

Tablas de frecuencia

UdeCataluña

Tablas de frecuencia

Los estadísticos básicos asociados a variables cualitativas son los referentes a los conteos de sus niveles., las frecuencias absolutas y relativas. Y junto con ellos la realización de gráficos como histogramas o gráficos de cajas. Para el análisis de variables cuantitativas por lo general es necesario agrupar los datos, este proceso facilita su análisis, más adelante se hablará de conceptos importantes para su realización.

Las tablas de frecuencia, permiten ver el comportamiento de los datos, y cuantas veces se repite un evento. Estos análisis son muy importantes para el control de la calidad, para ver defectos en procesos o en productos. Y es un paso obligatorio en la creación de histogramas

1. Frecuencias absolutas y relativas

Frecuencia absoluta de un valor de la variable estadística X , es el número de veces que aparece ese valor en el estudio. Se suele denotar por n_i a la frecuencia absoluta del valor $X = x_i$ de la variable X . Dada una muestra de N elementos, la suma de todas las frecuencias absolutas debe dar el total de la muestra estudiada N .

Frecuencia relativa es el cociente entre la frecuencia absoluta y el tamaño de la muestra (N). Es decir:

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

Para variables cualitativas la moda o modas son los niveles con mayor frecuencia.

Frecuencia absoluta acumulada: (N_i) se refiere al total de las frecuencias absolutas para todos los eventos iguales o anteriores a un cierto valor en la tabla.

Frecuencia absoluta acumulada: (F_i) es el cociente entre la frecuencia absoluta acumulada y el tamaño de la muestra N

$$F_i = \frac{N_i}{N} = \frac{N_i}{\sum_i n_i}$$

2. Datos agrupados

Cuando se analiza una gran cantidad de datos y estos tienen un rango amplio debe considerarse agrupar los datos en intervalos equidistantes de la variable en cuestión, esto facilita su manejo y su visualización al realizar los histogramas.

Regla de Sturges: Es una regla practica que debe considerarse para encontrar el número de clases, N es el número de

$$k = 1 + \log_2(R)$$

Clase: Es cada una de los intervalos diseñados para agrupar un conjunto de datos, se caracteriza por un límite inferior y superior, además por una marca de clase que es el promedio del límite inferior y superior.

Amplitud: Es la distancia de los intervalos que definen a cada clase, se calcula mediante la siguiente expresión, donde R : Rango de los datos y k : número de clases.

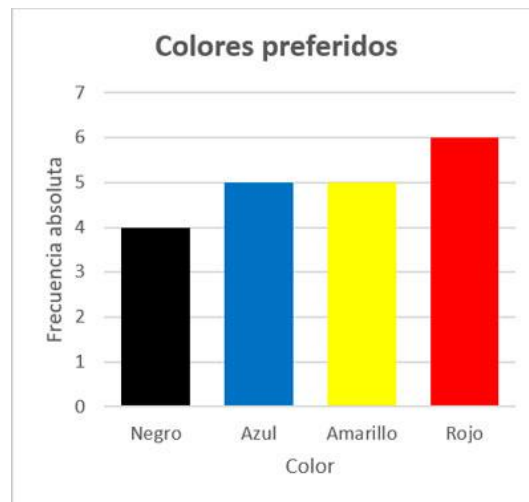
$$A = \frac{R}{k} = \frac{\max - \min}{k}$$

Para datos agrupados es posible hallar los 4 tipos de frecuencia mencionadas en el inciso anterior.

3. Gráficos útiles para el análisis

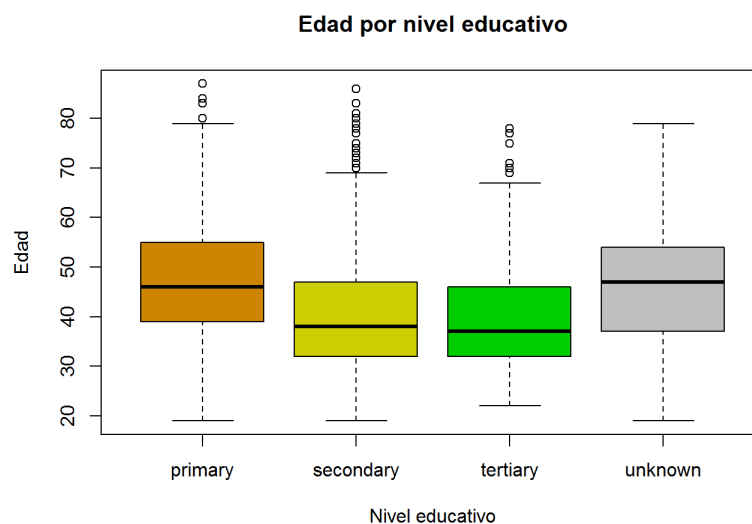
Diagramas de barras

Los diagramas de barras permiten una rápida visualización de los niveles de determinada cantidad de datos. Dependiendo de la distribución que se les otorgue, se podrán o no analizar pronta y claramente los datos. Se genera en R studio empleando la función `barplot()`.



Diagramas de cajas

El diagrama de caja es una representación visual que describe al mismo tiempo varias características importantes de datos, tales como el centro, la dispersión, la desviación de simetría y la identificación de observaciones que se alejan de manera poco usual del resto de los datos. Se genera en R studio empleando la función `boxplot()`.



Referencias

1. <https://www.rdocumentation.org/>

UdeCataluña

© U de Cataluña, 2020

Todos los derechos reservados. Prohibida la reproducción total o parcial sin permiso o autorización de la Universidad, Bogotá - Colombia.