

# Apuntes Stats

Oskar Denis Siodmok

9 de enero de 2021



# 1 Introducción

## 1.1. Conceptos Básicos

- Individuos o elementos: Contienen la información a estudiar.
- Población: Conjunto de individuos o elementos que presentan la variable a estudiar.
- Muestra: Subconjunto representativo de una población.
- Variables: Propiedades de los elementos de la población que tendrán sus respectivos valores
- Clases: Conjunto de valores que cumplen una propiedad. Por definición un valor solo puede pertenecer a una clase (por ejemplo, intervalos en variables continuas).
- Parámetro: Será una función que operará con las diferentes variables y valores sobre la población con un propósito.
- Estadístico: Será una aproximación al parámetro a partir de una muestra.

### 1.1.1. Variables Estadísticas

Las variables estadísticas, ya definidas en el anterior apartado, se denotarán mediante una letra mayúscula (por lo general con  $X$  o  $Y$ ). Podrán tomar cualquier valor de cualquier conjunto. El dominio de la variable será el conjunto de todos los posibles valores de dicha variable.

### 1.1.2. Tipos de variables

- Variables Cuantitativas: Se expresan en cantidades numéricas o cualquier otro sistema que se pueda ordenar. A su vez se clasifican en:
  - Variables Discretas: Toman valores concretos de conjuntos finitos o infinitos.
  - Variables Continuas: Toman valores de conjuntos infinitos y no concretos (el dominio son valores continuos, como todos los reales en un intervalo).

Para muchas variables resulta complicado distinguir el tipo. Por ejemplo, aunque la altura matemáticamente sea continua, en la vida real nadie va a determinar una altura por encima de 2 decimales.

- Variables Cualitativas: No se pueden medir, solo clasificar. Un tipo concreto de este tipo de variables son las *Variables Ordinales*, las cuales pese a no tener un valor numérico si que pueden tener relaciones de orden.

### 1.1.3. Representación de datos

Esta se puede realizar de varias formas:

- Tablas y Gráficos: representan información de forma rápida y visual.
- Medidas Descriptivas: describen la información de forma numérica.

## 1.2. Tablas de Frecuencias

Se pueden realizar sobre cualquier conjunto de datos y sobre cualquier variable. Por ejemplo, dada una población de  $n$  individuos que presenta la variable  $X$  se obtienen las clases  $c = \{c_1, c_2, \dots, c_k\}$  posibles. En este caso,  $n_i$  hará referencia al número de observaciones para  $i \in \{1, \dots, k\} \subseteq \mathbb{N}$ . De esta forma,  $n = \sum_i n_i$  será el número total de observaciones de nuestra variable, independientemente de la clase de cada observación. Por otro lado, la frecuencia relativa de una variable representará la frecuencia de una variable ( $n_i$ ) sobre 1 y se calculará como  $\frac{n_i}{n}$ .

Clase $c_i$	$c_1$	$c_2$	$\dots$	$c_k$	
Freq. Absoluta $n_i$	$n_1$	$n_2$	$\dots$	$n_k$	$n = \sum_i n_i$
Freq. Relativa $f_i$	$f_1$	$f_2$	$\dots$	$f_k$	$f = \sum_i f_i$

Se podrían añadir las frecuencias acumuladas, las cuales se van sumando a los datos anteriores. Siendo  $N_i$  la frecuencia absoluta acumulada y  $F_i$  la frecuencia relativa acumulada:

Clase $c_i$	$c_1$	$c_2$	$\dots$	$c_k$	
Freq. Absoluta $n_i$	$n_1$	$n_2$	$\dots$	$n_k$	$n = \sum_i n_i$
Freq. Relativa $f_i$	$f_1$	$f_2$	$\dots$	$f_k$	$f = 1$
F. abs. acum. $N_i = \sum_{j=1}^i n_j$	$N_1$	$N_2$	$\dots$	$N_k$	$N = N_k = n$
F. rel. acum $F_i = \sum_{j=1}^i f_j$	$F_1$	$F_2$	$\dots$	$F_k$	$F = F_k = 1$

Si la variable es cualitativa entonces las clases serán nominales. Si la variable es discretas las clases serán valores numéricos dentro del rango y si es continua serán intervalos  $(l_{i-1}, l_i] \forall i$ . Si las clases son intervalos habrá varios parámetros y conceptos de interés:

- Amplitud del intervalo:  $a_i = l_i - l_{i-1}$ .
- Marca de clase: Será el valor representativo del intervalo. Por ejemplo, el punto medio:  $c_i = \frac{l_i + l_{i-1}}{2}$ . Podrá representarse en la segunda columna o fila de la tabla.
- Número de intervalos: Se realizará mediante aproximación pues se pueden plantear los intervalos que se quiera. Una elección típica es:

$$k = \begin{cases} \sqrt{n} \\ 1 + \log_2(n) \end{cases}, n \text{ no es muy grande}$$

- Normalización: Cuando el intervalo tiene muchos decimales conviene redondear hacia arriba para que los datos sean más legibles. Notar que si se redondea a la baja se altera el número de intervalos, por lo cual no conviene.
- Límites de los intervalos: Por definición de clase, un valor de una variable solo puede pertenecer a una clase. Debido a ello hay que tener en cuenta los límites de los intervalos prestando atención a que ningún par de clases contenga valores repetidos.

## 1.3. Gráficos

### 1.3.1. Diagrama de Barras

En el eje  $X$  se representan las clases y en el  $Y$  las frecuencias absolutas o relativas (ya que son proporcionales la escala se mantendría). Busca en google como son que no tengo ningún conjunto de datos interesante. Si el diagrama de barras se hace a partir de una variable continua y sus intervalos entonces será un histograma. Ante intervalos de mimas amplitud, el gráfico resultante será proporcional a su correspondiente histograma de frecuencias relativas, esta característica se tendrá que cumplir siempre. Otra vez, hay que tener cuidado con el número de intervalos seleccionados para que el histograma muestre la información de forma óptima. Según la forma del histograma puede ser:

- Distribución unimodal simétrica: Los datos tienen forma de campana de gauss.
- Distribución bimodal simétrica: Hay cierta simetría respecto al eje que divide los datos por la mitad pero no es necesariamente gaussiana.

- Distribución asimétrica a la derecha: Hay escasez de datos a la derecha.
- Distribución asimétrica a la izquierda: Hay escasez de datos a la izquierda.

### 1.3.2. Diagrama de Sectores

Se divide un círculo en sectores proporcionales a las frecuencias. El ángulo de cada clase se calcularía con una simple regla de tres:

$$\frac{n}{n_i} = \frac{2\pi}{x_i}$$

## 1.4. Medidas Descriptivas de Centralización

Las medias descriptivas de centralización aproximan el valor central respecto al cual los datos se ordenan. De entre estas medidas destacan la media, moda y mediana.

### 1.4.1. Media Aritmética

Esta medida será la suma total de todos los datos dividida por el número total de observaciones.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Donde  $x_i$  hará referencia a cada observación. En el caso de que los datos se den en forma de tabla de frecuencia:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k x_i n_i = \sum_{i=1}^k x_i f_i$$

En el caso de ser una variable continua solo habrá que cambiar  $x_i$  por el número de observaciones dentro del intervalo, o sea,  $c_i$ . Dependiendo de la  $a_i$ , o sea, la amplitud de los intervalos, habrá una mayor o menor pérdida de precisión. Por otro lado, la linealidad de la media será una función  $\bar{y} = a + b\bar{x}$  de la misma forma que  $Y = a + bX$ .

Este parámetro presenta una serie de inconvenientes:

- No se puede calcular con variables cualitativas o nominales.
- Es muy sensible a los valores extremos.
- Centraliza los datos de forma subóptima ante distribuciones de datos asimétricas.
- Ante variables continuas y tablas de frecuencia su valor depende de los intervalos.
- Ante una variable discreta, la media puede no pertenecer al dominio de la variable.

### 1.4.2. Moda

Se trata del valor mas frecuente dentro de las observaciones. En el caso de que se trabaje con una variable continua se considerará el intervalo modal como al intervalo de mayor frecuencia. Este parámetro tiene la ventaja de ser muy fácil de calcular pero puede no ser única.

### 1.4.3. Mediana

La mediana será el parámetro que divida al número de observaciones al 50 % por los dos lados, o sea, será la observación central de todas las observaciones. Ante  $n$  observaciones:

$$M_e = \begin{cases} x_{(n+1)/2} & \Longleftrightarrow n \text{ impar} \\ \frac{x_{n/2} + x_{(n/2)+1}}{2} & \Longleftrightarrow n \text{ par} \end{cases}$$

Si los datos se ordenan en una tabla de frecuencia,  $M_e$  será el primer valor con  $F_i \geq 1$  o con  $N_i \geq n$ . Entre las propiedades de este parámetro destaca:

- No le afectan las observaciones extremas pues solo depende del orden de los valores de la variable. Debido a esto la mediana se usa mucho en distribuciones asimétricas.

- No se puede calcular con variables cualitativas o nominales, al igual que la media.
- Su mayor defecto es que tiene propiedades muy complicadas en las que se profundizará más adelante. Debido a esta abstracción es difícil de usar en inferencia estadística.

## 1.5. Medidas Descriptivas de Posición

Una posición genérica es un cuantil, al igual que la mediana informa sobre el valor medio de todas las observaciones, un cuantil informa sobre la posición en concreto. De esta forma,  $p \in \{n \in \mathbb{R} : 0 < n < 100\}$  será un percentil e informará sobre el porcentaje de datos que tendrá por debajo de si mismo denotado como  $P_p$ . A partir de estos percentiles se pueden definir los cuartiles.

- Primer Cuartil:  $Q_1 = P_{25}$ .
- Segundo Cuartil:  $Q_2 = P_{50} = M_e$ .
- Tercer Cuartil:  $Q_3 = P_{75}$

Estos ofrecen información muy genérica. Por ejemplo, para un conjunto de notas, si  $Q_3 = 6$  sabremos que un 7.79 se encuentra entre el 25 % mejores notas de la clase. De la misma forma se definirán los deciles como  $P_p$  con  $p \in \{10n : n \in [1, 9] \subseteq \mathbb{N}\}$ .

## 1.6. Medidas Descriptivas de Dispersión

Ante datos muy dispersos las medidas de centralización pierden su efectividad. Para esto existen las medidas de dispersión las cuales trabajarán con la sensibilidad de los datos.

### 1.6.1. Rango

Se define como  $R = \max\{x_1, \dots, x_n\} - \min\{x_1, \dots, x_n\}$ . Este parámetro es fácil de calcular y comparte unidades con los datos. Entre las desventajas de esta medida destacan:

- No se utilizan todos los datos (solo 2).
- Puede variar mucho ante datos extremos.
- Ante el aumento de observaciones el rango o aumenta o se queda igual, nunca baja.

De la misma forma se puede definir el rango intercuartílico como  $RQ = Q_3 - Q_1$ . Tanto el rango intercuartílico como el rango general son de gran importancia para realizar diagramas de cajas.

### 1.6.2. Diagramas de Cajas

Este sintetizará múltiples parámetros de posición y, a su vez, de dispersión. Presentan información sobre la simetría y los datos atípicos de la dispersión. El proceso de creación de un diagrama de cajas es bastante más específico que los métodos del resto de tipo de gráficos. El proceso se puede sintetizar en múltiples pasos:

- Primero se realiza un rectángulo que pasa por los cuartiles. Este rectángulo contendrá una línea perpendicular que lo dividirá en dos y representará la mediana. De esta forma, dentro de la caja quedaría representado el 50 % de los datos.
- A continuación se representarán 2 líneas perpendiculares al eje de los datos a  $1.5RQ$  de los 2 cuartiles, tanto a la derecha como a la izquierda.
- Para terminar la representación de estos límites se conectarán las líneas a la caja con línea punteada. Así, cualquier dato fuera de estos límites se podría considerar atípico y se representaría como puntos concretos.

### 1.6.3. Desviación Media

Otra forma bastante eficiente de calcular que tan atípico es un valor  $x_i$  es con la diferencia de este con la media. De esta forma  $x_i - \bar{x} \forall i$  serían las desviaciones de cada valor respecto a la media aritmética. Una forma de condensar toda esta información sobre cada observación podría ser con la media de todas las desviaciones:

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

El problema aparece teniendo en cuenta que en función de si la observación está a la derecha o a la izquierda el valor de la desviación es o positivo o negativo, de forma que la desviación media se acercaría a 0, o sea,  $D_m \approx 0$ . Solucionar este problema es tan sencillo como usar el valor absoluto de la diferencia para cada dato. Finalmente, la desviación media se definirá como:

$$D_m = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

### 1.6.4. Varianza y Desviación Típica

La esencia será la misma que la de  $D_m$  pero para forzar el símbolo positivo se usará el cuadrado de la diferencia. Esto implicará que el parámetro será mucho más sensible.

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Hay que tener en cuenta que la magnitud de este parámetro será el cuadrado de la magnitud de los datos. Para solucionar esto aparece el parámetro de *Desviación típica*, la cual se definirá como la raíz de la varianza:

$$s = \sqrt{s^2}$$

Además, con un simple desarrollo se comprueba que:

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2$$

De la misma forma que la media, para una tabla de frecuencia:

$$s^2 = \frac{1}{n} \sum_{i=1}^k x_i^2 n_i - \bar{x}^2 = \sum_{i=1}^k x_i^2 f_i - \bar{x}^2$$

Entre las propiedades de estas medidas destaca:

- Las dos son sensibles a los valores extremos.
- Se comprueba que para  $s$  mínimo el 75 % de los datos se encuentra en el intervalo  $(\bar{x} - 2s, \bar{x} + 2s)$ .
- No se puede usar para variables nominales (de la misma forma que la media).

Por último, cabría destacar la cuasi-varianza la cual se diferencia de la varianza por dividir el resultado entre  $n - 1$  y no  $n$ . Este valor tiene mejores propiedades que la varianza para estimar por lo cual se usa mucho más en Inferencia Estadística.

$$s_c^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{ns^2}{n-1}$$

### 1.6.5. Tipificación

Este proceso consistirá en convertir una variable a una media de 0 y una desviación estándar  $s_z = 1$ . Esta nueva variable se llamará variable tipificada y se denotará como:

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

Tipificar a variables con estas propiedades es útil para poder hacer comparaciones entre diferentes conjuntos de datos con diferentes dominios

### 1.6.6. Coeficiente de Variación

$$CV = \frac{s}{\bar{x}} 100$$

Esta medida se usa para comparar variables entre 2 conjuntos de datos con diferente media o unidades. También sirve para determinar si la media es consistente con una varianza o para comprobar la variabilidad entre grupos de datos tomados por personas distintas. Todo esto se debe a que este parámetro se mide en porcentajes. Hay que tener cuidado con la tipificación, pues una media de 0 podría implicar una operación ilegal.

## 1.7. Medidas de Forma

Al estudiar un conjunto de datos también resulta interesante ver si estos se distribuyen siguiendo una simetría o no. Una vez conocido este valor vendría bien preguntarse si el histograma es más o menos apuntado, característica que se medirá a partir de la frecuencia de la normal. De esta forma, si una variable es continua simétrica y unimodal entonces la media, mediana y moda coincidirán.

Además, si se presenta una asimetría hay que considerar si es positiva o negativa, esto implicaría frecuencias más altas a la izquierda o a la derecha, respectivamente. Los diferentes parámetros de forma son:

- Momento de orden  $p$  para  $p \in \mathbb{N}$ :  $\mu_p = \frac{1}{n} \sum_{i=1}^n x_i^p$ .
- Momento central de orden  $p$  para  $p \in \mathbb{N}$ :  $m_p = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^p$ .
- Coeficiente de asimetría:  $\gamma_1 = \frac{m_3}{m_2 \sqrt{m_2}} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{s^3}$ .
- Coeficiente de curtosis o apuntamiento:  $\gamma_2 = \frac{m_4}{s^4} - 3$ .

Siendo este tercer parámetro el más importante, pues nos informa sobre la asimetría de forma directa. Si  $\gamma > 0 \rightarrow$  asimetría positiva y si  $\gamma < 0 \rightarrow$  asimetría negativa. De la misma forma,  $\gamma_2$  cuantifica que tan apuntada es una distribución. La referencia de este parámetro es el apuntamiento de la distribución gaussiana o normal para la cual  $\frac{m_4}{s^4} = 3$ . De esta forma, si  $\gamma > 0$  la distribución será leptocúrtica (más apuntada que la normal) y si  $\gamma < 0$  platicúrtica (más aplastada que la normal).



## 2 Datos Bivariantes

### 2.1. Distribución de dos variables

Dada una población estadística de  $n$  individuos y 2 variables  $X$  e  $Y$  se define:

- Frecuencia Total: Número total de individuos  $n$ .
- Frecuencia absoluta del par  $(x_i, y_j)$ : Se refiere al número de observaciones total para cada observación de las 2 variables a la vez. Se denota por  $n_{ij}$ .
- Frecuencia relativa del par  $(x_i, y_j)$ : Exactamente igual que las frecuencias relativas para una variables. Se denota como  $f_{ij} = \frac{n_{ij}}{n}$ .

### 2.2. Tablas de frecuencia

Estos pares de observaciones para las 2 variables se pueden organizar en tablas de frecuencia bivariantes. En estas tablas se pueden mostrar tanto frecuencias absolutas como relativas. En el caso de que las variables sean cualitativas, entonces la tabla se denominará como tabla de contingencia.

$X \setminus Y$	$y_1$	$y_2$	$\dots$	$y_l$	
$x_1$	$n_{11}$	$n_{12}$	$\dots$	$n_{1l}$	$n_{1.}$
$x_2$	$n_{21}$	$n_{22}$	$\dots$	$n_{2l}$	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$x_k$	$n_{k1}$	$n_{k2}$	$\dots$	$n_{kl}$	$n_{k.}$
	$n_{.1}$	$n_{.2}$	$\dots$	$n_{.l}$	$n$

### 2.3. Distribuciones Marginales

Una frecuencia absoluta o relativa marginal se refiere al número de veces que se repite un  $x_i$  sin tener en cuenta el valor de  $Y$ , como si con una distribución de una sola variable se estuviera trabajando. De esta forma, en base a la tabla del apartado anterior:

$$n_{i.} = \sum_{j=1}^l n_{ij}, \quad n_{.j} = \sum_{i=1}^k n_{ij}, \quad f_{i.} = \frac{n_{i.}}{n}, \quad f_{.j} = \frac{n_{.j}}{n}$$

### 2.4. Distribuciones Condicionales

Una distribución de  $Y$  sabiendo que  $X = X_i$  se denota como  $(Y|X = x_i)$ . La misma notación se usa para distribuciones de  $X$  con un  $Y$  específico.

$Y X = x_i$	$y_1$	$y_2$	$\dots$	$y_l$	
$n_{ij}$	$n_{i1}$	$n_{i2}$	$\dots$	$n_{il}$	$n_{i.}$

### 2.5. Variables Independientes

Dos variables son estadísticamente independientes entre ellas si la varianza de una de ellas no afecta a los valores de la otra. De esta forma, ocurre que la distribución de una de las variables condicionada por la otra no afecta a los valores de esos valores. Además,  $X, Y$  independientes  $\iff f_{ij} = f_{i.}f_{.j}$ .

## 2.6. Gráficos

Estos datos se pueden representar de diferentes formas, por ejemplo, si el dominio de una de las variables corresponde a solo 2 valores, entonces se podría hacer 2 representaciones gráficas de la segunda variable condicionada por los 2 valores de la primera. Son de interés los diagramas de barras, los cuales se pueden representar apilados o agrupados.

Para datos multivariantes aparece un tipo de representación nueva: los diagramas de dispersión. Para datos bivariantes, por ejemplo, se usa cada pareja de observaciones  $(x_i, y_i)$  como un punto del plano. Una vez representados todos los puntos, se pueden buscar relaciones entre las variables de forma bastante cómoda.

## 2.7. Medidas Descriptivas de Dependencia Lineal

Dos variables serán dependientes linealmente cuando el aumento de una implica el aumento o disminución de la otra. Existen diferentes medidas para calcular esta dependencia.

### 2.7.1. Covarianza

Se define como:

$$s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

El razonamiento para esta cuenta es el mismo que el de la varianza para una variable. Ahora bien, se dirá que si las variables no están relacionadas entonces la covarianza tiende a cero, en cambio, si la varianza es negativa la relación entre las variables será negativa. Haciendo un poco de álgebra para la expresión de la covarianza se puede llegar a que:

$$s_{xy} = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y}$$

Expresión que, computacionalmente, es mucho más óptima.

Finalmente, hay que tener en cuenta que una covarianza de 0 no implica necesariamente una independencia en los datos. Por ejemplo, si  $X$  depende de  $Y$  de forma parabólica, la covarianza se acercará a 0, pero los datos estarán claramente relacionados.

### 2.7.2. Vector de Medias y Matriz de Covarianza

Será una medida de interés el vector de medias, el cual, como indica, contendrá los valores de las medias de todas las variables.

Por otro lado, la matriz de covarianza se define como:

$$S = \begin{pmatrix} s_x^2 & s_{xy} \\ s_{xy} & s_y^2 \end{pmatrix}$$

Donde, recordando,  $s_{x_i}^2$  hace referencia a la varianza de una variable  $X_i$ .

### 2.7.3. Coeficiente de Correlación Lineal

Aunque la covarianza sea una buena medida para medir correlaciones, está se ve afectada por las unidades: si una variable es dependiente del otra al 100% pero están relacionadas por una recta de mucha pendiente, entonces la covarianza será muy grande. Para evitar esto hace falta plantear una nueva medida que informe solamente del nivel de dependencia de las variables, esta es el coeficiente de correlación lineal, que se define como:

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

De esta forma, si la correlación es absoluta,  $r_{xy} = (1 \vee -1)$ . En caso contrario, si la correlación lineal es nula,  $r_{xy} \rightarrow 0$ . Hay que tener en cuenta que, en el caso de que una de las variables sea constante, independientemente del valor de la otra, la correlación lineal será de 0, aunque gráficamente parezca que existe una correlación. Finalmente, hay que tener en cuenta que un coeficiente de correlación cercano a 1 no implica una relación de causalidad entre las variables, podría tratarse solamente de una coincidencia. A este fenómeno se le conoce como correlación espuria.

### 2.7.4. Recta de Regresión Lineal

Una forma interesante de expresar una relación lineal es, redundantemente, con una recta. Esta recta a su vez servirá de modelo para poder predecir resultados de observaciones de una sola variable. Esta recta se representa como:

$$Y = a + bX - \epsilon$$

Donde  $a$  y  $b$  son los coeficientes de la regresión y  $\epsilon$  es el error.

Una forma óptima de calcular esta recta y sus parámetros es minimizando el error, o sea, que la distancia de cada uno de los puntos a la recta sea la mínima posible. Así, para cada par de datos  $(x_i, y_i)$ , el error de esa observación respecto a la recta será:

$$\epsilon_i = y_i - (a + bx_i)$$

#### Ajusto por Mínimos Cuadrados

Como se ha mencionado en el anterior párrafo, el punto de la cuenta de la recta será minimizar el error según los valores de  $a$  y  $b$ . O sea, se quiere minimizar la función:

$$E(a, b) = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - a - bx_i)^2$$

Haciendo las cuentas se obtiene que:

$$b = \frac{s_{xy}}{s_x^2}, \quad a = \bar{y} - b\bar{x}$$

Una vez el modelo es óptimo, se puede usar la recta  $Y = a + bX$  para hacer otras predicciones. Hay que considerar que es peligroso hacer observaciones fuera del rango de las variables. Se cumple también que, una vez optimizado el modelo, la media de los errores tiende a 0.

#### Varianza Residual y Coeficiente de determinación

Dada una recta de regresión lineal ajustada mediante mínimos cuadrados, la varianza residual se define como:

$$s_\epsilon^2 = \frac{1}{n} \sum_{i=1}^n \epsilon_i^2 = s_y^2(1 - r_{xy}^2)$$

Donde aparece el coeficiente de determinación  $R^2 = r_{xy}^2$ . Este valor hace referencia a la proporción de la variación total en la variable dependiente  $Y$ .



# 3 Probabilidad

## 3.1. Conceptos Básicos

- Experimento Aleatorio: el resultado impredecible pertenece a un conjunto de posibles resultados.
- Suceso Elemental: elemento del conjunto de posibles resultados.
- Espacio muestral: conjunto de resultados  $E = \{e_1, e_2, \dots, e_n\}$ .
- Suceso aleatorio: subconjunto de  $E$ .
- Suceso seguro: siempre se verifica y coincide con  $E$ .
- Suceso imposible: nunca se verifica ( $\emptyset \subset E$ ).
- Suceso contrario de  $A$ :  $A^c = \bar{A} = \{e \in E : e \notin A\}$ .

## 3.2. Operaciones básicas de sucesos

Aparte de las típicas operaciones de conjuntos, es de interés la diferencia simétrica:

$$A \triangle B = (A - B) \cup (B - A) = (A \cup B) - (B \cap A)$$

## 3.3. Definición de Probabilidad

$P$  es una función de probabilidad sobre  $E$  si se verifican 3 axiomas:

1.

$$\begin{aligned} P : E &\longrightarrow [0, 1] \subset \mathbb{R} \\ A &\longmapsto 0 \leq P(A) \leq 1 \end{aligned}$$

2.

$$P(E) = 1$$

3.

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \iff A_i \cap A_j = \emptyset \forall A_i, A_j \in A$$

### Interpretación Clásica de Probabilidad

La Regla de Laplace es una interpretación clásica de la probabilidad:

$$P(A) = \frac{\# \text{ Casos favorables de } A}{\# \text{ Casos posibles}}$$

Solo para resultado verosímiles.

### Interpretación Frecuentista de la Probabilidad

Se aplican a experimentos con repeticiones indefinidas con condiciones similares.

$$P(A) = \lim_{n \rightarrow \infty} \frac{A_n}{n}$$

Se puede demostrar que esta definición cumple los axiomas de la probabilidad. YA que en la práctica tender al infinito es imposible, simplemente se hace un número de observaciones suficientes como para poder estimar la probabilidad.

### 3.4. Probabilidad como Grado de Confianza

Las ventajas de usar la probabilidad como modelo de ocurrencia son que el modelo mejora con el tiempo (modelo evolutivo) y que se puede aplicar ante la no opinión sobre la ocurrencia de un suceso.

### 3.5. Probabilidad Condicionada e Independencia de Sucesos

Se define la probabilidad de  $A$  condicionada por la advertencia de  $B$  como:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, P(B) \neq 0$$

- $P(A|B) = P(A) \vee P(B|A) = P(B) \iff (A, B)$  independientes.
- $P(A \cap B) = P(A)P(B) \iff (A, B)$  independientes.
- $P(\bar{A}|B) = 1 - P(A|B)$ .

### 3.6. Teorema de la Probabilidad Compuesta

Para una colección  $A_1, A_2, \dots, A_n \subset E$  de sucesos aleatorios se verifica:

$$P(A) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P(A_n|A_1 \cap \dots \cap A_{n-1})$$

### 3.7. Teorema de la Probabilidad Total

Dada una colección  $A_1, A_2, \dots, A_n \subset E$  que es un sistema exhaustivo y mutuamente excluyente se verifica:

$$\forall B \subset E : P(B) = \sum_{i=1}^n P(B|A_i)P(A_i)$$

Donde una colección es un sistema exhaustivo y excluyente si se verifica:

$$\bigcup_{i=1}^n A_i = E \wedge A_i \cap A_j = \emptyset \forall i \neq j$$

### 3.8. Teorema de Bayes

Dado un sistema exhaustivo y excluyente  $A_1, A_2, \dots, A_n \subset E$  y sea  $B \subset E$  un suceso entonces se verifica:

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^n P(B|A_i)P(A_i)} \quad \forall j \in \{1, \dots, n\}$$

Este teorema se puede probar fácilmente mediante la definición de probabilidad condicionada y el teorema de la probabilidad total.

## 4 Variables Aleatorias

### 4.1. Definición

Una variable aleatoria es una función tal que:

$$\begin{aligned} X : E &\longrightarrow \mathbb{R} \\ e &\longmapsto X(e) \end{aligned}$$

O sea, a cada suceso elemental  $e \in E$  se le asocia un valor real mediante  $X(e)$ .  $X$  se considera una variable aleatoria pues se aplica sobre sucesos aleatorios. Aun así, los valores que toma la variable no son aleatorios, están definidos de forma precisa para cada suceso  $e$ . Además:

$$P(X = x) = P(\{e \in E : X(e) = x\})$$

$$P(X \in A) = P(\{e \in E : X(e) \in A\})$$

Finalmente, la distribución de una variable aleatoria  $X$  se da a partir de calcular las probabilidades de todos los posibles valores de  $X$ , o sea, de su dominio.

### 4.2. Clasificación de variables

- Variable Aleatoria Discreta: Solo puede tomar valores de un conjunto finito o infinito numerable. Por ejemplo:  $X : E \rightarrow \mathbb{N}$ .
- Variable Aleatoria Continua: Toma infinitos valores no numerables. Por ejemplo:  $X : E \rightarrow \mathbb{R}$ .

### 4.3. Variables Aleatorias Discretas

#### 4.3.1. Función de Masa de Probabilidad

Dada una  $X : E \rightarrow \mathbb{N}$ , se define una función de masa de probabilidad,  $f(x_i)$ , como la probabilidad de que  $X$  tome el valor  $x_i$ :

$$\begin{aligned} f : \mathbb{N} &\longrightarrow [0, 1] \\ x_i &\longmapsto f(x_i) = P(X = x_i) = P(\{e \in E : X(e) = x_i\}) \end{aligned}$$

$$x_i \notin \text{dom}(X) \longrightarrow f(x_i) = 0$$

A partir de los axiomas de la probabilidad y siendo  $\text{dom}(X) = \{x_1, x_2, \dots, x_k\}$  se deduce que:

- $f(x_i) \geq 0 \forall i \in \{1, \dots, k\}$ .
- $\sum_{i=1}^k f(x_i) = \sum_{i=1}^k P(X = x_i) = 1$ .

Resulta cómodo, para la representación de funciones de masa, usar un diagrama de barras.

### 4.3.2. Función de Distribución

$$\begin{aligned} F : \mathbb{R} &\longrightarrow [0, 1] \\ x &\longmapsto F(x) = P(X \leq x) = P(\{e \in E : X(e) \leq x\}) \end{aligned}$$

Propiedades:

- $F$  es monótona no decreciente.
- $F$  es continua por la derecha.
- $\lim_{x \rightarrow -\infty} F(x) = 0.$
- $\lim_{x \rightarrow +\infty} F(x) = 1.$
- $P(X > x) = 1 - F(x).$
- $P(x_i < X \leq x_j) = F(x_j) - F(x_i).$

Esta función es escalonada en cada intervalo  $[x_k, x_{k+1})$  y tiene discontinuidades en los saltos de intervalos (o sea,  $k = 1, 2, \dots$ ). El salto entre intervalos representa la probabilidad del valor en ese punto.

### 4.3.3. Momentos de una variable aleatoria discreta

Estos, equivalentes a las medidas descriptivas, son:

- Esperanza (Media) de una Variable Aleatoria:  $E(X) = \mu = \sum x_i f(x_i).$
- Varianza de una Variable Aleatoria:  $Var(X) = \sigma^2 = \sum (x_i - \mu)^2 f(x_i) = \sum x_i^2 f(x_i) - \mu^2.$  Además, se observa que:  $Var(X) = E(X^2) - E(X)^2.$
- Desviación Típica de una Variable Aleatoria:  $sd(X) = \sigma = \sqrt{Var(X)}.$

## 4.4. Desigualdad de Chebycheff

Para  $X$  una distribución con media  $\mu$  y desviación típica  $\sigma$ , la probabilidad de obtener un valor que se desvie de la media mínimo  $k\sigma$  veces es de  $1/k^2$ . Esto es:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \equiv P(|X - \mu| < k\sigma) > 1 - \frac{1}{k^2}$$

Es así como la desviación típica se usapara medir la dispersión. Este teorema ofrece un rango límite ( $1/k^2$ ) que en muchas ocasiones puede resultar obvio. Para mejores estimaciones se usa otras hipótesis.

## 4.5. Distribución de Bernoulli

Un experimento consiste en observar la ocurrencia de un suceso.  $p$  será la probabilidad de que ocurra y  $q$  será  $1 - p$ , o sea, la probabilidad de fracaso. Este experimento se podría definir mediante una variable aleatoria discreta como  $X = 1$  si el suceso ocurre y  $X = 0$  si no lo hace. Así, se dirá que  $X \sim \text{Bernoulli}(p)$  si

$$X = \begin{cases} 1 & \text{donde } P(X = 1) = p \\ 0 & \text{donde } P(X = 0) = q \end{cases}$$

### 4.5.1. Funciones de una variable aleatoria de Bernoulli

- Función de masa de probabilidad de la variable:

$$\begin{aligned} f(1) &= P(X = 1) = p \\ f(0) &= P(X = 0) = q \end{aligned}$$



- Función de distribución de la variable:

$$F(x) = \begin{cases} 0 & \Longleftrightarrow x < 0 \\ q & \Longleftrightarrow 0 \leq x < 1 \\ 1 & \Longleftrightarrow x \geq 1 \end{cases}$$

- Media y Varianza de la variable:

$$\begin{aligned} E[X] &= \mu = p \\ Var[X] &= \sigma^2 = pq \end{aligned}$$

## 4.6. Distribución Binomial

Un experimento consiste en considerar  $n$  repeticiones independientes. Los resultados posibles son el fracaso y el éxito.  $p$  será la probabilidad de éxito en una repetición. La variable aleatoria discreta se definirá como  $X$  = número de éxitos obtenidos en  $n$  pruebas. Así, se dirá que  $X$  sigue una distribución binomial  $X \sim B(n, p)$ .

- Función de masa de probabilidad de la variable:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x \in \{0, 1, 2, \dots, n\}$$

- Media y Varianza de la variable:

$$\begin{aligned} E[X] &= \mu = np \\ Var[X] &= \sigma^2 = np(1-p) \end{aligned}$$

Se cumple que si  $p = 0.5$  entonces la distribución binomial es simétrica. Si  $p > 0.5$  entonces será asimétrica negativa. En el caso contrario, será asimétrica positiva.

## 4.7. Distribución Geométrica

Un experimento consistirá en estudiar repeticiones independientes con dos posibles resultados. Estos son el fracaso y el éxito.  $p$  será la probabilidad de éxito para una repetición y es de interés el instante del primer éxito. La variable aleatoria discreta se definirá como  $X$  = repetición en la que se observa el primer éxito. Así, se dirá que  $X$  sigue una distribución geométrica  $X \sim G(p)$ .

- Función de masa de la variable:

$$P(X = x) = p(1-p)^{x-1}, \quad x \in \{1, 2, \dots\}$$

- Media y Varianza de la variable:

$$\begin{aligned} E[X] &= \mu = \frac{1}{p} \\ Var[X] &= \sigma^2 = \frac{1}{p^2} \end{aligned}$$

Obviamente, si el éxito se produce en  $k$ , entonces habrá  $k - 1$  fracasos. Por lo general, además, se cumple que  $P(X > x_j | X > x_i) = P(X > x_j - x_i)$ . Debido a esto se dice que la distribución geométrica no tiene memoria.

## 4.8. Distribución de Poisson

Estas distribuciones estudiarán el número de ocurrencias de un suceso dentro de un intervalo temporal o espacial. Se denotará por  $\lambda$  a la intensidad de ocurrencias y esta será conocida y dependiente de la longitud del intervalo, a parte de proporcional. La variable aleatoria discreta se definirá como  $X$  = número de observaciones de un suceso en un intervalo dado. Así, se podrá decir que  $X$  sigue una distribución de Poisson  $X \sim Po(\lambda)$ .

- Función de masa de probabilidad de la variable:

$$P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x \in \{0, 1, 2, \dots\}$$

- Media y Varianza de la variable:

$$\begin{aligned} E[X] &= \mu = \lambda \\ Var[X] &= \sigma^2 = \lambda \end{aligned}$$

La distribución de Poisson puede interpretarse como una aproximación a la binomial como  $\lambda = np$ . Se considera la aproximación válida para  $n \geq 20$  con  $p \leq 0.05$ . Si  $n \geq 100$  y  $np \leq 10$  la aproximación es muy buena.

## 4.9. Variables Aleatorias Continuas

Cuando se trata de una variable discreta y aunque los valores tomen valores infinitos numerables, cada valor tiene una probabilidad de modo  $f(x_i) = P(X = x_i)$ . De esta forma:

$$\sum_{i=1}^{\infty} f(x_i) = \sum_{i=1}^{\infty} P(X = x_i) = 1$$

Ahora bien, en una variable aleatoria continua no se puede hacer una suma infinita, debido a que los valores posibles de la variable son infinitos no numerables. De esta forma, mientras que para sumas infinitas se usan sumatorios, para sumas infinitesimales se usan integrales, que van a ser la herramienta clave de estas variables.

### 4.9.1. Función de Densidad

La función de densidad de una variable aleatoria continua  $X$  es  $f : \mathbb{R} \rightarrow \mathbb{R}$ . Esta es integrable y no negativa. Se cumple que:

- $f(x) \geq 0$ .
- $\int_{-\infty}^{\infty} f(x)dx = 1$ .
- $P(a \leq X \leq b) = \int_a^b f(x)dx$ ,  $a < b$ .

Debido a la naturaleza infinitesimal de estas variables se cumple que la probabilidad de un punto concreto es nula, por definición. Debido a esto mismo, es indiferente si los límites de los valores de la variable están incluidos o no:

$$P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b)$$

### 4.9.2. Función de Distribución

Para una v.a.  $X$  se define la función de densidad como:

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt, \quad \forall x \in \mathbb{R}$$

Propiedades:

- $P(a < X \leq b) = F(b) - F(a)$ .
- $F$  es monótona no decreciente.
- $F(-\infty) = 0$ .
- $F(\infty) = 1$ .
- $F$  es continua por la derecha.
- $F'(x) = f(x)$ .

### 4.9.3. Momento de una variable aleatoria continua

- Esperanza (Media) de una V.A.:  $E(X) = \mu = \int_{-\infty}^{\infty} x f(x)dx$ .
- Varianza de una V.A.:  $Var(X) = \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$ .
- Desviación típica de una V.A.:  $\sigma = \sqrt{Var(X)}$ .

## 4.10. Distribución Uniforme

Esta distribución se denota como  $X \sim U(a, b)$  para una función de densidad

$$f(x) = \begin{cases} \frac{1}{b-a} & , x \in (a, b) \\ 0 & , x \notin (a, b) \end{cases}$$

Con función de distribución:

$$F(X) = \begin{cases} 0 & , x \leq a \\ \frac{x-a}{b-a} & , a \in (a, b) \\ 1 & , x \geq b \end{cases}$$

Donde la media y varianza son:

$$\mu = \frac{a+b}{2}, \sigma^2 = \frac{(b-a)^2}{12}$$

## 4.11. Distribución Exponencial

$$X \sim \text{Exp}(\alpha), \alpha > 0 \iff f(x) = \begin{cases} \alpha e^{-\alpha x} & , x > 0 \\ 0 & , x \leq 0 \end{cases}$$

Se observa que la distribución exponencial es equivalente a la distribución geométrica para una variable discreta. Esta distribución es de interés para describir procesos en los que se interesa conocer el tiempo que pasa hasta que ocurren. Eso se da asumiendo que el proceso no tiene memoria (no depende de lo ocurrido anteriormente). La media y varianza de una distribución de este tipo son:

$$\mu = \frac{1}{\alpha}, \text{Var}(X) = \frac{1}{\alpha^2}$$

Con una función de distribución tal que:

$$F(x) = \begin{cases} 0 & , x \leq 0 \\ 1 - e^{-\alpha x} & , x > 0 \end{cases}$$

Por lo general, si  $X \sim \text{Exp}(\lambda)$  entonces

$$P(X < x_j | X > x_i) = P(X > x_j - x_i)$$

Es por esto que se dice que la distribución exponencial no tiene memoria.

## 4.12. Distribución Normal

Esta distribución seguramente sea la más importante de la estadística. Esto se debe a:

1. Es común la medición de magnitudes con errores normales.
2. Debido a la naturaleza de la distribución, los cálculos resultan sencillos.
3. Gracias al Teorema Central del Límite muchas distribuciones se pueden aproximar a la distribución normal, con muestras lo suficientemente grandes.

De esta forma, se dice que

$$X \sim N(\mu, \sigma^2), f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}$$

Se cumple que:

- $\int_{-\infty}^{\infty} f(x)dx = 1.$
- $E(X) = \int_{-\infty}^{\infty} xf(x)dx = \mu.$
- $\text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx = \sigma^2.$

### 4.13. Distribución Normal Estándar

Se trata de la distribución normal con media 0 y varianza 1, o sea  $Z \sim N(0, 1)$ . En este caso, la función para todo  $z$  en los reales es:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Con función de distribución:

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^z f(t) dt$$

El problema es que la cuenta de esta integral resulta extremadamente tediosa. Debido a esto se usan tablas con muchos valores de  $\Phi(z)$  para  $z \geq 0$ .