Super Nota

Nombre del Alumno: ***Scarlet Alegría Sánchez***

Nombre del tema: ***Actividad II***

Parcial: ***I***

Nombre de la Materia: ***Estadística Descriptiva***

Nombre del profesor: ***Andrés Alejandro Reyes Molina***

Nombre de la Licenciatura: ***Licenciatura en contaduría pública y finanzas***

Cuatrimestre: ***3er***

Nomenclatura del grupo: ***LCF26SSC0124-A***

Clave de la materia: ***LCF315***

UNIDAD 3

ESTADISTICOS DE FORMA DE LA DISTRIBUCION

3. 1 ASIMETRIA

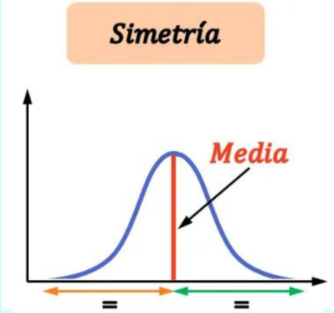
Las medidas de asimetría son indicadores que permiten establecer el grado de simetría (o asimetría) que presenta una distribución de la probabilidad de una variable aleatoria sin tener que hacer su representación gráfica.

Eje de simetría

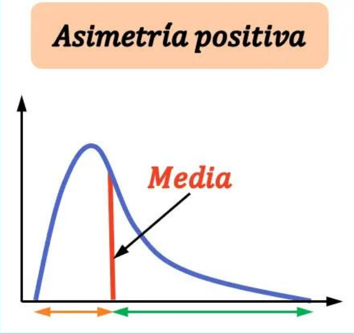
Una recta paralela al eje de ordenadas que pasa por la media de la distribución.

Pueden ocurrir tres casos:

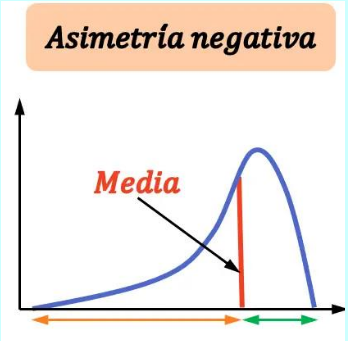
Si la distribución es simétrica, existe el mismo número de valores a la derecha que la izquierda de la media, por tanto, el mismo numero de desviaciones con signo negativo que con signo positivo. En este caso coincide la moda, la media y la mediana, la distribución se adapta a la forma de la campana de Gauss



Decimos que hay asimetría positiva (o la derecha) si la "cola" a la derecha de la media es mas larga que la de la izquierda, es decir que hay valores mas separados de la media a la derecha.



Diremos que hay asimetría negativa (o la izquierda) si la "cola" a la izquierda de la media es mas larga que la de la derecha, es decir, si hay valores mas separados de la media a la izquierda.

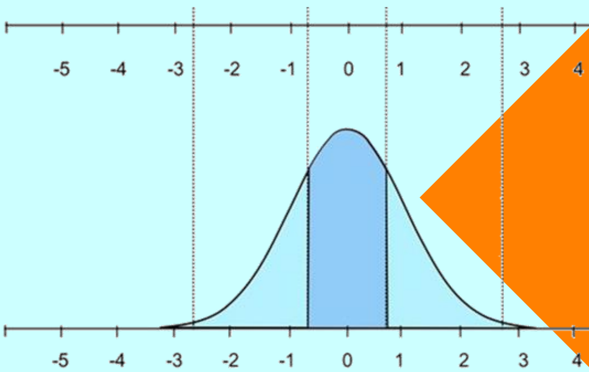


3. 2 APUNTAMIENTO O CURTOSIS

La curtosis es una medida estadística que determina el grado de concentración que presentan los valores de una variable alrededor de la zona central de la distribución de frecuencias.

También se le conoce como......

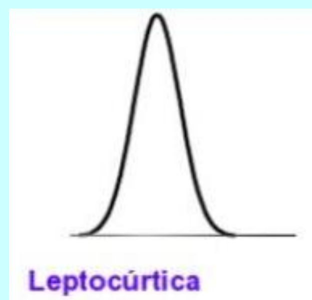
Medida de apuntamiento



Existen algunas variables que presentan un mayor grado de concentración (menor dispersión) de los valores en torno a su media y otras, por el contrario, presentan un menor grado de concentración (mayor dispersión) de sus valores en torno a su valor central.

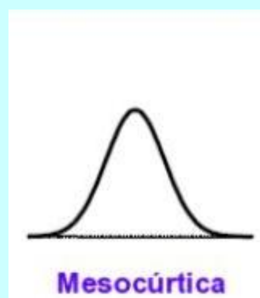
Leptocurtica

Existe una gran concentración de los valores en torno a su media ($g_2>3$)



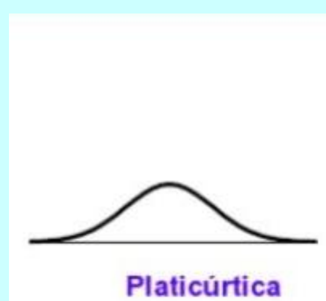
Mesocurtica

Existe una concentración normal de los valores en torno a su media ($g_2=3$)



Platicurtica

Existe una baja concentración de los valores en torno a su media ($g_2<3$)



3. 3 ESTADÍSTICOS DE POSICIÓN INDIVIDUAL

Los estadísticos introducidos en este tema indican la posición relativa de las puntuaciones individuales respecto de su grupo. Por ello, son especialmente adecuadas para comparar datos de individuos respecto de grupos no similares en tendencia central, variabilidad u otras características, y respecto de variables con escalas diferentes.

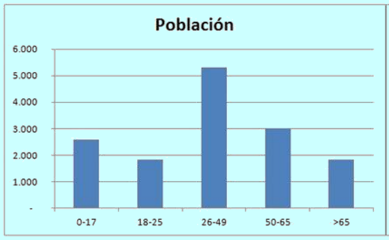
EJEMPLO:

Nos dice un amigo que les han pasado a todos los trabajadores de su empresa un test de aptitudes verbales y que ha obtenido una puntuación igual a 134; a continuación, sin más detalles, nos pregunta que si esa puntuación significa que es bueno o malo en aptitudes verbales. ¿qué información adicional podría sernos de utilidad a fin de poder ofrecerle algún tipo de valoración de esa puntuación?

EN ESTE CASO SE TRATA DE ESTADÍSTICOS QUE NOS VA A OFRECER INFORMACIÓN SOBRE UN VALOR CONCRETO EN RELACIÓN A LA POSICIÓN QUE OCUPA DENTRO DE UN CONJUNTO DE VALORES OBSERVADOS, Y CON ESTO NOS VA A PERMITIR ESTABLECER UNA INTERPRETACIÓN RELATIVA DE LOS VALORES OBSERVADOS

3. 4 LOS PORCENTAJES ACUMULADOS

El porcentaje acumulado (%a) de un valor concreto de una variable es el porcentaje de casos que obtienen un valor inferior o igual a ese en la variable en cuestión, información que puede obtenerse directamente a partir de la distribución de frecuencias correspondiente a esa variable.



Edad	Población
0-17	2.592
18-25	1.834
26-49	5.314
50-65	3.012
>65	1.839

Nota	Alumnos
Suspenso	7
Aprobado	12
Notable	14
Sobresaliente	4

Puede incluir una columna o una fila en el informe que muestre un total acumulado. El total acumulado se puede expresar como un valor numérico o un porcentaje. En informes de Repórter, se puede calcular un total acumulado para más de una categoría.

EJEMPLO DE OBTENCION DE PORCENTAJES

Sea la siguiente distribución de frecuencias de las puntuaciones en un test de inteligencia administrado a una muestra de 250 personas.

¿Cuál sería el porcentaje acumulado correspondiente a una puntuación de 97 en ese test?, ¿cómo la interpretaríamos? - ¿Y si la puntuación fuese igual a 103? - ¿Y si fuese igual a 91? => interpolación del %a

CI	n_i	n_a	p_i	p_a	$\%_a$
89	1	1	0,004	0,004	0,4
90	2	3	0,008	0,012	1,2
92	3	6	0,012	0,024	2,4
93	5	11	0,02	0,044	4,4
94	8	19	0,032	0,076	7,6
95	10	29	0,04	0,116	11,6
96	14	43	0,056	0,172	17,2
97	17	60	0,068	0,24	24
98	24	84	0,096	0,336	33,6
99	29	113	0,116	0,452	45,2
100	36	149	0,144	0,596	59,6
101	33	182	0,132	0,728	72,8
102	26	208	0,104	0,832	83,2
103	19	227	0,076	0,908	90,8
104	12	239	0,048	0,956	95,6
105	7	246	0,028	0,984	98,4
107	2	248	0,008	0,992	99,2
110	1	249	0,004	0,996	99,6
114	1	250	0,004	1	100
	250				

Nota: En modo Repórter, si la categoría seleccionada incluye un cálculo existente, el valor del cálculo se incluye en el total acumulado.

- Cálculo de los totales acumulados como valores numéricos. Se pueden mostrar totales acumulados como valores numéricos.
- Cálculos totales acumulados como valores de porcentaje se pueden mostrar totales acumulados como valores de porcentaje

3. 5 LAS PUNTUACIONES TIPICAS

Las puntuaciones directas (puntuaciones de un sujeto en un test, etc.) son los primeros datos de los que habitualmente disponemos, pero la comparación de las puntuaciones directas de un mismo sujeto en dos variables puede llevarnos a confusión. De hecho, conocida una puntuación directa no sabemos si se trata de un valor alto o bajo porque esto depende del promedio del grupo.

Si a una puntuación directa Xi le restamos la media de su grupo obtenemos una puntuación diferencial o de diferencia, que representamos por xi (minúscula):

$$x_i = X_i - \bar{X}$$

Las puntuaciones diferenciales nos indican si la puntuación coincide con la media de su grupo, es inferior o es superior a ella. Tienen las siguientes propiedades:

SU MEDIA ES CERO: $\bar{x} = 0$

LA VARIANZA DE LAS PUNTUACIONES DIFERENCIALES ES IGUAL A LA VARIANZA DE LAS PUNTUACIONES DIRECTAS.

Por tanto, al restar a las puntuaciones directas su media hemos obtenido una nueva escala con media 0 y con idéntica varianza a las puntuaciones directas.

DATO

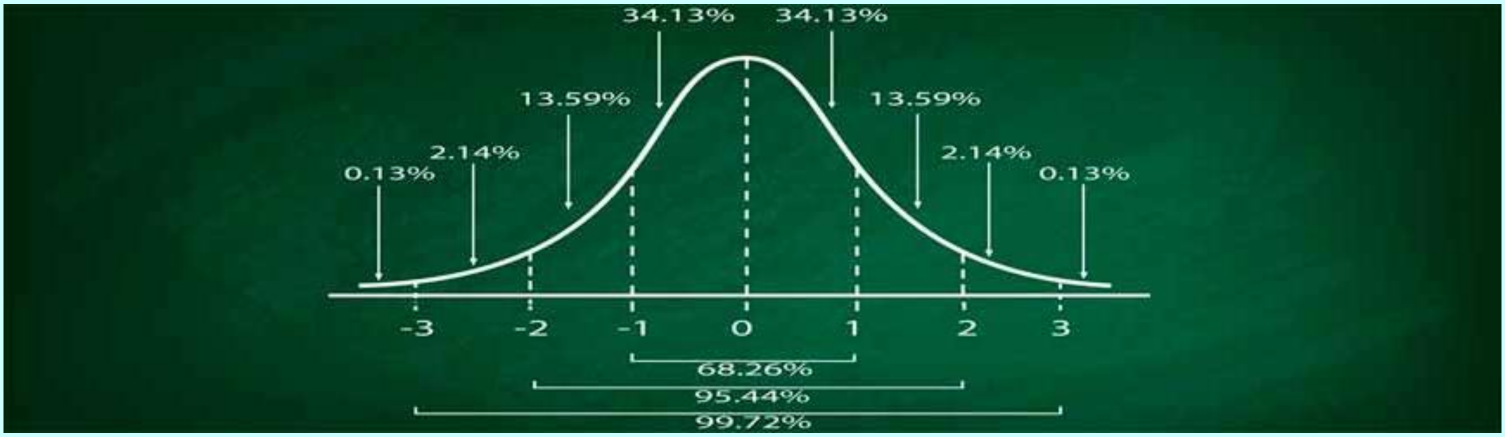
Al proceso de obtener puntuaciones típicas se llama tipificación, y las puntuaciones se denominan también ``tipificadas``.

Las puntuaciones típicas tienen las siguientes propiedades:

SU MEDIA ES CERO:

SU VARIANZA ES IGUAL A 1:

Las puntuaciones típicas reflejan las relaciones entre las puntuaciones con independencia de la unidad de medida. Así, permiten hacer comparaciones entre distintos grupos e incluso entre distintas variables.



3. 6 MEDIDAS DE POSICIÓN

Son medidas estadísticas cuyo valor representa el valor del dato que se encuentra en el centro de la distribución de frecuencia, por lo que también se les llama "Medidas de Tendencia Central ".

Pero estas medidas de posición de una distribución de frecuencias han de cumplir determinadas condiciones para que lean verdaderamente representativas de la variable a la que resumen.

se describen las medidas de posición más comunes utilizadas en estadística, como lo son:

Cuartiles:

Hay 3 cuartiles que dividen a una distribución en 4 partes iguales: primero, segundo y tercer cuartil.

CUARTILES (Q1, Q2, Q3)

Primer cuartil (Q1): Aquel valor de una serie que supera al 25% de los datos y es superado por el 75% restante. Formula de Q1 para series de Datos Agrupados en Clase.

$$Q_1 = L_i + \frac{\frac{\sum f_i}{4} - f_{aa}}{f_i} * I_c$$

Segundo cuartil (Q2): Coincide, es idéntico o similar al valor de la Mediana (Q2 = Md). Es decir, supera y es superado por el 50% de los valores de una Serie.

Tercer cuartil (Q3): Aquel valor, termino o dato que supera al 75% y es superado por el 25% de los datos restantes de la Serie.

$$Q_3 = L_i + \frac{\frac{3\sum f_i}{4} - f_{aa}}{f_i} * I_c$$

Deciles:

Hay 9 deciles que la dividen en 10 partes iguales: (primero al noveno decil).

DECILES

(D1, D9)

Primer decil (D1):

El primer decil es aquel valor de una serie que supera a 1/10 parte de los datos y es superado por las 9/10 partes restantes (respectivamente, hablando en porcentajes, supera al 10% y es superado por el 90% restante).

$$D_1 = L_i + \frac{\frac{*\sum f_i}{10} - f_{aa}}{f_i} * I_c$$

Noveno decil (D3):

El D9 (noveno decil) supera al 90% y es superado por el 10% restante. Como se observa, son formulas parecidas a la del cálculo de la Mediana, cambiando solamente las respectivas posiciones de las medidas.

$$D_9 = L_i + \frac{\frac{*9\sum f_i}{10} - f_{aa}}{f_i} * I_c$$

Percentiles:

Hay 99 percentiles que dividen a una serie en 100 partes iguales: (primero a los noventa y nueve percentiles).

PERCENTIL

(P1, P99)

Primer percentil (P1):

El primer percentil supera al uno por ciento de los valores y es superado por el noventa y nueve por ciento restante.

$$P_1 = L_i + \frac{\frac{*\sum f_i}{100} - f_{aa}}{f_i} * I_c$$

Noventa y nueve percentiles (P99):

El P99 (noventa y nueve percentiles) supera al 99% de los datos y es superado a su vez por el 1% restante.

$$P_{99} = L_i + \frac{\frac{*99\sum f_i}{100} - f_{aa}}{f_i} * I_c$$

Para determinar estas medidas se aplicará el principio de la mediana; así, el primer cuartil cereal valor por debajo del cual se encuentra el 25 por ciento de los datos; bajo el tercer cuartil se encuentra el 75 por ciento; el 80 decil será el valor por encima del cual estará el 20 por ciento de los datos, etc. Como se observa, todas estas medidas no son sino casos particulares del percentil ya que el primer cuartil no es sino el 25° percentil, el tercer cuartil el 75° percentil, el cuarto decil el 40° percentil, etc.

3. 7 LAS ESCALAS DERIVADAS

se han propuesto algunas transformaciones lineales de las puntuaciones típicas que pretenden hacerlas más intuitivamente interpretables.

Todas estas escalas derivadas de la escala de las puntuaciones típicas se basan en una transformación genérica del tipo:

$$D_i = a\, z_i + b,$$

Consecuencia inmediata de este tipo de transformación, las nuevas puntuaciones D pasarán de tener una media 0 y una desviación típica 1, a una media b y una desviación típica a.

Diversas propuestas de escala derivadas han sido planteadas sin que haya generalizado el uso concreto de ninguna de ellas. Entre las que han tenido más repercusión las siguientes:

- La escala T $\longrightarrow$ $T_i = 10 \cdot z_i + 50$ $\quad (\overline{T} = 50 \quad s_T = 10)$

- La escala S o de estaninos $\longrightarrow$ $S_i = 2 \cdot z_i + 5$ $\quad (\overline{S} = 5 \quad s_S = 2)$

- La escala CI $\longrightarrow$ $CI_i = 15 \cdot z_i + 100$ $\quad (\overline{CI} = 100 \quad s_{CI} = 15)$

3.8 ORGANIZACIÓN Y REPRESENTACIÓN GRÁFICA DE DATOS MULTIVARIADOS

Se describen una serie de procedimientos asociados al tratamiento conjunto de dos o más variables, los cuales van a permitir extraer diversas facetas de la información compartida por esas variables. En bastantes momentos se va a ceñir esta exposición al caso bivariado (dos variables) por ser más sencillo en su presentación y por tratarse, con frecuencia, del caso particular más simple del modo general de abordar el problema a nivel multivariado (dos o más variables)

1.- LA DISTRIBUCIÓN CONJUNTA MULTIVARIADA

La distribución conjunta multivariada De modo análogo a lo que se planteó para el caso univariado en el tema, un resumen básico de la información de un grupo de o más variables consiste en la distribución conjunta de frecuencias de las mismas, la cual se basa en el conteo del número de casos (frecuencias) que presentan las distintas combinaciones de valores que a nivel empírico se hayan dado para esas variables.

2.-LAS MODALIDADES DE UNA DISTRIBUCIÓN CONJUNTA DE FRECUENCIAS

Consisten, no en los valores de una variable concreta, sino en todas las posibles combinaciones de los valores de las variables que se consideren excepto aquellas combinaciones que no se hayan presentado a nivel empírico y que por tanto no tiene sentido incluir en la distribución de frecuencias.

La distribución conjunta de frecuencias relativas o proporciones (**p i**) y la de porcentajes (**% i**) pueden obtenerse a partir de las frecuencias absolutas dividiendo cada frecuencia absoluta entre el nº de casos (n) y multiplicando las frecuencias relativas por cien, respectivamente.

EL ORDENAMIENTO DE LAS MODALIDADES EN UNA DISTRIBUCIÓN CONJUNTA DE FRECUENCIAS

El ordenamiento de las modalidades en una distribución conjunta de frecuencias carece de sentido, si bien, se suelen situar en orden alfabético/numérico creciente a fin de poder localizar más fácilmente cualquier combinación de valores de las variables. La obtención de las frecuencias acumuladas, ya sean absolutas, relativas o porcentajes, carece también aquí de sentido dado que las modalidades de la distribución no representan un continuo –al igual que ocurría con las distribuciones de frecuencias de las variables categóricas.

3.9 LA DISTRIBUCIÓN CONJUNTA MULTIVARIADA

La distribución conjunta multivariada describe las probabilidades de diferentes combinaciones de resultados para múltiples variables aleatorias que están relacionadas. En otras palabras, muestra cómo se comportan varias variables juntas.

Variables aleatorias
La imposibilidad de ejercer o gozar de la capacidad de obrar se conoce como incapacidad.

Distribucion conjunta
Especifica la probabilidad de que cada variable tome un valor particular.

Variables independiente
Si el valor de una variable no afecta la probabilidad de las otras, se dice que son independientes.

Distribucion marginal
Es la distribución de probabilidad de una sola variable, ignorando las demás.

La distribución conjunta multivariada es fundamental en estadística y probabilidad, permitiendo analizar relaciones entre múltiples variables y modelar fenómenos complejos. Se utiliza en:

Análisis multivariante
Un área de la estadística que estudia la observación y análisis simultáneo de múltiples variables.

Predicción
Permite estimar el valor de una variable basado en el valor de otras.

Machine learning
Se utiliza para entrenar modelos que aprenden patrones y relaciones complejas entre datos.

Simulación
Permite generar datos sintéticos que siguen patrones observados en datos reales.

3.10 LA TABLA DE CONTINGENCIA

Es una tabla de doble entrada en que cada lado de la tabla contiene las modalidades de una variable. En las casillas interiores de la tabla aparecen las frecuencias conjuntas (ya sean absolutas, relativas o porcentajes) de la combinación de los valores fila y columna correspondientes.

EJEMPLO:

Se llevó a cabo un estudio para evaluar si el estado de ánimo de los mayores de 65 años podía verse influido por el hecho de vivir en una residencia geriátrica o no. Se recogieron datos de una muestra de 500 personas de las variables "Estado de ánimo": negativo (−); neutro (±); positivo (+) y "Vivir en residencia" (Sí; No). La distribución conjunta de frecuencias es la siguiente:

	−	±	+
Sí	48	42	60
No	70	105	175

¿Cómo se ha construido esa tabla de contingencia? Realizando, a partir de la matriz de datos original, un recuento del nº de casos que presentan cada combinación de par de valores.

Caso	Residencia	Estado ánimo
1	S	−
2	N	±
3	S	−
4	S	+
...	...	...
500	N	±

También a partir de esa tabla de datos, como ya se ha visto, es posible obtener la distribución de cada variable por separado (= distribuciones marginales):

Residencia (X)

X_i	n_i	p_i
Sí	150	0.30
No	350	0.70
	500	1

Estado ánimo (Y)

Y_i	n_i	p_i
−	118	0.236
±	147	0.294
+	235	0.470
	500	1

La distribución conjunta de ambas variables:

X_i, Y_i	n_i	p_i
Sí ∩ −	48	0.096
Sí ∩ ±	42	0.084
Sí ∩ +	60	0.120
No ∩ −	70	0.140
No ∩ ±	105	0.210
No ∩ +	175	0.350
	500	1

En las tablas de contingencia es habitual incluir en los laterales las sumas de las celdas de filas y columnas => distribuciones marginales (= distribución de cada variable por separado)

	−	±	+	Total
Sí	48	42	60	150
No	70	105	175	350
Total	118	147	235	500

3.11 CALCULO PARA UNA DISTRIBUCIÓN DE FRECUENCIA

Para el cálculo de esta medida en datos agrupados en una distribución de frecuencia, se utiliza el mismo procedimiento estudiado para el cálculo de la Mediana, e; cual es:

1. Se efectúa la columna de las frecuencias acumuladas.
2. Se determina la posición del término cuyo valor se pretende calcular, en caso de ser el primer cuartil será, $\frac{1*\sum f_i}{4}$ si fuese el 95° centil $\frac{95*\sum f_i}{100}$ .... Etc.
3. Se verifica cual es la clase que lo contiene; para ello se utiliza la columna de las frecuencias acumuladas.
4. Se hace la diferencia entre el número que representa el orden de posición cuyo valor se pretende calcular y la frecuencia acumulada de la clase anterior a la que lo contiene.
5. Se calcula la medida solicitada de acuerdo a la siguiente fórmula:

$$P = 1_i + \frac{P - f_{a-1}}{f_i} * I_c$$

Donde:

L i: límite inferior de la clase que lo contiene: valor que representa la posición de la medida.

fi: la frecuencia de la clase que contiene la medida solicitada.

fa-l: frecuencia acumulada anterior a la que contiene la medida solicitada.

Ic: intervalo de clase.

3.12 REPRESENTACIÓN GRÁFICA EN EL ANÁLISIS DE DATOS

La representación gráfica en el análisis de datos es una herramienta esencial para visualizar y comprender conjuntos de datos complejos. Permite identificar patrones, tendencias y relaciones que podrían pasar desapercibidos en datos tabulares, facilitando la interpretación y comunicación de resultados.

se aconseja que la presentación de datos numéricos se haga habitualmente por medio de tablas, en ocasiones un diagrama o un gráfico pueden ayudarnos a representar de un modo más eficiente nuestros datos.

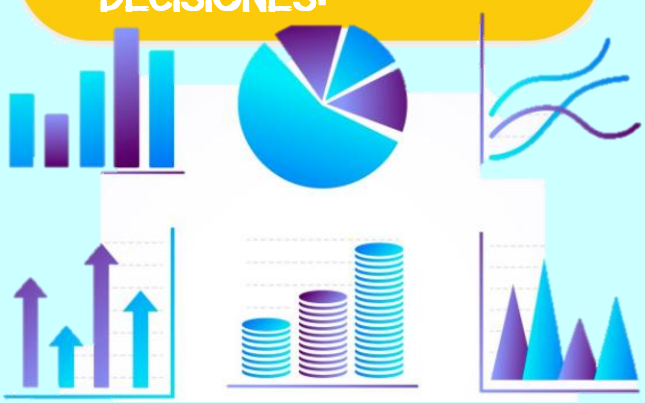
Beneficios de la representación gráfica:

COMPRENSIÓN RÁPIDA: Las imágenes visuales permiten una comprensión más rápida y fácil de la información que los datos tabulares.

IDENTIFICACIÓN DE PATRONES: Los gráficos ayudan a identificar patrones, tendencias y relaciones que podrían ser difíciles de detectar en datos sin procesar.

COMUNICACIÓN EFECTIVA: Los gráficos son una herramienta poderosa para comunicar hallazgos a una audiencia amplia y diversa.

FACILITA LA TOMA DE DECISIONES: Al comprender mejor los datos, se facilita la toma de decisiones informadas.



3.13 EL CASO DE DOS VARIABLES CATEGORICAS

En estadística, el análisis de dos variables categóricas implica examinar la relación entre dos variables que tienen categorías como valores.

Se utilizan tablas de contingencia, gráficos de barras agrupadas o apiladas para visualizar la relación, y pruebas estadísticas como la chi-cuadrado para determinar si existe una asociación significativa entre las variables.

Análisis de dos variables categóricas:

Variables categóricas:

Son variables que representan datos que se pueden dividir en categorías o grupos, como el estado civil (soltero, casado, divorciado)

Edad	Satisfacción en el trabajo: A	B	C	D	E	$n_{i\bullet}$
< 25	10 $e_{11} = 33,75$ $g_{11} = -12,16$	10 $e_{12} = 26,25$ $g_{12} = -9,65$	20 $e_{13} = 24,37$ $g_{13} = -3,95$	40 $e_{14} = 26,25$ $g_{14} = 16,85$	70 $e_{15} = 39,37$ $g_{15} = 40,28$	150 (150) (31,37)
25 - 36	20 $e_{21} = 21,37$ $g_{21} = -1,33$	10 $e_{22} = 16,62$ $g_{22} = -5,08$	15 $e_{23} = 15,44$ $g_{23} = -0,43$	20 $e_{24} = 16,62$ $g_{24} = 3,7$	30 $e_{25} = 24,94$ $g_{25} = 5,54$	95 (95) (2,4)
> 36	60 $e_{31} = 34,87$ $g_{31} = 32,56$	50 $e_{32} = 27,12$ $g_{32} = 30,59$	30 $e_{33} = 25,19$ $g_{33} = 5,24$	10 $e_{34} = 27,12$ $g_{34} = -9,98$	5 $e_{35} = 40,69$ $g_{35} = -10,48$	155 (155) (47,93)
$n_{\bullet j}$	90	70	65	70	105	400 (81,7)

Tablas de contingencia:

También llamadas tablas de doble entrada, son una herramienta para organizar y visualizar la relación entre dos variables categóricas. Muestran la frecuencia de cada combinación de categorías.

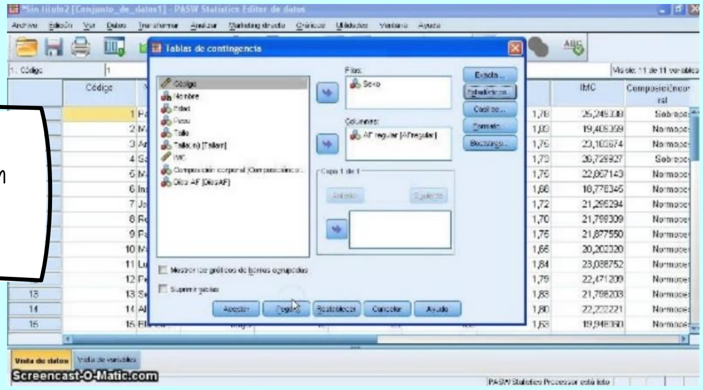
Género \ Inclinación profesional	Lenguas Extranjeras	Licenciatura en Matemáticas	Enfermería	Total
Masculino (M)	$\frac{7}{50} = 0,14$	$\frac{15}{50} = 0,3$	$\frac{5}{50} = 0,1$	$\frac{27}{50} = 0,54$
Femenino (F)	$\frac{8}{50} = 0,16$	$\frac{5}{50} = 0,1$	$\frac{10}{50} = 0,2$	$\frac{23}{50} = 0,46$
Total	$\frac{15}{50} = 0,3$	$\frac{20}{50} = 0,4$	$\frac{15}{50} = 0,3$	$\frac{50}{50} = 1$

Gráficos:

Se pueden utilizar gráficos de barras agrupadas o apiladas para representar la relación entre las variables categóricas. En un gráfico de barras agrupadas, cada categoría de una variable se representa con un grupo de barras, y cada barra dentro del grupo representa una categoría de la otra variable. En un gráfico de barras apiladas, cada barra representa una categoría de una variable, y las barras se dividen en segmentos que representan las categorías de la otra variable.

Prueba de independencia chi-cuadrado:

Es una prueba estadística que se utiliza para determinar si existe una asociación significativa entre dos variables categóricas. Compara las frecuencias observadas en la tabla de contingencia con las frecuencias esperadas si las variables fueran independientes.



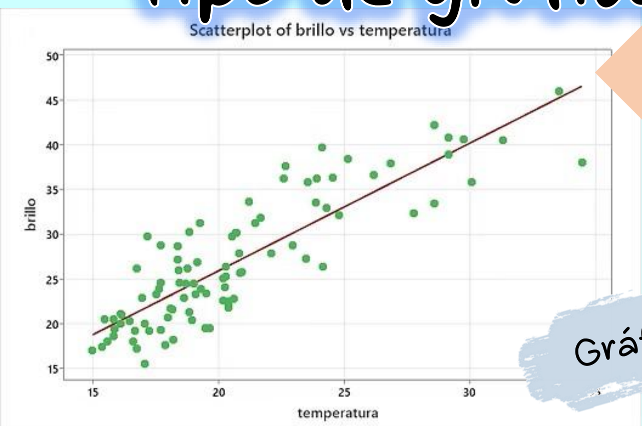
3.14 EL CASO DE DOS VARIABLES CUANTITATIVAS

Busca determinar si existe una relación entre ellas y, en caso afirmativo, cuál es la naturaleza de esa relación. Esto implica examinar la covariación de las variables, medir la fuerza y dirección de la asociación y, a menudo, representarlas gráficamente para visualizar su comportamiento.

ANALISIS DE DOS VARIABLES CUANTITATIVAS

Cuando se observan dos características cuantitativas (es decir, medibles) en cada individuo de una población, se considera que se está estudiando una variable estadística bidimensional, representada como (X, Y). El objetivo principal es analizar la relación entre estas dos variables, examinando si una afecta a la otra y cómo lo hace.

Tipo de grafico:



Una representación gráfica de los datos donde cada punto representa un par de valores (X, Y). Permite visualizar la relación entre las variables y detectar posibles patrones o tendencias.

Gráfico de dispersión

PASOS COMUNES EN EL ANÁLISIS:

- Recolección de datos: Obtener pares de valores (X, Y) para cada individuo.
- Cálculo de medidas de resumen: Calcular la media, desviación estándar y otras medidas descriptivas para cada variable.
- Análisis de la covariación: Determinar si las variables tienden a cambiar juntas y en qué dirección.
- Cálculo del coeficiente de correlación: Cuantificar la fuerza y dirección de la relación lineal.
- Representación gráfica: Crear un diagrama de dispersión para visualizar la relación.
- Interpretación: Analizar los resultados y sacar conclusiones sobre la relación entre las variables.

3.15 EL CASO DE UNA VARIABLE CATEGÓRICA Y UNA VARIABLE CUANTITATIVA

El diagrama de dispersión también puede ser aplicado en la representación conjunta de la distribución de frecuencias absolutas de una variable categórica y una variable cuantitativa. A este tipo de gráfico se le denomina en algunos textos como diagrama de puntos y es habitual que aparezca representada la variable categórica (cualitativa) en el eje de abscisas y la variable numérica (cuantitativa) en el eje de ordenadas.

EJEMPLO:

Comparar el rendimiento académico (variable cuantitativa) entre estudiantes que asisten a diferentes tipos de escuelas (variable categórica, por ejemplo, escuela pública vs. escuela privada).

En resumen:

Al analizar una variable categórica junto con una variable cuantitativa, se pueden obtener ideas valiosas sobre la relación entre diferentes grupos y sus características numéricas.

También se pueden usar gráficos, como diagramas de caja o histogramas, para visualizar la distribución de la altura para cada género y facilitar la comparación.