EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA

SEP DGETI

GUIA DE ESTUDIO DE PROBABILIDAD Y ESTADÍSTICA 2023

SECRETARÍA DE EDUCACIÓN PÚBLICA
SUBSECRETARÍA DE EDUCACIÓN MEDIA SUPERIOR
CENTRO DE ESTUDIOS TECNOLÓGICOS industrial y de servicios No. 5
"Gertrudis Bocanegra"

GUIA DE ESTUDIOS
SEMESTRE FEBRERO JULIO 2022

GUIA DE ESTUDIO DE PROBABILIDAD Y ESTADÍSTICA 2022

Contenido

1.- ¿Qué es estadística?

- **Es la ciencia que se encarga de recolectar, organizar, resumir y analizar datos para después obtener conclusiones a partir de ellos.**

Proviene de la palabra alemana *statistik* y significa “ciencia del Estado”

2.- ¿En cuántas ramas se divide la estadística?

- **Se divide en dos ramas**

Descriptiva → **Comprende el conjunto de herramientas para recolectar, presentar y caracterizar un conjunto de datos.**

Inferencial → **Se refiere al conjunto de técnicas mediante las cuales es posible hacer conjeturas o generalizaciones, a partir de información parcial.**

POBLACIÓN Y MUESTRA.

3.- ¿Qué es la población?

- **Es la totalidad de elementos o unidades que comparten alguna o algunas características en común.**

4.- ¿A qué se le llama elemento?

- **Los elementos o unidades son las entidades de las cuales se obtiene la información.**

5.- ¿A qué se le llama muestra?

- **Es un subconjunto de elementos o unidades de la población de interés, seleccionado mediante algún método o elemento.**

6.- ¿A qué se le llama variable en estadística?

- **Se define como variable a la característica, cualidad o atributo de interés de las unidades de una población o una muestra.**

7.- ¿Qué tipo de variables existen en estadística?

-**Cualitativas**: Cuyos posibles valores son únicamente categorías, En este tipo de variables se ubican las categorías nominal y ordinal.

-Nominales: Se utiliza para clasificar a la población en diferentes categorías, en las cuales no existe una relación de orden o jerarquía.

Ejemplo: un ser humano puede ser de sexo masculino o femenino; con los números se pueden realizar operaciones de suma, resta, multiplicación o división; los ciudadanos pueden tener como estado civil: soltero o casado.

-Cuantitativas: Cuyos posibles valores son números, en ésta se ubican las categorías nominal y ordinal.

-Ordinales: La escala ordinal se caracteriza porque las modalidades que integran a la misma tienen un orden entre ellas. Se distingue porque no se define la distancia entre las observaciones.

Ejemplo: México es un país que consume altas cantidades de refresco, pero al entrevistar a una persona acerca de la frecuencia con que consume esta bebida, se pueden tener respuestas como: regularmente, algunas veces, pocas veces, casi nunca.

-Cardinales: Son las más complejas su variable operacional es una escala cardinal que se caracteriza por las diferencias iguales entre dos de sus puntos y son iguales entre sí.

-Discretas: Si se cumple que sus valores pueden ser números enteros.

-Continuas: Pueden tener valores en el campo de los números reales.

8.- ¿De cuántas partes consta una tabla estadística?

-**Encabezado:** Son las descripciones de las filas y columnas de un cuadro estadístico, el encabezado se ubica en la parte superior del cuerpo del cuadro, indica las variables y sus categorías o valores, también puede indicar un periodo de tiempo.

-Cuerpo: Es el contenido numérico del cuadro, es la parte donde se colocan los datos correspondientes a las características o variables indicados en el encabezado o en los conceptos, es decir, presenta la distribución de los elementos según la clasificación en categorías de las variables

-Pie: Se usa para aclarar algunos términos o siglas, y también para indicar que elementos están o no incluidos en algunos de los conceptos del cuadro.

9.- ¿Qué es la distribución de frecuencia?

- **Es el número de veces que se observa una variable con un cierto valor y en las tablas se representa con la literal *f*.**

  Dicho más sencillo. Es el número de veces que se repite un cierto valor en un conjunto de datos.

10.- ¿Qué tipo de distribuciones se pueden realizar?

-Frecuencia acumulada: La frecuencia acumulada es la suma de las frecuencias absolutas de todos los valores inferiores o iguales al valor considerado.
La frecuencia acumulada se representa por F_i.

-Frecuencia relativa: Se conoce como frecuencia relativa al cociente que resulta de dividir la frecuencia simple (f) entre el número (n) de elementos de la muestra y se representa con la literal f_r. Su fórmula es f_r.=f/n.

-Frecuencia relativa acumulada: Es la suma progresiva de las frecuencias relativas observadas y se representan con el literal Fr

11.- ¿A qué nos referimos con datos agrupados?

- **Son aquellos datos que pertenecen a un tamaño de muestra mayor a 20 o más elementos, por lo que para ser analizados, requieren ser agrupados en clases a partir de ciertas características.**

12.- ¿Qué es clase en estadística?

- **Una clase es un intervalo abierto por la izquierda y cerrado por la derecha en el que se encuentran contenidos una proporción de todos los valores observados y tiene la forma (a, b)**

13.- ¿Qué es intervalo en clase?

- **Rango utilizado para dividir el conjunto de posibles valores numéricos al trabajar con grandes cantidades de datos. Por ejemplo, si los valores están entre 1 y 100, se podrían definir grupos por medio de los intervalos 1-25, 26-50, 51-75, 76-100 cuando el intervalo de la clase es 25.**

14.- ¿Qué es marca de clase?

- **La marca de clase es el punto medio de cada intervalo, la marca de clase es el valor que representa a todo el intervalo para el cálculo de algunos parámetros como la media aritmética o la desviación típica. Se representa por** c_i **o** x_i**.**

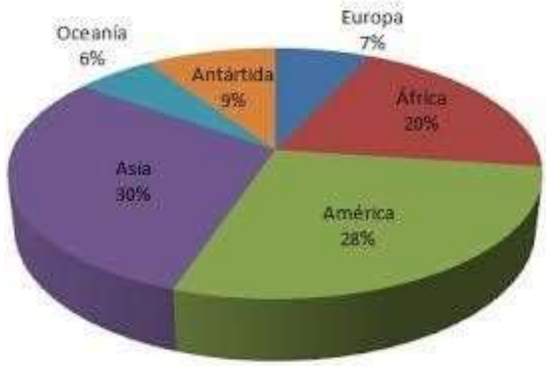
15.- ¿Qué es una gráfica?

- **Gráfica o gráfico es el conjunto de elementos o signos que permiten la interpretación de cualquier elemento o cosa. Es un conjunto de puntos x, y, que se plasman en coordenadas cartesianas, y sirven para analizar el comportamiento de un proceso que se está llevando a cabo.**

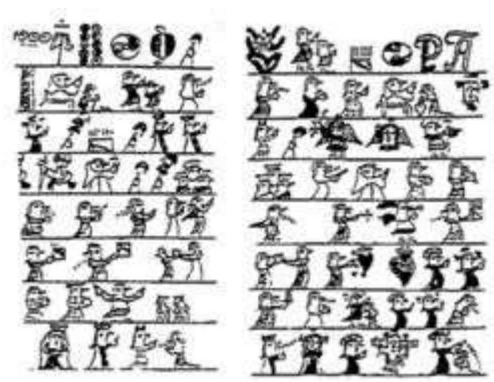
16.- ¿Cuáles tipos de graficas se pueden realizar en estadística?

-Grafica de barras: Se utiliza para representar tablas de frecuencias cuando se tienen más de 6 valores para las categorías nominales, por tanto, se puede considerar como un gráfico alternativo al diagrama circular.

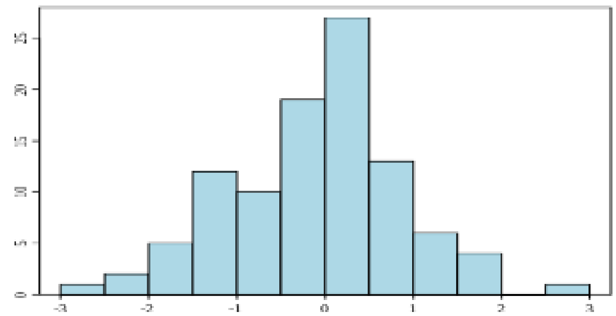
-Grafica circular: Se utiliza cuando la variable a graficar se ubica en la escala nominal u orinal, además se recomienda que a la variable en cuestión no se le asignen más de 6 valores. Este tipo de gráfica, como su nombre lo indica, utiliza un círculo como base para su elaboración y esta asociada con valores porcentuales.



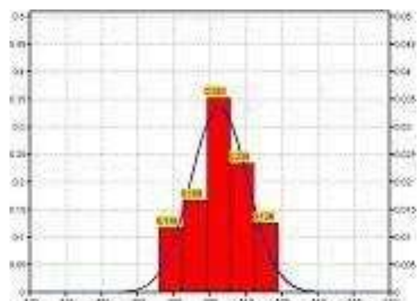
-Pictograma: Un pictograma es un signo icónico dibujado y no lingüístico que representa figurativamente, de forma más o menos realista, un objeto real o significado. En agrupaciones es precursor o antecedente de los sistemas de escritura propiamente dichos. Las historietas o cómics y los chistes gráficos sin texto son también pictogramas. Se distinguen de los ideogramas en que estos son más esquemáticos, resumidos y abstractos; los pictogramas son más concretos.
Con su nombre se suelen denominar los signos de los sistemas no alfabéticos basados en dibujos significativos.



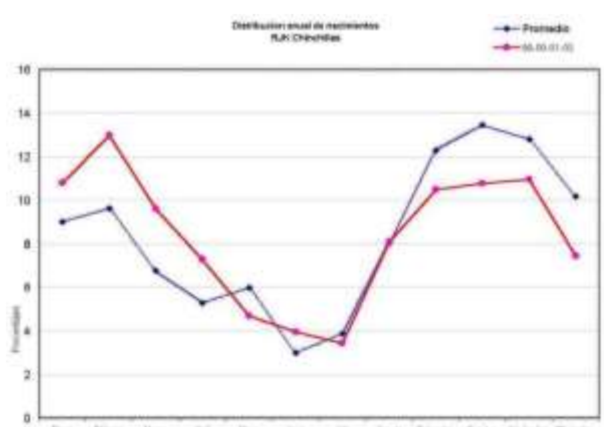
-Histograma: Es una representación gráfica de una variable en forma de barras, donde la superficie de cada barra es proporcional a la frecuencia de los valores representados, ya sea en forma diferencial o acumulada. Sirven para obtener una "primera vista" general, o panorama, de la distribución de la población, o la muestra, respecto a una característica, cuantitativa y continua, de la misma y que es de interés para el observador (como la longitud o la masa). De esta manera ofrece una visión en grupo permitiendo observar una preferencia, o tendencia, por parte de la muestra o población por ubicarse hacia una determinada región de valores dentro del espectro de valores posibles (sean infinitos o no) que pueda adquirir la característica.



-Polígono de frecuencias: Polígono de frecuencia es el nombre que recibe una clase de gráfico que se crea a partir de un histograma de frecuencia. Estos histogramas emplean columnas verticales para reflejar frecuencias): el polígono de frecuencia es realizado uniendo los puntos de mayor altura de estas columnas.



-Gráfico de líneas: Los gráficos de líneas muestran una serie como un conjunto de puntos conectados mediante una sola línea. Los gráficos de líneas se usan para representar grandes cantidades de datos que tienen lugar durante un período continuado de tiempo.



17.- ¿Qué son las medidas de tendencia central?

Las **medidas de tendencia central (media, mediana y moda)** sirven como puntos de referencia para interpretar las calificaciones que se obtienen en una prueba.

El propósito de las medidas de tendencia central es:

- Mostrar en qué lugar se ubica la persona promedio o típica del grupo.
- Sirve como un método para comparar o interpretar cualquier puntaje en relación con el puntaje central o típico.
- Sirve como un método para comparar el puntaje obtenido por una misma persona en dos diferentes ocasiones.
- Sirve como un método para comparar los resultados medios obtenidos por dos o más grupos.

Las medidas de tendencia central más comunes son:

- La **media aritmética**: comúnmente conocida como **media o promedio**. Se representa por medio de una letra **M** o por una **X** con una línea en la parte superior.
- La **mediana**: la cual es el puntaje que se ubica en el centro de una distribución. Se representa como **Md**.
- La **moda**: que es el puntaje que se presenta con mayor frecuencia en una distribución. Se representa **Mo**.

De estas tres medidas de tendencia central, la **media** es reconocida como la mejor y más útil. Sin embargo, cuando en una distribución se presentan casos cuyos puntajes son muy bajos o muy altos respecto al resto del grupo, es recomendable utilizar la mediana o la moda. (Porque dadas las características de la media, esta es afectada por los valores extremos).

La media es considerada como la mejor medida de tendencia central, por las siguientes razones:

- Los puntajes contribuyen de manera proporcional al hacer el cómputo de la media.
- Es la medida de tendencia central más conocida y utilizada.
- Las medias de dos o más distribuciones pueden ser fácilmente promediadas mientras que las medianas y las modas de las distribuciones no se promedian.

- La media se utiliza en procesos y técnicas estadísticas más complejas mientras que la mediana y la moda en muy pocos casos.

18.- Define la Media Aritmética

Media aritmética $(\bar{x})$ o promedio

Es aquella medida que se obtiene al **dividir la suma de todos los valores de una variable por la frecuencia total**. En palabras más simples, corresponde a la suma de un conjunto de datos dividida por el número total de dichos datos.

$$\bar{x} = \frac{\text{suma de todos los valores}}{\text{cantidad total de datos}} = \frac{x_1 + x_2 + x_3 + x_4 + \ldots + x_n}{N}$$

Ejemplo 1:

En matemáticas, un alumno tiene las siguientes notas: **4, 7, 7, 2, 5, 3**
n = 6 (número total de datos)

$$\bar{x} = \frac{4+7+7+2+5+3}{6} = \frac{28}{6} = 4{,}8$$

La **media aritmética** de las notas de esa asignatura es 4,8. Este número representa el **promedio**.

Ejemplo 2:

Cuando se tienen muchos datos es más conveniente agruparlos en una tabla de frecuencias y luego calcular la media aritmética. El siguiente cuadro con las medidas de 63 varas de pino lo ilustra.

Largo (en m)	**Frecuencia absoluta**	**Largo por Frecuencia absoluta**
5	10	5 . 10 = 50
6	15	6 . 15 = 90
7	20	7 . 20 = 140
8	12	8 . 12 = 96
9	6	9 . 6 = 54
	Frecuencia total = 63	**430**

$$\bar{x} = \frac{430}{63} = 6{,}825$$

Se debe recordar que la **frecuencia absoluta** indica cuántas veces se repite cada valor, por lo tanto, la tabla es una manera más corta de anotar los datos (si la frecuencia absoluta es 10, significa que el valor a que corresponde se repite 10 veces).

- -x=∑ *fx/n* Serie de datos o datos simples.

Una **serie de datos** es un conjunto de valores, numéricos o no numéricos, generalmente ligados a una secuencia temporal. Por ejemplo:

- serie de datos pluviométricos diarios: estos datos son directamente observados en las estaciones pluviométricas o meteorológicas, pueden también ser llamadas series primarias;
- serie de datos pluviométricos mensuales, se trata de una elaboración a partir de los datos primarios observados día a día en las estaciones de medición;
- serie de resultados numéricos obtenidos en un ensayo de laboratorio, en este caso puede no tener una relación secuencial.

En el caso que esté ligado a una serie temporal, se le denomina serie de tiempo.

A partir de los años 1980, se hace cada vez más frecuente la utilización de redes de sensoreamiento remoto, generando de esta forma una serie temporal de observaciones con un intervalo de tiempo que pueden ser muy pequeños. Por ejemplo, en una red de sensoreamiento remoto de precipitaciones, el barrido de los sensores puede hacerse a cada uno en pocos minutos.

Ejemplo de serie de datos secundarios

De la serie anterior, se pueden calcular los totales mensuales, para este caso son:
144; 129; 104; 100; 73; 31; 26; 56; 115; 144; 152; 158

lo que constituye una serie de datos secundarios, qué para el ejemplo considerado, son las precipitaciones mensuales en la estación pluviométrica en el año considerado.

Características de una serie de datos numéricos

Las principales características estadísticas de una serie de datos numéricos son:

- **Media**; en el ejemplo considerado **media anual de precipitaciones diarias**

$$M_d = \sum_{i=1}^{365} \frac{p_i}{365}$$

Donde:

M_d= media diaria anual en mm.

p_i= precipitación del día **i** en mm.

Para el caso considerado en el ejemplo sería: $M_d = \frac{1323}{365} = 3.38$ mm/día

- **Desviación típica "σ"**

$$\sigma = \sqrt{\frac{1}{365}\sum_{i=1}^{365}(p_i - M_d)^2}$$

- -x= ∑ *fx*/n   Distribución de frecuencias
  Método corto o tradicional

Una **distribución de frecuencias** o **tabla de frecuencias** es una **ordenación** en forma de **tabla** de los **datos estadísticos**, asignando a cada **dato** su **frecuencia correspondiente**.

Tipos de frecuencia

Frecuencia absoluta

La **frecuencia absoluta** es el **número de veces** que aparece un determinado **valor** en un estudio estadístico.

Se representa por **f_i**.

La **suma de las frecuencias absolutas** es igual al número total de datos, que se representa por **N**.

$$f_1 + f_2 + f_3 + \ldots + f_n = N$$

Para indicar resumidamente estas sumas se utiliza la letra griega **Σ** (sigma mayúscula) que se lee suma o sumatoria.

$$\sum_{i=1}^{i=n} f_i = N$$

Frecuencia relativa

La **frecuencia relativa** es el **cociente** entre la **frecuencia absoluta** de un determinado valor y el **número total de datos**.

Se puede expresar en tantos por ciento y se representa por $\mathbf{n_i}$.

$$n_i = \frac{f_i}{N}$$

La suma de las frecuencias relativas es igual a 1.

Frecuencia acumulada

La **frecuencia acumulada** es la **suma de las frecuencias absolutas** de todos los **valores inferiores o iguales** al **valor** considerado.
Se representa por $\mathbf{F_i}$.

Frecuencia relativa acumulada

La **frecuencia relativa acumulada** es el **cociente** entre la **frecuencia acumulada** de un determinado **valor** y el **número total de datos**. Se puede expresar en tantos por ciento.

Ejemplo

Durante el mes de julio, en una ciudad se han registrado las siguientes temperaturas máximas:

32, 31, 28, 29, 33, 32, 31, 30, 31, 31, 27, 28, 29, 30, 32, 31, 31, 30, 30, 29, 29, 30, 30, 31, 30, 31, 34, 33, 33, 29, 29.

En la primera columna de la tabla colocamos la variable ordenada de menor a mayor, en la segunda hacemos el recuento y en la tercera anotamos la frecuencia absoluta.

x_i	Recuento	f_i	F_i	n_i	N_i
27	I	1	1	0.032	0.032
28	II	2	3	0.065	0.097
29	~~IIII~~ I	6	9	0.194	0.290
30	~~IIII~~ II	7	16	0.226	0.516
31	~~IIII~~ III	8	24	0.258	0.774
32	III	3	27	0.097	0.871
33	III	3	30	0.097	0.968
34	I	1	31	0.032	1
		31		1	

Este tipo de **tablas de frecuencias** se utiliza con **variables discretas**.

Distribución de frecuencias agrupadas

La **distribución de frecuencias agrupadas** o **tabla con datos agrupados** se emplea si las **variables** toman un **número grande de valores** o la **variable es continua**.

Se **agrupan** los **valores** en **intervalos** que tengan la **misma amplitud** denominados **clases**. A cada **clase** se le asigna su **frecuencia correspondiente**.

Límites de la clase

Cada **clase** está **delimitada** por el **límite inferior de la clase** y el **límite superior de la clase**.

Amplitud de la clase

La **amplitud de la clase** es la **diferencia** entre el **límite superior e inferior** de la **clase**.

Marca de clase

La **marca de clase** es el **punto medio** de cada **intervalo** y es el **valor** que representa a todo el **intervalo** para el **cálculo** de algunos **parámetros**.

Construcción de una tabla de datos agrupados

3, 15, 24, 28, 33, 35, 38, 42, 43, 38, 36, 34, 29, 25, 17, 7, 34, 36, 39, 44, 31, 26, 20, 11, 13, 22, 27, 47, 39, 37, 34, 32, 35, 28, 38, 41, 48, 15, 32, 13.

1º se localizan los valores menor y mayor de la distribución. En este caso son 3 y 48.

2º Se restan y se busca un número entero un poco mayor que la diferencia y que sea divisible por el número de intervalos de queramos poner.

Es conveniente que el número de intervalos oscile entre 6 y 15.

En este caso, 48 - 3 = 45, incrementamos el número hasta 50 : 5 = 10 intervalos. Se forman los intervalos teniendo presente que el límite inferior de una clase pertenece al intervalo, pero el límite superior no pertenece intervalo, se cuenta en el siguiente intervalo.

	c_i	f_i	F_i	n_i	N_i
[0, 5)	2.5	1	1	0.025	0.025
[5, 10)	7.5	1	2	0.025	0.050
[10, 15)	12.5	3	5	0.075	0.125
[15, 20)	17.5	3	8	0.075	0.200
[20, 25)	22.5	3	11	0.075	0.2775
[25, 30)	27.5	6	17	0.150	0.425
[30, 35)	32.5	7	24	0.175	0.600
[35, 40)	37.5	10	34	0.250	0.850
[40, 45)	42.5	4	38	0.100	0.950
[45, 50)	47.5	2	40	0.050	1
		40		1	

- -x= A+(∑fd/n) Método por desviación

La **desviación estándar o desviación típica** es la **raíz cuadrada de la varianza**.

Es decir, la raíz cuadrada de la media de los cuadrados de las puntuaciones de desviación

La **desviación estándar** se representa por **σ**.

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{N}} \qquad \sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{N}}$$

Desviación estándar para datos agrupados

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 f_1 + (x_2 - \bar{x})^2 f_2 + \ldots + (x_n - \bar{x})^2 f_n}{N}} \qquad \sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 f_i}{N}}$$

Para simplificar el cálculo vamos o utilizar las siguientes expresiones que son equivalentes a las anteriores.

$$\sigma = \sqrt{\frac{x_1^2 + x_2^2 + \ldots + x_n^2}{N} - \bar{x}^2} \quad \sigma = \sqrt{\sum_{i=1}^{n} \frac{x_i^2 f_i}{N} - \bar{x}^2}$$

Desviación estándar para datos agrupados

$$\sigma = \sqrt{\frac{x_1^2 f + x_2^2 f_2 + \ldots + x_n^2 f_n}{N} - \bar{x}^2} \quad \sigma = \sqrt{\sum_{i=1}^{n} \frac{x_i^2}{N} - \bar{x}^2}$$

Ejercicios

Calcular la **desviación estándar** de la distribución:

9, 3, 8, 8, 9, 8, 9, 18

$$\bar{x} = \frac{9+3+8+8+9+8+9+18}{8} = 9$$

$$\sigma = \sqrt{\frac{(9-9)^2 + (3-9)^2 + (8-9)^2 + (8-9)^2 + (9-9)^2 + (8-9)^2 + (9-9)^2 + (18-9)^2}{8}} = 3.87$$

Calcular la desviación típica de la distribución de la tabla:

	x_i	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
[10, 20)	15	1	15	225
[20, 30)	25	8	200	5000
[30,40)	35	10	350	12 250
[40, 50)	45	9	405	18 225
[50, 60)	55	8	440	24 200
[60,70)	65	4	260	16 900
[70, 80)	75	2	150	11 250
		42	1 820	88 050

$$\bar{x} = \frac{1820}{42} = 43.33$$

$$\sigma = \sqrt{\frac{88050}{42} - 43.33^2} = 14.797$$

Propiedades de la desviación estándar

1 La desviación estándar será siempre un valor positivo o cero, en el caso de que las puntuaciones sean iguales.

2 Si a todos los valores de la variable se les suma un número la desviación estándar no varía.

3 Si todos los valores de la variable se multiplican por un número la desviación estándar queda multiplicada por dicho número.

4 Si tenemos varias distribuciones con la misma media y conocemos sus respectivas desviaciones estándar se puede calcular la desviación estándar total.
Si todas las muestras tienen el mismo tamaño:

$$\sigma = \sqrt{\frac{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2}{n}}$$

Si las muestras tienen distinto tamaño:

$$\sigma = \sqrt{\frac{k_1 \cdot \sigma_1^2 + k_2 \cdot \sigma_2^2 + \ldots + k_n \cdot \sigma_n^2}{k_1 + k_2 + \ldots + k_n}}$$

Observaciones sobre desviación la estándar

1 La **desviación estándar**, al igual que la media y la varianza, es un índice muy sensible a las puntuaciones extremas.

2 En los casos que no se pueda hallar la media tampoco será posible hallar la **desviación estándar**.

3 Cuanta más pequeña sea la **desviación estándar** mayor será la **concentración de datos** alrededor de la **media**.

- -x=A+(∑fu/n)a Método clave

19.- Define la Mediana.

Mediana (Med)

Para reconocer la mediana, es necesario tener ordenados los valores sea de mayor a menor o lo contrario. Usted divide el total de casos (N) entre dos, y el valor resultante corresponde al número del caso que representa la mediana de la distribución.

Es el **valor central** de un conjunto de valores **ordenados** en forma creciente o decreciente. Dicho en otras palabras, la Mediana corresponde al valor que deja igual número de valores antes y después de él en un conjunto de datos agrupados.

Según el número de valores que se tengan se pueden presentar dos casos:
Si el número de valores es impar, la Mediana corresponderá al **valor central** de dicho conjunto de datos.

Si el número de valores es par, la Mediana corresponderá al promedio de los dos valores centrales (los valores centrales se suman y se dividen por 2).

Ejemplo 1:

Se tienen los siguientes datos: 5, 4, 8, 10, 9, 1, 2

Al ordenarlos en forma creciente, es decir de menor a mayor, se tiene: 1, 2, 4, **5**, 8, 9, 10

El 5 corresponde a la Med, porque es el valor central en este conjunto de datos impares.

Ejemplo 2:

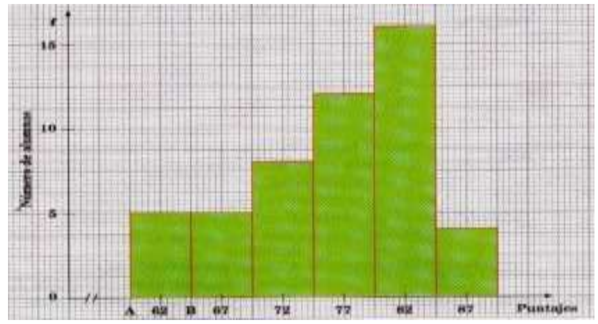
El siguiente conjunto de datos está ordenado en forma decreciente, de mayor a menor, y corresponde a un conjunto de valores pares, por lo tanto, la Med será el promedio de los valores centrales.

21, 19, 18, 15, **13, 11**, 10, 9, 5, 3

$$\text{Med} = \frac{13+11}{2} = \frac{24}{2} = 12$$

Ejemplo 3:

Interpretando el gráfico de barras podemos deducir que:



5 alumnos obtienen puntaje de 62
5 alumnos obtienen puntaje de 67
8 alumnos obtienen puntaje de 72
12 alumnos obtienen puntaje de 77
16 alumnos obtienen puntaje de 82
4 alumnos obtienen puntaje de 87
Lo que hace un total de 50 alumnos
Sabemos que la mediana se obtiene haciendo

$$\text{Med} = \frac{50+1}{2} = \frac{51}{2} = 25{,}5$$

Lo cual significa que la mediana se ubica en la posición intermedia entre los alumnos 25 y 26 (cuyo promedio es 25,5), lo cual vemos en el siguiente cuadro:

puntaje	alumnos
62	1
62	2
62	3
62	4
62	5
67	6
67	7
67	8
67	9
67	10
72	11
72	12
72	13
72	14
72	15
72	16
72	17
72	18
77	19
77	20
77	21
77	22
77	23
77	24
77	25
77	26
77	27
77	28
77	29

77	30
82	31
82	32
82	33
82	34
82	35
82	36
82	37
82	38
82	39
82	40
82	41
82	42
82	43
82	44
82	45
82	46
87	47
87	48
87	49
87	50

El alumno 25 obtuvo puntaje de 77
El alumno 26 obtuvo puntaje de 77
Entonces, como el total de alumnos es par debemos promediar esos puntajes:

$$\mathbf{Med} = \frac{77 + 77}{2} = \frac{144}{2} = 77$$

La mediana es 77, lo cual significa que 25 alumnos obtuvieron puntaje desde 77 hacia abajo (alumnos 25 hasta el 1 en el cuadro) y 25 alumnos obtuvieron puntaje de 77 hacia arriba (alumnos 26 hasta el 50 en el cuadro).

- Md= (n+1/2) serie de datos.

En tablas de distribuciones de frecuencias el método anterior se aplica para obtener la clase mediana y del intervalo correspondiente se toma el límite inferior: fa corresponde a la suma de las frecuencias que se encuentran por encima de la frecuencia de la clase mediana; ***fm*** a la frecuencia de la clase mediana.

- Md=Li+n/2 – fa/fm) a Distribución de frecuencias.

20.- Define Moda.

Moda (Mo)

Es la medida que indica cual dato tiene la **mayor frecuencia** en un conjunto de datos; o sea, cual se repite más.

Ejemplo 1:

Determinar la moda en el siguiente conjunto de datos que corresponden a las edades de niñas de un Jardín Infantil.

5, 7, **3, 3**, 7, 8, **3**, 5, 9, 5, **3**, 4, **3**

La edad que más se repite es 3, por lo tanto, la **Moda es 3 (Mo = 3)**
Ejemplo 2:

20, 12, 14, 23, 78, 56, 96

En este conjunto de datos **no** existe ningún valor que se repita, por lo tanto, este conjunto de valores **no tiene** moda.

En distribución de frecuencias se emplea la siguiente fórmula:

- Mod = Li+(d1/d1+d2)a

Donde Li es el límite inferior de la clase mediana, d1 es la diferencia de la frecuencia de la clase mediana con la frecuencia inmediata superior, y d2 con la inmediata inferior.

MEDIDAS DE DISPESIÓN

21.- Define Rango.

El Rango es la medida que resulta de la diferencia entre los dos valores extremos de un conjunto de datos.

Rango = Valor $_{\text{máximo}}$ – Valor $_{\text{mínimo}}$

22.- Define Varianza

La varianza es una medida de dispersión que indica la forma en que se encuentran esparcidos los valores de las observaciones respecto de la media aritmética o media.

- ∑f(x-x-)2/n Serie de datos
- ∑f(Xi-X-)2/n Distribución de frecuencias

23.- ¿Cómo se obtiene la varianza estándar?

- ∑f(X-X-)2/n

La varianza (que es el cuadrado de la desviación estándar: **σ^2**) se define así:
Es la media de las diferencias con la media **elevadas al cuadrado**.

En otras palabras, sigue estos pasos:

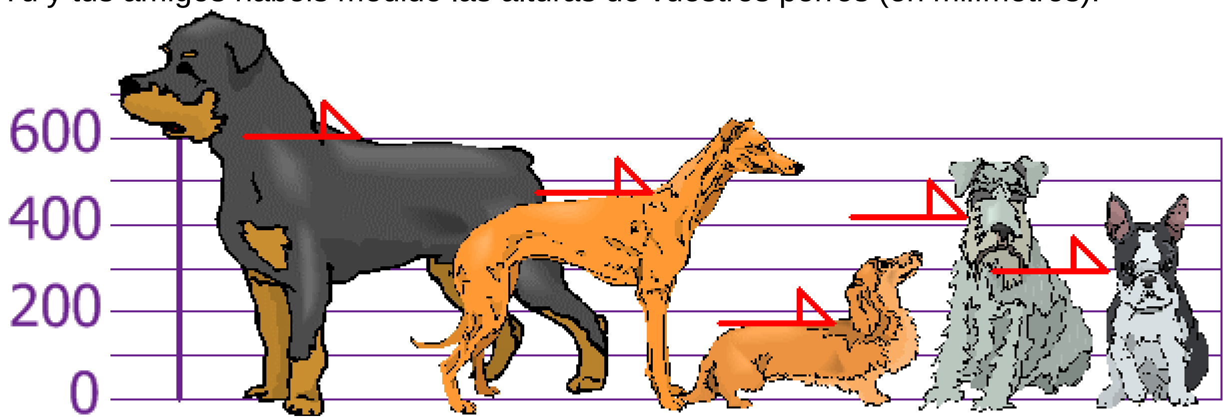
1. Calcula la media (el promedio de los números).

2. Ahora, por cada número resta la media y eleva el resultado al cuadrado (la diferencia elevada al cuadrado).

3. Ahora calcula la media de esas diferencias al cuadrado. (¿Por qué al cuadrado?)

Ejemplo

Tú y tus amigos habéis medido las alturas de vuestros perros (en milímetros):

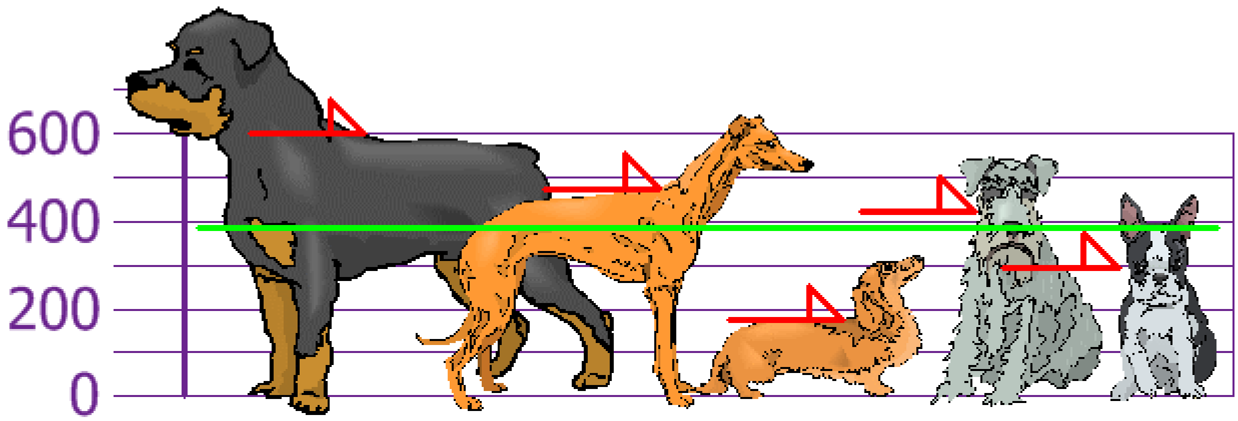


Las alturas (de los hombros) son: 600mm, 470mm, 170mm, 430mm y 300mm.
Calcula la media, la varianza y la desviación estándar.

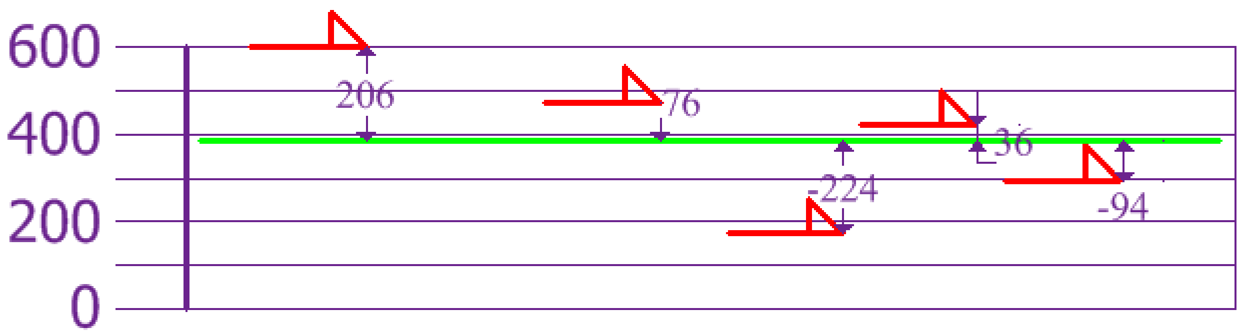
Respuesta:

$$\text{Media} = \frac{600 + 470 + 170 + 430 + 300}{5} = \frac{1970}{5} = 394$$

Así que la altura media es 394 mm. Vamos a dibujar esto en el gráfico



Ahora calculamos la diferencia de cada altura con la media:

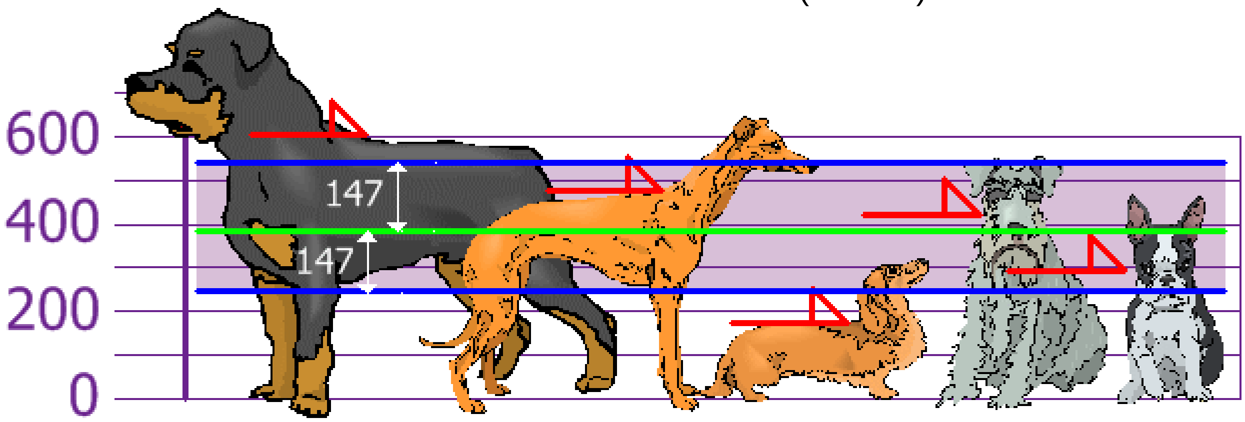


Para calcular la varianza, toma cada diferencia, elévala al cuadrado, y haz la media:

$$\text{Varianza: } \sigma^2 = \frac{206^2 + 76^2 + (-224)^2 + 36^2 + (-94)^2}{5} = \frac{108{,}520}{5} = 21{,}704$$

Así que la varianza es 21,704.
Y la desviación estándar es la raíz de la varianza, así que:

Desviación estándar: **σ = √21,704 = 147**
y lo bueno de la desviación estándar es que es útil: ahora veremos qué alturas están a distancia menos de la desviación estándar (147mm) de la media:



Así que usando la desviación estándar tenemos una manera "estándar" de saber qué es normal, o extra grande o extra pequeño.

*Nota: ¿por qué *al cuadrado*?

Elevar cada diferencia al cuadrado hace que todos los números sean positivos (para evitar que los números negativos reduzcan la varianza)
Y también hacen que las diferencias grandes se destaquen. Por ejemplo 100^2=10,000 es mucho más grande que 50^2=2,500.

Pero elevarlas al cuadrado hace que la respuesta sea muy grande, así que lo deshacemos (con la raíz cuadrada) y así la desviación estándar es mucho más útil.

24.- ¿Qué es el coeficiente de variación?

- V= /x

El **coeficiente de variación** es la relación entre la **desviación típica** de una muestra y su **media**.

$$C.V = \frac{\sigma}{\bar{x}}$$

El **coeficiente de variación** se suele expresar en **porcentajes**:

$$C.V = \frac{\sigma}{\bar{x}} \cdot 100$$

El **coeficiente de variación** permite comparar las **dispersiones** de dos distribuciones distintas, siempre que sus **medias** sean **positivas**.

Se calcula para cada una de las distribuciones y los valores que se obtienen se comparan entre sí.

La **mayor dispersión** corresponderá al valor del **coeficiente de variación mayor**.

Ejercicio:

Una distribución tiene x = 140 y σ = 28.28 y otra x = 150 y σ = 24. ¿Cuál de las dos presenta mayor dispersión?

$$C.V_1 = \frac{28.28}{140} \cdot 100 = 20.2\%$$

$$C.V_2 = \frac{24}{150} \cdot 100 = 16\%$$

La primera distribución presenta mayor dispersión.

PROBLEMAS PROPUESTOS

1.- ¿Cuál es el rango de las edades? 15, 16, 13, 19, 22, 17, 19, 14, 20, 18 años.

2.- ¿Cuál es el rango de las siguientes estaturas? 140, 162, 148, 172, 184, 162, 153, 162, 183, 172 cm.

3.- ¿Cuál es el rango de los siguientes pesos 83, 76, 54, 62, 78, 50, 59 kg.?

4.- Calcular la desviación media del siguiente grupo de edades 18, 15, 23, 14, 15, 18, 22, 15, 26, 34 años.

5.- Calcular la desviación media del siguiente grupo de datos (estaturas) 170, 175, 150, 145, 155, 170, 185, 190, 150, 160.

6.- Calcular la desviación típica de los pesos en Kg. De un grupo de estudiantes 65, 54, 46, 85, 70, 75, 62, 53, 55, 80, 90, 75, 85, 68, 52.

7.- Calcular la varianza de las calificaciones de 10 estudiantes: 6, 9, 10, 7, 8, 6, 9, 7, 6, 7.

8.- Calcular el coeficiente de variación del número de integrantes por familia 2, 5, 7, 4, 3, 8, 7, 4, 9, 6.

9.- Cuál es el coeficiente de variación de las estaturas de 12 estudiantes: 150, 164, 176, 155, 185, 182, 178, 184, 168, 152, 144, 186.

10.- Cuál es el coeficiente de variación de los pesos en Kg. De 10 niños de 1 año: 6, 8, 5, 7, 4, 9, 8, 9, 10, 6 kg.

TABLAS

01.- Calcular la desviación media del siguiente cuadro estadístico.

Edad	F	x	Fx	x-x	x-x	f x-x
18 - 19	6	18.5	111	-15.38	-15.38	92.28
20 -21	8	20.5	164		-17.38	139.04
22 -23	15	22.5	337.5		-19.38	290.7
24 -25	20	24.5	490		-21.38	427.6
26 -26	13	26	338		-22.88	297.44
28 -29	4	28.5	114		-25.38	101.52
a=2	66	140.5	206.5			1348.58

02.- Calcular la desviación media del siguiente cuadro es estaturas:

Estaturas	f	x	Fx	x-x	x-x	f x-x
140 - 144	4	142	568	-20.23	-20.23	80.92
145 - 149	12	147	1,764		-15.23	182.76
150 - 154	18	152	2,736		-10.23	184.14
155 -159	26	157	4,082		-5.23	135.98
160 - 164	34	162	5,508		-0.23	7.82
165 - 169	20	167	3,340		4.77	95.4
170 - 174	19	172	3,268		9.77	185.63
175 -179	13	177	2,301		14.77	192.01
180 - 184	6	182	1,092		-156.23	937.38
a=5	152	1,458	24,659			2002.04

03.- Calcular la desviación típica del siguiente cuadro:

Pesos Kg	f	x	Fx	x-x	x-x2	f x-x2
40-44	3	42	126	-24.83	616.52	1849.56
45-49	8	47	376	-19.83	393.22	3145.76
50-54	12	52	624	-14.83	219.92	2639.04
55-59	17	57	969	-9.83	96.62	1642.54
60-64	28	62	1736	-4.83	23.32	652.96
65-69	46	67	3082	0.17	0.02	0.92
70-74	32	72	2304	5.17	26.72	855.04
75-79	18	77	1386	-10.17	103.42	1861.56
80-84	15	82	1230	-15.17	230.12	3451.8
85-89	4	87	348	-20.17	406.82	1627.28
90-94	2	92	184	-25.17	633.52	1267.04
a=5	185	737	12,365			18993.5
						102.667568
				Desviación típica		10.13

04.- Calcular la desviación típica del cuadro siguiente:

Calif.	f	x	fx	x-x	x-x2	f x-x2
1.-2.	10	1.5	15	-4.18	17.47	174.7
3.-4	26	3.5	91	-2.18	4.75	123.5
5.-6	32	5.5	176	-0.18	0.32	10.24
7.-8	28	7.5	210	-1.82	3.31	92.68
9.-10	14	9.5	133	-3.82	14.59	204.26
a=2	110	27.5	625			605.38
						5.50345455
				Desviación típica		2.34

05.- Calcular el coeficiente de variación:

Estaturas	f	x	fx	x-x	x-x2	f x-x2
150-159	12	154.5	1854	-18.78	352.68	4232.16
160-169	15	164.5	2467.5	-8.78	77.08	1156.2
170-179	23	174.5	4013.5	-1.22	1.48	34.04
180-189	18	184.5	3321	-11.22	125.88	2265.84
190-199	6	194.5	1167	-21.22	450.28	2701.68
A=10	74	872.5	12,823			10,390

06.- Calcular el coeficiente de variación del cuadro siguiente:

Pesos	f	x	fx	x-x	x-x2	f x-x2
50-59	6	54.5	327	-20.92	437.64	2625.84
60-69	13	64.5	838.5	-10.92	119.24	1550.12
70-79	23	74.5	1713.5	-0.92	0.84	19.32
80-89	15	84.5	1267.5	-9.08	82.44	1236.6
90-99	8	94.5	756	-19.08	364.04	2912.32
A=10	65	372.5	4,903			8344.2

07.- Calcular el coeficiente de variación

Int. Fam.	f	x	fx	x-x	x-x2	f x-x2
0.-1	8	0.5	4	-4.75	22.56	180.48
2.-3	18	2.5	45	-2.75	7.56	136.08
4.-5	36	4.5	162	-0.75	0.56	20.16
6.-7	30	6.5	195	-1.25	1.56	46.8
8.-9	24	8.5	204	-3.25	10.56	253.44
a=2	116	22.5	610			636.96

08.- Calcular la desviación media, la desviación típica y el coeficiente de variación.

Edad	f	x	fx	x-x	X2	f x-x	x-x2	f x-x2
13-17	8	15	120	-15.55	225	105	210	1680
18-22	16	20	320	-10.55	400	300	380	6080
23-27	24	25	600	-5.55	576	575	551	13224
28-32	38	30	1140	-0.55	1444	1110	1414	53732
33-37	30	35	1050	-4.45	900	1015	865	25950
38-42	18	40	720	-9.45	1600	680	1560	28080
43-47	10	45	450	-14.45	100	405	55	550
a=5	144	210	4400			4190	5035	129296

09.- Calcular la desviación media, la desviación típica y el coeficiente de variación.

Estatura	f	x	fx	x-x	X2	f x-x	x-x2	f x-x2
150-154	12	152	1824	-16.96	12104	1672	11952	11940
155-159	16	157	2512	-11.96	24649	2355	24492	24476
160-164	24	162	3888	-6.96	26244	3726	26082	26058
165-169	32	167	5344	-1.96	27889	5177	27722	27690
170-174	30	172	5160	-3.04	29584	4988	29412	29382
175-179	18	177	3186	-8.04	31329	3009	31152	31134
180-184	16	182	2912	-13.04	33124	2730	32942	32926

185-189	10	187	1870	-18.04	34969	1683	34782	34772
a=5	158	1356	26696		219892	25340	218536	218378

14.- Calcular la desviación media, desviación típica, varianza y coeficiente de variación

Edad	f	x	fx	x-x	x-x	f x-x	x-x2	f x-x2
10-12	6	11	66	-6.2	-6.2	-55	38.44	27.56
13-15	8	14	112	-3.2	-3.2	-98	10.24	101.76
16-18	15	17	255	-0.2	-0.2	-238	0.04	254.96
19-21	9	20	180	-2.8	-2.8	-160	7.84	172.16
22-24	7	23	161	-5.8	-5.8	-138	33.64	127.36
a=3	45	85	774				90.2	683.8

15.- Calcular la desviación media, desviación típica, la varianza y el coeficiente de variación

Estatura	f	x	fx	x-x	x-x	f x-x	x-x2	f x-x2
150-154	6	152	912	-25.53	-25.53	31.53	651.78	260.22
155-159	10	157	1570		-20.53	30.53	421.48	1148.52
160-164	15	162	2430		-15.53	30.53	241.18	2188.82
165-169	18	167	3006		-10.53	28.53	110.88	2895.12
170-174	24	172	4128		-5.53	29.53	30.58	4097.42
175-179	30	177	5310		-0.53	30.53	0.28	5309.72
180-184	20	182	3640		-4.47	24.47	19.98	3620.02
185-189	19	187	3553		-9.47	28.47	89.68	3463.32
190-194	16	192	3072		-14.47	30.47	209.38	2862.62
195-199	12	197	2364		-19.47	31.47	379.08	1984.92
200-204	8	202	1616		-24.47	32.47	1054.3	561.7
a=5	178	1947	31601			328.53	3208.6	28392.4

GUIA DE ESTUDIO DE PROBABILIDAD

1.- Define probabilidad.

La **probabilidad** es un método por el cual se obtiene la frecuencia de un acontecimiento determinado mediante la realización de experimentos aleatorios, de los que se conocen todos los resultados posibles, bajo condiciones *suficientemente* estables. La probabilidad es un evento o suceso que puede ser improbable,probable o seguro.

La teoría de la probabilidad se usa extensamente en áreas como la estadística, la física, la matemática, las ciencias y la filosofía para sacar conclusiones sobre la probabilidad discreta de sucesos potenciales y la mecánica subyacente discreta de sistemas complejos, por lo tanto es la rama de las matemáticas que estudia, mide o determina a los experimentos o fenómenos aleatorios.

2.- Define espacio muestral y menciona su representación.

Es el conjunto de todos los posibles resultados de una **experiencia aleatoria**, lo representaremos por **E** (o bien por la letra griega **Ω**).
Espacio muestral de una moneda:
E = {C, X}.
Espacio muestral de un dado:
E = {1, 2, 3, 4, 5, 6}.
Suceso aleatorio
Suceso aleatorio es cualquier subconjunto del espacio muestral.
Por ejemplo al tirar un dado un suceso sería que saliera par, otro, obtener múltiplo de 3, y otro, sacar 5.

3.- Cuál es el principio fundamental del conteo.

El principio básico o fundamental de conteo se puede utilizar para determinar los posibles resultados cuando hay dos o más características que pueden variar.

Ejemplo: El helado puede venir en un cono o una tasa y los sabores son chocolate, fresa y vainilla.

```
                 / tasa de chocolate
    / chocolate <
   /             \ cono de chocolate
  /
 /         / tasa de fresa
<-- fresa <
 \         \ cono de fresa
  \
   \            / tasa de vainilla
    \ vainilla <
                \ cono de vainilla
```

El diagrama anterior se llama diagrama de árbol y muestra todas las posibilidades. El diagrama de árbol también se puede ordenar de otra forma. Ambos diagramas tienen un total de 6 resultados.

```
             / tasa de chocolate
            /
    / tasa <-- tasa de fresa
```

```
   /        \
  /          \ tasa de vainilla
 /
<
 \
  \          / cono de chocolate
   \        /
    \ cono <-- cono de fresa
            \
             \ cono de vainilla
```

Para determinar la cantidad total de resultados, multiplica la cantidad de posibilidades de la primera característica por la cantidad de posibilidades de la segunda característica. En el ejemplo anterior, multiplica 3 por 2 para obtener 6 posibles resultados.

Si hay más de dos resultados, continúa multiplicando las posibilidades para determinar el total de resultados.

4.- En un estudio de economía de combustible se prueban 3 autos de ***carreras con 5 diferentes marcas de gasolina en 7 sitios de pruebas utilizándose 2 pilotos. ¿Cuántas son todas posibilidades existentes que reflejen la totalidad del espacio muestal?.***

5.- Una prueba consta de 3 preguntas en las que solo se puede contestar ***"verdadero" o "falso", cuántas maneras diferentes existen para contestar la prueba.***

6.- Realiza la operación "5!"

5!= 5 x 4 x 3 x 2 x 1= 120

7.- Realiza la operación (2!*1!*5!).

8.-Con las letras de la palabra "RELOJ" cuantas permutaciones con repetición de 2 letras del total las cinco pueden hacerse.

9.- A que es igual la formula nPr= n!/(n-r)!

Formula de la función factorial

10.- Sea A=(a, b, c, d) ¿cuántas ordenaciones sin repetición se pueden hacer?

(b, c, d, a) – (c, d, a, b) – (d, a, b, c)

11.- Si tomamos el mismo conjunto A= (a, b, c, d) ¿Cuántos subconjuntos de 2 elementos cada uno se pueden obtener?

12.- Define Evento elemental.

Se llama evento elemental al que solamente contiene un elemento del espacio muestral y se representa con una letra minúscula.

De acuerdo con lo anterior, si se lanza un dado hay seis eventos elementales, que son {1}, {2}, {3}, {4}, {5}, {6} y el conjunto de los seis números es el espacio muestral del experimento,

13.- Define Evento seguro.

Un evento seguro es cuando se tiene el 100 % de probabilidad de que ocurra, por ejemplo, el agua a presión normal y "limpia" hervirá a los 100 °C, es lo que yo conozco como un evento determinístico, también puede ocurrir que un evento tenga varios resultados posibles, como lanzar un dado y decimos que hay un 100 % de que el número que quede en la cara de arriba al lanzarlo sea un 1 o 2 o 3 o 4 o 5 o 6, es decir planteamos como "evento", todas las posibilidades que se puedan dar.

14.- Qué son los eventos mutuamente excluyentes.

Los eventos mutuamente excluyentes son dos resultados de un evento que *no pueden ocurrir al mismo tiempo*.

- Sacar una carta de un mazo estándar y que salga un as y un rey son eventos mutuamente excluyentes, ya que no pueden ocurrir los dos al mismo tiempo.

- Sin embargo, sacar una carta roja y rey no son eventos mutuamente excluyentes, ya que puedes sacar perfectamente un rey rojo.

15.- Define la probabilidad clásica de un evento E.

La probabilidad clásica de un evento E, que denotaremos por P(E), se define como el número de eventos elementales que componen al evento E, entre el número de eventos elementales que componen el espacio maestral:

16.- Cuál es la probabilidad de caiga un número par en la acción de lanzar un dado.

Evento elemental

17.- Cuál es la probabilidad de que al alzar dos dados simultáneamente la suma de los puntos en sus caras sea igual a 4 (cuatro).

Evento mutuamente excluyente

18.- Según los axiomas de probabilidad cual es la condición de que un evento cualquiera ocurra con toda certeza.

El evento seguro

BIBLIOGRAFIA

http://www.mathematicsdictionary.com/spanish/vmd/full/c/classinterval.htm
http://es.slideshare.net/pbacelis/datos-agrupados-y-no
Estadística: Medidas de tendencia central
Serie de datos - Wikipedia, la enciclopedia libre
Distribución de frecuencias
La desviación estándar o desviación típica es la raíz cuadrada de la varianza.
http://www.disfrutalasmatematicas.com/datos/desviacion-estandar.html
... vía Definición ABC http://www.definicionabc.com/general/probabilidad.php
Leer más: http://www.monografias.com/trabajos69/teoria-elemental-probabilidad/teoria-elemental-probabilidad.shtml#ixzz45pjcWwTJ