Estadistica Computacional

🅰nalisis de datos

Braulio Fuentes - Diego Quezada

Temario

- Medidas de tendencia central.
- Medidas de dispersion.
- Medidas de localizacion.
- Medidas de forma.
- Medidas de asociacion de variables.
- Independencia estadistica.

Medidas de tendencia central

Media aritmetica

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \overline{x} = \frac{1}{n}\sum_{i=1}^{n} f_i X_i$$

Propiedades interesantes:

1. $mean(\{kx_i\})$ = $k \cdot mean(\{x_i\})$
2. $mean(\{x_i + c\})$ = $mean(\{x_i\}) + c$

Mediana

- Para **datos no agrupados ordinales o cuantitativos** se deben ordenar los datos de menor a mayor y tomar el del medio. Si n es par entonces se toma el promedio de los dos centrales.
- Para **datos agrupados ordinales o cuantitativos** la clase mediana es aquella que acomula al menos el 50% de los datos.
- Para **datos cuantitativos** se puede obtener mediante una linda formula. 👽

Moda

- Para **datos no agrupados** es el dato con mayor frecuencia.
- Para **datos agrupados** la clase modal es aquella con mayor frecuncia absoluta.
- Para **datos cuantitativos** se puede obtener la mediana mediante una linda formula. 👽

> Unica medida para datos nominales. 😢

Medidas de dispersion

- Indican la precision de los datos.

Varianza

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \qquad S_n^2 = \sum_{i=1}^{k} f_i(X_i - \overline{x})^2$$

> Una posible interpretacion es el error cuadratico medio.

Propiedades interesantes:

1. $var(\{kx_i\})$ = $k^2 \cdot var(\{x_i\})$
2. $var(\{x_i + c\})$ = $var(\{x_i\})$

Desviacion estandar

$$S_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2} \qquad S_n = \sqrt{\sum_{i=1}^{k} f_i(X_i - \overline{x})^2}$$

Propiedades analogas se pueden inferir tomando la raiz de las propiedades de la varianza.

Desviacion media

$$MD = \frac{1}{n}\sum_{i=1}^{n}|x_i - \overline{x}| \qquad MD = \sum_{i=1}^{k} f_i|X_i - \overline{x}|$$

Coeficiente de variacion

- Indica el nivel de homogeneidad de los datos ❤️
- Permite **comparar dipersiones** de dos distribuciones distintas.

$$CV = \frac{S}{\overline{x}}$$

Medidas de localizacion

- Dividen la muestra en partes iguales.
- Permiten clasificar a un dato dentro de una determinada categoria.
- Las principales son los cuartiles y los percentiles.
- Permiten generar (o motivan) otras medidas de dispersion como el IQR (la cajita del boxplot ❤️).

Medidas de forma

Medidas de asimetria

Indican el grado de simetria de la distribucion de probabilidad de los datos.

Coeficiente de Asimetria de Bowley-Yule

$$IS = \frac{Q_1 + Q_3 - 2Q_2}{IQR}$$

Coeficiente de asimetria de Fisher 👀

$$\gamma_1 = \frac{\overline{m_3}}{S^3}$$

Achatamiento

Curtosis o coeficiente de apuntamiento

Complementa la informacion que brinda la varianza

$$\gamma_2 = \frac{\overline{m_4}}{S^4} - 3$$

> De donde viene el 3 ❓

Medidas de asociacion de variables

- Indican si dos variables aleatorias tienen alguna relacion y en que grado.

Covarianza

Para **datos no agrupados**:

$$Cov(x,y) = \frac{1}{n}\sum_{i}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

Para **datos agrupados**:

$$Cov(x,y) = \sum_{i}^{k} f_{ij}(X_i - \overline{x})(Y_i - \overline{y})$$

Indica si existe una relacion directamente proporcional o indirectamente proporcional.

Correlacion de Pearson

- Indica el **grado de relacion lineal** entre x e y.
- Si r_{xy} es cercano a 0, hay una relacion lineal debil o no existe tal relacion.
- Si r_{xy} es cercano a 1, hay una relacion lineal fuerte directamente prorpocional.
- Si r_{xy} es cercano a -1, hay una relacion lineal fuerte indirectamente prorpocional.

$$r_{xy} = \frac{Cov(x, y)}{S_x S_y}$$

Independencia estadistica

- Independencia implica covarianza (y por lo tanto correlacion) igual a 0.
- Correlacion **no** implica dependencia.
- x e y son independientes ssi:

$$f_r(x_i, y_j) = f_r(x_i) \cdot f_r(y_j), \; \forall i, j$$

> Basta que un par (i, j) no cumpla la igualdad para garantizar que **no hay** independencia.

- La ecuacion anterior (regla de oro) se puede dejar en funcion de la frecuencia absoluta.

$$\begin{pmatrix} & B_1 & B_2 & \cdots & B_s & TOTAL \\ A_1 & f_{11} & f_{12} & \cdots & f_{1s} & f_{1\cdot} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ A_r & f_{r1} & f_{r2} & \cdots & f_{rs} & f_{r\cdot} \\ TOTAL & f_{\cdot 1} & f_{\cdot 2} & \cdots & f_{\cdot s} & f_{\cdot\cdot} \end{pmatrix}$$