

ACTIVIDAD II



SUPER NOTA

ESTADÍSTICOS DE FORMA DE LA DISTRIBUCIÓN

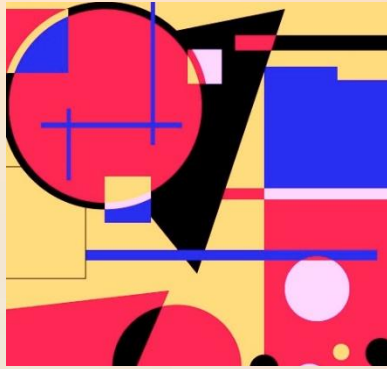


JUAN CARLOS DE LEÓN MORALES
ADMINISTRACIÓN Y ESTRATEGIAS
DE NEGOCIOS

Asimetría

1

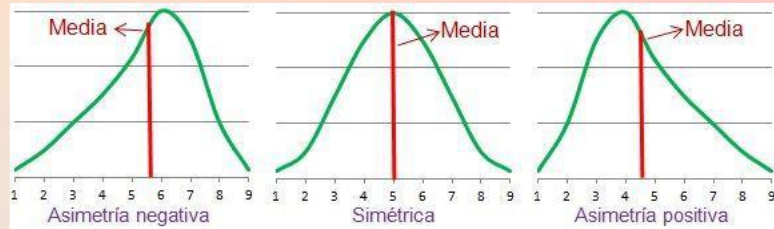
Son indicadores que permiten establecer el grado de simetría (o asimetría) que presenta una distribución de probabilidad de una variable aleatoria sin tener que hacer su representación gráfica.



2

Tipos

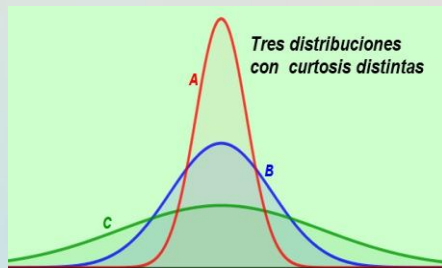
- Asimetría negativa: se alarga para valores inferiores.
- Simétrica: se adapta a la forma de la campana.
- Asimetría positiva: se alarga para valores superiores.



1

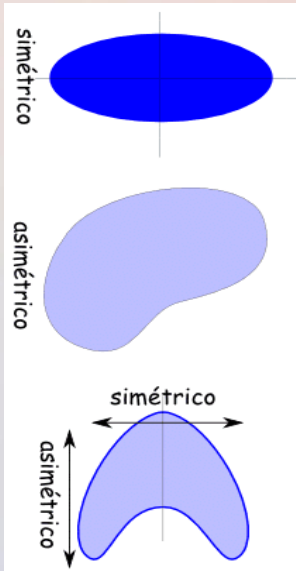
Apuntamiento o Curtosis

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.



3

Utilidad

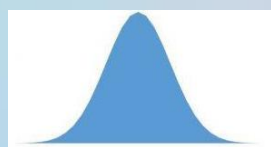


La asimetría resulta útil en muchos campos. Muchos modelos simplistas asumen una distribución normal, esto es, simétrica en torno a la media. La distribución normal tiene una asimetría cero.

2

Tipos

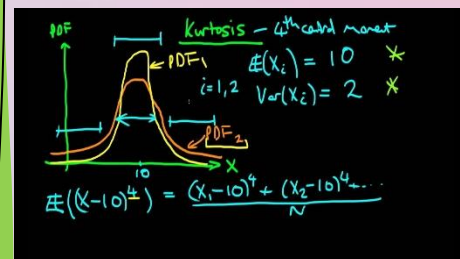
- ✚ Leptocúrtica: gran concentración de los valores ($g_2 > 3$).
- ✚ Mesocúrtica: concentración normal de los valores ($g_2 = 3$).
- ✚ Platicúrtica: baja concentración de los valores ($g_2 < 3$).



3

Pruebas

Si la altura media de la clase es 1,72 cm, lo más normal es que las alturas del resto de los alumnos estén en torno a este valor (con cierto grado de variabilidad, pero sin ser esta demasiado grande).



1

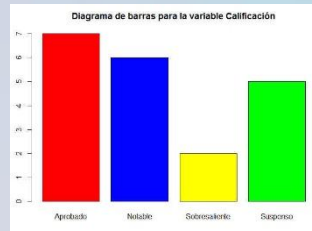
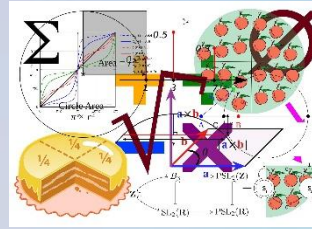
Estadísticos De Posición Individual

En cuanto que se trata de estadísticos que ofrecen información sobre la posición de un valor respecto a un grupo de referencia, nos van a permitir establecer una interpretación relativa de los valores observados.



2

Ejemplo

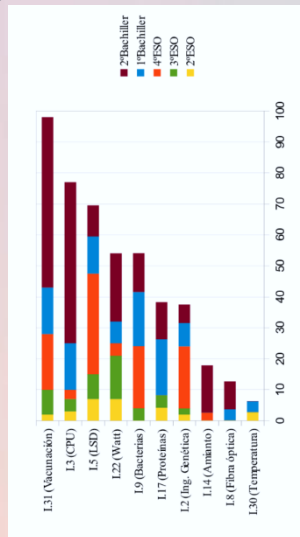


Nos dice un amigo que les han pasado a todos los trabajadores de su empresa un test de aptitudes verbales y que ha obtenido una puntuación igual a 134; a continuación, sin más detalles, nos pregunta que si esa puntuación significa que es bueno o malo en aptitudes verbales.

1

Los Porcentajes Acumulados.

El porcentaje acumulado (%) 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.



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.

Propiedades

2

Produits	Nb de Ventes	Pourcentage	Cumul
Produit A	12	16,22%	16,22%
Produit B	8	10,81%	27,03%
Produit C	15	20,27%	47,30%
Produit D	18	24,32%	71,62%
Produit E	7	9,46%	81,08%
Produit F	4	5,41%	86,49%
Produit G	10	13,51%	100,00%

3

Ejemplo

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

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

4

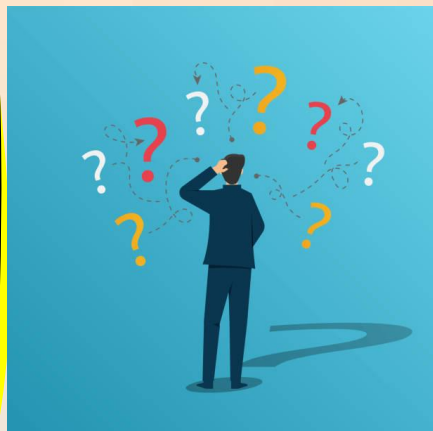
En modo Reporter, si la categoría seleccionada incluye un cálculo existente, el valor del cálculo se incluye en el total acumulado.



1

Las Puntuaciones Típicas

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.



2

Propiedades

- Su media es cero: $\bar{x} = 0$
- La varianza de las puntuaciones diferenciales es igual a la varianza de las puntuaciones directas.

- 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.

1

Medidas De Posición

Las medidas de posición nos facilitan información sobre la serie de datos que estamos analizando.



2

Centro

En vez de manejar todos los datos sobre las variables, tarea que puede ser pesada, podemos caracterizar su distribución de frecuencias mediante algunos valores numéricos, eligiendo como resumen de los datos un valor central alrededor del cual se encuentran distribuidos los valores de la variable.



3

Las Mas Comunes

- Cuartiles: Hay 3 cuartiles que dividen a una distribución en 4 partes iguales: primero, segundo y tercer cuartil.
- Deciles: Hay 9 deciles que la dividen en 10 partes iguales: (primero al noveno decil).
- Percentiles: Hay 99 percentiles que dividen a una serie en 100 partes iguales: (primero al noventa y nueve percentiles).



1

Las Escalas Derivadas

Una dificultad con las puntuaciones típicas se puede plantear a la hora de comunicar resultados debido a las posiciones decimales y valores negativos inherentes a las mismas.



3

Muestra

Todas estas escalas derivadas de la escala de las puntuaciones típicas se basan en una transformación genérica del tipo: $d_i = a \cdot z_i + b$.



2

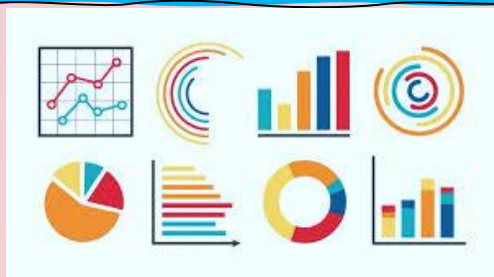
Escalas

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 $T \rightarrow T_i = 10 \cdot z_i + 50 \quad (\bar{T} = 50 \quad s_T = 10)$
- La escala S o de estatinos $\rightarrow S_i = 2 \cdot z_i + 5 \quad (\bar{S} = 5 \quad s_S = 2)$
- La escala $CI \rightarrow CI_i = 15 \cdot z_i + 100 \quad (\bar{CI} = 100 \quad s_{CI} = 15)$

Organización Y Representación Gráfica De Datos Multivariados

Tras abordar en temas previos el tratamiento individualizado (univariado) de las variables, en éste y temas sucesivos 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.



1

2

Uso

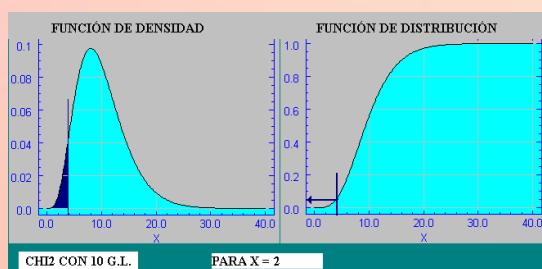


La distribución conjunta multivariada De modo análogo a lo que se planteó para el caso univariado en el tema, un resumen básico de la información de un grupo de o más variables consiste en la distribución conjunta de frecuencias de las mismas.

1

La Distribución Conjunta Multivariada

Un resumen básico de la información de un grupo de 2 o más variables 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

Ejemplo

Se obtuvo información en una muestra de 71 sujetos de las 3 variables siguientes: sexo (1: hombre; 2: mujer); n° de parejas estables a lo largo de los últimos 5 años; y situación emocional actual (1: satisfactoria; 2: ni satisfactoria ni insatisfactoria; 3: insatisfactoria).

ID	Sexo	Num parejas	Sit actual
1	1	1	3
2	1	4	2
3	2	1	1
4	2	2	1
5	2	1	3
6	1	0	1
7	2	3	2
...	...	...	...
71	1	1	1

La Tabla De Contingencia

1

En el caso de dos variables, una forma muy conveniente de visualizar la distribución conjunta de ambas es en forma de tabla de contingencia, esto es, una tabla de doble entrada en que cada lado de la tabla contiene las modalidades de una variable.

TABLA DE CONTINGENCIA PARA VARIABLES CUALITATIVAS

Estado Civil	Genero		Total
	Masculino	Femenino	
Soltero	10	8	18
Casado	5	13	18
Divorciado	12	11	23
Total	27	32	59

3

Organización

La organización de los datos de la anterior tabla en forma de distribución conjunta de frecuencias (absolutas) quedaría como sigue, donde $X = \text{Sexo}$; $Y = \text{Num_parejas}$ y $Z = \text{Sit_actual}$.

X_i, Y_i, Z_i	n_i
1,0,1	4
1,0,2	3
1,0,3	2
1,1,1	12
1,1,2	8
1,1,3	6
1,2,1	5
1,2,2	1
1,2,3	2
1,4,2	1
2,0,1	6
...	...
2,3,2	1
	71

Ejemplo

2

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 ($\pm$); positivo (+) y “vivir en residencia” (sí; no). La distribución conjunta de frecuencias es la siguiente:

	-	$\pm$	+
Sí	48	42	60
No	70	105	175

3

Nota

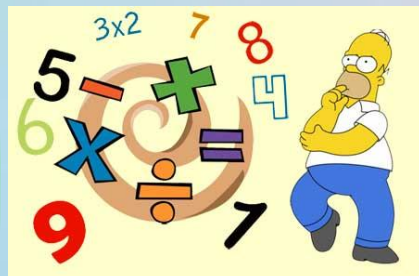
En las tablas de contingencia es habitual incluir en los laterales las sumas de las celdas de filas y columnas => distribuciones marginales (= distribución de cada variable por separado).

	-	$\pm$	+	Total
Sí	48	42	60	150
No	70	105	175	350
Total	118	147	235	500

Cálculo Para Una Distribución De Frecuencia

1

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.



2

Cálculo

Se determina la posición del término cuyo valor se pretende calcular, en caso de ser el primer cuartil será, si fuese el 95° centil ... etc.

$$\frac{1 * \sum f_i}{4}$$

$$\frac{95 * \sum f_i}{100}$$

3

Ejemplo

Determinación del primer cuartil, el cuartil textil, el séptimo decil y el 30° percentil.

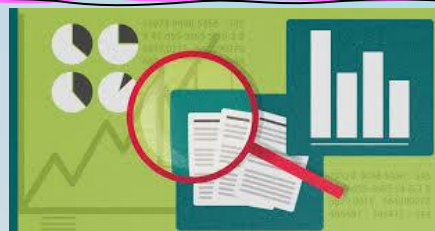
Formula:

$$P = \left[f_{a-1} + \frac{f_i(P - L_i)}{lc} \right] \frac{100}{N}$$

1

Representación Gráfica En El Análisis De Datos

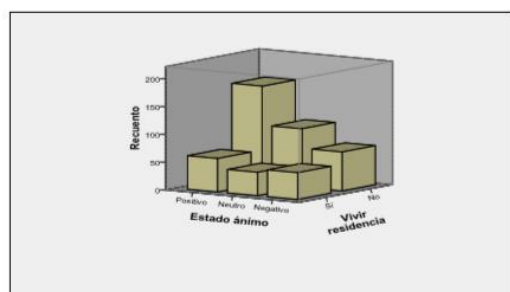
Se presentan a continuación una serie de gráficos orientados a mostrar datos multivariados, si bien, la mayoría se ciñen al caso de 2 variables. Se diferencia su presentación en función del tipo de variables al que van dirigidos.



1

El Caso De Dos Variables Categóricas

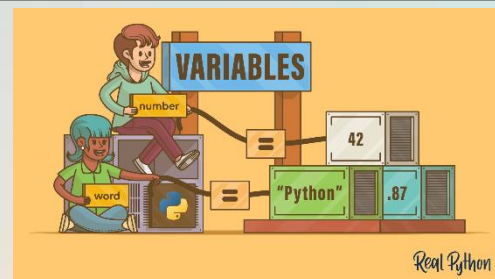
El diagrama de barras tridimensional o 3-D Ejemplos de diagrama de barras 3-D con la distribución conjunta de frecuencias absolutas de “Estado de ánimo” y “Vivir residencia”, intercambiando la posición de ambas variables.



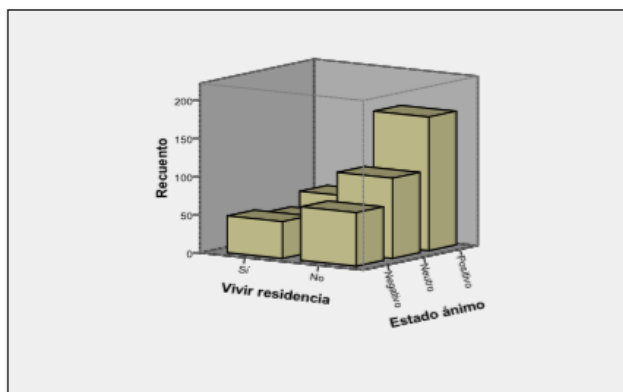
2

Variantes

Aunque 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.



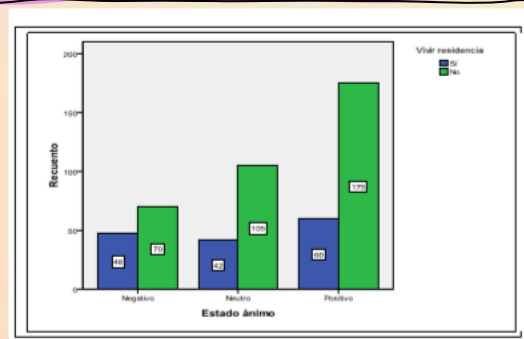
2



3

Diagrama De Barras

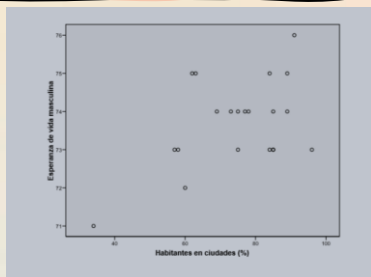
El diagrama de barras Ejemplos de diagrama de barras de la distribución conjunta de frecuencias absolutas de “Estado de ánimo” y “Vivir residencia” intercambiando la posición de ambas variables.



1

El Caso De Dos Variables Cuantitativas

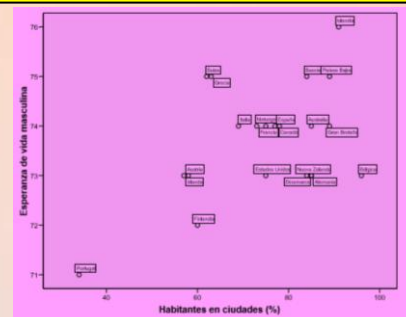
Ejemplo de diagrama de dispersión obtenido a partir de los datos de una muestra de 8 fumadores en las variables “N° de cigarrillos que, en promedio, se fuma al día” y “Puntuación en un test de ansiedad [0, ..., 30]”. Se muestran también los datos a partir de los que ha sido obtenido el mismo con el programa SPSS:



2

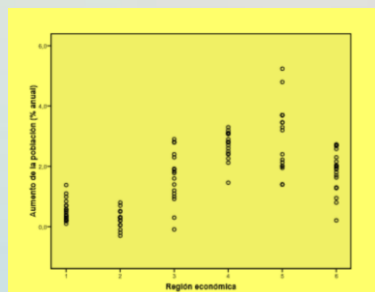
Ejemplo

A continuación, se muestra el mismo diagrama de dispersión con los puntos etiquetados, a fin de facilitar la posición de los casos (en este ejemplo, países) en la representación gráfica.



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.



2

Ejemplo

Europa oriental; 3: Asia/Pacífico; 4: África; 5: Oriente Medio; 6: América latina] y “% anual de crecimiento de la población” obtenida a partir de los datos recogidos para un total de 109 países de todo el mundo (N = 109):

1	Azerbaijón	Oriente Medio	1.4
2	Afganistán	Asia / Pacífico	2.8
3	Alemania	OCDE	.4
4	Arabia Saudí	Oriente Medio	3.2
5	Argentina	América Latina	1.3
6	Armenia	Oriente Medio	1.4
7	Australia	OCDE	1.4
8	Austria	OCDE	.2
9	Bahrein	Oriente Medio	2.4
10	Bangladesh	Asia / Pacífico	2.4
11	Barbados	América Latina	.2
12	Bélgica	OCDE	.2
13	Bielorusia	Europa Oriental	.3
14	Bolivia	América Latina	2.7
15	Bosnia	Europa Oriental	.7
16	Botswana	África	2.7
17	Brasil	América Latina	1.3
18	Bulgaria	Europa Oriental	-.2
19	Burkina Faso	África	2.8

Bibliografía

<https://plataformaeducativauds.com.mx/assets/docs/libro/LAN/f0795cf0998fe289a1811a9520e172fc-LC-LAN303%20ESTADISTICA%20DESCRIPTIVA.pdf>

Pág. 70 a la 99.