



Nombre de alumnos: Fernanda
Patricia Hernández Díaz

Nombre del profesor: Rosario Gómez

Nombre del trabajo: Porcentajes
acumulados y gráficas

Materia: Estadística Descriptiva

Grado: 2^o Cuatrimestre

Grupo: LPS19SSCO9-A

Porcentajes acumulados y gráficas

Porcentajes acumulados

Puede incluir una columna o una fila en el informe que muestre un total acumulado.

se puede expresar como un valor numérico o un porcentaje.

Modo Reporter

Se puede calcular un total acumulado para más de una categoría.

Modo Explorar

Selecciona una categoría sobre la que calcular un total acumulado como porcentaje numérico o acumulativo del valor base.

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.

Puntuación directa

Una puntuación directa no sabemos si se trata de un valor alto o bajo porque esto depende del promedio del grupo.

Puntuaciones diferenciales

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

Puntuaciones típicas

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

Propiedades

- Si media es cero
- Si varianza es igual a 1

Porcentajes acumulados y gráficas

Porcentajes acumulados

Puede incluir una columna o una fila en el informe que muestre un total acumulado.

se puede expresar como un valor numérico o un porcentaje.

Modo Reporter

Se puede calcular un total acumulado para más de una categoría.

Modo Explorer

Selecciona una categoría sobre la que calcular un total acumulado como porcentaje numérico o acumulativo del valor base.

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.

Puntuación directa

Una puntuación directa no sabemos si se trata de un valor alto o bajo porque esto depende del promedio del grupo.

Puntuaciones diferenciales

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

Puntuaciones típicas

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

Propiedades

- Su media es cero
- Su varianza es igual a 1

Porcentajes acumulados y gráficas

Medidas de tendencia central

Son indicadores usados para señalar que porcentaje de datos dentro de una distribución de frecuencias superan estas escalas.

Cuyo valor representa el valor del dato que se encuentra en el centro de la distribución de frecuencia.

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-dec).

Percentiles

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

Representación gráfica

Aunque es aconsejable que la presentación de estos números 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.

Análisis descriptivo

Primer paso consiste en presentar esa información de forma que ésta se pueda visualizar de una manera más sistemática y resumida.

Variables categóricas, como el sexo, estado civil, profesión, etc., se quiere conocer la frecuencia y el porcentaje del total de casos que "caen" en cada categoría.

Diagramas de barras

Son similares a los gráficos de sectores. Se representan tantas barras como categorías tiene la variable, de modo que la altura de cada una de ellas sea proporcional a la frecuencia o porcentaje de casos en cada clase.

Histograma

Para variables numéricas continuas, tales como la edad, la tensión arterial o el índice de masa corporal.

Se divide el rango de valores de la variable en intervalos de igual amplitud, representando sobre cada intervalo un rectángulo que tiene a este segmento como base.

Un diagrama de barras o un histograma son gráficos sencillos que permiten comprobar, de un modo bastante visual, la simetría y el "apuntamiento" de la distribución de una variable y, por lo tanto, verificar su desviación de la normalidad.

Comparación de dos o más grupos

Comparación de variables continuas en dos o más grupos se realiza habitualmente en términos de su valor medio, por medio del test t de Student.

Análisis de la varianza o métodos no paramétricos equivalentes, y así se ha de reflejar en el tipo de gráfico utilizado.

Gráficos de líneas

Los gráficos de líneas pueden resultar también especialmente interesantes, sobre todo cuando interesa estudiar tendencias a lo largo del tiempo

Relación entre dos variables numéricas

método de análisis adecuado es el estudio de la correlación.

Los coeficientes de correlación (Pearson, Spearman) valoran hasta qué punto el valor de una de las variables aumenta o disminuye cuando crece el valor de la otra

Otros gráficos

Diagramas de líneas superpuestos para visualizar los resultados de un análisis de la varianza con dos factores.

Diagrama de dispersión es el método adecuado para valorar el resultado de un modelo de regresión logística.

La mayoría de los paquetes estadísticos (SPSS, STATGRAPHICS, S-PLUS, EGRET, ...) ofrecen grandes posibilidades en este sentido.

Porcentajes acumulados y gráficos

EJERCICIO

Un grupo de investigadores pertenecientes a la secretaría de seguridad pública, tomó una muestra aleatoria de las velocidades (km/h) registradas por 30 vehículos en el trayecto Pichucalco-Villahermosa, con el fin de establecer nuevos límites máximos de velocidad para una carretera. La muestra arrojó los datos siguientes: 90, 99, 104, 99, 119, 98, 95, 112, 95, 120, 100, 90, 116, 96, 114, 108, 98, 118, 100, 106, 114, 100, 112, 106, 100, 115, 111, 105, 114, 97

•**Construye una distribución de frecuencia para velocidades**, que tenga frecuencia absoluta acumulada, frecuencia relativa, frecuencia relativa acumulada y marca de clase así como también un histograma.

Intervalo de clase	f	fa	fr	fra	MC	%	% acumulados
[90-95)	2	2	0.07	0.07	92.5	7	7
[95-100)	8	10	0.27	0.34	97.5	27	34
[100-105)	5	15	0.17	0.51	102.5	17	51
[105-110)	4	19	0.13	0.64	107.5	13	64
[110-115)	6	25	0.2	0.84	112.5	20	84
[115-120]	5	30	0.16	1	117.5	16	100
Total	30		1			100	

Operación

$$K = 1 + 3.3 \log(30) = 1 + 3.3 (1.4771212547) = 1 + 4.87 = 5.87 \approx 6$$

$$Fr = f/N$$

$$Fr = 2/30 = 0.07$$

$$Fr = 8/30 = 0.27$$

$$Fr = 5/30 = 0.17$$

$$Fr = 4/30 = 0.13$$

$$Fr = 6/30 = 0.2$$

$$Fr = 5/30 = 0.16$$

ESTADÍSTICOS DE ASOCIACIÓN ENTRE VARIABLES.

Introducción

Para empezar, para **determinar la relación entre dos variables cuantitativas** en un grupo; debe seguir los objetivos de dicho análisis:

- a. **Determinar si las dos variables están correlacionadas**, es decir si los valores de una variable tienden a ser más altos o más bajos para valores más altos o más bajos de la otra variable.
- b. Poder **predecir el valor de una variable** dado un valor determinado de la otra variable.
- c. **Valorar el nivel de concordancia** entre los valores de las dos variables.

También está el método de correlación, ya que dicho cálculo es el primer paso para determinar la relación entre las variables.

El **coeficiente de correlación** posee las siguientes **características**:

- a. El valor del coeficiente de correlación **es independiente** de cualquier unidad usada para medir las variables.
- b. El valor del coeficiente de correlación **se altera de forma importante** ante la presencia de un valor extremo, como

sucede con la desviación típica. Ante estas situaciones conviene realizar una transformación de datos que cambia la escala de medición y modera el efecto de valores extremos (como la transformación logarítmica).

c. El coeficiente de correlación mide solo la relación con una línea recta. Dos variables pueden tener una relación curvilínea fuerte, a pesar de que su correlación sea pequeña. Por tanto cuando analicemos las relaciones entre dos variables debemos representarlas gráficamente y posteriormente calcular el coeficiente de correlación.

d. El coeficiente de correlación no se debe extrapolar más allá del rango de valores observado de las variables a estudio ya que la relación existente entre X e Y puede cambiar fuera de dicho rango.

e. La correlación no implica causalidad. La causalidad es un juicio de valor que requiere más información que un simple valor cuantitativo de un coeficiente de correlación.

Además, el coeficiente de correlación de Pearson (r) puede calcularse en cualquier grupo de datos, sin embargo la validez del test de hipótesis sobre la correlación entre las variables requiere en sentido estricto lo siguiente:

a) que las dos variables procedan de una muestra aleatoria de individuos.

b) que al menos una de las variables tenga una distribución normal en la población de la cual la muestra procede.

REGRESION LINEAL

Analizamos que el término regresión se utilizó por primera vez en el estudio de variables antropométricas: al comparar la estatura de padres e hijos, donde resultó que los hijos cuyos padres tenían una estatura muy superior al valor medio, tendían a igualarse a éste, mientras que aquellos cuyos padres eran muy bajos tendían a reducir su diferencia respecto a la estatura media; es decir, "regresaban" al promedio.

Y también el término línea se emplea para distinguirlo del resto de técnicas de regresión, que emplean modelos basados en cualquier clase de función matemática.

El modelo lineal relaciona la variable dependiente Y con K variables explícitas ($k = 1, \dots, K$), o cualquier transformación de éstas que generen un hiperplano de parámetros desconocidos: donde es la perturbación aleatoria que recoge todos aquellos factores de la realidad no controlables u observables y que por tanto se asocian con el azar, y es la que

confiere al modelo su carácter estocástico.

El problema de la regresión consiste en elegir unos valores determinados para los parámetros desconocidos, de modo que la ecuación quede completamente especificada. Para ello se necesita un conjunto de observaciones.

Los valores escogidos como estimadores de los parámetros, son los coeficientes de regresión sin que se pueda garantizar que coincidan con parámetros reales del proceso generador.

Los valores son por su parte estimaciones o errores de la perturbación aleatoria.