datos, un estadígrafo?
 - Según estudios, se producen más accidentes en el centro de Santiago, a 35 km/h que a 65 km/h.
 - En una muestra de 250 empleados, se obtuvo un sueldo promedio de \$150.000
 - La tasa de nacimiento en el país aumentó en 5% con relación al mes precedente.
 - Las edades son 85, 36, 57, 24
 - Se sabe que el 55% de las personas en Chile son varones.
 - Según datos de años anteriores se estima que la temperatura máxima de este año aumentará en un 5%.

2. De los siguientes enunciados ¿cuál probablemente usa la estadística descriptiva y cuál, la estadística inferencial?
 - Un médico general estudia la relación entre el consumo de cigarrillo y las enfermedades del corazón.
 - Un economista registra el crecimiento de la población en un área determinada.
 - Se desea establecer el promedio de bateo de un equipo determinado.
 - Un profesor de expresión oral emplea diferentes métodos con cada uno de sus 2 cursos. Al final del curso compara las calificaciones con el fin de establecer cual método es más efectivo.

3. Clasificar las siguientes variables en: continua, discreta, nominal, ordinal:
 - n° de alumnos por carrera
 - comuna en que viven los alumnos del curso de estadística
 - color de ojos de un grupo de niños
 - monto de pagos por concepto de aranceles en la universidad

- sumas posibles de los números obtenidos al lanzar dos dados
- clasificación de los pernos en un local según sus diámetros
- peso del contenido de un paquete de cereal
- monto de la venta de un artículo en \$
- valor de venta de las acciones
- n° de acciones vendidas
- nivel de atención en el Banco
- nivel de educacional
- AFP a que pertenece un individuo
- edad
- clasificación de la edad en: niño, joven, adulto y adulto mayor

4. De cada una de las siguientes situaciones responda las preguntas que se plantean:

- Un fabricante de medicamentos desea conocer la producción de personas cuya hipertensión (presión alta) puede ser controlada con un nuevo producto fabricado por la compañía. En un estudio a un grupo de 13.000 individuos hipertensos, se encontró que el 80% de ellos controló su presión con el nuevo medicamento.
- Según una encuesta realizada a 500 adultos mayores de la comuna de Santiago, reveló que en promedio realizan 6 visitas anuales al consultorio. En vista de los resultados el ministerio de salud deberá aumentar los recursos en un 10%.

- a) ¿Cuál es la población?
- b) ¿Cuál es la muestra?
- c) Identifique el parámetro de interés
- d) Identifique el estadígrafo y su valor
- e) ¿Se conoce el valor del parámetro?

5. Conteste V ó F

- a) ___ La estadística descriptiva es el estudio de una muestra que permite hacer proyecciones o estimaciones acerca de la población de la cual procede.
- b) ___ Un parámetro es una medida calculada de alguna característica de una población.
- c) ___ Abrir una caja de manzanas y contar los que están en mal estado es un ejemplo de dato numérico continuo.
- d) ___ En una muestra aleatoria simple todos tiene la misma posibilidad de ser seleccionados.
- e) ___ No tiene mayor importancia el criterio que se tome para determinar a cual intervalo pertenece un elemento cuyo valor coincida con el limite de una clase.
- f) ___ La suma de las frecuencias absolutas es siempre igual a 1.
- g) ___ Mientras mayor es el número de intervalos elegidos para la formación de una distribución de frecuencias, menor es la exactitud de los estadígrafos que se calculan.
- h) ___ $H_i > h_i$ (para todo i).
- i) ___ La marca de clase debe ser siempre un número entero y positivo.

6. Completa las siguientes frases.

- a) La estadística que analiza los datos y los describe es _____.
- b) Por medio de una investigación se recolectan los _____.
- c) Por razones de costo y del tiempo que se gastaría en encuestar a todos los elementos de una _____, se recurre al _____.
- d) Para obtener una _____ aleatoria de la población, cada elemento debe tener _____ oportunidad de ser _____.

7. ¿En que nivel de medición se puede expresar cada una de estas variables?.
Razone su respuesta:

- a) Los estudiantes califican a su profesor de estadística en una escala de: horrible, no tan malo, bueno, magnifico, dios griego.
- b) Los estudiantes de una universidad se clasifican por especialidades, como marketing, dirección, contabilidad, etc.
- c) Los estudiantes se clasifican por especialidades con ayuda de los valores 1, 2, 3, 4, y 5.
- d) Agrupar mediciones de líquidos en pinta, cuarto y galón.
- e) Edades de los usuarios.

Sesión 2

Tema: Tablas de frecuencia.

I. **Objetivos de la sesión:** aprender a agrupar los datos para facilitar los cálculos de los estadígrafos.

II. Tema:

Cuando los datos estadísticos de que se dispone son numerosos, es difícil realizar cálculos sobre ellos. Por esta razón se organizan en tablas de manera de facilitar el trabajo.

Una tabla de frecuencia es la ordenación de la información obtenida de una muestra, en el estudio de una sola variable.

Cuando se dispone de un gran número de datos, es útil distribuirlos en categorías dentro de una tabla para facilitar el análisis. Se explicara con un ejemplo:

1. Veamos el caso de una variable discreta, pero antes se mencionaran las siguientes notaciones:

Ejemplo: en una encuesta de presupuesto familiar, se ha obtenido la siguiente información respecto al n° de hijos en 2 familias.

Variable $x = n^\circ$ de hijos
Los datos son los siguientes:

3, 1, 2, 0, 3, 2, 1, 1, 3, 3, 2, 4, 2, 2, 0, 2, 1, 3, 4, 2, 3
↓
 x_1 x_2 x_{21}

Vemos que la variable x toma valores entre 0 y 4, es decir, existen en este grupo 5 categorías o clases.

Contamos el número de familias en cada categoría y formamos la tabla.

Tabla 1.

Categorías X_i	Frec. Absoluta f_i	Frec. Absoluta Acum. F_i	Frec. Relativa h_i	Frec. Rel. Acum. H_i
0	$f_1 = 2$	$f_1 = 2$	$f_1/n = 0.095$	$h_1 = 0.095$
1	$f_2 = 4$	$f_1 + f_2 = 6$	$f_2/n = 0.190$	$h_1 + h_2 = 0.286$
2	$f_3 = 7$	$f_1 + f_2 + f_3 = 13$	$f_3/n = 0.333$	$h_1 + h_2 + h_3 = 0.619$
3	$f_4 = 6$	$f_1 + \dots + f_4 = 19$	$f_4/n = 0.285$	$h_1 + \dots + h_4 = 0.904$
4	$f_5 = 2$	$f_1 + \dots + f_5 = 21$	$f_5/n = 0,095$	$h_1 + \dots + h_5 = 1,000$
Total	$n = 21$		1.000	

n = tamaño de la muestra

X_i = valor de la variable en el individuo i

f_i = frecuencia absoluta: nº de veces que se repite la variable en la categoría i

F_i = frecuencia absoluta acumulada. Indica el nº de individuos hasta la categoría i

h_i = frecuencia relativa. Porcentaje de la categoría respecto del total, se obtiene dividiendo la frecuencia de la clase por el tamaño de la muestra.

H_i = frecuencia relativa acumulada. Porcentaje acumulado

Observamos algunos detalles importantes:

- i) n es la suma de la columna f_i , es decir, siempre debe dar como resultado el tamaño de la muestra.
- ii) En la columna de frecuencia absoluta acumulada se va sumando los valores de la columna f_i , por lo tanto el último valor debe ser igual a n .
- iii) La columna frecuencia relativa (h_i) representa en % de familias en cada categoría. Por ejemplo, en las categorías con 3 hijos a un 28.5% de familias. Esta columna debe sumar 1.
- iv) La H_i acumula los valores de la frecuencia relativa, por lo tanto el último valor debe ser 1. Ejemplo H_4 : el 90.4% de las familias encuestadas tienen a los más 3 hijos.

2. En el caso de analizar una variable continua, la tabla de frecuencia cambia sólo en el comienzo. También sé vera en un ejemplo:

Salarios semanales de 40 personas en miles de pesos.

90	62	102	85	92	106	110	95	105	112
108	86	110	68	118	99	98	74	91	80
80	100	79	93	93	104	77	106	98	73
95	85	91	83	67	119	108	115	74	88

Efectuemos previamente los siguientes pasos.

- i) Se busca el valor mínimo y el valor máximo $X_{\min} = 62$ $X_{\max} = 119$
- ii) Se calcula el rango: $119 - 62 = 57$.
Rango: en todo conjunto de valores estadísticos hay valores extremos: el menor de todos y el mayor de todos; la diferencia entre estos valores extremos se llama rango.
- iii) La cantidad de intervalos no debe ser menor de 5 ni mayor de 18. Por lo general tiene el mismo ancho. Una forma de calcular el nº de intervalos para generar la tabla de frecuencias es mediante la siguiente formula:
 $k = 1 + 3.322 \times \log(40) = 6.322$ usamos $k \approx 6$
- iv) Se calcula la amplitud de cada intervalo $c = \text{rango} / k = 57 / 6 = 9.5 \approx 10$
- v) Se construye la tabla:

Tabla 2.

Intervalos $Y_{i-1} - Y_{i+1}$	Marca de clase Y_1	Frec. Absoluta f_i	F_i	h_i	H_i
[60 – 70)	65	3	3	0,075	0,075
[70 – 80)	75	5	8	0,125	0,20
[80 – 90)	85	7	15	0,175	0,375
[90 – 100)	95	11	26	0,275	0,62
[100 – 110)	105	8	34	0,20	0,85
[110 – 120)	115	6	40	0,15	1,00
Total		40		1,00	

El resto de las columnas se forman de la misma manera que la tabla 1.

n = tamaño de la muestra

X_i = valor de la variable en el individuo i

f_i = frecuencia absoluta: nº de veces que se repite la variable en la categoría i

F_i = frecuencia absoluta acumulada. Indica el nº de individuos hasta la categoría i

h_i = frecuencia relativa. Porcentaje de la categoría respecto del total, se obtiene dividiendo la frecuencia de la clase por el tamaño de la muestra.

H_i = frecuencia relativa acumulada. Porcentaje acumulado

Y_i = marca de clase: su valor es igual a la mitad de la suma de los límites inferior y superior del intervalo de clase. En todos los análisis estadísticos se supone que el valor de la marca de una clase es el valor que corresponde asignar a cada uno de los elementos ubicados en ese intervalo.

C = amplitud del intervalo: la diferencia entre los límites reales de un intervalo.

Y_{i-1} = límite inferior del intervalo

Y_{i+1} = límite superior del intervalo

III. Actividad previa: leer textos indicados en bibliografía, referentes al tema.

IV. Actividad post sesión: realizar la guía de ejercicios que a continuación se presenta

Ejercicios.

1. En una cierta ciudad se ha tomado una muestra representativa del total de familias que en ella viven y se ha anotado el número de hijos de cada una. Los valores de esta variable son los siguientes:

0	1	0	4	2	2	1	2	3	2
3	2	1	3	4	2	2	3	2	1

- a) Diga que tipo de datos son estos.
- b) Construya una tabla de frecuencias correspondiente a este ejercicio.
2. Se visitaron 25 empresas citrícolas de una cierta zona y en cada una se anotó la cantidad de plantas atacadas por un cierto hongo, de lo cuál resultaron los siguientes datos:

15	20	25	15	18	16	17
18	20	18	18	18	19	16
17	19	16	17	17	17	19
18	19	18	15			

- a) Diga que tipo de datos son estos.
- b) Construya una tabla de frecuencias correspondiente a este ejercicio.
3. Determine para el ejercicio 1 la frecuencia relativa de familias con 2 hijos o menos y la frecuencia relativa de familias que tienen más de 2 hijos y no más de 4.
4. Determine para el ejercicio 2 la frecuencia relativa de empresas citrícolas que tienen 18 o menos plantas atacadas por el hongo. Calcule también la frecuencia relativa de empresas citrícolas que tienen no menos de 18 plantas atacadas por el hongo.
5. El gerente de personal de una compañía registró el número de días que sus 50 empleados habían tomado como licencia por enfermedad.

10	35	12	8	44	6	15	20	5	7
5	11	17	8	4	7	25	9	2	10
12	12	3	10	9	3	5	16	31	9

0	4	7	11	3	18	2	10	6	22
2	9	8	29	6	4	7	10	0	1

- a) Identificar la variable y su clasificación.
- b) Construir una tabla de frecuencia apropiada a estos datos.

6. Dada la siguiente tabla de frecuencias:

Clases	f_i
9 – 12	2
12 – 15	1
15 – 18	2
18 – 21	5
21 – 24	8
24 – 27	11
27 – 30	7
30 – 33	4
33 – 36	5
36 – 39	2

- a) Determine las marcas de clases y la amplitud de las clases.
- b) Determine las frecuencias absolutas acumuladas.
- c) Determine las frecuencias relativas.
- d) Determine las frecuencias relativas acumuladas.

7. En la tabla de frecuencia que se da a continuación faltan datos, complétela.

Valores	f_i	F_i	h_i	H_i
0		2		
1		5		
2		9		
3		14		0.7
4			0,2	
5				
Totales				

8. Completa la información que falta en la siguiente tabla.

$Y_{i-1} - Y_i$	Y_i	f_i	F_i	h_i	H_i
- 100		2			
	150		7		
				0,2	
					0,8
			30		
Total					

9. En la tabla de frecuencias siguiente faltan algunos datos, complétela:

Clases	Y_i	f_i	F_i	h_i	H_i
20 - 24					0,1
24 - 28					0,25
- 32			11		0,55
32 -					0,85
- 40					1,0
Total					

10. De la siguiente tabla, interpreta algunos valores:

Los datos muestran los resultados obtenidos por un grupo de estudiantes en una prueba de habilidad de lectura,

$X_{i-1} - X_{i+1}$	X_i	f_i	F_i	h_i	H_i
[32 - 35)	33,5	5	5	0,04	0,04
35 - 38	36,5	12	17	0,11	0,15
38 - 41	39,5	18	35	0,16	0,31
41 - 44	42,5	19	54	0,17	0,48
44 - 47	45,5	26	80	0,23	0,71
47 - 50	48,5	19	99	0,17	0,88
50 - 53	51,5	13	112	0,12	1,00
		n = 112		1,00	

¿Cómo interpretas los números en negrita?

Sesión 3

Tema: Gráficos.

I. Objetivos de la sesión: que los alumnos sean capaces de representar las variables con diferentes modelos de gráficos.

II. Tema:

La entrega de información utilizando gráficos y dibujos es un método funcional que no solo sirve para presentar datos sino también para expresar ideas que se desean destacar.

Los gráficos son las representaciones visuales de los datos en donde se evidencian fundamentalmente 3 características:

- forma
- acumulación o tendencia
- dispersión o variabilidad

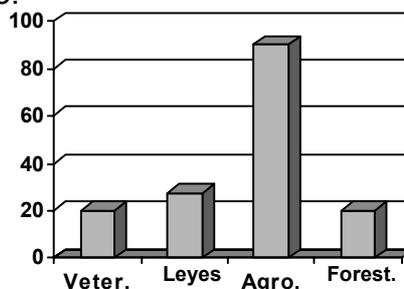
Los gráficos no deben considerarse como sustitutos de un análisis estadísticos, sino más bien como una ayuda visual del comportamiento de los datos.

Existen diferentes tipos de gráficos:

- barras
- histogramas y polígonos
- histogramas y ojivas
- circulares
- barras subliminales
- pictogramas

1. Barras:

Se construye sobre el sistema de ejes cartesianos. Es un procedimiento gráfico para representar los datos nominales u ordinales. Para cada categoría se traza una barra vertical en que la altura es la frecuencia absoluta de la categoría. El ancho de la barra es arbitrario.

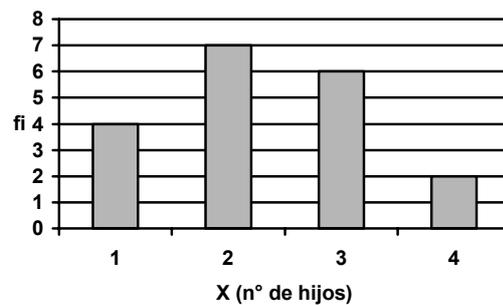


También se utiliza si la variable en estudio es numérica discreta.

Ej. en el ejemplo del n° de hijos en 21 familias.

X_i	f_i
0	2
1	4
2	7
3	6
4	2
	21

El gráfico de barras sería

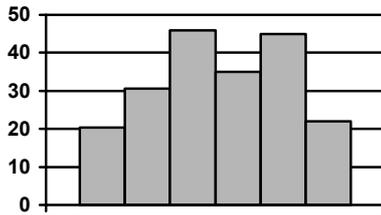


2. Histogramas y polígonos de frecuencia:

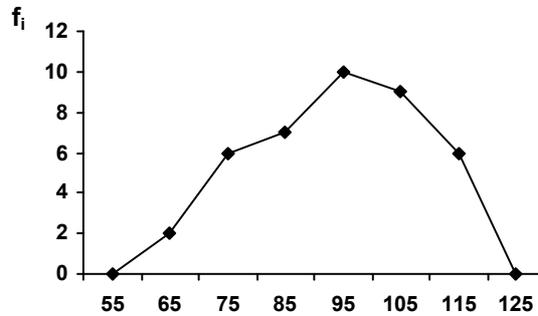
Se construyen sobre el sistema de coordenadas cartesianas. Se utiliza cuando la variable en estudio es continua o esta agrupada en una tabla de frecuencia con intervalos en cada categoría.

En el eje X se identifica la variable en estudio y en el eje Y se gráfica la frecuencia absoluta o la frecuencia relativa. Consiste en una serie de rectángulos en donde su altura depende del valor de cada frecuencia.

Cada categoría de la variable se representa por una barra. El ancho de cada barra depende de la amplitud del intervalo.



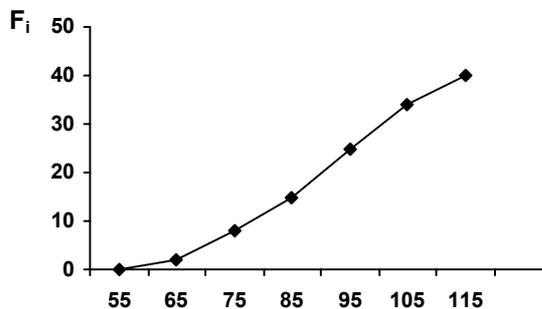
El polígono se gráfica uniendo la punta superior de cada barra por segmento de recta. Para que el polígono quede cerrado se considera un punto en la recta horizontal, antes y después de las anotadas.



El polígono se dibuja midiendo los puntos medios de cada barra, que corresponde a la marca de clase.

2. Histogramas y ojivas:

También se gráfica la columna de frecuencia absoluta acumulada. El gráfico siempre será en forma ascendente. Se vera en un ejemplo usando la tabla 2.



La ojiva se dibuja midiendo segmentos de recta en la parte superior de cada barra, y no se cierra.

4. Gráfico circular:

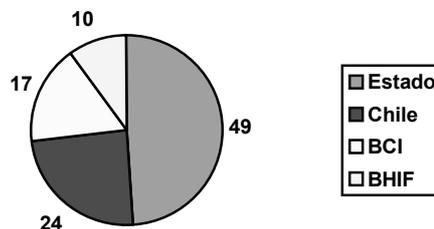
Esta es otra forma de representar los datos, en especial cuando se trata de cualidades. En un gráfico dibujado dentro de un círculo.

Es necesario en primer lugar calcular el porcentaje de cada categoría respecto del total y luego repartir proporcionalmente estos porcentajes en los 360° del círculo.

Ejemplo:

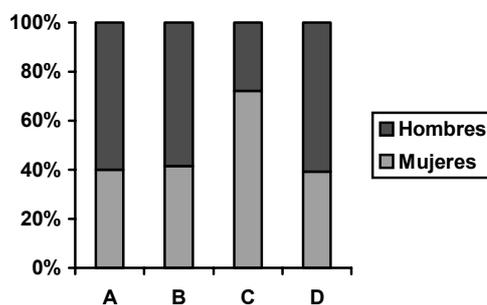
Para transformar la frecuencia relativa a grados sexagesimales se aplica el siguiente método: $\frac{100\%}{48,8} = \frac{360^\circ}{x^\circ} \Rightarrow x = 175,7^\circ$

Bancos	N° de clientes	%	Grados
Estado	10.000	48.8	175.7
Chile	5.000	24.3	87.5
BCI	3.500	17.7	61.5
BHIF	2.000	9.8	35.3
	20.500	100.0	360.0



5. Barras subliminales:

Es un gráfico de barras muy apropiado para comprobar subdivisiones en la variable. Por ejemplo: % de estudiantes en diferentes carreras, separadas por sexo. Cada barra es un 100%.

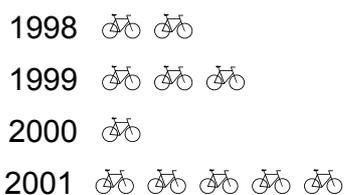


6. Pictogramas:

Un pictograma es la representación de datos estadísticos por medio de símbolos que por su forma sugieren la naturaleza del dato. Por ejemplo: producción de bicicletas (en miles.)

Año	Producción (miles de bicicletas)
1998	2004
1999	3000
2000	1000
2001	5025

El gráfico es el siguiente:



III. Actividad previa: leer textos indicados en bibliografía, referentes al tema.

IV. Actividad post sesión: realizar los gráficos de las tablas de frecuencia de la guía de ejercicios de la sesión 2.

Ejercicios.

1. En una industria el informe de contabilidad muestra que la producción fue de \$62.600.000 y los gastos así: de administración \$11.160.000, de materiales y energéticos \$15.650.000, salarios y prestaciones \$18.780.000. Elabore con los datos un diagrama circular.
2. Utilizando un diagrama lineal represente la deuda externa de América Latina cuyos valores en millones de US\$ son:

1973	55.4	1978	141.6
1974	68.5	1979	169.2
1975	82.9	1980	207.1
1976	98.3	1981	279.1
1977	119.1	1982	312.0

3. La siguiente tabla presenta la distribución de frecuencias del número de kilómetros recorridos hasta que se produjo la primera falla grave de motor en cierto modelo de vehículo.

Distancia recorrida (en miles de kms.)	Frecuencia
0 – 20	8
20 – 40	13
40 – 60	18
60 – 80	27
80 – 100	36
100 – 120	48
120 – 140	35
140 – 160	18
160 – 180	4
180 – 200	4

- a) Representar mediante un histograma la distribución de frecuencia.
- b) Señalar en el histograma la región correspondiente a vehículos que no han sufrido averías serias en los 100.000 kilómetros. ¿Qué proporción representa?.
- c) Hallar aproximadamente el valor del kilometraje para el que han fallado la mitad de los vehículos.

d) ¿Qué forma tiene la distribución?

4. Las siguientes medidas corresponden a las alturas de 50 niños.

1,56	1,59	1,63	1,62	1,65
1,61	1,59	1,51	1,62	1,62
1,53	1,49	1,57	1,54	1,53
1,59	1,58	1,57	1,47	1,64
1,55	1,59	1,53	1,56	1,53
1,47	1,57	1,60	1,54	1,56
1,50	1,62	1,59	1,62	1,54
1,68	1,52	1,62	1,59	1,49
1,65	1,53	1,59	1,56	1,54
1,58	1,52	1,63	1,56	1,62

- Construir una distribución de frecuencias absolutas y relativas.
- Obtener las correspondientes distribuciones de frecuencias acumuladas.
- Representar las distribuciones anteriores mediante histogramas.
- Dibujar los correspondientes polígonos de frecuencias.
- Hallar a partir del polígono de frecuencias acumuladas la proporción de observaciones entre 1,59 y 1,62 ambas inclusive.
- ¿Qué conclusiones puede extraerse?

5. En una empresa se ha llevado a cabo un estudio sobre sueldos que se resumen en la siguiente tabla.

sueldos	25 - 50	50 - 75	75-100	100-125	125-150	150-175	175-200	200-225	225-250
Hombres	202	231	687	833	375	192	101	38	21
Mujeres	99	202	246	121	96	35	21	8	4

Construir un grafico de barras subliminales.

Sesión 4

Tema: Estadígrafos de posición.

I. **Objetivos de la sesión:** que el alumno aprenda a calcular e interpretar los diferentes estadígrafos de posición.

II. **Tema:**

1. Estadígrafos de posición:

Los estadígrafos de posición o de tendencia central son valores que se ubican al centro de un conjunto de datos ordenados según su magnitud. Existen varios de estos estadígrafos y se estudiarán de a uno, considerando si los datos se tienen no tabulados o tabulados

1.1 Media aritmética o promedio:

- **datos no tabulados:** se define como el cociente que se obtiene al dividir la suma de los valores de la variable por el n° de observaciones.

Su fórmula es la siguiente:

$$\bar{X} = \frac{\sum_{i=1}^{n=10} Xi}{n}$$

Ej. Edades de las personas

10, 18, 21, 32, 17, 27, 28, 20, 35, 31

n = 10

$$\bar{X} = \frac{19 + 18 + 21 + 32 + \dots + 31}{10} = \frac{248}{10} = 24.8$$

- **datos tabulados:** se usara los 40 datos de sueldo (tabla 2) en este caso se usa la siguiente fórmula:

$$\bar{X} = \frac{\sum_{i=1}^k Y_i \cdot n_i}{n}$$

$$\bar{X} = \frac{65 \cdot 3 + 75 \cdot 5 + 85 \cdot 7 + \dots + 115 \cdot 6}{40} = \frac{3740}{40} = 93.5$$

También se llama media ponderada.

Si se sumara los 4 datos sueltos, el resultado daría $3697/40 = 92,4$
Esta diferencia se produce porque al estar tabulados se pierde el valor real de cada dato, ya que se encasilla en un intervalo y es reemplazado por la marca de clase.

Ej. Un inversionista tiene 1.200 acciones cuyo valor promedio es \$34 y 800 acciones cuyo valor promedio es \$45. El valor promedio de las 2.000 acciones es:

$$\bar{X} = \frac{34 \cdot 1.200 + 45 \cdot 800}{1.200 + 800} = \frac{76.800}{2.000} = 38,4$$

Propiedades de la media aritmética.

a) La media de una constante es igual a la constante

$$\bar{X} = \frac{\sum_{i=1}^n K}{n} = K$$

Ej. Si los datos son 5, 5, 5, 5, 5, 5, 5 n=7

$$\bar{X} = \frac{5 + 5 + 5 + 5 + 5 + 5 + 5}{7} = \frac{7 \cdot 5}{7} = 5$$

- b) La media de los valores que son multiplicados por una constante es igual a la constante por la media de los datos.

$$\bar{X} = \frac{\sum_{i=1}^n X_i \cdot K}{n} = \frac{K \sum_{i=1}^n X_i}{n}$$

Ej. 10 sueldos en miles de personas.

100, 120, 119, 121, 122, 150, 135, 125, 145, 105

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{10} = \frac{100 + 120 + 119 + \dots + 105}{10} = \frac{1242}{10} = 124,2$$

Supongamos que se produce un aumento de un 10% a todos los sueldos ¿cómo varía el promedio?

$$\begin{aligned} \frac{\sum_{i=1}^n X_i \cdot 1,1}{10} &= \frac{100 \cdot 1,1 + 120 \cdot 1,1 + 119 \cdot 1,1 + \dots + 1,1 \cdot 105}{10} \\ &= 1,1 \left(\frac{100 + 120 + \dots + 105}{10} \right) \\ &= 1,1 \cdot \frac{\sum_{i=1}^n X_i}{n} \\ &= 1,1 \cdot 124,2 \\ &= 136,62 \end{aligned}$$

- c) La suma de las desviaciones de cada dato respecto al promedio de siempre cero.

$$\sum_{i=1}^5 (X_i - \bar{X}) = 0$$

Ej. Supongamos los datos de edades 25, 35, 24, 32, 28 n = 5

Luego $\bar{X} = 29$, entonces

$$\begin{aligned}
\sum_{i=1}^5 (X_i - \bar{X}) &= (26 - 29) + (35 - 29) + (24 - 29) + (32 - 29) \\
&= (26 + 35 + 24 + 32 + 28) - (29 + 29 + 29 + 29 + 29) \\
&= 145 - 145 \\
&= 0
\end{aligned}$$

Otra forma

$$\begin{aligned}
\sum_{i=1}^5 (X_i - \bar{X}) &= \sum_{i=1}^5 X_i - \sum_{i=1}^5 \bar{X} \\
&= \sum_{i=1}^5 X_i - n \cdot \bar{X} \\
&= \sum_{i=1}^5 X_i - n \cdot \frac{\sum_{i=1}^5 X_i}{n} \\
&= \sum_{i=1}^5 X_i - \sum_{i=1}^5 X_i \\
&= 0
\end{aligned}$$

1.2. Mediana:

Es otra medida de posición o tendencia central. Se define como aquel valor de la variable que supera la mitad de las observaciones y a su vez es superado por la otra mitad de ellas. Por esta razón, se la considera como el valor central, ya que se divide a los datos en 2 grupos (las observaciones deben estar ordenadas de mayor a menor).

- **Datos no tabulados:** se ordenan las observaciones de menor a mayor y se ubica el valor central. Si la constante de datos (n) es par, se promedian los 2 valores centrales. En cambio, si n es impar habrá solo un valor en el centro.

Ej. 3, 4, 4, 5, 6, 8, 8, 8, 10 n = 9 (n impar)

↓
Mediana

5, 5, 7, 9, 11, 12, 15, 18 n = 8 (n par)

$$\frac{9 + 11}{2} = 10 \rightarrow \text{mediana}$$

- **Datos tabulados:** si los datos están tabulados no es posible individualizar el valor de la mediana, pero si es factible determinar el intervalo donde se encuentra.

La fórmula para encontrar la mediana es:

$$Me = Li + C \left(\frac{\frac{n}{2} - F_{i-1}}{f_i} \right)$$

donde L_i = límite inferior del intervalo (Y_{i-1})
 C = amplitud del intervalo
 F_{i-1} = frecuencia absoluta acumulada anterior
 f_i = frecuencia absoluta del intervalo

Para encontrar el intervalo donde está “Me”, se debe calcular $n/2$ y buscar en la columna F_i , el valor que sugiere $n/2$.

Ej.: tabla 2

$$\frac{n}{2} = \frac{40}{2} = 20$$

En la columna F_i , el 4° intervalo supera a 20. El cálculo de Me es el siguiente:

$$Me \equiv 90 + 10 \left(\frac{\frac{40}{2} - 15}{11} \right)$$

$$Me = 90 + 10 \left(\frac{20 - 15}{11} \right)$$

$$Me = 94,54$$

(\$94.540 es el sueldo medio)

Su interpretación es el 50% inferior de la muestra, gana menos de \$94.540 y el otro 50% gana mas de \$94.540

1.3 Moda:

Es un estadígrafo que puede definirse como el valor mas frecuente o el valor de la variable que presenta la mayor frecuencia absoluta.

- **Datos no tabulados:** se busca el valor más repetido

Ej. 3, 5, 5, 5, 3, 4, 7 Mo = 5
6, 3, 3, 6, 5, 8, 4 Mo = 3 y 6

- **Datos tabulados:** en este caso no es fácil visualizar el valor con mayor frecuencia ya que se encuentra dentro de un intervalo. Para buscar la moda se usa la siguiente fórmula:

$$Mo = Li + C \left(\frac{\Delta 1}{\Delta 1 + \Delta 2} \right)$$

donde

L_i = límite inferior del intervalo ($Y_i - 1$)
 C = amplitud del intervalo
 $\Delta 1$ = diferencia entre f_i y f_{i-1}
 $\Delta 2$ = diferencia entre f_i y f_{i+1}

Ej. Tabla 2

$$Mo = 90 + 10 \cdot \left[\frac{11 - 7}{(11 - 7) + (11 - 8)} \right]$$

$$Mo = 90 + 10 \cdot \left[\frac{4}{4 + 3} \right]$$

$$Mo = 90 + 10 \cdot (0,57)$$

$$Mo = 95,71 \quad (\$95.710 \text{ valor mas repetido})$$

Características de la media, mediana y la moda.

Hemos aprendido que la media es una medida de tendencia central en la que la suma de las desviaciones respecto a ella es cero; que la mediana es la medida de tendencia central que divide el área de un histograma o el área bajo la curva de distribución en 2 partes iguales de manera que el número de observaciones por

debajo de la mediana es igual al número de observaciones por encima de ella, y que la moda es el valor que presenta la mayor frecuencia.

Nos referiremos a algunas de sus características y las relaciones mas importantes que hay entre ellas:

De las 3 medidas de tendencia central, la media aritmética es muy sensible a los valores extremos, en tanto que la mediana y la moda no son afectadas por los valores de los extremos. Por ejemplo, en la serie 3, 5, 7, 7, 8 la media aritmética es igual a 6; si cambiamos el valor extremo 8 por 18 se tiene la serie 3, 5, 7, 7, 18 cuya media aritmética es igual a 8 es decir la media varió de 6 a 8.

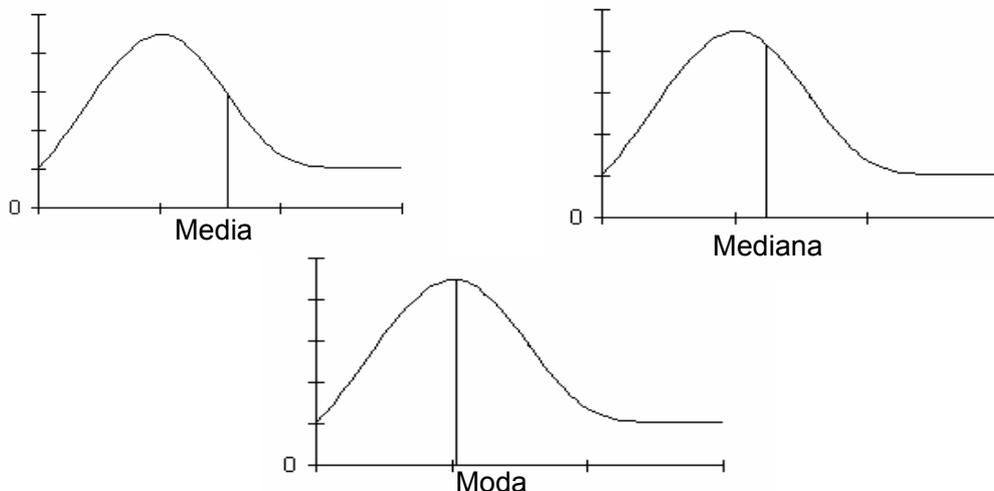
La mediana por ser insensible a los valores extremos no cambia en ninguna de las 2 series manteniendo su valor 7. La moda en ambas series es 7 por ser el valor mas frecuente.

Debido a la gran sensibilidad de la media aritmética a lo valores extremos, a veces resulta que su valor produce efectos engañosos. Por ejemplo, si se esta estudiando el ingreso de un grupo de personas que tiene los siguientes valores como sueldos en miles de pesos: 320, 400, 400, 400, 450, 500, 550, 2000, 2900.

Media = 880
Mediana = 450
Moda = 400

Obsérvese que solo 2 tienen ingresos altos y las 7 restantes tienen ingresos menores o iguales a 550 o sea que en este caso la media resulto atípica. La mediana y la moda resultan mas representativas para esta distribución. El conocimiento de las 3 medidas da una buena apreciación de la distribución de los valores pero si se debe hacer una apreciación con una sola de las medidas es mejor usar la mediana.

Para ilustrar estas diferencias se han diseñado los siguientes gráficos.



La media aritmética es un punto de equilibrio (semejante a un centro de gravedad).

La mediana tiene la propiedad de que su ordenada divide el área bajo la curva en dos partes iguales.

La moda es la abscisa correspondiente a la mayor ordenada o pico de la curva.

1.4 Cuartiles y deciles:

En forma similar a la mediana, pueden definirse otros estadígrafos, llamados cuartiles o deciles, que cumplen con la condición de superar a no más de cierto porcentaje de los datos y de ser superado, a los más por porcentaje complementario de las observaciones.

- **Datos no tabulados:**

Ej. 3, 3, 5, 7 | 8, 9, 9, 10 n = 8

$\begin{array}{c} | \\ Q_1 \\ \hline \\ | \\ Q_2 \\ \hline \\ | \\ Q_3 \end{array}$

Si separamos los datos en 4 grandes grupos, cada uno será un cuartil.

$$Q_1 = \frac{3 + 5}{2} = 4 \qquad Q_2 = \frac{7 + 8}{2} = 7,5 \qquad Q_3 = 9$$

La interpretación, por ejemplo del Q_1 es: el valor 4 supera el 25% de los datos y es superado por el 75% de las observaciones restantes (ordenadas de mayor a menor).

- **Datos tabulados:** se usa la misma fórmula que se ocupó en la mediana, con una pequeña modificación.

$$Q_1 = Li + C \left(\frac{\frac{i \cdot n}{4} - F_{i-1}}{f_i} \right)$$

donde i = es el n° del cuartil deseado

Ej. Tabla 2: calcular el Q_1 .

Primero calculamos $\frac{i \cdot n}{4} = \frac{1 \cdot 40}{4} = 10$

Buscamos en la columna N_i el valor que supere el 10, en este caso es el 3^{er} intervalo (en ese intervalo se busca el cuartil 1)

$$Q_1 = 80 + 10 \cdot \left[\frac{10 - 8}{7} \right]$$

$$Q_1 = 80 + 10 \cdot (0,28)$$

$$Q_1 = 82,85$$

El 25% de las personas de la muestra gana menos de \$82.850 y el 75% superior gana mas de \$82.850 (ordenados 2 sueldos de >a<)

Usando el mismo ejemplo, calculamos el decil 8.

$$D_8 = Li + C \left(\frac{\frac{8 \cdot n}{10} + F_{i-1}}{f_i} \right)$$

calculamos en primer lugar $\frac{8 \cdot n}{10} = \frac{8 \cdot 40}{10} = 32$

y se busca en la columna N_i el valor que supere 32. En este caso el 5° intervalo sirve.

$$D_8 = 100 + 10 \cdot \left[\frac{32 - 26}{8} \right]$$

$$D_8 = 100 + 10 \cdot (0,75)$$

$$D_8 = 107,5$$

La interpretación es: el 80% de las personas ganan menos de \$107.500 y el 20% restante gana mas de \$107.500.

Rango intercuartil.

Al estudiar el rango, vimos que era muy influenciado por los valores extremos; para eliminar la influencia de los extremos en estadística se suele analizar la situación del intermedio de la distribución y a esto se refiere el rango intercuartil que es la diferencia entre el tercer cuartil Q_3 y el primero Q_1 .

$$\text{Rango intercuartil} = Q = Q_3 - Q_1$$

Rango semi-intercuartílico o desviación cuartil. Es la mitad del rango intercuartílico; designándolo por Q_D se tiene:

$$Q_D = \frac{Q_3 - Q_1}{2}$$

A pesar de que el rango intercuartil y la desviación cuartil, como medidas de la variabilidad de las observaciones son mas adecuadas que el rango, presentan varios inconvenientes que demeritan su uso. Así:

1. No toma en consideración todos los valores de la distribución y puede ocurrir que los valores inferiores a Q_1 o superiores a Q_2 estén muy compactos o muy dispersos, y el valor de Q sería el mismo.
2. No es posible, conociendo solo Q , hacer la ubicación precisa de una observación dentro de la distribución.
3. Al igual que la mediana, que es el segundo cuartil, no tiene propiedades que les permitan intervenir en las relaciones matemáticas que utiliza la estadística.

III. Actividad previa: leer textos indicados en bibliografía, referentes al tema.

IV. Actividad post sesión: realizar los ejercicios de la guía que va a continuación.

Ejercicios

1. El precio de 100 artículo es \$185,7 en promedio, los artículos se dividen en dos grupos de precios promedios \$175,8 y \$197,8. ¿Cuántos artículos hay en cada grupo?.
2. Dada la siguiente tabla.

Salario diario (miles \$)	N° de obreros
0 – 3,0	10
3,0 – 4,0	16
4,0 – 5,0	35
5,0 – 6,0	26
6,0 – 7,0	13

- a) ¿Cuál es el salario máximo que ganan diariamente el 30% de obreros con sueldos más bajos?.
 - b) ¿Qué % de obreros ganan mas de \$5.500?
3. En una población hay 350 individuos con valores en la primera decil ¿cuántos individuos de la población entre percentil 18 y la percentil 45?
 4. Se prueban 2 tratamientos A y B para controlar un virus que ataca la hoja del tabaco. La hoja pierde valor comercial mientras mayor sea el número de lesiones por hoja producida por el virus. La medición del número de lesiones por hoja dio la siguiente tabla de frecuencia:

N° de lesiones /hoja	Frecuencia A	Frecuencia B
0	90	130
1	60	100
2	40	50
3	60	20
4	20	40
5	10	60
Total	280	400

- a) ¿Cuál tratamiento es mejor para obtener más hojas con 0 lesiones?.
 - b) Calcule medidas estadísticas que le permitan comprar en la mejor forma la efectividad de ambos tratamientos y coméntelas.
 - c) Construya un gráfico, que compare adecuadamente la efectividad de los tratamientos A y B.
5. La siguiente tabla muestra los tiempos de reacción (en seg.) de 250 perros sometidos a anestésico inyectado:

Tiempo reacción (seg.)	Frecuencia
[120 – 170)	20
[170 - 220)	35
[220 – 270)	85
[270 – 320)	50
[320 – 370)	30
[370- 420)	20
[420- 470)	10
Total	250

- a) Calcule las medidas estadísticas que representen mejor esta información e interprételas.
- b) ¿A los cuántos seg. reaccionara el 15% de los perros mas sensibles al anestésicos?.
- c) ¿Cuántos de los 250 perros reaccionaran después de 5 minutos?

6. Un grupo de 200 personas viaja en 2 aviones, el primero de ellos lleva a 150 personas.
El peso promedio de las 200 personas es de 72,5 Kg. Los del segundo avión pesan en promedio 3,8 Kg. menos que los del primer avión. ¿cuál es el peso promedio de los pasajeros de cada uno de los aviones?.
7. La renta semanal media de los trabajadores de una fábrica es de \$80.000, siendo \$95.000 para los administrativos y de \$70.000 para los obreros. Calcule el porcentaje de administrativos y de obreros que tiene la fábrica, si en total suman 120.
8. La tabla siguiente representa la distribución de frecuencias de las vidas medias de 400 ampollitas probadas en la empresa XXX.

Vida Media (horas)	N ampollitas
300 – 399	14
400 – 499	46
500 – 599	58
600 – 699	76
700 – 799	68
800 – 899	62
900 – 999	48
1000 – 1099	22
1100 – 1199	6

Determinar:

- Limite inferior de la quinta clase.
- Marca de clase de la tercera clase.
- La frecuencia de la cuarta clase.
- Porcentaje de ampollitas cuya vida media es de al menos 500 horas, pero menos de 1000 horas.
- Número de ampollitas cuya vida media es superior a 600 horas.

Construir:

- Tabla de frecuencia.
- Histograma, Polígono de frecuencias y ojiva de porcentajes.

h) ¿Sobre cuantas horas se encuentra el 35% de ampolletas de mayor duración?

i) Calcule media y mediana.

j) Si la vida media se incrementa en 24%, calcular los nuevos promedios y desviación típica.

9. Los gastos de publicidad son un componente significativo en el costo de los bienes que se venden. La lista de abajo es una distribución de frecuencia que muestra los gastos en publicidad de 60 empresas de manufacturas.

Gastos en publicidad en millones de \$	Nº de empresas
25 – 35	6
35 – 45	10
45 – 55	21
55 – 65	16
65 – 75	8

a) Si las empresas que gastaron menos de 45 millones de pesos aumentan sus gastos en 13%, ¿cómo se altera el promedio de gastos en publicidad?

b) ¿Qué porcentaje de empresas gastas mas de 58 millones de pesos en publicidad?

10. La siguiente información representa la distribución de los gastos en alimentación que realizaron un conjunto de familias de Santiago durante el año 2001.

Gastos en miles de pesos	Nº de familias
150 – 250	15
250 – 350	27
350 – 450	32
450 – 550	21
550 – 650	10

- a) Para el presente año se espera que tal gasto se incremente en un 8%, mas un gasto fijo por cada familia de 25 mil pesos. ¿Cómo varia el promedio del gasto en alimentación de las familias estudiadas en el año 2002 con respecto al año 2001?.
- b) A las familias que gastaron durante el año 2001 menos de 270 mil pesos se les dará en bono de 45 mil pesos a cada una y a las restantes se les dará un bono de 28 mil pesos a cada una. ¿Cuál es el nuevo promedio?.
- c) Se sabe que la relación de los ingresos de este grupo familiar y los gastos en alimentación durante el año 2001 estuvo dada por:

$$I = 1.5 G + 75$$
 Se desea saber cual es el promedio de los ingresos durante el año 2001.
- d) ¿Cuál es el gasto en alimentación que divide la muestra en partes iguales?.
- e) ¿Qué porcentaje de familias que gastaron menos de 300 mil pesos?.

11. En un análisis de las llamadas telefónicas que salían a diario de una oficina, se determinó que 64 llamadas tenían un promedio de 2,3 minutos; 47 llamadas de entre 3 a 10 minutos, promediaron 6,1 minutos y 4 llamadas de mas de 10 minutos demoraron en promedio 20,6 minutos. ¿cuál es el promedio d la duración de estas llamadas?.

12. En un año de biología hay 20 alumnos de 1^{er} año, 18 de 2^o año y 12 de 3^{er} año. Los promedios de notas en los alumnos de 1^{er} y 2^o son 68 y 75 respectivamente. ¿Cuál es el promedio de los alumnos de 3^{er} año, si el promedio de toda la clase fue de 74,84 puntos?.

13. La siguiente tabla corresponde al monto de compras realizadas por la cartera de cliente de una empresa en el 2001.

Ventas x (miles\$)	Nº de clientes
1,0 – 2,5	14
2,5 – 4,0	18
4,0 – 5,5	22
5,5 – 7,0	33
7,0 – 8,5	24
8,5 – 10,0	9

- a) Suponiendo que las compras del año 2002 están dadas por la función $y = 1,2x + 30$. Calcular el promedio y varianza para el año 2002.
- b) ¿Qué porcentaje de clientes compro mas de 7,5 (miles de \$) el año 2001?.
- c) ¿Cuál es le monto máximo del 25% de los clientes que compraron menos el año 2002?.

Sesión 5

Tema: Estadígrafos de dispersión.

I. **Objetivos de la sesión:** que el alumno aprenda a calcular e interpretar los diferentes estadígrafos de dispersión.

II. **Tema:**

1. Estadígrafos de dispersión.

La idea de dispersión se relaciona con la mayor o menor concentración de los datos en torno a un valor central, generalmente la media aritmética.

1.1 Varianza:

Es la media aritmética de los cuadrados de las desviaciones. Se denota por S^2 . Este valor cuantifica el grado de dispersión o separación de los valores de la distribución con respecto a la media aritmética. A mayor dispersión mayor valor de la varianza, a menor dispersión menor valor de la varianza.

- **Datos no tabulados:** tomamos como ejemplo los siguientes datos.

5, 8, 6, 7, 5, 6, 5

en primer lugar, se debe calcular el promedio aritmético:

$$\bar{X} = \frac{5 + 8 + \dots + 5}{7} = \frac{42}{7} = 6$$

La fórmula para calcular la varianza es

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

$$S^2 = \frac{(5 - 6)^2 + (8 - 6)^2 + (6 - 6)^2 + \dots + (5 - 6)^2}{7}$$

$$S^2 = \frac{1 + 4 + 0 + \dots + 1}{7} = \frac{8}{7} = 1,14$$

se puede calcular la varianza por una fórmula mas reducida que la anterior:

$$S^2 = \frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2$$

En ejemplo:

X	5	4	6	8	8	6	5	
X ²	25	16	36	64	64	36	25	= 266

$$S^2 = \frac{266}{7} - 6^2 = 2,0$$

- **Datos tabulados:** en los de tener los datos en una tabla de frecuencia, el calculo de la varianza se hace a través de la siguiente fórmula.

$$S^2 = \frac{\sum_{i=1}^n (Y_i - \bar{X})^2 \cdot f_i}{n}$$

o bien es reducida.

$$S^2 = \frac{\sum_{i=1}^n Y_i^2 \cdot f_i}{n} - \bar{X}^2$$

Para el ejemplo se usara la tabla 2, que se vuelve a repetir en este capítulo:

Intervalos Y _{i-1} – Y _{i+1}	Y _i	f _i	Y _i ²	Y _i ² · f _i
[60 – 70)	65	3	4225	12675
[70 – 80)	75	5	5625	28125
[80 – 90)	85	7	7225	50575
[90 – 100)	95	11	9025	99275
[100 – 110)	105	8	11025	88200
[110 – 120)	115	6	13225	72350
		40		351200

$$S^2 = \frac{351200}{40} - (93,5)^2$$

$$S^2 = 8780 - 8742,25$$

$$S^2 = 37,75$$

- Propiedades de la varianza.

1) La varianza es siempre un valor positivo

$$S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 \geq 0$$

2) Sea $Y = a \pm x$, entonces $V(Y) = V(a \pm x) = V(x)$. Si a una variable se le suma o resta una constante, la varianza permanece igual.

3) Sea $Y = a \cdot x$, entonces $V(Y) = V(a \cdot x) = a^2 V(x)$. Si una variable se le multiplica por una constante, la varianza cambia multiplicándose por la constante al cuadrado.

Ejemplo: 80 empleados de una compañía tienen un sueldo promedio de \$125.000 y una varianza de \$12.000. Si reciben un reajuste del 20%, calcular la nueva varianza.

Si se realizara un reajuste del 20% la constante será 1,2

$$\begin{aligned} V(x \cdot 1,2) &= (1,2)^2 \cdot V(x) \\ &= 1,44 \cdot 12.000 \\ &= \$17.280 \end{aligned}$$

2.2 Desviación estándar:

Se designa la varianza por la letra S y se define como la raíz de la varianza.

$$S = \sqrt{S^2}$$

En el ejemplo de la tabla 2 la desviación estándar es $S = \sqrt{204} = 14,28$

La desviación estándar es mas usada que la varianza. Una de sus utilidades es medir la concentración de los datos respecto a la media aritmética.

Para distribuciones normales:

- el 68,27% de los datos están comprendidos en el rango $\bar{X} - S, \bar{X} + S$
- el 95,45% de los datos están comprendidos en el rango $\bar{X} - 2S, \bar{X} + 2S$
- el 99,73% de los datos están comprendidos en el rango $\bar{X} - 3S, \bar{X} + 3S$

2.3 Coeficiente de variación:

Las medidas de dispersión que se han estudiado anteriormente son medidas absolutas y se expresan en las mismas unidades con las que se mide la variable.

Si se necesita comparar dos o más grupos de datos medidos con diferentes unidades, por lo general, no es posible la comparación utilizando la dispersión absoluta. Por ejemplo, una serie de precios en dolares con una serie de precios en pesos.

Para estos casos se usa la dispersión relativa:

$$\text{Dispersión relativa} = \frac{\text{dispersión absoluta}}{\text{Media}}$$

Si en el caso particular de usar la desviación estándar (S) como dispersión absoluta y la media aritmética (\bar{X}), recibe el nombre de coeficiente de variación:

$$C_v = \frac{S}{\bar{X}} \cdot 100$$

Donde S = desviación estándar
 \bar{X} = promedio aritmético

Ejemplo: Supongamos que un grupo de profesionales de una empresa A tienen un sueldo promedio de \$500.000 y una varianza de \$12.000. En otra empresa, B, otro grupo de profesionales tienen un sueldo de \$600.000 y una desviación de \$18.000.

$$C_A = \frac{12.000}{500.000} \cdot 100 = 2,4$$

$$C_B = \frac{18.000}{600.000} \cdot 100 = 3,0 \longrightarrow \text{presenta mayor dispersión}$$

Ejemplo: Un curso dio un examen de estadísticas y otro un examen de álgebra con los siguientes resultados:

Estadística

Álgebra

$$\bar{X} = 78$$

$$S = 8,0$$

$$\bar{X} = 73$$

$$S = 7,6$$

¿ Qué curso tuvo mayor dispersión?

$$COEF(estadística) = \frac{8}{78} = 0,1025$$

$$COEF(álgebra) = \frac{7,6}{73} = 0,1041$$

El curso que dio el examen de álgebra tuvo mayor dispersión.

2.4 Desviación media.

La desviación media es una medida de dispersión bastante objetiva, es decir mientras mas dispersos están los datos mayor es la desviación media, pero no muestra si están bajo \bar{X} , ya que tomamos su valor absoluto.

- **Datos no tabulados:**

$$DM = \frac{\sum_{i=1}^n |X - \bar{X}|}{n}$$

- **Datos tabulados:**

$$DM = \frac{\sum_{i=1}^n |X - \bar{X}| \cdot f_i}{n}$$

En el ejemplo para la tabla 2 el calculo de DM es el siguiente, sabiendo que $\bar{X} = 93,5$

Intervalos $Y_{i-1} - Y_{i+1}$	Marca de clase Y_i	Frec. Absoluta f_i	$ X_i - \bar{X} $	$ X_i - \bar{X} \cdot f_i$
[60 – 70)	65	3	28,5	85,5
[70 – 80)	75	5	18,5	92,5
[80 – 90)	85	7	8,5	59,5
[90 – 100)	95	10	1,5	15,0
[100 – 110)	105	9	11,5	103,5
[110 – 120)	115	6	21,5	129,0
Total		40		485

$$DM = \frac{485}{40} = 12,125$$

2.5 Variable normalizada o estandarizada.

A diferencia del coeficiente de variación, que compara dos grupos, esta variable sirve para comparar un individuo dentro de su mismo grupo. Su fórmula es:

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

Ejemplo: Un curso dio un examen de estadísticas y otro un examen de álgebra con los siguientes resultados:

$$\begin{array}{l} \text{Estadística} \\ \bar{X} = 78 \\ S = 8,0 \end{array}$$

$$\begin{array}{l} \text{Álgebra} \\ \bar{X} = 73 \\ S = 7,6 \end{array}$$

Un estudiante que dio ambos exámenes sacó 75 puntos en estadísticas y 71 en álgebra. ¿En qué asignatura tuvo mejor resultado respecto a su curso.?

$$Z(\text{estadística}) = \frac{75 - 78}{8} = -0,375$$

$$Z(\text{álgebra}) = \frac{71 - 73}{7,6} = -0,263$$

En álgebra el alumno tuvo menor dispersión respecto al promedio del curso

III. Actividad previa: leer textos indicados en bibliografía, referentes al tema.

IV. Actividad post sesión: realizar la guía de ejercicios que a continuación se presenta.

Ejercicios.

1. Supongamos que un grupo de profesionales en un país A tienen un salario promedio de US\$26.888 y varianza US\$14.400. En un país B otro grupo de profesionales con iguales características reciben un salario promedio de US\$8.570 con desviación estándar de US\$80. ¿Cuál grupo de salarios presenta una menor variabilidad?.
2. En un inventario realizado en la bodega de un almacén se encontraron 200 artículos que fueron importados a diferentes precios (en dólares)

X_i	f_i
20,5	20
32,0	30
48,6	50
50,0	60
60,4	40
	200

- a) Calcular la desviación estándar.
 - b) Calcular la desviación media.
 - c) Calcular coeficiente de variación.
3. En el primer semestre de este año 30 empresas tuvieron en promedio \$374 millones en gastos con una varianza de \$80 millones. Por un error cada una de las empresas no contabilizó \$7 millones en los gastos. Corregir el promedio y la varianza.
 4. 80 empleados de una compañía tienen un salario promedio de \$125.000 y una varianza \$12.000. Si reciben un reajuste del 20%. Calcular el nuevo promedio y varianza.
 5. Se administra un antibiótico al ganado para combatir cierta enfermedad, el peso (en gramos) del antibiótico depende del peso del animal, el cual debe ser medido con mucha precisión, puesto que una sobredosis puede ser perjudicial para el animal. A continuación se muestra la distribución de frecuencia del peso de las dosis.

Peso (gramos)	f_i
15 – 20	7
20 – 25	25
25 – 30	31
30 – 35	20
35 – 40	11

- a) Calcular los estadígrafos de posición y dispersión que le parezcan adecuados (no todos), explique su decisión.
- b) Investigadores afirman que una dosis con peso mayor o igual a 30 gr. sería peligroso. Según la información de que dispone, ¿qué porcentaje de la dosis se clasifica como peligrosa?.
- c) Construya histograma y polígono de frecuencias asociado a los datos.
6. Use sus conocimientos para completar las siguientes frases.
- a) Si calculo la media de los valores absolutos de las desviaciones de las observaciones respecto a la media de ellas, obtenga el valor de _____.
- b) Si calculo la media del cuadrado de las desviaciones de las observaciones respecto a la media aritmética de ellas, obtengo el valor de _____.
- c) Si debo comparar el grado de variabilidad de dos series de observaciones, debo utilizar dispersiones _____ que se obtienen dividiendo la _____ por la _____.
- d) Si mido la estatura y el peso de una serie de alumnos, para determinar cual de las dos series de valores tiene mayor grado de variaciones debo utilizar medidas de _____.
- e) Una variable normalizada o _____ o simplemente calificación _____ indicada a cuantas unidades de desviación estándar esta una puntuación _____.
- f) Si $\bar{X} = 21$ y $s = 3$, la puntuación $X = 18$ esta _____.
- g) José obtuvo calificación $Z = -0,82$ y Luis $-0,78$, entonces el resultado obtenido por _____ es mejor que el obtenido por _____.

7. Un postulante presento examen de admisión a dos universidades; en la universidad A obtuvo 325 puntos y en ella la calificación media fue de 305 puntos con desviación estándar de 26. En la universidad B obtuvo 210 puntos y en ella la calificación media fue de 195 puntos con desviación estándar de 18. Halle en que examen fue mejor el resultado.
8. En una prueba deportiva la media para varones es 140 puntos con una desviación estándar 24; para mujeres la media es 162 con una desviación 22. Ana y su hermano Juan participaron en el evento; Juan obtuvo 151 puntos y Ana 171. Hallar:
- a) Quien tuvo el mejor resultado.
 - b) El rango percentil de Juan y el rango de Ana.

Sesión 6

Tema: Combinatoria.

I. Objetivos de la sesión: aprender técnicas de conteo para ser aplicadas en el cálculo de las probabilidades.

II. Tema:

En muchos problemas de probabilidades debemos citar todas las alternativas posibles en una situación dada o por lo menos determinar cuántas posibilidades diferentes existen.

La combinatoria proporciona los procedimientos y formulas necesarias para contar las posibilidades que hay de elegir un conjunto de elementos con determinadas características.

1. Regla de la multiplicación.

Si una operación consta de 2 pasos, de los cuales el primero puede efectuarse de n_1 formas y para cada una de éstas el segundo puede realizarse de n_2 formas, entonces toda la operación se lleva a cabo de $n_1 \cdot n_2$ maneras.

Ej.: Supóngase que alguien desea viajar en 3 tipos de vehículos diferentes para un periodo de vacaciones, a una de 5 ciudades de la costa. Determine el n° de formas diferentes en que se logra esto.

Solución. - el vehículo en que se viajará se puede escoger de 3 formas diferentes
 $n_1 = 3$.
- las ciudades a visitar, se pueden escoger de 5 maneras diferentes
 $n_2 = 5$.

Entonces, el total de combinatoria posibles es $n_1 \cdot n_2 = 3 \cdot 5 = 15$ maneras diferentes de visitar una ciudad en un tipo de vehículo.

Ej.: ¿Cuántos almuerzos diferentes se pueden formar si se componen de: sopa, plato de fondo, postre y bebidas, pudiendo elegir entre 4 tipos de sopas, 3 clases de plato de fondo, 5 tipos de postres y 4 tipos de bebidas diferentes?.

Solución.

El n° de combinatorias posibles de elegir un almuerzo es:
 $4 \cdot 3 \cdot 5 \cdot 4 = 240$ formas diferentes.

2. Permutaciones sin elementos repetidos.

Al tomar todos los elementos de un conjunto finito y ordenarlos de todas las maneras posibles, tendremos una permutación. Es decir, con “n” elementos, se ordenan de todas las maneras, sin dejar ninguno fuera de la ordenación. Ej. Si se tiene las letras a, b y c, sus ordenaciones son:

abc, acb, bac, bca, cab, cba.

En total, 6 maneras de ordenar las 3 letras.

2.1 Para calcular sus valores, se efectúa el siguiente cálculo cuando la ordenación es forma lineal.

$$P_n = n! \quad \text{donde } n! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot n$$

Ej.: ordenar 7 personas en una fila.

$$P_7 = 7! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 = 5040 \text{ maneras.}$$

2.2 Si la ordenación de los n elementos es forma circular, su cálculo cambia a:

$$P_n = (n - 1)! \quad \text{donde } (n - 1)! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot (n - 1)$$

Ej.: ¿de cuántas maneras distintas se pueden sentar 7 personas en una mesa redonda?

$$(7 - 1)! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 = 720 \text{ maneras}$$

3. Permutaciones con elementos repetidas.

En variadas situaciones se deben ordenar n elementos dentro de los cuales hay algunos repetidos. Si se supone que de n elementos, hay p elementos iguales entre si, q elementos iguales entre si y r elementos iguales entre si. La forma de ordenar estos n elementos es:

$$P_{n(p,q,r)} = \frac{n!}{p!q!r!}$$

Ej.: ordenar en fila 7 fichas, de las cuales hay 2 fichas rojas, 4 azules y 1 amarilla.

Solución. Como deben ordenarse las 7 fichas, se concluye que es una permutación, pero con elementos repetidos.

$$P_{7(2,4,1)} = \frac{7!}{2!4!1!} = \frac{5040}{(2)(24)(1)} = 105 \text{ maneras}$$

4. Variaciones sin elementos repetidos.

Se dispone de un elemento, pero se escogen k (algunos) de ellos.

Se llama variación de orden k, a cada grupo que se pone de k elementos escogidos de un total, se tienen algún elemento distinto o al orden en que están ubicados es diferente.

Su cálculo es:

$$V_k^n = \frac{n!}{(n-k)!} \text{ con } n \geq k$$

Ej.: Se toman 4 nombres de los 24 miembros de un club para ocupar los cargos de presidente, vicepresidente, tesorero y secretario. ¿En cuántas formas diferentes puede hacerse?.

Solución. Del total de 24 personas, se escogen grupos de 4, donde el orden en que ocupen un cargo es importante en cada grupo. Por ejemplo, si se escoge a: Pedro, Juan, Diego y José, las combinatorias son:

Presidente	Vicepresidente	Tesorero	Secretario
Juan	Pedro	Diego	José
Pedro	Diego	Juan	José
José	Pedro	Juan	Diego
.	.	.	.
.	.	.	.
.	.	.	.
etc.			

Es decir, los 4 escogidos, se pueden formar varios grupos diferentes.

$$V_4^{24} = \frac{24!}{(24-4)!} = \frac{24!}{20!} = 255024 \text{ maneras diferentes}$$

5. Variaciones con elementos repetidos.

De un total de n elementos, se deben escoger k de ellos para formar un grupo, pero en este caso, algunos elementos están repetidos.

Cada grupo difiere de otro por los elementos que lo componen o por el orden en que están ubicados. Es decir, interesa el orden de los elementos dentro del grupo.

Su cálculo se efectúa a través de:

$$\overline{V}_k^n = n^k$$

Ej.: ¿Cuántos números de 6 cifras existen si se eligen de entre los dígitos 1, 2, 3, 4, 5, 6, 7 y todos comienzan con 3?

$$3 _ _ _ _ _ _ \\ \downarrow \\ \overline{V}_5^7 = 7^5 = 16.807 \text{ números diferentes.}$$

6. Combinaciones sin elementos repetidos.

De un total de n elementos se escogen k formando grupos de orden k. Cada grupo se diferencia de otro si tienen elementos distintos pero no por el orden en que están dispuestos.

Su cálculo se hace a través:

$$C_k^n = \frac{n!}{k!(n-k)!}$$

Ej.: De un grupo de 20 personas, se designa un comité de 3. ¿De cuántas maneras es posible hacerlo?

Solución. No interesa el orden en que se escojan las 3 personas, por lo tanto es una combinación.

$$C_3^{20} = \frac{20!}{3!(20-3)!}$$

7. Combinaciones con elementos repetidos.

De un total de n elementos se escogen k de ellos de los cuales algunos están repetidos. El orden en que están dispuestos los elementos dentro de un grupo no interesa.

Su cálculo es a través de:

$$C_k^{n+k-1} = \frac{(n+k-1)!}{k!(n-1)!}$$

Ej.: Se tienen las letras a, b y c. ¿De cuántas maneras se pueden escoger 2 letras, pudiendo repetirlas y sin importar el orden en que se dispongan?.

Solución. $C_2^{3+2-1} = C_2^4 = \frac{4!}{2!(4-2)!} = 6$ maneras

Elas son: aa, ab, ac, bb, bc, cc

Notar que la dupla ab, es la misma que si la escribimos ba, por lo tanto se cuenta una sola vez.

III. Actividad previa: leer textos indicados en bibliografía, referentes al tema.

IV. Actividad post sesión: realizar ejercicios que se presentan a continuación.

Ejercicios.

1. ¿Cuántos productos distintos se pueden formar con los dígitos 2, 3, 5 y 7? Si se usan:
 - a) 2 factores
 - b) 3 factores
2. ¿De cuántas maneras diferentes pueden colocarse 6 libros uno al lado de otro, en un estante?
3. Un ministerio consta de 25 consejeros y 10 secretarios. ¿Cuántos comités se pueden formar con 5 consejeros y 3 secretarios?
4. El profesor y 25 alumnos se sientan alrededor de una mesa redonda. Si el puesto del profesor es fijo. ¿De cuántas maneras se pueden distribuir los alumnos?
5. ¿En cuántas formas diferentes pueden producirse 6 lanzamientos al aire de una moneda, 2 caras y 4 cruces?
6. Una prueba de opción múltiple consta de 15 preguntas y cada una tiene 3 alternativas. ¿En cuántas formas diferentes puede marcar un estudiante su respuesta a estas preguntas?
7. ¿De cuántas maneras pueden distribuirse 10 personas en 10 asientos dispuestos en filas, si 4 de ellas deben estar siempre juntas?
8. Se fabricó una pulsera con forma de argolla, sin comienzo ni fin. Calcula de cuántas maneras se podrían haber ordenado los 7 colgantes que la ordenan.
9. ¿De cuántas maneras se pueden sentar 10 personas en un banco si hay 4 sitios disponibles?
10. Hay que colocar a 5 hombres y 4 mujeres en una fila de modo que las mujeres que ocupen los lugares pares. ¿De cuántas maneras puede hacerse?
11. Cuatro libros diferentes de matemáticas, 6 de física y 2 de química han de ser colocados en una estantería. ¿Cuántas ubicaciones distintas admiten si:
 - a) los libros de cada materia han de estar juntos, y
 - b) sólo los de matemáticas han de estar juntos?
12. Con 7 consonantes y 5 vocales, ¿cuántas palabras se pueden formar que tengan 4 consonantes distintas y 3 vocales distintas? Se admiten palabras sin significado.

Sesión 7

Tema: Probabilidad clásica.

I. **Objetivos de la sesión:** aprender a realizar cálculos de probabilidades según las leyes que las rigen.

II. **Tema:**

Con frecuencia es útil cuantificar la posibilidad de que se presente un resultado de un experimento aleatorio. “la posibilidad de que llueva hoy es de 30%” es una afirmación que refleja una creencia sobre la posibilidad de que llueva. La posibilidad de un resultado se cuantifica asignándole un número de intervalo $[0,1]$, o un porcentaje del 0 al 100%. Entre mas grande sea el número, mayor es la probabilidad del resultado. Un cero indica que el resultado no se presentara; un uno indica el resultado seguro.

A medida que se desarrolla cualquier rama de la ciencia, siempre surge en cierta etapa la necesidad de tener en consideración de las desviaciones accidentales y su influencia sobre el curso de los fenómenos por estudiar. A causa de eso, toda rama aplicada de la ciencia tiene que acudir a la Teoría de las probabilidades y cada día de amplían mas los campos de aplicación de los métodos estadísticos.

Las probabilidades pertenecen siempre a la ocurrencia a no de eventos, de manera que se explicaran algunos conceptos previos, dentro de los cuales está “evento”.

- **Experimento:** se denomina experimento a un proceso de observación o medición cualquiera. Por ejemplo, lanzar 2 dados y anotar sus resultados.
- **Espacio muestral:** es el conjunto de todos los posibles resultados de un experimento. Siguiendo con el ejemplo anterior, el espacio muestral del lanzamiento de los 2 dados, son todos los pares ordenados que se pueden formar con los 2 dados, es decir $(1,1)$, $(1,2)$, $(1,3)$,..., $(6,6)$. En total 36 pares. Se designa por E.
- **Evento:** es un subconjunto del espacio muestral. Son elementos del espacio muestral que cumplen con alguna condición dada. Por ejemplo, el evento “que los dados sumen 7”. Los pares ordenados que cumplen esta condición son: $(2,5)$, $(5,2)$, $(3,4)$ y $(4,3)$. Los eventos se anotan con letras del abecedario mayúsculas, ejemplo A, B, C, ..., etc.

- Probabilidad de un evento: Los eventos que designaron con las letras A, B, C,... tienen una probabilidad asociada a cada uno de ellos y se designa como P(A), P(B), P(C),..., etc.

La probabilidad de un evento es un número real no negativo y no superior a 1, que mide la posibilidad que suceda dicho evento.

Se cumple que:

- a) $P(A) \geq 0$
- b) $P(E) = 1$
- c) $0 \leq P(A) \leq 1$
- d) $P(A) + P(A^C) = 1$ donde A^C es el complemento de A.

La probabilidad de un suceso es uno de los conceptos fundamentales de la teoría de las probabilidades. Se puede decir que las probabilidades es un número que caracteriza la posibilidad de que se produzca un suceso.

Se llama probabilidad de un suceso, desde el punto de vista clásico, al cociente del número de resultados que son favorables al suceso y el número total de resultados elementales e igualmente posibles del experimento.

Una probabilidad se calcula como una frecuencia relativa, es decir:

$$P(A) = \frac{\text{N}^\circ \text{ de casos posibles del evento A}}{\text{N}^\circ \text{ de casos totales}}$$

Ej.: En un mazo de naipes barajado que contiene 4 ases, dentro de las 52 cartas, la probabilidad de obtener un as en una extracción (evento A) es:

$$P(A) = \frac{4 \text{ ases}}{52 \text{ cartas en total}} \\ = \frac{4}{52} = \frac{1}{13}$$

- Eventos mutuamente excluyentes: dos o más eventos son mutuamente si no puede ocurrir al mismo tiempo, es decir, la ocurrencia de un evento A impide automáticamente la ocurrencia de un evento B.

Ej.: Al extraer una carta de un naipe de 52 cartas. Sea A = sacar un as.
B = sacar un rey.

Podemos ver que A y B son mutuamente excluyentes, ya que al sacar solo una carta, ella puede ser el as o el rey pero no saldrán ambos.

En cambio si A = sacar un as
B = sacar un trébol

Se puede notar que A y B no son excluyentes, ya que al extraer una carta, ella puede ser un as y a la vez puede ser un trébol. Una no excluye a la otra.

- Eventos independientes: dos eventos son independientes cuando la ocurrencia de un evento no tiene ningún efecto en la ocurrencia de otro. La probabilidad de un evento no influye e la probabilidad del otro.

Ej.: Si se lanza un dado, se puede calcular la probabilidad que suceda un 5. Si se lanza por 2ª vez, la probabilidad, que suceda un 4 no tiene ninguna influencia anterior que ya salió. Es decir, el n° del 2º lanzamiento es independiente del que salió en el 1º lanzamiento.

- Reglas de la adición: las reglas de la adición se cumplen cuando la probabilidad que ocurra un evento u otro. Existen 3 casos a considerar:

- a) Si los eventos son mutuamente excluyentes, entonces:

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

- b) Si los eventos no son mutuamente excluyentes, entonces:

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

Si se extiende a 3 eventos, entonces:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C)$$

- c) Si los eventos son independientes, entonces:

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

Ej.: Se extrae una carta de un mazo de 52 naipes. ¿Cuál es la probabilidad que salga un as o un rey?

Solución. Sea A: un as
B: un rey

Como en este caso A y B son mutuamente excluyentes, entonces:

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) \\ &= \frac{4}{52} + \frac{4}{52} \\ &= \frac{2}{13} \end{aligned}$$

Ej.: Se extrae una carta de un mazo de 52 naipes. ¿Cuál es la probabilidad que salga un as o un trébol?

Solución. A: un as
B: un trébol

Los eventos A y B no son excluyentes, entonces:

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

$$= \frac{4}{52} + \frac{13}{52} - \frac{1}{52}$$

$$= \frac{4}{13}$$

- **Regla de la multiplicación:** se usa para calcular la probabilidad que suceda 2 eventos en forma conjunta, es decir, los 2 eventos a la vez. También hay que considerar 2 casos:

a) Si los eventos son no excluyentes, entonces:

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

b) Si los eventos son independientes, entonces:

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

Ej.: Si se lanzan 2 dados 2 veces seguidas. Calcular la probabilidad que en ambos lanzamientos, la suma de sus caras sea 7.

Solución. Cada lanzamiento es independiente del otro.

- Sea A: la suma es 7 en el 1^{er} lanzamiento \longrightarrow (1,6) (6,1) (2,5) (5,2) (3,4) (4,3)

$$P(A) = \frac{6}{36} = \frac{1}{6}$$

- Sea B: la suma es 7 en el 2^o lanzamiento.

$$P(B) = \frac{1}{6}$$

Entonces $P(A \cap B) = P(A) \cdot P(B)$

$$= \frac{1}{6} \cdot \frac{1}{6}$$

$$= \frac{1}{36}$$

- **Probabilidad condicional:** se dice que una probabilidad es condicional cuando al tener 2 eventos, ya se conoce el resultado de uno de ellos, es decir, se sabe con anterioridad lo que sucedió con uno de ellos. Se denota $P(A/B)$ y se lee "la probabilidad que suceda A, dado que ya sucedió B". Su cálculo dependerá si los eventos son independientes o no lo son.

a) $P(A/B) = \frac{P(A \cap B)}{P(B)}$ si A y B no son independientes.

b) $P(A/B) = P(A)$ si A y B son independientes.

Ej.: Supóngase que una oficina tiene 100 máquinas calculadoras. Algunas de esas máquinas son eléctricas (E) y otras son manuales (M). Además, algunas

son nuevas (N), mientras que otras son usadas (U). Una persona entra en la fabrica, escoge una máquina al azar y descubre que es nueva. ¿Cuál es la probabilidad que la máquina sea eléctrica?.

	E	M	Total
N	40	30	70
U	20	10	30
Total	60	40	100

Solución. Es una probabilidad condicional porque se sabe previamente que la máquina escogida es nueva. Entonces, se expresa como:

$$\begin{aligned}
 P(E / N) &= \frac{P(E \cap N)}{P(N)} \\
 &= \frac{40}{70} = \frac{4}{7}
 \end{aligned}$$

Ej.: Se lanzan 2 dados. Si la suma de sus caras es 7. ¿Cuál es la probabilidad que uno de los dados sea 2?.

Solución. Al igual que el ejemplo anterior, también es una probabilidad condicional, ya que se sabe de antemano que la suma de las caras dio 7, y ese dato pasa a ser la condición dada.

Sea A: que un dado sea 2

B: la suma de ambas caras es 7

Entonces:

$$P(A / B) = \frac{P(A \cap B)}{P(B)}$$

De todos los pares que sumen 7: (1,6) (6,1) (2,5) (5,2) (3,4) (4,3), solo 2 pares de ellos tienen un número 2 en uno de sus dados. Por lo tanto, solo los pares (2,5) (5,2) cumplen los 2 eventos mencionados, luego:

$$P(A / B) = \frac{2}{36} = \frac{1}{18}$$

- Probabilidad total: si un suceso A está participando en n sucesos excluyentes, la probabilidad de A resulta ser una suma de las probabilidades de los n sucesos.

Sea $A = B_1 + B_2 + \dots + B_n$

Donde $B_i \cap B_j = \emptyset \quad (i \neq j)$

Entonces:

$$P(A) = P(A/B_1) \cdot P(B_1) + P(A/B_2) \cdot P(B_2) + \dots + P(A/B_n) \cdot P(B_n)$$

Ej.: Supongamos que tenemos 3 cajas C_1, C_2, C_3 y que tienen igual probabilidad de ser escogidas. Cada caja tiene 10 bolitas entre blancas y negras. La caja C_1 tiene 3 bolitas blancas, la caja C_2 tiene 2 bolitas blancas y la caja C_3 tiene 5 bolitas blancas.

Se escoge una caja y luego una bolita. ¿Cuál es la probabilidad que salga blanca?

Solución. Sea A: sacar bolita blanca

Podemos ver que el evento A está particionado en 3 subgrupos, ya que hay bolitas blancas en cualquiera de las 3 cajas. Por lo tanto, se debe considerar que una bolita es blanca si se escoge de 3 lugares diferentes. entonces:

$$\begin{aligned} P(A) &= P(A/C_1) \cdot P(C_1) + P(A/C_2) \cdot P(C_2) + P(A/C_3) \cdot P(C_3) \\ &= \frac{2}{10} \cdot \frac{1}{3} + \frac{3}{10} \cdot \frac{1}{3} + \frac{5}{10} \cdot \frac{1}{3} \\ &= \frac{2}{30} + \frac{3}{30} + \frac{5}{30} \\ &= \frac{1}{3} \end{aligned}$$

El diagrama de árbol es un método para graficar eventos secuenciales o probabilidades totales, como se vera en el siguiente ejemplo:

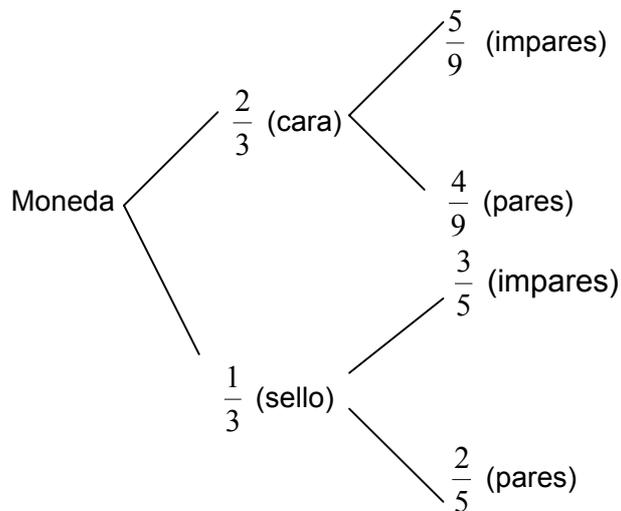
Ej.: Se lanza una moneda cargada, de modo que $P(\text{cara}) = \frac{2}{3}$ y $P(\text{sello}) = \frac{1}{3}$.

Si sale cara, se escoge un número al azar entre 1 y 9. Si sale sello se escoge un número al azar entre 1 y 5. Hallar la probabilidad de escoger un número par.

Solución. El evento A: escoger un n° par.

Puede suceder de 2 particiones posibles, dependiendo si sucedió cara o sello. Por tanto, es una probabilidad total.

En el siguiente diagrama muestra las probabilidades según el caso:



Luego:

$$\begin{aligned}
 P(A) &= P(A/C) \cdot P(C) + P(A/S) \cdot P(S) \\
 &= \frac{4}{9} \cdot \frac{2}{3} + \frac{2}{5} \cdot \frac{1}{3} \\
 &= \frac{8}{27} + \frac{2}{15} \\
 &= \frac{58}{135}
 \end{aligned}$$

- **Teorema de Bayes:** considerando nuevamente un evento A, particionado en n partes, en que cada una es excluyente de otra. El teorema de Bayes permite calcular la probabilidad de que el evento A haya sucedido dentro de una de las particiones, sabiendo que ya ocurrió.

Tomando el ejemplo de las 3 cajas C_1 , C_2 y C_3 , anteriormente, con la probabilidad total se calculó la probabilidad que una bolita extraída al azar sea blanca.

En este caso, la pregunta cambia de la siguiente manera: ¿cuál es la probabilidad que la bolita que se sacó provenga de la caja C_1 sabiendo que fue blanca?

Podemos observar que se conoce un antecedente a priori, por lo tanto pasa a ser una condición dada. El objetivo en este caso es conocer la partición en particular de donde provino la bolita blanca.

El teorema de Bayes se aplica de la siguiente manera:

$$P(C_1 / A) = \frac{P(A/C_1) \cdot P(C_1)}{P(A/C_1) \cdot P(C_1) + P(A/C_2) \cdot P(C_2) + P(A/C_3) \cdot P(C_3)}$$

Se puede observar que el denominador de esta expresión es la probabilidad total.

En el ejemplo:

$$P(C_1 / A) = \frac{\frac{2}{10} \cdot \frac{1}{3}}{\frac{2}{10} \cdot \frac{1}{3} + \frac{3}{10} \cdot \frac{1}{3} + \frac{5}{10} \cdot \frac{1}{3}} = \frac{\frac{2}{30}}{\frac{10}{30}} = \frac{1}{5}$$

Ej.: Dos cajas contienen clavos grandes y clavos pequeños. Suponga que la caja A contiene 60 clavos grandes y 40 pequeños. La caja B contiene 10 clavos grandes y 20 pequeños. Se extrae un clavo y resulta ser grande. ¿Cuál es la probabilidad que ase haya sacado de la caja B?

Solución. Sea G: el clavo es grande.

Se conoce de antemano el tipo de clavo y se considera como una condición dada.

$$P(B / G) = \frac{P(G / B) \cdot P(B)}{P(G / A) \cdot P(A) + P(G / B) \cdot P(B)}$$
$$= \frac{\frac{10}{30} \cdot \frac{1}{2}}{\frac{60}{100} \cdot \frac{1}{2} + \frac{10}{30} \cdot \frac{1}{2}} = \frac{\frac{1}{6}}{\frac{1}{3} + \frac{1}{3}} = \frac{1}{4}$$

III. Actividad previa: leer textos indicados en bibliografía, referentes al tema.

IV. Actividad post sesión: realizar la guía de ejercicios que a continuación se presenta.

Ejercicios.

1. Dos hombres H_1 y H_2 y 3 mujeres M_1 , M_2 , M_3 participan en un campeonato de ajedrez. Aquellos de igual sexo tiene la misma probabilidad de ganar. Pero los hombres tienen el doble de posibilidad de ganar que las mujeres. Encuentre la probabilidad que gane una mujer.
2. Un lote contiene 10 artículos buenos, 4 con defectos secundarios y 2 con defectos importantes. Se toma al azar un artículo. Determine la probabilidad que el artículo:
 - a) no tenga defectos
 - b) no tenga defectos importantes

Del mismo lote se extraen 2 artículos al azar. Determine la probabilidad que:

 - a) ambos sean buenos si se extrae uno después del otro
 - b) ambos tengan defectos importantes
3. Una compañía de artículos para computadores colocará un anuncio de su nuevo módem en una revista de computación. La compañía cree que el anuncio será leído por un 32% de los lectores de la revista y que el 2% de aquellos que lean el anuncio comprarán el módem. Encontrar la probabilidad de que el lector de la revista lea el anuncio y compre el módem.
4. Dos estudiantes A y B están inscritos en un curso. Si el estudiante A asiste a las clases el 80% de las veces y el estudiante B el 60%, y las ausencias de ellos son independientes. ¿Cuál es la probabilidad de que al menos uno de los 2 estudiantes esté en clases un día determinado?.
5. La probabilidad de que un doctor diagnostique de manera correcta una enfermedad particular es 0.7. Dado que el doctor hace un diagnóstico incorrecto, la probabilidad de que el paciente presente una demanda es 0.9. ¿Cuál es la probabilidad de que el doctor haga un diagnóstico incorrecto y el paciente lo demande?.
6. En una fábrica de pernos, las máquinas 1, 2, 3 fabrican el 25%, 35% y 40% respectivamente de la producción total. De lo que producen, el 5%, 4% y 2% respectivamente son defectuosos. Se escoge un perno al azar y se encuentra

que es defectuoso. ¿Cuál es la probabilidad que haya sido producido por la máquina 1?

7. Si se detiene a 3 personas, seleccionadas al azar, en una calle, ¿cuáles son las probabilidades de que:
 - a) todas ellas hayan nacido en viernes;
 - b) dos de ellas hayan nacido en viernes y otra en martes;
 - c) ninguna de ellas haya nacido en lunes?.

8. Determinar la probabilidad p de los siguientes sucesos:
 - a) Una tirada de un dado resulte par.
 - b) Al menos una cara en dos tiradas de una moneda.
 - c) Un as, el 10 de diamantes o el 2 de picas aparezcan al sacar una sola carta de una baraja francesa de 52 naipes.
 - d) La suma de dos dados sea 7.
 - e) Que aparezca sello en la próxima tirada de una moneda si han salido 56 caras en 100 tiradas previas.

9. De una baraja de 52 naipes, mezclados al azar, se sacan dos naipes. Hallar la probabilidad de que ambas sean ases si la primera carta:
 - a) se devuelve al mazo
 - b) si no se devuelve.

10. Tres joyeros idénticos tiene dos cajones cada uno. Cada cajón del primero contiene un reloj de oro, cada uno del segundo un reloj de plata y en un cajón del tercero hay un reloj de oro y en el otro uno de plata. Si se selecciona un joyero al azar, abrimos uno de los cajones y en el hay un reloj de plata, ¿cuál es la probabilidad de que en el otro cajón haya un reloj de plata? (Utilizar la fórmula de Bayes).

11. Se tienen dos armas A y B. La probabilidad de dar con ellas en el blanco son $P(A) = 0,6$ y $P(B) = 0,8$. se dispara con ambas armas, determinar la probabilidad de dar en el blanco con una sola de ellas.
12. Se escogen al azar 3 lámparas entre un total de 15, de las cuales 5 son defectuosa. Hallar la probabilidad que:
- ninguna sea defectuosa
 - una exactamente sea defectuosa
 - una por los menos sea defectuosa.
13. Solo el 60% de los estudiantes de la clase de estadística del profesor H pasaron la 1ª prueba. De quienes pasaron, el 80% estudiaron, el 20% de quienes no pasaron si estudiaron. ¿Debería usted estudiar para la prueba del profesor?.
14. Si se tiene tres resultados definidos en un espacio muestral S y se conoce que:
 $P(A) = 0,40$; $P(B) = 0,42$; $P(C) = 0,15$; $P(A/B) = 0$; $P(A/C) = 0$; $P(C/B) = 0,11$
Diga si:
- A y B son independientes.
 - A y C son mutuamente excluyentes.
 - B y C son independientes.

Sesión 8

Tema: Variable aleatoria.

I. Objetivos de la sesión: conocer el concepto de variable aleatoria y su función de probabilidad, tanto en variable continua como en discreta.

II. Tema:

1. Concepto de variable aleatoria.

El espacio muestral son todos los posibles resultados de un experimento aleatorio. Estos resultados pueden ser cuantitativos o cualitativos.

Por ejemplo, si en un experimento se lanza una moneda, sus resultados cualitativos son: cara, sello.

Es útil transformar estos resultados a medidas numéricas, para facilitar el estudio de su comportamiento aleatorio.

El concepto de variable aleatoria proporciona una medida para relacionar cualquier resultado con una medida numérica (real), es decir, la variable que asocia un número con el resultado de un experimento aleatorio se conoce como variable aleatoria.

Ej. Si el espacio muestral es $= \{\text{cara, sello}\}$. Se define "x" como la variable aleatoria donde $x(\text{cara}) = 0$ y $x(\text{sello}) = 1$. Transformándose así, los 2 resultados en puntos sobre la recta de los reales.

Otro ejemplo:

Si se lanzan 2 dados, el espacio muestral es $(1,1) (1,2) \dots (6,6)$. En total 36 pares como resultados. Si se define la variable aleatoria $x = \text{suma de las caras}$, entonces $x = 2,3,4, \dots 12$. En este caso hay variaciones pares asociados a un mismo valor, ya que varios suman un mismo número.

En tal caso, cada par tiene asociado algún valor numérico. Esto se muestra en la siguiente tabla:

Resultados.	X = suma de las caras.
(1,1)	2
(1,2) (2,1)	3
(1,3) (2,2) (3,1)	4
.	.
.	.
.	.
(6,6)	12

- Si la variable aleatoria x tiene valores enteros, entonces se dice que x es discreta.
- Si la variable aleatoria x tiene valores reales, entonces se dice que x es continua.

Ej.: se sacan 2 bolitas de manera sucesiva, sin reemplazo, de una caja que contiene 4 bolitas rojas y 3 bolitas negras. Se define la variable aleatoria $x = n^\circ$ de bolitas rojas, entonces

$$\begin{array}{cccc} \text{Espacio muestral} = \{RR, RB, BR, BB\} & & & \\ \downarrow & \downarrow & \downarrow & \downarrow \\ X = 2, & 1, & 1, & 0 \end{array}$$

Como x tiene valores enteros, x es variable aleatoria discreta. El conjunto de valores de x puede ser finito o infinito, como se muestra en el siguiente ejemplo:

Ej.: se lanza un dado, hasta que ocurre un n° 5 por 1^{ra} vez. (S = si, N = no)

$$\text{Espacio muestral} = \{S, NS, NNS, NNNS, \dots\}$$

Se define $x = n^\circ$ de lanzamientos necesarios hasta que sale un 5. Entonces $x = 1, 2, 3, 4, 5, \dots$ en este caso, x es infinito.

2. Distribución de probabilidades para variable aleatoria discreta.

Dada una variable aleatoria, se llama función de probabilidad a la expresión matemática que permite calcular la probabilidad asociada a cada valor de la variable.

En algunos casos la función de probabilidad no0 es posible expresarla con la ayuda de una función matemática, entonces es necesario calcularla directamente.

Distribución de probabilidad se refiere a la colección de valores de la variable y a la función de probabilidad de éstos.

Una función de probabilidad debe cumplir

$$1) p(x) \geq 0 \quad \forall x \in X$$

$$2) \sum p(x) = 1$$

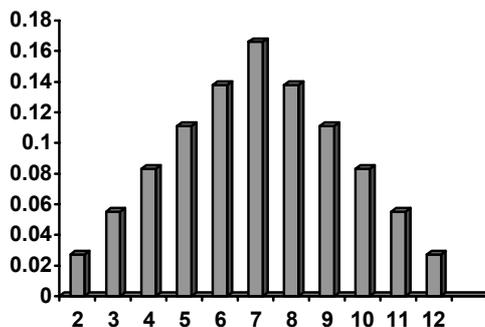
Se define la función probabilidad acumulativa de la variable x a la probabilidad que x sea menos o igual a un valor específico.

$$F(X) = P(X \leq x) = \sum^x p(x)$$

Ej.: Si se lanzan 2 dados y x es la suma de sus caras

Espacio Muestral	X	P(X = x)
(1,1)	2	1/36
(1,2) (2,1)	3	2/36
(1,3) (2,2) (3,1)	4	3/36
(1,4) (4,1) (2,3) (3,2)	5	4/36
.	.	.
.	.	.
.	.	.
(5,6) (6,5)	11	2/36
(6,6)	12	1/36
		$\sum P(X = x) = 1$

Podemos observar que cada valor de probabilidad es positivo y la suma es 1.



Si se pide: $P(X \leq 4) = P(x=2) + P(x=3) + P(x=4)$

$$\begin{aligned} F(4) &= \frac{1}{36} + \frac{2}{36} + \frac{3}{36} \\ &= \frac{1}{6} \end{aligned}$$

Si calculamos $F(12)$ notamos que su valor es 1 ya que acumula todos las probabilidades de la variable aleatoria.

Entonces, se cumple que:

1) $0 \leq F(x) \leq 1$

2) $F(x_i) \geq F(x_j)$ si $x_i > x_j$

3) $P(X \geq x) = 1 - F(x)$

4) $P(X = x) = F(x) - F(x - 1)$

5) $P(x_i \leq X \leq x_j) = F(x_j) - F(x_i)$

Ej.: Usando el mismo ejemplo anterior.

a) Calcular $P(X = 3)$

Si se observa la tabla $P(X = 3) = \frac{2}{36} = 1$ pero también se puede calcular:

$$\begin{aligned} P(X = 3) &= F(3) - F(2) \\ &= \frac{3}{36} - \frac{1}{36} \\ &= \frac{2}{36} \end{aligned}$$

donde $F(3) = \frac{1}{36} + \frac{2}{36}$

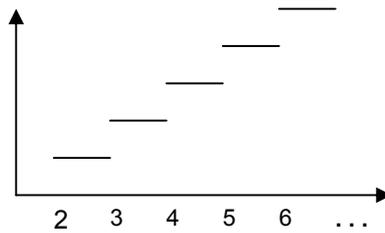
$$F(2) = \frac{1}{36}$$

b) Calcular $P(2 \leq X \leq 5) = F(5) - F(2)$

Donde $F(5) = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} + \frac{4}{36} = \frac{10}{36}$

$$F(2) = \frac{1}{36}$$

entonces $P(2 \leq X \leq 5) = \frac{10}{36} - \frac{1}{36} = \frac{9}{36}$



3. Distribución de probabilidad para variable aleatoria continua.

La distribución de probabilidades de una variable aleatoria que toma valores reales (continua), está caracterizada por una función $f(x)$ que recibe el nombre de “función de densidad de probabilidad”.

Esta función $f(x)$ no es la misma función de probabilidades que se usa para el caso discreto.

Para una variable x continua, la probabilidad que x tome un valor específico es cero, es decir, $P(X = x) = 0$, se considera entonces la variable dentro de un intervalo $a \leq X \leq b$.

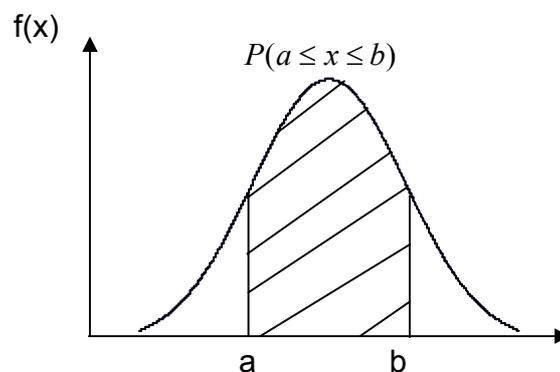
Si $f(x)$ es una función de densidad, entonces se cumple:

$$1) f(x) \geq 0 \quad -\infty < x < \infty$$

$$2) \int_{-\infty}^{\infty} f(x) dx = 1$$

$$3) P(a \leq x \leq b) = \int_a^b f(x) dx$$

El gráfico de la función densidad es una curva continua, cuya área bajo esta curva vale 1.



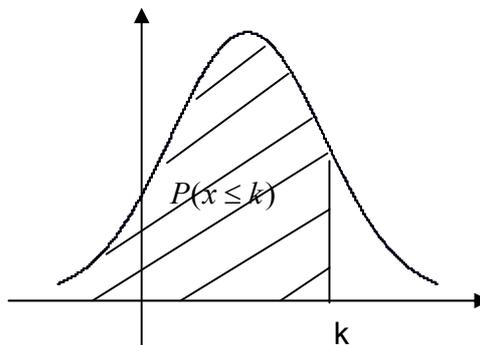
La probabilidad del intervalo $a \leq X \leq b$ es el área bajo la curva $f(x)$, acotada entre a y b .

La función distributiva acumulativa de una variable aleatoria continua es el área bajo la curva a la izquierda de un valor determinado.

$$F(X) = P(X \leq x) = \int_{-\infty}^x f(t) \cdot dt$$

Se cumple que:

- 1) $F(-\infty) = 0$
- 2) $F(\infty) = 1$
- 3) $P(a \leq X \leq b) = F(b) - F(a)$
- 4) $\frac{dF(x)}{dx} = f(x)$



Ej.: la variable continua que representa la proporción de accidentes automovilísticos fatales en E.E.U.U. tiene la siguiente función de densidad de probabilidad.

$$F(x) = \begin{cases} 42x(1-x)^5 & \text{si } 0 < x \leq 1 \\ 0 & \text{para otro valor} \end{cases}$$

¿Cuál es la probabilidad que no más de 25% de los accidentes automovilísticos sean fatales?.

Solución: se pide $P(x \leq 0,25)$

$$\begin{aligned} F(x) = P(X \leq x) &= \int_0^x 42x(1-x)^5 dx \\ &= 42 \left(\frac{x^2}{2} - \frac{5x^3}{3} + \frac{10x^4}{4} - \frac{10x^5}{5} + \frac{5x^6}{6} - \frac{x^7}{7} \right) \\ &= 21x^2 - 70x^3 + 105x^4 - 84x^5 + 35x^6 - 6x^7 \end{aligned}$$

entonces:

$$\begin{aligned} F(0,25) = P(X \leq 0,25) &= 21(0,25)^2 - 70(0,25)^3 + \dots - 6(0,25)^7 \\ &= 0,5551 \end{aligned}$$

Ej.: Sea x la variable aleatoria continua, de todos los valores posibles que puede tener la vida útil de un transmisor.

$$f(x) = \begin{cases} \frac{1}{2} e^{-\frac{x}{2}} & , x > 0 \\ 0 & , \text{en otro valor} \end{cases}$$

Calcular la probabilidad de que la vida útil de una batería determinada de este tipo sea menor que 200 o mayor que 400 horas.

Solución: Sea el evento A: x es menor que 2
B: x sea mayor que 4

Como A y B son mutuamente excluyentes

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) \\ &= \int_0^2 \frac{1}{2} e^{-\frac{x}{2}} dx + \int_4^{\infty} \frac{1}{2} e^{-\frac{x}{2}} dx \\ &= (1 - e^{-1}) + (e^{-2}) \\ &= 1 - 0,368 + 0,135 \\ &= 0,767 \end{aligned}$$

Para el mismo ejemplo, calcular la probabilidad de que una batería de este tipo dure mas de 300 horas dado que ya ha estado en uso durante mas de 200 horas.

Esto es: $P(X > 3 / X > 2) = \frac{P(X > 3 \wedge X > 2)}{P(X > 2)}$

Pero la intersección en $P(X > 3 \wedge X > 2)$ es $P(X > 3)$

Entonces:

$$P(X > 3 / X > 2) = \frac{P(X > 3)}{P(X > 2)}$$

$$\begin{aligned} & \frac{\int_3^{\infty} \frac{1}{2} e^{-\frac{x}{2}} dx}{\int_2^{\infty} \frac{1}{2} e^{-\frac{x}{2}} dx} \\ &= \frac{e^{-\frac{3}{2}}}{e^{-1}} \\ &= e^{-\frac{1}{2}} \\ &= 0,606 \end{aligned}$$

III. Actividad previa: leer textos indicados en bibliografía, referentes al tema.

IV. Actividad post sesión: realizar la guía de ejercicios que a continuación se presenta.

Ejercicios.

1. Un jugador lanza un dado no cargado. Si sale un n° primo gana dicho número en dólares, pero si no sale un n° primo, entonces pierde esa cantidad de dinero.

- a) Definir la variable aleatoria y sus valores.
- b) Calcular su función de probabilidad.

2. Un juego se juega como sigue: se tiran 3 dados no cargados y se permite que el apostador apueste 1 dólar a la ocurrencia de uno de los n° (del 1 al 6). Suponga que apuesta al 5.

Entonces, si sale 5 en 1 dado de los 3, gana 1 dólar. Si sale 5 en 2 de los 3 dados, gana 2 dólares y si sale 5 en los 3 dados, gana 3 dólares. Pero si no sale 5 en ningún dado, pierde 1 dólar.

- a) Encuentre la variable aleatoria y sus valores.
- b) Encuentre la función de probabilidad.

3. Suponga que una variable aleatoria x tiene una función de densidad de probabilidad expresada mediante:

$$f(x) \begin{cases} kx(1-x) & 0 \leq x \leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

- a) Calcular el valor de k en que hace a la ecuación una función de densidad de probabilidad.
- b) Calcular $P(0,4 \leq x \leq 1)$
- c) Calcular $P(X \leq 0,4 / x \leq 0,8)$
- d) Calcular $F(x) = P(X \leq k)$ y graficarla.

4. Hay 10 estudiantes inscritos en una clase de estadística, de entre los cuales 3 tienen 19 años, 4 tienen 20, 1 tiene 21, 1 tienen 24 y 1 tiene 26 años. De esta clase se seleccionarían 2 estudiantes al azar sin reemplazo. Sea X la edad promedio de los 2 estudiantes seleccionados. Derivar la función de probabilidad para X .

5. Un hombre tiene 4 llaves en su bolsillo. Como está oscuro, no puede ver cuál es la llave de su puerta, por lo que prueba cada una a la vez, hasta encontrar la correcta. Sea X el número de llaves que prueba (incluyendo la correcta) para abrir la puerta.
- a) ¿Cuál es la función de probabilidad para X ?
 - b) ¿Cuál es el valor de $P(X \leq 1)$?
 - c) ¿Cuál es el valor de $P(X \leq 3)$?
 - d) ¿Cuál es el valor de $P(X = 4)$?
6. Sea Z la variable aleatoria que representa el número de sellos en 2 lanzamientos de una moneda honrada. Hallar la distribución de probabilidad de Z .

Sesión 9

Tema: Distribución binomial.

I. Objetivos de la sesión: conocer los conceptos básicos para iniciar el tema de estadística descriptiva.

II. Tema:

La distribución binomial es una función de probabilidad aplicada a variable discreta, que cumplan algunas características en particular.

El proceso Bernoulli se aplica cuando una variable tiene solo dos posibles resultados: éxito o fracaso. Estos dos resultados son mutuamente excluyentes.

Se considera una distribución binomial como n repeticiones de un proceso del Bernoulli.

Por ejemplo, si se lanza una moneda, sus dos posibles resultados son cara o sello. Si el evento a medir es el que al lanzarla salga cara, su resultado es éxito (si sale cara) o fracaso (si sale sello).

Cuando este proceso se repite n veces y se cuenta la cantidad de veces que resultó el éxito en las n repeticiones, estamos frente a una distribución binomial.

Si x es una variable que se define como binomial, entonces ella cumple las siguientes características:

- a) x es el número de veces que sucede el éxito en n repeticiones.
- b) En cada repetición se miden dos resultados: éxito o fracaso.
- c) El rango de x varía entre cero y n , ya que puede suceder que en las n repeticiones no ocurra nunca el éxito o por el contrario, en las n repeticiones ocurre el éxito la totalidad de las veces.
- d) La probabilidad de que suceda éxito es constante en cada repetición.
- e) Los ensayos son independientes.

Su función de probabilidad es:

$$P(x = k) = \binom{n}{k} \cdot p^k (1 - p)^{n-k}$$

Donde n : es el número de repeticiones de en experimento.

k : es el número de veces que queremos que suceda éxito

P : es la probabilidad de éxito en una repetición

Como toda función de probabilidad, tiene una esperanza y varianza. Ella se calcular como:

$$E(x) = np \quad V(x) = np(1-p)$$

Ej.1: Si una maquina herramienta desgastada produce 1% partes defectuosas y las partes producidas son independientes, entonces el número promedio de partes defectuosas en una muestra de 25 es:

$$E(x) = 25(0.01) = 0.25$$

La varianza del número de partes defectuosas es:

$$V(x) = 25(0.01)(0.99) = 0.2475$$

Ej. 2: Si se lanza una moneda 10 de veces. ¿Cuál es la probabilidad que salga cara 6 veces?

Solución.

Sea x = número de veces que sucede cara.

$$n = 10$$

$$k = 6$$

$$p = 0,5 \text{ (probabilidad de cara)}$$

$$\text{rango de } x = 0,1,2,3\dots 10$$

Podemos ver que la variable cumple las características de una distribución binomial, por tanto, se usa su función para resolver el problema.

$$P(x = 6) = \binom{10}{6} (0,5)^6 (1 - 0,5)^4$$

Donde $\binom{10}{6}$ es un factor combinatorio que se calcula a través de factoriales.

$$P(x = 6) = \frac{10!}{6!(10-6)!} (0,5)^6 \cdot (1 - 0,5)^4$$

$$P(x = 6) = 210(0,5)^6 \cdot (1 - 0,5)^4$$

$$P(x = 6) = 0,205$$

Entonces, la probabilidad que salga cara 6 veces al lanzar la moneda 10 veces es 0.205 o lo mismo que decir, el 20.5% de posibilidades.

Usando el mismo ejemplo, supongamos que la pregunta hubiera sido ¿cuál es la probabilidad de que al lanzar la moneda 10 veces, salga cara, a lo más tres veces?

Se pide calcular $P(x \leq 3)$, esto es, la probabilidad acumulada. La función acumulada de esta distribución es:

$$P(x \leq j) = \sum_{k=0}^j \binom{n}{k} p^k (1-p)^{n-k}$$

Que es lo mismo que:

$$P(x \leq j) = P(x=0) + P(x=1) + P(x=2) + \dots + P(x=j)$$

En el ejemplo:

$$P(x \leq 3) = P(x=0) + P(x=1) + P(x=2) + P(x=3)$$

Donde:

$$P(x=0) = \binom{10}{0} (0,5)^0 (1-0,5)^{10-0} = 0,00097$$

$$P(x=1) = \binom{10}{1} (0,5)^1 (1-0,5)^{10-1} = 0,0097$$

$$P(x=2) = \binom{10}{2} (0,5)^2 (1-0,5)^{10-2} = 0,0439$$

$$P(x=3) = \binom{10}{3} (0,5)^3 (1-0,5)^{10-3} = 0,1172$$

Efectuando la suma de cada una de las probabilidades, resulta:

$$P(x \leq 3) = 0,17177$$

El 17.177% de posibilidad que en 10 lanzamientos de la moneda, salgan cara a lo más 3 veces.

El cálculo de una probabilidad acumulada resulta larga si se hace a través de los cálculos individuales como se mostró en el ejemplo anterior.

La probabilidad acumulada de una distribución binomial se encuentra trasladada en una tabla fácil de usar, para algunos valores de n, p y k

Ej. 3: Si la probabilidad de que un vendedor realice una venta a un cliente es 0.2

- ¿Cuál es la probabilidad que al atender a 7 clientes, 3 de ellos compren el producto que vende el vendedor?
- ¿Cuál es el número esperado de ventas efectivas en los 5 clientes?
- ¿Cuál es el número de ventas efectivas en los 7 clientes?.

Solución.

- $x = \text{N}^\circ$ de ventas efectivas
rango de $x = 0, 1, 2, 3, \dots, 7$
 $n = 7$
 $p = 0.2$

La variable x distribuye binomial, ya que cumple las características necesarias.

$$P(x = 3) = \binom{7}{3} (0,2)^3 (1 - 0,2)^{7-3} \\ = 0,1146$$

- $P(x > 5) = P(x = 6) + P(x = 7)$
 $= \binom{7}{6} (0,2)^6 (0,8)^1 + \binom{7}{7} (0,2)^7 (0,8)^0$
 $= 0,00037$
 $= 0,037\%$

- $E(x) = np$
 $= 7 \cdot (0,2)$
 $= 1,4$

Se espera de los 7 clientes, 1.4 de ellos compren lo que el vendedor les ofrece.

III. Actividad previa: leer textos indicados en bibliografía, referentes al tema.

IV. Actividad post sesión: realizar la guía de ejercicios que a continuación se presenta.

Ejercicios

1. Todos los días se seleccionan, de manera electiva, 15 unidades de un proceso de manufactura con el propósito de verificar el porcentaje de unidades es defectuosa en la producción. Se sabe que le 5 % de las unidades es defectuosa.
La gerencia ha decidido detener la producción si la muestra tiene dos o más defectuosa. ¿Cuál es la probabilidad de que n un día cualquiera la producción se detenga?.
2. Un club de autopistas, comienza una campaña telefónica con el propósito de aumentar el número de socios. Con base a la experiencia, se sabe que 1 de cada 20 llamados, se inscribe en el club. Si en un día se efectúan 25 llamados.
 - a) ¿Cuál es la probabilidad que por lo menos 2 de ellas se hagan socios?
 - b) ¿Cuál es la probabilidad de que 2 se hagan socios?
 - c) ¿Cuál es el promedio esperado en ese día de socios nuevos?
3. Supongamos que la probabilidad de un nacimiento aislado corresponda al sexo masculino es 0.54. Si se asume que partos diferentes son independientes con relación al sexo. ¿cuál es la probabilidad que una familia de 5 hijos tenga:
 - a) exactamente 3 varones;
 - b) por lo menos 1 varón?
4. Un examen de estadística consiste en 10 preguntas de alternativa múltiples, cada una con 5 posibles respuestas. Calcula la probabilidad de aprobar si todas las calificaciones mínima que se pide es del 60%.
5. Un ingeniero de seguridad afirma que solo el 40% de todos los trabajadores utilizan cascos de seguridad cuando almuerzan en el lugar de trabajo. Suponga que esta afirmación es cierta, encuentre la probabilidad que 4 de 6 trabajadores elegidos al azar, utilicen sus cascos mientras almuerzan en el lugar de trabajo.
6. Durante un año en particular, el 70% de las acciones ordinarias inscritas en la Bolsa de Valores de Nueva York elevaron su valor de mercado, mientras que el 30% restante permaneció sin cambios o redujo su valor de mercado. Al inicio del año, un servicio de consultoría accionaría eligió 10 emisiones de acciones

como “especialmente recomendables”. Si las 10 emisiones representan una selección aleatoria, ¿cuál es la probabilidad de que:

a) las 10 emisiones hayan elevado su valor de mercado;

b) al menos 8 emisiones hayan elevado su valor de mercado?.

7. Una secretaria que debe llegar a su trabajo todas las mañanas a las 8:00 se retarda 15 minutos o más el 20% de las veces. El presidente de la compañía, que no llega sino a las 10:00 a.m., llama ocasionalmente a la oficina entre 8:00 y 8:15 para dictar una carta. ¿cuál es la probabilidad de que, tres mañanas de las 6 en que el presidente llama, la secretaria no esté en la oficina?.

8. El archivo personal de una fábrica indica que el 10% de los empleados de la línea de montaje se retiran al año de haber sido contratados. Se acaba de contratar 10 empleados nuevos.

a) ¿Cuál es la probabilidad de que exactamente la mitad de ellos siga trabajando después de un año?.

b) ¿Cuál es la probabilidad de que todos sigan trabajando después de un año?.

c) ¿Cuál es la probabilidad de que 3 de los 10 se vayan antes de terminar el año?.

Sesión 10

Tema: Distribución Poisson.

I. Objetivos de la sesión: conocer una de las distribuciones de probabilidad para variables discretas.

II. Tema:

1. Distribución Poisson:

La distribución Poisson es otra función de probabilidad para variables discretas. La variable aleatoria representa el número de eventos independientes que suceden en un intervalo pudiendo esto ocurrir en forma aleatoria en cualquier momento dentro del intervalo.

El intervalo puede ser de tiempo, distancia, área, volumen o alguna unidad similar.

Algunos ejemplos típicos son el número de personas que llegan a una tienda de autoservicio en un tiempo determinado, el número de solicitudes de seguro procesados por una compañía en un período específico, el estudio de choques de vehículos en un determinado lapso de tiempo, número de usuarios que se conectan a internet durante un día, etc.

Además, ofrece una aproximación excelente a la función de probabilidad binomial, cuando “p” es pequeño y “n” grande. Defiere de la distribución binomial en que x es infinito.

Igual que la función binomial, la distribución Poisson tiene tabla para algunos valores.

Sea x una variable aleatoria discreta, si cumple que x representa el n° de eventos aleatorios independientes que ocurren a velocidad constante en un lapso de tiempo, entonces su función de probabilidad es:

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!} \quad \text{con } k = 0, 1, 2, \dots$$

A diferencia de la distribución binomial, esta variable toma valores discretos infinitos.

El parámetro λ es el n° promedio de ocurrencia del evento por unidad de tiempo o por intervalo de referencia.

La distribución Poisson tiene los siguientes requisitos:

- La variable x es el nº de ocurrencias de un suceso durante un intervalo de tiempo.
- Las ocurrencias deben ser aleatorias.
- Las ocurrencias deben ser uniformemente distribuidas dentro del intervalo empleado.

$E(x) = \lambda$	esperanza
$V(x) = \lambda$	varianza

Dado un intervalo de números reales, supóngase que el conteo de ocurrencia es aleatorio en dicho intervalo. Si éste puede dividirse en subintervalos pequeños, tales que:

- 1) La probabilidad de mas de una ocurrencia en el subintervalo es cero.
- 2) La probabilidad de una ocurrencia es la misma para cualquier subintervalo.
- 3) El conteo de ocurrencias en cada subintervalo es independiente de los demás intervalos

entonces el experimento aleatorio recibe el nombre de proceso Poisson.

Ej.: Esperanza de (x)

El gerente de una planta industrial planea comprar una máquina nueva de cada una de los tipos A y B. Por cada día de funcionamiento, el número de reparaciones “ x ” que necesita la máquina A es una variable aleatoria Poisson con promedio $0,10t$ (siendo t el tiempo, entonces, de funcionamiento diario). El número de reparaciones “ y ” para la máquina B es una variable aleatoria Poisson con promedio $0,12t$.

El costo diario de operaciones de A es $C_A(t) = 10t + 30x^2$ y para B es $C_B(t) = 10t + 30y^2$.

- a) ¿Cuál de las máquinas da el costo mínimo esperado si un día consiste en 10 horas?.
- b) ¿Cuál de las máquinas da el costo mínimo esperado si un día consiste en 20 horas?.

Solución.

- Costo esperado de A.

$$E(C_A^{(t)}) = 10t + 30E(x^2)$$

Pero $V(x) + E^2(x) = E(x^2)$

entonces

$$\begin{aligned} &= 10t + 30(V(x) + E^2(x)) \\ &= 10t + 30(0,10t + (0,10t)^2) \\ &= 10t + 30(0,10t + 0,01t^2) \\ &= 13t + 0,3t^2 \end{aligned}$$

$$\begin{aligned} \text{Si } t = 10 &\longrightarrow E(C_A^{(t)}) = 13 \cdot 10 + 0,3(10)^2 = 160 \\ t = 20 &\longrightarrow E(C_A^{(t)}) = 13 \cdot 20 + 0,3(20)^2 = 380 \end{aligned}$$

- Costo esperado de B.

$$\begin{aligned} E(C_B^{(t)}) &= 8t + 30E(y^2) \\ &= 8t + 30[V(y) + E^2(y)] \\ &= 8t + 30[0,12t + (0,12t)^2] \\ &= 8t + 30[0,12t + 0,144t^2] \\ &= 11,6 + 0,432t^2 \end{aligned}$$

$$\begin{aligned} \text{Si } t = 10 &\longrightarrow E(C_B^{(t)}) = 11,6 \cdot (10) + 0,432(10)^2 = 159,2 \\ t = 20 &\longrightarrow E(C_B^{(t)}) = 11,6 \cdot (20) + 0,432(20)^2 = 404,8 \end{aligned}$$

La máquina B tiene operación más económica durante costos periodos debido a su menor costo horario de funcionamiento. Sin embargo, para periodos largos, la máquina A, es mas económica porque tiene reparaciones menos frecuentes.

Se puede aplicar a variables que miden por ejemplo las siguientes situaciones:

- N° de accidentes por semana en una autopista.
- N° de personas que llegan a un banco a una hora determinada.
- N° de errores pro página que comete una mecanógrafa.

Ej.: Suponga que el n° de llamadas que llegan a una central telefónica es de 0,5 por minuto en promedio. Encuentre la probabilidad que:

a) En 1 minuto no lleguen llamadas.

X : n° de llamadas en un minuto.
 $\lambda = 0,5$ $x = 0, 1, 2, 3, \dots$

$$P(x = 0) = \frac{e^{-0,5} (0,5)^0}{0!} = 0,607$$

b) En un minuto lleguen mas de 3 llamadas.

$$\begin{aligned} P(x > 3) &= 1 - P(x \leq 3) \\ &= 1 - [P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3)] \\ &= 1 - e^{-0,5} \left[\frac{(0,5)^0}{0!} + \frac{(0,5)^1}{1!} + \frac{(0,5)^2}{2!} + \frac{(0,5)^3}{3!} \right] \\ &= 1 - 0,998 \\ &= 0,002 \end{aligned}$$

c) Cuantas llamadas se espera que lleguen en 5 minutos.

$$E(x) = 5 \cdot \lambda = 5 \cdot (0,5) = 2,5 \text{ llamadas.}$$

d) Que en 5 minutos lleguen mas de 2.

$$\lambda = 5(0,5) = 2,5 \text{ en 5 minutos}$$

$$\begin{aligned} P(x > 2) &= 1 - P(x \leq 2) \\ &= 1 - e^{-2,5} \left[\frac{(2,5)^0}{0!} + \frac{(2,5)^1}{1!} + \frac{(2,5)^2}{2!} \right] \\ &= 1 - 0,54 \\ &= 0,46 \end{aligned}$$

2. Distribución binomial como aproximación a la Poisson.

Ej.: En una concurrida intersección de tránsito, la probabilidad de que un automóvil tenga un accidente es $p = 10^{-4}$. Si pasan 1.000 automóviles, halla la probabilidad de que ocurran 2 o más accidentes.

$$\left. \begin{array}{l} X : \text{n}^\circ \text{ de accidentes} \\ n = 1.000 \end{array} \right\} \begin{array}{l} x = 0, 1, 2, 3, \dots, 1.000 \\ p = 10^{-4} \end{array} \Rightarrow \text{Poisson}$$

como p pequeño
n grande

$$\lambda = np = (1.000)(10^{-4}) = 0,1$$

$$\begin{aligned} P(x \geq 2) &= 1 - P(x < 2) = 1 - P(x \leq 1) \\ &= 1 - \left[\frac{e^{-0,1} (0,1)^0}{0!} + \frac{e^{-0,1} (0,1)^1}{1!} \right] \end{aligned}$$

Ej.: Poisson

1. Durante un experimento de laboratorio el n° promedio de partículas radiactivas que pasan a través de un contador en un milisegundo es 4. ¿Cuál es la probabilidad de que 6 partículas entren al contador en un milisegundo?.

$$\begin{aligned} k &= 6 & \lambda &= 4 \\ P(x = 6) &= \frac{e^{-4} 4^6}{6!} = 0,1042 \end{aligned}$$

2. El n° promedio de camiones tanque que llegan cada día a cierta ciudad portuaria es 10. Las instalaciones en el puerto pueden manejar a lo más 15 camiones tanque por día. ¿Cuál es la probabilidad de que en un día dado los camiones se tengan que regresar?.

$$\begin{aligned} \lambda &= 10 & k &= 15 \\ P(x > 15) &= 1 - P(x \leq 15) = 1 - 0,9513 = 0,0487 \Rightarrow 4,87\% \end{aligned}$$

Ej: Binomial a Poisson

En un proceso de fabricación donde se manufacturan productos de vidrio ocurren defectos o burbujas. Se sabe, en promedio, una de cada 1.000 de estos artículos que se producen tiene 1 o más burbujas. ¿Cuál es la probabilidad que en una muestra de 8.000 artículos tengan menos de 7 artículos con burbujas?.

$$n = 8.000 \text{ (grande)} \quad p = 0,001 \text{ (pequeño)}$$

$$\lambda = np \quad \lambda = 8$$

$$P(x < 7) = P(x \leq 6) = 0,3134$$

III. Actividad previa: leer textos indicados en bibliografía, referentes al tema.

IV. Actividad post sesión: realizar la guía de ejercicios que a continuación se presenta.

Ejercicios.

1. En un departamento de reparación de maquinarias se recibe un promedio de 5 llamadas de servicio por hora. ¿Cuál es la probabilidad de que en una hora determinada se reciban exactamente 3 llamadas de servicio?.
2. Para determinada industria manufacturera, el número de accidentes de trabajo es 3 por semana. Calcular la probabilidad de que en una semana no se presenten accidentes.
3. En una tienda de telas, un promedio de 12 personas por hora le hacen preguntas a un decorador. ¿Cuál es la probabilidad de que 10 o más personas se acerquen al decorador para hacerle preguntas en un periodo de 10 minutos?.
4. El número promedio de camiones que llegan en un día cualquiera a un depósito de camiones en cierta ciudad es 12. ¿Cuál es la probabilidad de que en un día dado lleguen menos de 9 camiones a este depósito?.
5. Un cargamento grande de libros contiene 2% de ello con defectos en la encuadernación. Determine la probabilidad de que entre 400 libros seleccionados al azar del cargamento:
 - a) exactamente tengan 10 defectos
 - b) al menos 10 tengan defectos.
6. La probabilidad que un individuo sufra una reacción secundaria por una inyección de un medicamento determinado es una en mil. Calcular la probabilidad que de un total de 2.000 individuos inyectados:
 - a) exactamente 3 tengan reacción secundaria
 - b) menos de 2 tengan reacción secundaria
 - c) 4 o 5 tengan reacción secundaria.
7. Se supone que el número de defectos en los rollos de tela de cierta industria textil es una variable aleatoria Poisson con una media de 0.1 defectos por metro cuadrado.

- a) ¿Cuál es la probabilidad de tener 2 defectos en un metro cuadrado de tela?
 - b) ¿Cuál es la probabilidad de tener un defecto en 10 metros cuadrados de tela?
 - c) ¿Cuál es la probabilidad de que no halla defectos en 20 metros cuadrados de tela?
8. Un profesor de estadística se da cuenta de que cuando programa una hora para que los estudiantes que necesiten ayuda acudan a su oficina, llega un promedio de dos estudiantes. Calcule la probabilidad de que en una hora de oficina seleccionada al azar, el número de estudiantes que llegue sea:
- a) Cero.
 - b) Dos.
 - c) Menos de cinco.

Sesión 11

Tema: Dist. Geométrica, hipergeométrica y exponencial.

I. **Objetivos de la sesión:** diferenciar entre distintas distribuciones según sus características.

II. **Tema:**

1. Distribución geométrica.

La distribución geométrica es una función de probabilidad para variables discretas que cumplan algunas características en particular, como la distribución binomial antes vista.

Supongamos que se efectúa un experimento, donde un evento A tiene 2 probabilidades de suceder: éxito o fracaso.

La variable x mide el número de repeticiones del experimento necesarias hasta que ocurra el primer éxito.

Las características son parecidas a las de una variable binomial:

- a) las repeticiones son independientes unas de otras
- b) la probabilidad que ocurra el éxito es constante en cada repetición
- c) cada experimento tiene solo 2 posibilidades: éxito o fracaso

En la distribución binomial x es el número de éxitos en un total de n repeticiones. En cambio en una distribución geométrica, x es el número de repeticiones necesarias hasta que ocurra el primer éxito.

La función de probabilidad de una distribución geométrica es:

$$P(x = k) = (1 - p)^{k-1} \cdot p^k$$

El rango de x= 1, 2, 3,... La variable x es discreta e infinita. Su valor esperado o esperanza y varianza es:

$$E(x) = \frac{1}{p} \qquad V(x) = \frac{1-p}{p^2}$$

Se ha definido una variable aleatoria geométrica como el número de ensayos realizados hasta obtener el primer éxito. Sin embargo, dado que los ensayos son independientes, el conteo del número de éstos hasta que se tiene el primer éxito puede comenzar en cualquier ensayo sin cambiar la distribución de la variable aleatoria.

La suposición para el uso de un modelo geométrico es que el sistema presuntamente no se desgastará. La probabilidad del error permanece constante para todos los ensayos. En este sentido, se dice que la distribución carece de memoria.

Ej.1: Hallar la probabilidad que en lanzamientos sucesivos de 2 dados no cargados, la suma sea 12 por primera vez en el quinto lanzamiento.

Solución. Sea x = número de lanzamientos necesarios hasta que los dados sumen 12.

Dos dados suman 12 si salen los números (6,6) en cada dado, entonces si p es la probabilidad que suceda esto: $p = \frac{1}{36}$

$$\begin{aligned} \text{Luego: } P(x=5) &= \left(1 - \frac{1}{36}\right)^4 \cdot \left(\frac{1}{36}\right) \\ &= 0,0027 \\ &= 2,7\% \end{aligned}$$

Ej.2: Una empresa de reclutamiento encuentra que el 30% de los aspirantes para determinado puesto en la industria tiene conocimientos avanzados de programación. Se entrevista uno a uno, para la cual se seleccionan al azar de entre un grupo. Calcular la probabilidad de que el primer aspirante con conocimientos avanzados en programación sea el 5º entrevistado.

Solución. Sea x = número de entrevistas necesarias hasta un primer éxito.
 $p = 0,3$

$$\begin{aligned} \text{entonces: } P(x=5) &= (1 - 0,3)^4 \cdot (0,3) \\ &= 0,072 \\ &= 7,2\% \end{aligned}$$

Ej.3: Refiriéndose al ejemplo anterior, supóngase que al primer aspirante que se encuentre se le ofrece el puesto y que el aspirante acepta. Si cada entrevista cuesta \$30.000, calcular el costo esperado y la desviación estándar del costo total de las entrevistas hasta ocupar el puesto.

Solución. Costo $c = 30.000x$

$$\begin{aligned} \text{a) } E(c) &= 30.000E(x) \\ &= 30.000 \cdot \frac{1}{(0,3)} \\ &= 100.000 \text{ pesos} \end{aligned}$$

$$\begin{aligned}
\text{b) } V(c) &= V(30.000x) \\
&= (30.000)^2 \cdot V(x) \\
&= (30.000)^2 \cdot \frac{(1-0,3)}{(0,3)^2} \\
&= 7 \cdot 10^9
\end{aligned}$$

entonces, la desviación estándar es $\sqrt{7 \cdot 10^9} = 83.666$ pesos

Así como la distribución geométrica calcula el número de experimentos necesarios hasta el primer éxito, existe otra distribución de probabilidad que mide el número de ensayos necesarios hasta obtener r éxitos. A esta distribución se le llama distribución binomial negativa y su función de probabilidad es la siguiente.

$$P(x = r) = \binom{x-1}{r-1} (1-p)^{x-r} p^r, \text{ con } x = r, r+1, r+2, \dots$$

Si X es una variable aleatoria binomial negativa entonces

$$E(x) = \frac{r}{p} \qquad V(x) = \frac{r(1-p)}{p^2}$$

2. Distribución hipergeométrica.

Se considera un conjunto de N elementos, de los cuales r son considerados buenos y los otros (N - r) son considerados defectuosos. En forma similar a la distribución binomial, nos interesa la variable x que mide el número de éxitos en n ensayos pero esta vez, los ensayos o experimentos son sin sustitución y la probabilidad de un éxito cualquiera no es constante.

Su función de probabilidad es:

$$P(x = k) = \frac{\binom{r}{k} \binom{N-r}{n-k}}{\binom{N}{n}} \quad \text{con } k = 0, 1, 2, 3, \dots$$

Su esperanza y varianza, de la distribución son:

$$E(x) = \frac{nr}{N} \qquad V(x) = \frac{nr(N-r)(N-n)}{N^2(N-1)}$$

Ej.1: Como parte de un estudio de la contaminación del aire, un inspector decide examinar la emisión de gases a 6 de los 24 camiones de carga de una compañía.

Si 4 de los camiones de la compañía emiten cantidades excesivas de contaminantes, ¿cuál es la probabilidad de que ninguno de ellos sea incluido en la muestra del inspector?

Solución: $N = 24, n = 6, r = 4$

Del total de 26 camiones, 4 de ellos emiten contaminantes y 20 no lo hacen. Se inspecciona una muestra de 6 camiones y se pide que ninguno de los 4 camiones se encuentre en la muestra.

La variable x =número de camiones contaminantes que se encuentran en la muestra, tiene una distribución hipergeométrica.

$$\begin{aligned} P(x = 0) &= \frac{\binom{4}{0} \binom{20}{6}}{\binom{24}{6}} \\ &= 0,2880 \\ &= 28,8\% \end{aligned}$$

Ej 2: Quince de los estudiantes de un grupo escolar están insatisfechos con el texto que se emplea. Si una muestra de 4 estudiantes se interroga sobre el libro de texto, determine la probabilidad de que:

- a) exactamente 3 estudiantes se muestren insatisfechos
- b) al menos 3 estudiantes se muestren insatisfechos.

Solución.

$N=20, r=15, n=4$

X : número de estudiantes insatisfechos en la muestra (tomada sin reemplazo)

$$\text{a) } P(x = 3) = \frac{\binom{15}{3} \binom{20-15}{4-3}}{\binom{20}{4}} = 0,4695$$

$$\begin{aligned} \text{b) } P(x \geq 3) &= 1 - P(x \leq 2) \\ &= 1 - [P(x = 0) + P(x = 1) + P(x = 2)] \\ &= 1 - [0,00103 + 0,03095 + 0,2167] \\ &= 0,75132 \end{aligned}$$

3. Distribución exponencial

La distribución exponencial mide el tiempo transcurrido hasta la primera ocurrencia del evento A. es una función de probabilidad para una variable continua.

Esta variable se aplica en un proceso de Poisson en que nos interesa el tiempo hasta el primer evento, el tiempo entre dos eventos sucesivos o el tiempo hasta que ocurra el primer evento después de un punto temporal seleccionado.

Si λ es el promedio de ocurrencia del evento A en un lapso de tiempo, la probabilidad de que el tiempo hasta el primer evento distribuye exponencial con función de probabilidad.

$$f(x) \begin{cases} \alpha \cdot e^{-\alpha x} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Su función densidad o distribución acumulada es:

$$P(x \leq t) = 1 - e^{-\alpha t}$$

Y su esperanza y varianza es:

$$E(x) = \frac{1}{\alpha} \quad V(x) = \frac{1}{\alpha^2}$$

El parámetro λ es el promedio de eventos en el tiempo de interés y se calcula en función de λ .

Ej.: Un barco llega cada 2 días a muelle ¿cuál es la probabilidad de que tras la partida de un barco, pasen 4 días antes de la llegada del siguiente?

Solución.

Vemos que el número de barcos que llegan en un lapso de tiempo es un proceso de Poisson. En este caso nos interesa el tiempo que transcurre hasta que llega el primer barco, después que se fue otro barco.

Promedio en 2 días = 1

Promedio en 1 día = 0.5

Promedio en un periodo de 4 días = $4 \cdot 0.5 = 2$

Entonces: $\alpha = 2$, luego:

$$\begin{aligned}
P(x > 4) &= 1 - P(x \leq 4) \\
&= 1 - [1 - e^{-2}] \\
&= 0,1353
\end{aligned}$$

Ej.: En un departamento de reparación de maquinaria se recibe un promedio de 5 llamadas por hora. Si nos ubicamos en un momento determinado del día, ¿cuál es la probabilidad de que la primera llamada llegue en el curso de media hora?

Solución.

Este es un proceso de Poisson con $\lambda=5$. Sea x el tiempo hasta que llega la primera llamada, entonces $x = 2,5$ (promedio por media hora).

$$\begin{aligned}
P(x > 0,5) &= 1 - e^{-2,5} \\
&= 1 - 0,08208 \\
&= 0,91792
\end{aligned}$$

Ej.: Supongamos que un fusible tiene una duración x que puede considerarse como una variable aleatoria continua con una distribución exponencial. El proceso de fabricación del fusible tiene una duración esperada de 100 horas con un costo C por fusible. Supongamos además que si un fusible dura menos de 200 horas, se carga una pérdida de K pesos en contra del fabricante. Calcula el costo esperado.

Solución.

$$\text{Costo} = \begin{cases} C & \text{si } x > 200 \\ C + K & \text{si } x \leq 200 \end{cases} \quad \alpha = \frac{1}{100}$$

$$\begin{aligned}
E(C) &= C \cdot P(x > 200) + (C + K) \cdot P(x \leq 200) \\
&= C \cdot e^{-\left(\frac{1}{100} \cdot 200\right)} + (C + K) \cdot \left[1 - e^{-\left(\frac{1}{100} \cdot 200\right)}\right] \\
&= C \cdot e^{-2} + (C + K) \cdot [1 - e^{-2}] \\
&= C \cdot e^{-2} + C - C \cdot e^{-2} + K - K \cdot e^{-2} \\
&= C + K - K \cdot e^{-2} \\
&= C + K \cdot (1 - e^{-2})
\end{aligned}$$

III. Actividad previa: leer textos indicados en bibliografía, referentes al tema.

IV. Actividad post sesión: realizar la guía de ejercicios que a continuación se presenta.

Ejercicios.

1. Un promedio de 6 personas hace uso de una caja bancaria automática durante el horario de hora de mayor congestión de una tienda departamental.
 - a) ¿Cuál es la probabilidad de que pasen al menos 10 minutos entre las llegadas de dos clientes?
 - b) ¿Cuál es la probabilidad de que, tras la salida de un cliente, no llegue otro al menos 20 min?
 - c) ¿Cuál es la probabilidad de que un segundo cliente llegue en el curso de un minuto después de que un primer de que un cliente inicia una transacción bancaria?
2. Supongamos que el manuscrito de un libro de texto tiene 50 errores, o errores de dedo, en las 500 paginas de material y que los errores se distribuyen aleatoriamente a lo largo del texto. Al iniciar el lector de pruebas de un capítulo específico. ¿Qué probabilidad hay de que el primer error de ese capítulo aparece dentro de las primeras 5 páginas?. ¿Ocurra después de las primeras 5 paginas?.
3. Si el número de automovilistas que corren a lata velocidad (que un radar recta por hora) en cierta localidad de la ruta I-10 es una variable aleatoria de Poisson con $\lambda = 8.4$, ¿cuál es la probabilidad de tener un tiempo de espera menor de 10 minuto entre automovilista sucesivos que circulan a alta velocidad?.
4. El recorrido (en miles de millas) que obtienen los propietarios de automóviles con cierto tiempo de llanta radial es una variable aleatoria que tiene una distribución exponencial con $\theta = 40$. Obtenga las probabilidades de una de estas llantas dure:
 - a) cuando menos 20000 millas
 - b) cuando menos 30000 millas.
5. Entre los 120 aspirantes para ocupar un empleo, sólo 80 están realmente calificados para hacerlo. Si se selecciona al azar cinco de estos aspirantes para realizar una entrevista "a fondo", determine la probabilidad de que sólo dos de los cinco estén calificados para el trabajo mediante el uso de la distribución hipergeométrica.

6. Si la probabilidad de que cierto examen de una reacción positiva igual a 0.4.
¿Cuál es la probabilidad de que ocurran menos de 5 reacciones negativas antes de la primera positiva?.

7. Un cargamento de 80 alarmas contra robo contiene cuatro que están defectuosas. Si se selecciona al azar tres de estas y se envían a un cliente, determine la probabilidad de que el cliente reciba exactamente una unidad defectuosa mediante el uso de la fórmula de la distribución hipergeométrica.

8. Si la probabilidad es 0.75 de que un solicitante de licencia de manejo apruebe el examen de manejo en un ensayo dado, ¿cuál es la probabilidad de que un solicitante pase la prueba en un cuarto ensayo?.

9. En una red de computadores, el acceso de los usuarios al sistema puede considerarse como un proceso Poisson con una media de 25 accesos por hora. ¿Cuál es la probabilidad de que no haya accesos en un intervalo de 6 minutos?.

10. De los 15 altos ejecutivos de un negocio de importaciones y exportaciones, se seleccionan 12 para ser enviados al Japón a estudiar un nuevo proceso de producción. Ocho de los ejecutivos ya tienen algo de entrenamiento en el proceso, ¿cuál es la probabilidad de que 5 de los enviados tengan algo de conocimiento sobre el proceso antes de partir al lejano oriente?.

11. Un estado efectúa una lotería en la que se escogen al azar y sin reemplazo seis números de entre cuarenta. Un jugador selecciona seis números antes de que el estado tome la muestra.
 - a) ¿Cuál es la probabilidad de que los seis números seleccionados por el jugador coincidan con los seis números que hay en la muestra del estado?
 - b) ¿Cuál es la probabilidad de que cuatro de los seis números seleccionados por el jugador aparezcan en la muestra del estado?
 - c) Si el jugador participa en la lotería cada semana ¿Cuál es el número esperado de semanas que deben transcurrir para que los seis números que escoge el jugador coincidan con los que hay en la muestra tomada por el estado?

Sesión 12

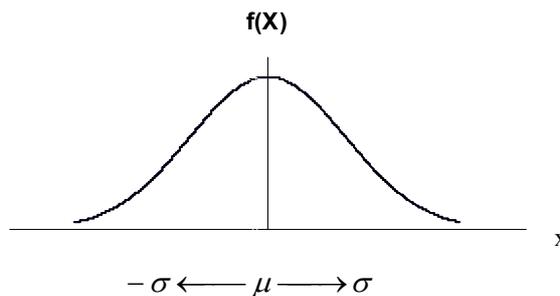
Tema: Distribución normal.

I. **Objetivos de la sesión:** conocer una de las distribuciones más usada dentro de las probabilidades.

II. **Tema:**

La más utilizada de las distribuciones continuas es la distribución normal. Esta distribución con un gráfico como campana se centra en el valor promedio, denominado μ y su dispersión respecto a él, se denomina σ , llamada desviación estándar.

La función normal describe de forma aproximada muchos fenómenos que suceden en la naturaleza tales como la estatura de las personas, el coeficiente intelectual en los niños, etc.



Su función de densidad es:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{con } -\infty < x < \infty \quad (\text{rango de } x)$$

Muchas mediciones que se hacen en forma natural tienden a tener distribuciones de frecuencia relativa que se asemejan mucho a la curva normal, probablemente porque la naturaleza tiende a promediar los efectos de las distribuciones variables que intervienen en una respuesta determinada.

En contraste, las vidas de los organismos biológicos o la duración de partes electrónicas, tienden a distribuciones de frecuencias relativa que no son normales. Esto se debe a que tienen un comportamiento extremo, no promedio.

Una propiedad importante de la distribución normal, es que cualquier función lineal de una variable aleatoria distribuida normal, también tiene distribución normal, es decir, si $x \sim N$.

Entonces $y = ax + b$, también $y \sim N$.

Se denota $x \sim N(\mu, \sigma^2)$

Si $y = ax + b \Rightarrow y \sim N(a\mu + b; a^2\sigma^2)$

Puede existir un n° infinito de distribuciones normales posibles, cada una con su propia media y su desviación estándar ya que obviamente no se puede analizar un n° tan grande de posibilidades, es necesario convertir todas estas distribuciones normales a la forma estándar. Esta conversión se hace a través de la variable Z.

$$Z = \frac{x - \mu}{\sigma}$$

Donde Z es una variable aleatoria normal con media cero y varianza 1.

La variable Z se define como el n° de desviaciones estándar a las una observación está de la media.

Ej.: Un curso tiene un promedio de 67 pulgada de estatura, con desviación de 2 pulgadas.

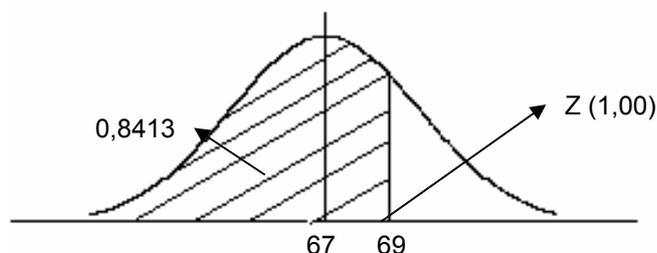
Si un alumno tiene una estatura de 63 pulgadas entonces $Z = \frac{63 - 67}{2} = -2$, esto quiere decir que su estatura esta a 2 desviación estándar por debajo de la media.

Estandarizar una distribución normal permite de terminar mas fácilmente la probabilidad de que ocurre cierto evento.

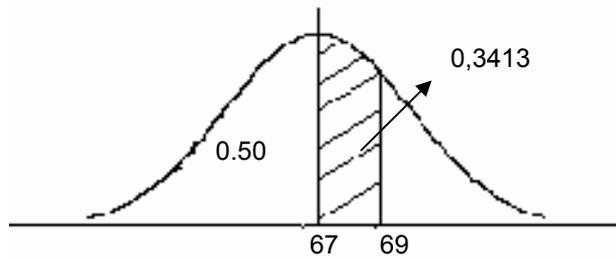
La probabilidad de un evento esta relacionado con el área bajo la curva normal estándar en el punto Z.

Ej.: $Z = \frac{69 - 67}{2} = 1,00$

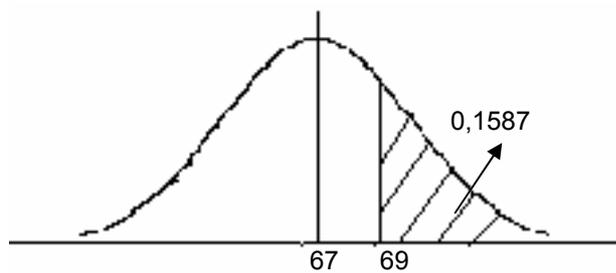
Si buscamos en la tabla estandarizada el valor 1,00 encontramos que el área bajo la curva es 0,8413.



Es la probabilidad que una persona mida menos de 69 pulgadas.



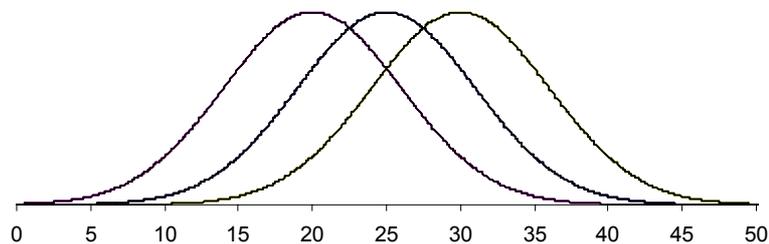
Es la probabilidad que una persona mide entre 67 y 69 pulgadas.



Es la probabilidad que un alumno mida mas de 69 pulgadas.

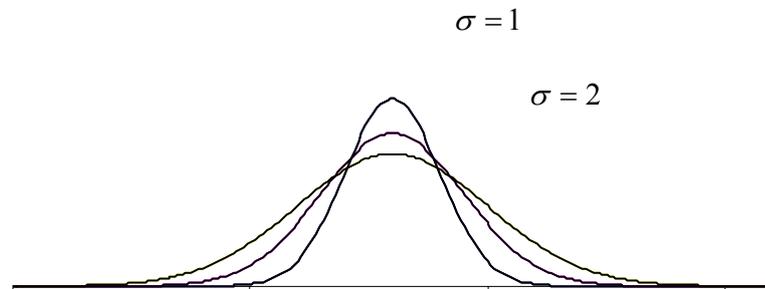
Una curva normal queda completamente determinada una vez que se especifica μ y σ^2 .

Ej.: si $\mu = 25$ $\sigma = 2$
 $\mu = 20$
 $\mu = 30$

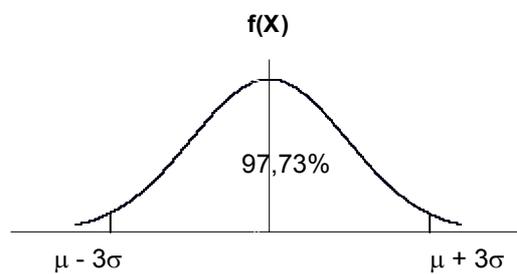


Este gráfico muestra la curva normal manteniendo la misma variabilidad pero distinto promedio

Ej.: El siguiente gráfico a diferencia del anterior muestra distintas curvas normales con el mismo promedio y diferentes variabilidades. Mientras más alta es la variabilidad la curva se aplana.



La simetría se da con la recta vertical $x = \mu$ (también es la mediana)
 El 99,73% de las observaciones quedan incluidas en el intervalo $\mu - 3\sigma, \mu + 3\sigma$



Es decir: $P(\mu - 3\sigma < x < \mu + 3\sigma) = 0,9973$

$P(\mu - \sigma < x < \mu + \sigma) = 0,6845$

$P(\mu - 2\sigma < x < \mu + 2\sigma) = 0,9545$

Cualquier probabilidad es el área bajo la curva.

Tabulación de una normal (μ, σ^2) a una normal estándar $(0,1)$:

La normal con parámetros $\mu = 0$ y $\sigma = 1$ es la única que esta tabulada, es decir, el valor de su probabilidad está calculado y representados en una tabla de sencillo uso.

El área permanece constante por lo tanto tiene igual probabilidad si se estandariza.

Como la tabla está calculada para valores $x < k$ es decir acumulados desde $-\infty$ hasta k , si se necesita calcular $x > k$, basta calcular su complemento, por ejemplo si queremos,

$$P(x > 2) = 1 - P(x < 2)$$

Después de haber invertido la probabilidad ahora se debe estandarizar.

Ej.: Una empresa que fabrica y embotella jugo de manzana tiene una máquina automática que llena la botella de 16 onzas. Sin embargo, hay cierta variación en la cantidad de liquido que llega a la botella. La desviación es de 1 onza. La cantidad vaciada en las botellas distribuye normal.

- a) Calcular la probabilidad que la máquina vacíe mas de 17 onzas de liquido en cualquier botella.

$$x \sim N(16, 1^2)$$

$$\begin{aligned} P(x > 17) &= P\left(Z > \frac{17 - \mu}{\sigma}\right) \\ &= P\left(Z > \frac{17 - 16}{1}\right) \\ &= P(Z > 1) \\ &= 1 - P(Z \leq 1) \\ &= 1 - 0,8413 \\ &= 0,1587 \end{aligned}$$

- b) Suponga que otra máquina trabaja de tal manera que las onzas que sirve tiene un promedio igual a la indicación del dial para "cantidad de liquido", pero que tiene una desviación de 1,2 onzas. Calcular la posición adecuada del dial para que solo se rebase la cantidad de 17 onzas en el 5% de las botellas.

$$P(x > 17) = 0,05$$

$$P\left(Z > \frac{17 - \mu}{1,2}\right) = 0,05$$

$$1 - P\left(Z < \frac{17 - \mu}{1,2}\right) = 0,05$$

$$P\left(Z < \frac{17 - \mu}{1,2}\right) = 0,95$$

Entonces:

$$1,64 = \frac{17 - \mu}{1,2}$$

$$1,968 = 17 - \mu$$

$$\mu = 17 - 1,968$$

$$\mu = 15,032$$

III. Actividad previa: leer textos indicados en bibliografía, referentes al tema.

IV. Actividad post sesión: realizar la guía de ejercicios que a continuación se presenta.

Ejercicios.

1. Determine el área bajo la curva normal estandarizada entre los siguientes valores de z :

- | | |
|----------------------------|--------------------------------|
| a) $z = 1,10$ y $z = 1,75$ | f) $z = -1,0$ y $z = -0,5$ |
| b) $z = 0,52$ y $z = 1,55$ | g) $z = -0,8$ y $z = 0,35$ |
| c) $z = -1,0$ y $z = 1,0$ | h) $z = 2,0$ y $z = \infty$ |
| d) $z = 1,50$ y $z = 1,55$ | i) $z = -\infty$ y $z = -2,0$ |
| e) $z = -2,0$ y $z = 2,0$ | j) $z = -\infty$ y $z = -0,48$ |

2. Supóngase que x tiene distribución Normal con media igual a 12 y desviación estándar igual a 2,56. Determine la probabilidad de que:

- a) x sea mayor que 13,3
- b) x sea menor que 15,2
- c) x tenga valores entre 12,7 y 14,8
- d) x tenga valores menores o iguales a 11,5

3. Los estudiantes de cierta escuela secundaria tienen un coeficiente intelectual promedio de 106 y varianza de 256. Suponga $\sim N$, hallar la probabilidad que un estudiante tenga un coeficiente intelectual:

- a) igual o menor que 98
- b) igual o menor que 13
- c) igual o mayor de 127
- d) entre 94 y 118

4. Una máquina despachadora de refrescos esta ajustada para servir en promedio 200 ml. por vasos. Si la cantidad es distribución normal con desviación de 15 ml.

- a) que fracción de los vasos contendrá mas de 224 ml.
- b) ¿cuál es la probabilidad que un vaso contenga entre 19,1 y 209 ml.?
- c) bajo que valor se obtiene es 25% mas pequeño de los refrescos?
5. Si los limites de peso considerados normales (sanos) se diferencian como aquellos pesos que alrededor del promedio incluyen al 90% de las personas sanas:
- a) ¿desde que peso se clasificaría a un individuo obeso?
- b) ¿desde que peso se clasificaría a un individuo delgado?.
6. Una variable Y sigue una distribución aproximadamente normal con desviación estándar de 10. Dado que la probabilidad de obtener valores de Y mayores de 70 es 0,1151. ¿Cuál es el valor medio de Y ?
7. Sea x una variable que se distribuye aproximadamente normal con media igual a 120. Si la probabilidad que x sea mayor de 125 es 0,0951. ¿Cuál es su varianza?.
8. Supóngase que la cantidad de radiación cósmica a que se expone una persona cuando vuela en jet por Estados Unidos es una variable aleatoria que tiene una distribución normal con una media de 4,35 mrem y una desviación estándar de 0,59 mrem. ¿Cuál es la probabilidad de que una persona esté expuesta a más de 5,20 mrem de radiación cósmica en un vuelo de este tipo?.
9. El diámetro del corazón de un grupo de individuos se distribuyen normal con media 8 cm. y desviación estándar 0,8 cm. Se ha observado que el individuo que presenta un diámetro de corazón superior en 10% al promedio experimenta molestias y cuando el tamaño excede de 9,6 cm. tiene altas probabilidades de sufrir un infarto. ¿Cuántos de 800 individuos de este grupo:
- a) no debería sufrir molestias?
- b) debería sentir molestias sin correr riesgo de sufrir infarto?
10. Si la variable largo de baya de una uva sultanina se distribuye normal (25,16) encuentre la probabilidad que:

- a) Una baya sea mayor que 32 mm.
- b) Una baya mida entre 21 y 17 mm.
- c) Una baya mida menos de 16 mm.
- d) Una baya mida mas de 29 mm.

11. El tiempo necesario para armar cierta unidad es una variable aleatoria normalmente distribuida con una media de 30 min. y desviación estándar igual a 2 min. Determinar el tiempo de armado de manera tal que la probabilidad de exceder este sea de 0,02.

12. Una universidad espera recibir, para el siguiente año escolar, 16.000 solicitudes de ingreso al primer año de licenciatura. Se supone que las calificaciones obtenidas por los aspirantes en la prueba SAT se puede calcular, de manera adecuada, por una distribución normal con media igual a 950 y desviación estándar igual a 100. Si la universidad decide admitir al 25% de todos los aspirantes que obtengan las calificaciones mas altas en la prueba SAT, ¿cuál es la mínima calificación que es necesario obtener en esta prueba, para ser admitido en la universidad?.

13. La vida promedio de cierto tipo de motor pequeño es de 10 años con una desviación estándar de 2 años. El fabricante repone sin cargo todos los motores que fallen dentro del periodo de garantía. Si está dispuesto a reponer sólo el 3% de los motores que falle, ¿qué tan larga deberá ser la garantía que otorgue?. Suponga que las vidas de los motores siguen una distribución normal.

14. Las longitudes de las sardinas recibidas por cierta enlatadora tienen una media de 4,62 pulgadas y una desviación estándar de 0,23 pulgadas.

- a) ¿Qué porcentaje de todas estas sardinas son mayores de 5 pulgadas?.
- b) ¿Qué porcentaje de las sardinas miden 4,35 y 4,85 pulgadas de longitud?.