



Estadística I

Mario
Francisco

Conceptos
generales

Distribuciones
de frecuencias

Representaciones
gráficas

Medidas
características

Part I

Descripción estadística de una variable



Conceptos generales

Estadística I

Mario
Francisco

Conceptos
generales

Distribuciones
de frecuencias

Representaciones
gráficas

Medidas
características

- El objeto de cualquier investigación estadística es la toma de información acerca de los individuos de cierto colectivo llamado **población estadística**.
- Cada elemento de una población se denomina **individuo** o **unidad estadística**.
- Las poblaciones estadísticas se clasifican en poblaciones **finitas** e **infinitas**, de acuerdo con el número de individuos incluidos en las mismas.
- El proceso de toma de información acerca de los individuos de una población puede realizarse mediante la elaboración de un **censo** o mediante la extracción de una **muestra**.



Conceptos generales

Estadística I

Mario
Francisco

Conceptos generales

Distribuciones
de frecuencias

Representaciones
gráficas

Medidas
características

- Si una muestra es representativa de una población, se pueden deducir importantes conclusiones acerca de ésta a partir del análisis de la información muestral. La parte de la Estadística que trata de las condiciones bajo las cuales tales inferencias son válidas se llama **Estadística Inductiva** o **Inferencial**.
- La parte de la Estadística que trata solamente de describir y analizar un grupo dado de datos sin sacar conclusiones o inferencias acerca de la población que los ha generado se llama **Estadística Descriptiva** o **Deductiva**.



Conceptos generales

Estadística I

Mario
Francisco

Conceptos
generales

Distribuciones
de frecuencias

Representaciones
gráficas

Medidas
características

Variables estadísticas

- Llamaremos **variable estadística** a cada una de las características consideradas con el propósito de describir a cada individuo de la muestra.
- Cada una de las variables estadísticas consideradas para la descripción de los individuos de la muestra puede presentar distintas **modalidades** o estados.
- Atendiendo a la naturaleza de las modalidades de las variables, éstas pueden clasificarse en **variables cualitativas** y **variables cuantitativas**. Estas últimas se clasifican en **variables discretas**, que son aquellas que toman un número finito o infinito numerable de valores distintos, y **variables continuas**, que son aquellas que pueden tomar cualquier valor en un intervalo de valores dado.



Distribuciones de frecuencias

Estadística I

Mario
Francisco

Conceptos
generales

Distribuciones
de frecuencias

Representaciones
gráficas

Medidas
características

- Sean x_1, \dots, x_n una muestra de una variable X que presenta las modalidades c_1, \dots, c_k , y sea n_i el número de individuos de la muestra con la modalidad c_i , $i = 1, \dots, k$.
- El número n_i se conoce como **frecuencia absoluta** de la modalidad c_i .

- La **frecuencia relativa** de c_i es $f_i = n_i/n$ ($i = 1, 2, \dots, k$).
- La **frecuencia absoluta acumulada** se define como

$$N_i = n_1 + n_2 + \dots + n_i = \sum_{j=1}^i n_j \quad (i = 1, 2, \dots, k)$$

- La **frecuencia relativa acumulada** se define como

$$F_i = f_1 + f_2 + \dots + f_i = \sum_{j=1}^i f_j = \frac{N_i}{n} \quad (i = 1, 2, \dots, k).$$



Distribuciones de frecuencias

Estadística I

Mario
Francisco

Conceptos
generales

Distribuciones
de frecuencias

Representaciones
gráficas

Medidas
características

Propiedades

1 $0 \leq n_i \leq n; \quad 0 \leq f_i \leq 1 \quad (i = 1, 2, \dots, k).$

2 $N_k = n_1 + n_2 + \dots + n_k = n; \quad F_k = f_1 + f_2 + \dots + f_k = 1.$



Distribuciones de frecuencias

Estadística I

Mario
Francisco

Conceptos
generales

Distribuciones
de frecuencias

Representaciones
gráficas

Medidas
características

Modalidad	Frecuencia absoluta	Frecuencia relativa	Fr. absoluta acumulada	Fr. relativa acumulada
c_1	n_1	f_1	N_1	F_1
c_2	n_2	f_2	N_2	F_2
\vdots	\vdots	\vdots	\vdots	\vdots
c_i	n_i	f_i	N_i	F_i
\vdots	\vdots	\vdots	\vdots	\vdots
c_k	n_k	f_k	N_k	F_k
Total	n	1		



Distribuciones de frecuencias

Estadística I

Mario
Francisco

Conceptos
generales

Distribuciones
de frecuencias

Representaciones
gráficas

Medidas
características

Construcción de las modalidades

- Si la variable es cualitativa, se tomarán como modalidades las distintas respuestas observadas en la muestra.
- Si la variable es discreta (que tome pocos valores distintos), las modalidades coincidirán con los distintos valores medidos en la muestra
- Si la variable es continua (o bien discreta, pero que toma muchos valores distintos), se tomarán como modalidades los intervalos de clase, tomándose como frecuencia absoluta de cada modalidad el número de observaciones agrupadas en el intervalo correspondiente.



Distribuciones de frecuencias

Estadística I

Mario
Francisco

Conceptos
generales

Distribuciones
de frecuencias

Representaciones
gráficas

Medidas
características

Intervalos de clase

- Si $e_0 < e_1 < \dots < e_{i-1} < e_i < \dots < e_k$ son los extremos de los k intervalos de clase contruidos, la frecuencia absoluta del i -ésimo intervalo es $n_i =$ **número de observaciones x_i tales que $e_{i-1} < x_i \leq e_i$** , siendo su **amplitud $a_i = e_i - e_{i-1}$** , para $i = 1, 2, \dots, k$.
- Una vez contruidos los intervalos de clase se elige un representante en cada uno de ellos, llamado **marca de clase**, que usualmente es tomado como el punto medio del intervalo

$$c_i = \frac{e_{i-1} + e_i}{2} \quad (i = 1, 2, \dots, k).$$



Distribuciones de frecuencias

Estadística I

Mario
Francisco

Conceptos
generales

Distribuciones
de frecuencias

Representaciones
gráficas

Medidas
características

Intervalos de clase

- *¿Cuántos intervalos construir?* Se recomienda construir tantos intervalos como el número entero (entre 5 y 20) más próximo a \sqrt{n} .
- *¿Intervalos de la misma amplitud o de amplitudes distintas?* En general construiremos intervalos de la misma amplitud.
- *¿Qué valor elegir como extremo inferior del primer intervalo (e_0)?* Se toma como e_0 un valor “un poco menor” que el mínimo de la muestra.



Representación gráfica de variables cualitativas

- **Diagrama de barras.** En un sistema de ejes de coordenadas se representa en el eje de abscisas las modalidades de la variable y en el eje de ordenadas las frecuencias (n_i o f_i). A continuación, sobre cada modalidad se levanta un rectángulo o barra de altura igual a la frecuencia (absoluta o relativa) representada, siendo todos los rectángulos de igual base.
- **Diagrama de sectores.** Se trata de repartir un círculo de radio arbitrario en sectores proporcionales a la frecuencia de cada modalidad.



Representaciones gráficas

Estadística I

Mario
Francisco

Conceptos
generales

Distribuciones
de frecuencias

Representaciones
gráficas

Medidas
características

Representación gráfica de variables cuantitativas discretas

- **Diagrama de barras.** Se construye de la forma anteriormente descrita para variables cualitativas, pudiendo sustituirse las barras por segmentos de recta.
- **Diagrama acumulativo de frecuencias.** Su construcción se realiza representando en un sistema de ejes de coordenadas los puntos (c_i, N_i) o (c_i, F_i) , según se representen frecuencias absolutas o relativas acumuladas, uniéndose a continuación de forma escalonada los puntos representados.



Representaciones gráficas

Estadística I

Mario
Francisco

Conceptos
generales

Distribuciones
de frecuencias

Representaciones
gráficas

Medidas
características

Representación gráfica de variables cuantitativas continuas

- **Histograma.** Es el equivalente continuo del diagrama de barras. Partiendo de un sistema de ejes de coordenadas, se representan en el eje de abscisas los extremos e_i de los intervalos de clase y a continuación, tomando como base de cada rectángulo la amplitud a_i del intervalo de clase correspondiente, se levanta sobre cada uno de los intervalos un rectángulo de altura $h_i = n_i/a_i$ o $h_i = f_i/a_i$, según se representen frecuencias absolutas o relativas.
- Uniendo, mediante segmentos de recta, los puntos medios de las bases superiores de cada rectángulo del histograma se obtiene la representación gráfica llamada **polígono de frecuencias**.



Representaciones gráficas

Estadística I

Mario
Francisco

Conceptos
generales

Distribuciones
de frecuencias

Representaciones
gráficas

Medidas
características

Representación gráfica de variables cuantitativas continuas

- **histograma móvil**. Para cada punto x se define

$$h(x) = \frac{\text{n}^\circ \text{ de observaciones } x_i \text{ en } [x - a/2, x + a/2]}{a \cdot n}$$

se representan gráficamente los pares $(x, h(x))$ para un número grande de valores x seleccionados, el gráfico resultante corresponde a una curva más “suave”.

- **Polígono acumulativo de frecuencias**. Es el resultado que se obtiene al unir mediante segmentos de recta los puntos (e_i, N_i) o (e_i, F_i) , según se trate de frecuencias absolutas o relativas acumuladas, representados en un sistema de ejes de coordenadas.



Representaciones gráficas

Estadística I

Mario
Francisco

Conceptos
generales

Distribuciones
de frecuencias

Representaciones
gráficas

Medidas
características

Diagrama de tallo y hojas

- Se utiliza a menudo para la descripción de variables cuantitativas (discretas o continuas), presenta la particularidad de permitir visualizar globalmente la distribución de frecuencias manteniendo la individualidad de los datos.
- Se redondean los datos a dos o tres cifras significativas, tomándose como **tallos** la primera o las dos primeras cifras de cada dato y como **hojas** las últimas cifras de cada dato.
- A continuación, separados por una línea vertical, se dispondrán los tallos a la izquierda y las hojas a la derecha del tallo correspondiente. De esta manera cada tallo, que se representa una sola vez, define una clase y el número de hojas representa la frecuencia de dicha clase.



Medidas características. Medidas de posición

Estadística I

Mario Francisco

Conceptos generales

Distribuciones de frecuencias

Representaciones gráficas

Medidas características

Medidas de posición central

- **Media aritmética.** Se define la **media aritmética** (\bar{x}) de una variable cuantitativa X como el valor numérico

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

- **Propiedades:**

- 1 $\min(x_i) \leq \bar{x} \leq \max(x_i)$.
- 2 Si $y_i = a + bx_i$ ($i = 1, 2, \dots, n$), entonces $\bar{y} = a + b\bar{x}$.
- 3 $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}) = 0$
- 4 $\bar{x} = \arg \min \sum_{i=1}^n (x_i - a)^2$



Medidas características. Medidas de posición

Estadística I

Mario
Francisco

Conceptos
generales

Distribuciones
de frecuencias

Representaciones
gráficas

Medidas
características

Medidas de posición central

- **Media truncada.** Un inconveniente que presenta la media aritmética es su sensibilidad a la presencia de observaciones atípicas (**outliers**) en la muestra. Para evitar este tipo de problema se utilizan las **medias truncadas**, que consisten en medias aritméticas calculadas con un porcentaje central de los datos.
- **Media recortada.** Consiste en la media aritmética de una modificación de los datos originales: un porcentaje central permanece sin modificar, cada uno de los datos de los menores excluidos se sustituye por el menor de los datos del porcentaje central no modificado y cada uno de los datos de los mayores excluidos se reemplaza por el mayor de los datos del porcentaje central no modificado.



Medidas características. Medidas de posición

Estadística I

Mario
Francisco

Conceptos
generales

Distribuciones
de frecuencias

Representaciones
gráficas

Medidas
características

Medidas de posición central

- **Media cuadrática (Q).**

$$Q = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}} = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}$$

- **Media geométrica (G).** Supuesto que $x_i > 0$, $1 \leq i \leq n$, se define como $G = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} = \prod_{i=1}^n x_i^{1/n}$.
- **Media armónica (H).** Si $x_i > 0$, $1 \leq i \leq n$, es

$$H = \frac{n}{1/x_1 + \dots + 1/x_n} = \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i} \right)^{-1}$$



Medidas características. Medidas de posición

Estadística I

Mario
Francisco

Conceptos
generales

Distribuciones
de frecuencias

Representaciones
gráficas

Medidas
características

Medidas de posición central

- **Mediana (M_e)**. Es la medida de posición central que, supuestos los datos ordenados de menor a mayor, deja igual número de valores a su izquierda que a su derecha.
- Si el número de datos es par, se toma como mediana la media aritmética de los dos valores centrales. Si el número de datos es impar, se toma como mediana el valor central.
- En el supuesto de que los datos hayan sido agrupados en intervalos de clase, se calcula el **intervalo mediano** (el intervalo de clase que contiene a la mediana), eligiéndose como mediana un representante de dicho intervalo (por ejemplo, la marca de clase).
- La mediana es una medida robusta, esto es, poco sensible a la presencia de observaciones atípicas en la muestra.



Medidas características. Medidas de posición

Estadística I

Mario
Francisco

Conceptos
generales

Distribuciones
de frecuencias

Representaciones
gráficas

Medidas
características

Medidas de posición central

- **Moda (M_0)**. Si X es una variable discreta, se define como el valor más frecuente de la variable.
- Si X es una variable continua, no tiene sentido hablar del valor más frecuente, por lo que se introduce el concepto de **intervalo modal**, que se define como aquel intervalo de clase al que le corresponde mayor altura en el histograma de frecuencias. Finalmente, se toma como **moda** un representante del intervalo modal (por ejemplo, la marca de clase).
- Debe observarse que, de acuerdo con la definición, la moda puede no ser única, en cuyo caso tendremos **distribuciones multimodales** (con 2 o más modas).



Medidas características. Medidas de posición

Estadística I

Mario
Francisco

Conceptos
generales

Distribuciones
de frecuencias

Representaciones
gráficas

Medidas
características

Otras medidas de posición

- **Cuantiles.** Supuestas ordenadas las observaciones de menor a mayor, se define el **cuantil de orden p** ($0 < p < 1$) como el valor q_p que deja a lo sumo np observaciones a su izquierda y a lo sumo $n(1 - p)$ observaciones a su derecha.
- **Cuartiles.** Son los cuantiles de órdenes $1/4$ (cuartil de primer orden o primer cuartil Q_1), $1/2$ (cuartil de segundo orden o **mediana**) y $3/4$ (cuartil de tercer orden o tercer cuartil Q_3).
- **Deciles.** Son los cuantiles de órdenes $r/10$ ($r = 1, \dots, 9$), y dividen el conjunto de observaciones en diez partes de igual frecuencia.
- **Percentiles.** Son los cuantiles de órdenes $r/100$ ($r = 1, \dots, 99$).



Medidas características

Estadística I

Mario
Francisco

Conceptos
generales

Distribuciones
de frecuencias

Representaciones
gráficas

Medidas
características

Diagramas de caja

Para la construcción de esta representación, se calculan previamente la mediana, los cuartiles Q_1 y Q_3 y los valores extremos LI y LS , donde

$$LI = \min \{x_i / x_i \geq Q_1 - 1'5(Q_3 - Q_1)\}$$

$$LS = \max \{x_i / x_i \leq Q_3 + 1'5(Q_3 - Q_1)\}$$

Las observaciones que caen fuera del intervalo (LI, LS) se consideran datos atípicos (*outliers*).



Medidas características

Estadística I

Mario
Francisco

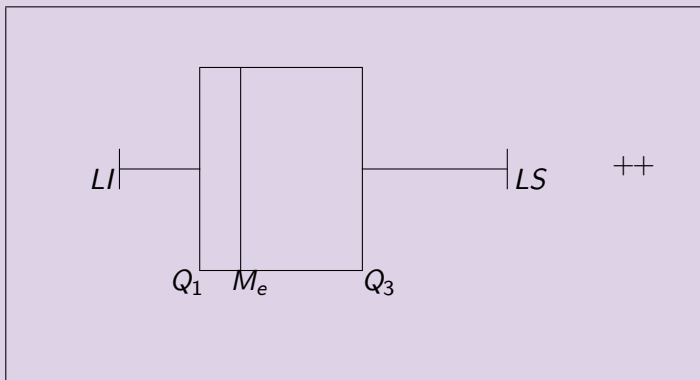
Conceptos
generales

Distribuciones
de frecuencias

Representaciones
gráficas

Medidas
características

Diagramas de caja. Las observaciones representadas con + son outliers.





Medidas características. Medidas de dispersión

Estadística I

Mario
Francisco

Conceptos
generales

Distribuciones
de frecuencias

Representaciones
gráficas

Medidas
características

Medidas de dispersión absoluta

- **Varianza.** $s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$.
- **Desviación típica.** $s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$
- **Propiedades:**
 - 1 La varianza y la desviación típica toman siempre valores no negativos.
 - 2 Si $y_i = a + bx_i$ ($i = 1, 2, \dots, n$), se tiene que $s_y^2 = b^2 s_x^2$ y $s_y = |b|s_x$.
 - 3 $s^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 = \overline{x^2} - \bar{x}^2$.



Medidas características

Estadística I

Mario
Francisco

Conceptos
generales

Distribuciones
de frecuencias

Representaciones
gráficas

Medidas
características

Desigualdad de Tchebychev

En el intervalo de centro la media aritmética y radio k veces la desviación típica están comprendidas como mínimo el $100(1 - 1/k^2)\%$ de las observaciones.

Variables tipificadas

Se define la **variable tipificada** de una variable estadística X como la variable (Z) que resulta de restarle su media aritmética y dividir por su desviación típica:

$$Z = \frac{X - \bar{x}}{s}$$

obteniéndose de esta manera una variable adimensional de media cero y desviación típica unidad ($\bar{z} = 0$, $s_z = 1$).



Medidas de dispersión absoluta

- **Recorrido o rango.** Es la diferencia entre los valores extremos:

$$R = \max(x_i) - \min(x_i)$$

- **Recorrido intercuartílico.** Es la diferencia entre los cuartiles de tercer y primer órdenes:

$$RI = Q_3 - Q_1$$



Medidas características. Medidas de dispersión

Estadística I

Mario
Francisco

Conceptos
generales

Distribuciones
de frecuencias

Representaciones
gráficas

Medidas
características

Medidas de dispersión relativas

- **Recorrido relativo.** Supuesto que $\bar{x} > 0$, se define como el cociente entre el recorrido y la media

$$RR = \frac{R}{\bar{x}}$$

- **Coefficiente de variación.** Es la medida de dispersión relativa más popular y, supuesto que $\bar{x} > 0$, se define como el cociente entre la desviación típica y la media

$$CV = \frac{s}{\bar{x}}$$



Medidas características

Estadística I

Mario
Francisco

Conceptos
generales

Distribuciones
de frecuencias

Representaciones
gráficas

Medidas
características

Momentos

- **Momento con respecto al origen** de orden r ($r \geq 0$) es el número

$$a_r = \frac{1}{n} \sum_{i=1}^n x_i^r$$

- **Momento central** o **con respecto a la media** de orden r ($r \geq 0$) es el número

$$m_r = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^r$$

Así, por ejemplo, $a_1 = \bar{x}$ y $m_2 = s^2$.

- Se tiene que $m_r = \sum_{k=0}^r (-1)^k \binom{r}{k} a_{r-k} a_1^k$.



Medidas características. Medidas de forma

Estadística I

Mario
Francisco

Conceptos
generales

Distribuciones
de frecuencias

Representaciones
gráficas

Medidas
características

Medidas de asimetría

- **Coefficiente de asimetría de Pearson.** Para distribuciones unimodales, mide si las observaciones están dispuestas simétricamente con respecto a la moda. $v = \frac{\bar{x} - M_0}{s}$
- **Coefficiente de asimetría de Fisher.** Mide si las observaciones están dispuestas simétricamente con respecto a la media y es obtenible aunque la distribución no sea unimodal. $g_1 = \frac{m_3}{s^3}$
- Si una distribución es **simétrica o insesgada** $v \simeq 0$ ($\bar{x} \simeq M_0$) y $g_1 \simeq 0$. Si $v > 0$ ($\bar{x} > M_0$) o $g_1 > 0$, diremos que la distribución es **sesgada a la derecha**; diremos que la distribución es **sesgada a la izquierda** si $v < 0$ ($\bar{x} < M_0$) o $g_1 < 0$.



Medidas características. Medidas de apuntamiento o curtosis

Estadística I

Mario Francisco

Conceptos generales

Distribuciones de frecuencias

Representaciones gráficas

Medidas características

Coefficiente de apuntamiento de Fisher

- Mide el grado de apuntamiento de una distribución con respecto al modelo *normal* de referencia. Es una medida adimensional definida según la expresión:

$$g_2 = \frac{m_4}{s^4} - 3$$

- Diremos que una distribución es **platicúrtica** (más aplastada que el modelo normal) si $g_2 < 0$, **mesocúrtica** si $g_2 \simeq 0$ y **leptocúrtica** (más apuntada que el modelo normal) si $g_2 > 0$.