

16. Estadística descriptiva

16.1. Los distintos tipos de media

Dado un conjunto de datos, en forma de vector, pretendemos dar **un número que resuma dicha información**. Con **Mean[v]** tenemos la **media aritmética**. El comando **Total[v]** nos permite sumar todos los elementos y **Lenght[v]** nos dice cuantos tiene.

Además de la media aritmética, existen muchas otras: La **media armónica (HarmonicMean[v])**, la **media geométrica (GeometricMean[v])**, la **media contraarmónica (ContraharmonicMean[v])** y la **media cuadrática (RootMeanSquare[v])**. Sus formulas son:

$$h) n : \left(\sum \frac{1}{x_i} \right) \quad g) \left(\prod x_i \right)^{1/n} \quad c) \frac{\sum x_i^2}{\sum x_i} \quad r) \sqrt{\frac{1}{n} \sum x_i^2}$$

Ejercicios

Calcula distintas medias con los datos

21,22,21,23,22,21,23,23,21,21,21,24,22

Nuevas funciones

Mean, Total, Lenght, HarmonicMean, GeometricMean, ContraharmonicMean, RootMeanSquare.

16.2. Mediana y moda

La **mediana** de un conjunto de datos es el dato que se encuentra en el punto medio, una vez ordenados los datos. Si el número de datos es par, entonces es la media de los dos datos centrales. Con **Sort[v]** podemos ordenar los datos del vector y extraer los valores con el doble corchete. También podemos emplear **Median[v]** que nos proporciona directamente la mediana.

El dato que más se repite (el más común) es la **moda**. En Mathematica tenemos el comando **Commonest[v]** para calcular la moda.

Ejercicios

Calcular la mediana de los datos anteriores. Ordenarlos y ver que coincide con el valor situado en el punto medio. Calcular la moda.

Nuevas funciones

Sort, Median, Commonest.

16.3. Máximo, mínimo y cuantiles

El **máximo** de los datos se obtiene con **Max[v]** y el **mínimo** con **Min[v]**. Restando ambos podemos obtener el **rango**. Los tres **cuantiles** con **Quartiles[v]**. El **rango intercuartílico** (la resta del tercer y primer cuartil) con **InterquartileRange[v]**. Finalmente los **percentiles** se calculan con **Quantile[v,p]**, donde además de los datos, debemos facilitar el percentil, como un número entre 0 y 1.

Ejercicios

Calcula el máximo, el mínimo, el rango, los cuantiles, el rango intercuartílico y algún percentil de los datos anteriores.

Nuevas funciones

Max, Min, Quartiles, InterquartileRange, Quantile.

16.4. Varianza y desviación típica

La **varianza** (**Variance[v]**) de un conjunto de datos es:

$$\frac{\sum (x_i - \bar{x})^2}{n - 1}$$

La varianza que calcula Mathematica es la que esta dividida en el número de datos **menos 1**, que muchas veces se encuentra en la literatura con el nombre de cuasivarianza.

La función (**StandardDeviation[v]**) nos da **desviación estándar** y se calcula aplicando la raíz cuadrada a la varianza.

La **desviación media** (**MeanDeviation[v]**) se calcula:

$$\frac{1}{n} \sum |x_i - \bar{x}|$$

Ejercicios

Calcula los parámetros al conjunto de datos.

Nuevas funciones

Variance, StandardDeviation, MeanDeviation.

16.5. Asimetría, kurtosis y momentos

La asimetría y la kurtosis son una medida de la posible asimetría de los datos y del posible apuntamiento de la distribución. Los comandos son **Skewness[v]** y **Kurtosis[v]**.

Para calcular el momento de orden r tenemos el comando **Moment[v,r]** y para el momento central **CentralMoment[v,r]**.

Ejercicios

Calcula la asimetría, la kurtosis y algún momento de los datos.

Nuevas funciones

Skewness, Kurtosis, Moment, CentralMoment.

16.6. Covarianza y correlación

Dados dos vectores de datos u y v de la misma dimensión, el comando **Correlation[u,v]** calcula la correlación entre ambas variables y **Covariance[u,v]** calcula la covarianza.

Ejercicios

Construir dos vectores aleatorios de la misma dimensión y calcular la correlación y la covarianza.

Nuevas funciones

Correlation, Covariance.

17. Gráficos estadísticos

17.1. Tablas de frecuencias discretas

Si tenemos una distribución discreta, la función **Tally[v]** nos permite hacer una tabla de frecuencias. El resultado es una matriz, cuya primera columna son los datos y la segunda columna son las frecuencias absolutas. El comando **Sort[v]** nos ordena la tabla. Con los dobles corchetes podemos extraer la segunda columna. Para dibujar el diagrama de barras utilizamos **BarChart[v]** y para el diagrama de sectores **PieChart[v]** y sus versiones tridimensionales.

Ejercicios

Genera un vector de 100 números aleatorios entre el 1 y el 6. Haz la tabla de frecuencias, ordénala y representa el diagrama de barras y el de sectores asociado.

Nuevas funciones

Tally, BarChart, PieChart, BarChart3D, PieChart3D.