

Estadística

DESCRIPTIVA E INFERENCIAL

Javier Blanco, Analista de Datos

www.linkedin.com/in/jblancop

ÍNDICE DE CONTENIDOS

I. Síntesis de la información	5
I.1. Distribución de los datos	5
I.2. Medidas centrales	6
I.2.1. Media aritmética	6
I.2.2. Mediana	6
I.2.3. Moda	7
I.2.4. Comparación entre media, mediana y moda	7
I.3. Medidas de posición	7
I.4. Medidas de dispersión	8
I.4.1. Varianza y desviación típica	8
I.4.2. Recorridos	8
I.4.3. Coeficiente de variación	8
I.5. Normalización de variables	9
II. Análisis bivariado	10
II.1. Distribución conjunta de dos caracteres	10
II.1.1. Distribuciones marginales	10
II.1.2. Distribuciones condicionadas	11
II.2. Independencia	11
II.2.1. Covarianza	11
II.2.2. Coeficiente de correlación	12
II.2.3. Coeficiente chi-cuadrado	12
III. Ajuste y regresión bidimensional	13
III.1. Mínimos cuadrados	13
III.1.1. Caso lineal	13
III.2. Bondad de ajuste	14
III.2.1. Coeficiente de determinación	14
IV. Combinatoria	16
IV.1. Variaciones	16
IV.1.1. Con repetición	16
IV.1.2. Sin repetición	16
IV.2. Permutaciones	16
IV.2.1. Sin repetición	16
IV.2.2. Con repetición	16
IV.3. Combinaciones	17
IV.3.1. Sin repetición	17
IV.3.2. Con repetición	17
V. Probabilidad	18
V.1. Conjuntos	18
V.1.1. Operaciones entre conjuntos	18
V.1.2. Conjuntos característicos	18
V.2. Álgebra de sucesos	18

V.2.1. Espacio muestral	19
V.2.2. Relaciones entre sucesos	19
V.3. Definición de probabilidad	19
V.4. Propiedades de la función de probabilidad	19
V.5. Probabilidad condicionada	20
V.5.1. Independencia	20
V.5.2. Teorema de la probabilidad total	20
V.5.3. Teorema de Bayes	21
VI. Variable aleatoria	22
VI.1. Caracterización de variables aleatorias	22
VI.1.1. Caso discreto	22
VI.1.2. Caso continuo	24
VI.2. Esperanza matemática	26
VII. Modelos probabilísticos	27
VII.1. Distribución binomial	27
VII.2. Distribución de Poisson	27
VII.3. Distribución exponencial	28
VII.4. Distribución gamma	29
VII.5. Distribución uniforme	29
VII.6. Distribución normal	30
VII.6.1. Normal tipo	31
VII.6.2. Teorema central del límite	32
VII.6.3. Distribuciones derivadas de la normal	32
VII.7. Distribución de Weibull	33
VIII. Inferencia estadística	34
VIII.1. Procedimientos inferenciales	34
VIII.2. Obtención de la información	34
VIII.2.1. Muestreo aleatorio	34
IX. Estimación puntual	36
IX.1. Estadísticos	36
IX.2. Estimación en poblaciones normales	36
IX.2.1. Distribución de la media muestral	36
IX.2.2. Distribución de la varianza muestral	36
IX.2.3. Distribución de la proporción muestral	37
IX.3. Propiedades de los estadísticos	37
IX.3.1. Insensatez	37
IX.3.2. Eficiencia	38
IX.3.3. Error cuadrático medio	38
IX.3.4. Consistencia	38
X. Intervalos de confianza	39
X.1. Longitud del intervalo	39
X.2. Método del pivote	40
X.3. Intervalos en poblaciones normales	40

X.3.1. Determinación de la media conocida la varianza	41
X.3.2. Determinación de la media desconocida la varianza	41
X.3.3. Determinación de la varianza conocida la media	41
X.3.4. Determinación de la varianza desconocida la media	42
X.4. Método asintótico	42
X.4.1. Determinación de la proporción	42
X.5. Determinación del tamaño muestral	43
XI. Contraste de hipótesis	44
XI.1. Clasificación de los contrastes	44
XI.1.1. Contrastes paramétricos	44
XI.2. Errores de un contraste	45
XI.3. Realización de contrastes paramétricos	45
XI.3.1. Nivel crítico	45
XII. Contrastes no paramétricos	47
XII.1. Análisis de la estructura de la población	47
XII.1.1. Contraste chi-cuadrado de Pearson	47
XII.1.2. Contraste de Kolmogorov-Smirnov	47
XII.1.3. Contraste de Shapiro-Wilk	48
XII. 2. Contrastes de localización y escala	48
XII.2.1. Contraste T de Wilcoxon	49

I. SÍNTESIS DE LA INFORMACIÓN

Los datos, observaciones resultantes de un experimento, constituyen la materia prima de la Estadística, y su descripción y clasificación se pueden hacer según diferentes enfoques, aunque el más básico es aquel que los divide en:

- **Atributos:** De tipo cualitativo; por ejemplo, la raza.
- **Variables:** De tipo cuantitativo o cardinal, expresables numéricamente; por ejemplo, la edad.

De manera genérica, tanto a un atributo como a una variable se les puede denominar carácter.

A su vez, los atributos pueden ser:

- **Ordinales:** Los valores siguen un orden natural, como, por ejemplo, en un atributo que mida la calidad: bueno, regular y malo.
- **Nominales:** Sólo admiten un orden arbitrario, como el alfabético y el dicotómico (atributos binarios).

Las variables, por otro lado, pueden ser:

- **Discretas:** No toman valores intermedios entre dos consecutivos; por ejemplo, el número de personas que votan por un determinado partido.
- **Continuas:** Pueden tomar cualquier valor dentro de un intervalo; por ejemplo, la estatura de ese grupo de personas.

En la práctica, todas las variables son discretas debido a las limitaciones de los aparatos de medida, pero se mantiene la distinción teórica. También es posible convertir una variable discreta en continua y viceversa mediante la agrupación de sus valores en intervalos.

I.1. Distribución de los datos

La organización de los datos va a depender del número de observaciones distintas y de su grado de repetición; así:

- Cuando se tiene un número pequeño de observaciones casi todas distintas, éste se puede dar por extensión.
- Cuando el número es grande, pero hay pocas distintas, se recurre a una tabla de frecuencias.
- Cuando el número es grande y la mayoría son distintas, las observaciones se agrupan en intervalos para los que a su vez se da una frecuencia.

En cualquier caso, se obtiene una distribución de frecuencias, que se puede representar esquemáticamente como sigue:

x_i	n_i	N_i	f_i	F_i
x_1	n_1	N_1	f_1	F_1
x_2	n_2	N_2	f_2	F_2
...
x_r	n_r	$N_r = n$	f_n	$F_r = 1$

x_i : Valores (o intervalos de valores)

n_i : Frecuencias absolutas

N_i : Frecuencias absolutas acumuladas

n : Número total de observaciones

f_i : Frecuencias relativas

F_i : Frecuencias relativas acumuladas

X es el carácter que representa genéricamente la distribución y toma r valores representados por x_i , siendo n_i la frecuencia absoluta de dicho valor.

N_i es la frecuencia absoluta acumulada, que se obtiene como $\sum_{j=1}^i n_j$ —es decir, el sumatorio de todas las frecuencias absolutas desde n_1 hasta n_i , siendo n_r la última posible, lo que resultaría en N_r —. Por otro lado, n , que equivale a N_r , representa el número total de observaciones y es igual a $\sum_{i=1}^r n_i$. Conviene hacer notar la diferencia con r , que sería el número de observaciones distintas; por ejemplo, una distribución consistente en tres cincos tiene $r = 1$ y $n = 3$.

Finalmente, f_i es la frecuencia relativa, definida como n_i/n , y F_i es la frecuencia relativa acumulada, que viene dada por $\sum_{j=1}^i f_j$ —procedimiento análogo al del cálculo de N_i —.

I.2. Medidas centrales

Tras organizar los datos según una distribución de frecuencias se procede a su resumen mediante una serie de medidas que ofrecen una representación numérica de dicha distribución.

I.2.1. Media aritmética

Se define de la siguiente manera:

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

x_i : Valores (o marcas de clase en el caso de intervalos)

n_i : Frecuencias absolutas

n : Número total de observaciones

f_i : Frecuencias relativas

Para aquellas distribuciones que se pueden dar por extensión:

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

La media es una medida óptima de representación dado que:

- Para su obtención se utilizan todas las observaciones —es decir, la máxima información disponible— y su valor, expresado en la misma unidad que los datos, está comprendido entre el mínimo y el máximo de la distribución.
- La suma de las desviaciones de los valores de la distribución respecto a la media es cero.
- Verifica las propiedades de traslación y homotecia: $Y = a \cdot X + b \Rightarrow \bar{y} = a \cdot \bar{x} + b$
- Es el valor ϕ que hace mínima la expresión $\sum_{i=1}^r (x_i - \phi)^2 \cdot n_i$; este mínimo es, precisamente, la varianza —una medida de dispersión— de X .

I.2.2. Mediana

Es un valor que, previa ordenación, deja la mitad de las observaciones a cada uno de sus lados; es decir, el 50 % de los datos son menores o iguales que la mediana, mientras que el otro 50 % es igual o superior.

Si el número de datos es impar, la mediana es el valor que queda en el centro, mientras que si es par se suele calcular como la media aritmética de los dos elementos centrales, aunque desde un punto de vista matemático sería correcto elegir cualquiera de ellos dos —o incluso cualquiera entre ellos dos—.

Como la media, es un valor comprendido entre el mínimo y el máximo de la distribución y viene expresado en las mismas unidades que los datos.

I.2.3. Moda

En términos absolutos, es el valor que más veces se repite en la distribución, mientras que aquellos valores que tengan frecuencias mayores que los adyacentes son modas relativas.

Al igual que media y mediana, es un valor comprendido entre el mínimo y el máximo de la distribución y viene expresado en las mismas unidades que los datos.

I.2.4. Comparación entre media, mediana y moda

La moda es la medida central más flexible, puesto que es aplicable incluso a atributos nominales, mientras que la mediana lo es a los ordinales y la media sólo a las variables.

En principio, la media es la mejor de las medidas para la representación de las variables al aprovechar toda la información contenida en los datos, mientras que la mediana se basa en su ordenación y la moda sólo en sus frecuencias.

La moda es muy inestable: un pequeño cambio en los valores puede hacerla variar de un extremo a otro; la mediana es insensible al valor concreto de los datos y puede permanecer constante aunque se altere enormemente la magnitud de las observaciones extremas, mientras que la media no y por ello puede considerarse el centro de gravedad de la distribución.

I.3. Medidas de posición

Los cuantiles r de orden k dividen la distribución en k partes iguales; es decir, con el mismo número de elementos. Los cuantiles más destacados son:

- **Cuartiles (Q):** Hay tres (r), que dividen la distribución en cuatro (k) partes iguales.
- **Deciles (D):** Hay nueve (r), que la dividen en diez (k).
- **Percentiles (P):** Hay noventa y nueve (r), que la dividen en cien (k).

Q_2 , D_5 y P_{50} equivalen a la mediana de la distribución.

De manera general, el cuantil r de orden k es el valor x_i cuya frecuencia absoluta acumulada es la inmediatamente superior a:

$$N_i = \frac{r}{k} \cdot n$$

N_i : Frecuencia absoluta acumulada

r : Posición del cuantil

k : Orden del cuantil

n : Número total de observaciones

I.4. Medidas de dispersión

Indican el nivel de concentración de los datos analizados y, en consecuencia, la bondad de los promedios calculados.

I.4.1. Varianza y desviación típica

La varianza es la media de las desviaciones cuadráticas de las observaciones con respecto a la media muestral; es decir, estima la distancia que existe entre cada observación y la media —centro de gravedad de la distribución—, ya que cuanto mayor sea esa distancia en promedio mayor será la dispersión —la falta de concentración— de los datos:

$$S_X^2 = \overline{(X - \bar{X})^2} = \frac{\sum_{i=1}^r (x_i - \bar{x})^2 \cdot n_i}{n} = \sum_{i=1}^r (x_i - \bar{x})^2 \cdot f_i$$

x_i : Valores (o marcas de clase en el caso de intervalos)

\bar{x} : Media aritmética

n_i : Frecuencias absolutas

n : Número total de observaciones

f_i : Frecuencias relativas

Cabe distinguir entre \bar{X} , la media muestral, y \bar{x} , que sería cada una de las realizaciones de \bar{X} .

La desviación típica es simplemente la raíz cuadrática positiva de la varianza:

$$S_X = +\sqrt{S_X^2}$$

La varianza es de más fácil manejo matemático mientras que la desviación típica es de interpretación más sencilla, puesto que está en las mismas unidades que la variable medida. El valor de ambas es siempre positivo y sólo pueden ser cero —dispersión nula— en caso de que todas las observaciones coincidan con la media muestral, supuesto conocido como representatividad absoluta de la media.

Al estar tan íntimamente relacionadas con la media, verifican también las propiedades de traslación y homotecia, de manera que:

$$Y = a \cdot X + b \Rightarrow S_Y^2 = a^2 \cdot S_X^2$$

I.4.2. Recorridos

Bajo este epígrafe se agrupan dos medidas diferentes:

- **Rango:** Se calcula como la diferencia entre los valores extremos. Es fácil de estimar, pero puede ofrecer una visión distorsionada de la dispersión de la distribución en caso de que haya valores atípicos aislados.
- **Recorrido intercuartílico:** Se calcula como la diferencia entre el tercer y el primer cuartil, tal que $R_I = Q_3 - Q_1$. Al centrarse en el 50 % de los valores centrales evita que observaciones atípicas extremas den una visión sesgada de la variabilidad de la distribución.

I.4.3. Coeficiente de variación

Las anteriores medidas de dispersión son absolutas, por lo que no tienen en cuenta la magnitud de las observaciones ni permiten hacer comparaciones entre variables.

Por contra, el coeficiente de variación es una medida relativa y adimensional, puesto que se calcula como el cociente entre la desviación típica y la media aritmética —en valor absoluto—, que se miden en las mismas unidades:

$$CV = \frac{S_x}{|\bar{x}|}$$

Así, tiene en cuenta el rango de valores en que se mueve la variable y permite comparar la dispersión de varias distribuciones. Además, es indicativo de la fiabilidad de la media; un valor del coeficiente de 0,5 es el límite superior para admitir que ésta representa aceptablemente al conjunto de la distribución.

I.5. Normalización de variables

También llamada tipificación, es un tipo de transformación que hace más regular la distribución y por tanto facilita su estudio y sobre todo su comparación. Como se verá más adelante, resulta esencial para la inferencia estadística.

A partir de una variable X de media \bar{x} y desviación típica S , se puede obtener una variable normalizada o tipificada Z como sigue:

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

Z , que debido al cociente carece de unidades, tiene media 0 y desviación típica 1; dado que la primera puede considerarse el centro de gravedad de la distribución y la segunda, su escala, al tipificar varias variables se consigue centrarlas en el mismo punto y dotarlas de la misma escala, lo que hace posible comparar a individuos de diferentes distribuciones.

II. ANÁLISIS BIVARIADO

El estudio simultáneo de dos caracteres permite determinar si existe relación entre ellos y de qué intensidad, si bien este análisis no indica cuál es la causa y cuál el efecto.

II.1. Distribución conjunta de dos caracteres

El estudio de dos caracteres de una misma población implica la obtención de pares de valores (X, Y) para cada individuo, que se pueden organizar en una tabla de frecuencias de doble entrada o bidimensional como la que sigue:

X, Y	y_1	...	y_j	...	y_s	$f(X)$
x_1	n_{11}	...	n_{1j}	...	n_{1s}	$n_{1.}$
...
x_i	n_{i1}	...	n_{ij}	...	n_{is}	$n_{i.}$
...
x_r	n_{r1}	...	n_{rj}	...	n_{rs}	$n_{r.}$
$f(Y)$	$n_{.1}$...	$n_{.j}$...	$n_{.s}$	n

x_i, y_i : Valores (o intervalos de valores)
 n_{ij} : Frecuencias absolutas del par (X, Y)
 $n_{i.}$: Frecuencias marginales absolutas de X
 $n_{.j}$: Frecuencias marginales absolutas de Y
 n : Número total de observaciones

Si X toma r valores e Y toma s , se define la frecuencia absoluta del par, n_{ij} , como el número de veces que aparece dicho par, siendo $\sum_{i=1}^r \sum_{j=1}^s n_{ij} = n$. Por otro lado, $f_{ij} = n_{ij}/n$ es la frecuencia relativa del par, y se verifica que $\sum_{i=1}^r \sum_{j=1}^s f_{ij} = 1$.

Si la distribución es de atributos, la tabla se llama de contingencia, mientras que si es de variables, de correlación.

De forma intuitiva, la acumulación de mayores frecuencias alrededor de una de las diagonales es indicativa de la existencia de relación entre ambos caracteres.

II.1.1. Distribuciones marginales

En la tabla anterior se han sumado las frecuencias fila a fila y columna a columna y el resultado se ha situado en los márgenes (en color azul claro), de manera que la última columna constituye la distribución marginal de X y la última fila, la de Y :

$$n_{i.} = \sum_{j=1}^s n_{ij} \text{ y } \sum_{i=1}^r f_{i.} = 1$$

$$n_{.j} = \sum_{i=1}^r n_{ij} \text{ y } \sum_{j=1}^s f_{.j} = 1$$

$$n \equiv n_{..} = \sum_{i=1}^r \sum_{j=1}^s n_{ij} \text{ y } \sum_{i=1}^r \sum_{j=1}^s f_{ij} = 1$$

$n_{i.}$: Frecuencias marginales absolutas de X
 n_{ij} : Frecuencias absolutas del par (X, Y)
 $f_{i.}$: Frecuencias marginales relativas de X
 $n_{.j}$: Frecuencias marginales absolutas de Y
 $f_{.j}$: Frecuencias marginales relativas de Y
 n : Número total de observaciones

f_{ij} : Frecuencias relativas del par (X, Y)

En cierto modo, las frecuencias marginales podrían considerarse frecuencias absolutas —o relativas— acumuladas, pero tomadas fila a fila y columna a columna.

Es conveniente reseñar que al considerar la distribución marginal de X se está valorando la distribución de esta variable sin tener en cuenta —es decir, independientemente de— cómo se distribuye la otra, Y , y viceversa.

II.1.2. Distribuciones condicionadas

Supone considerar la distribución de una variable para un valor fijo de la otra, dado que poseer información previa de Y puede afectar a X —y viceversa—:

$$f_{i|j} = \frac{n_{ij}}{n_{.j}} \quad \forall i = 1, 2, \dots, r$$
$$f_{j|i} = \frac{n_{ij}}{n_{i.}} \quad \forall j = 1, 2, \dots, s$$

Así, la distribución de X condicionada a Y sería una serie de frecuencias relativas (cuyo sumatorio equivaldría a uno), resultante del cociente entre la frecuencia absoluta del par (X, Y) y la frecuencia marginal de Y para cualquier valor de i (es decir, cualquier fila) y un único valor de j (el correspondiente a dicha columna). La explicación sería análoga para Y .

II.2. Independencia

El grado de dependencia entre variables depende de la cantidad de información que sobre una variable arroja el tener conocimiento de la otra; así, pueden establecerse tres categorías:

- **Dependencia funcional:** Información total; propia del cálculo: $y = f(x)$.
- **Independencia:** Información nula.
- **Dependencia estadística:** Información parcial.

X es independiente de Y si la frecuencia condicionada coincide con la marginal:

$$f_{i|j} = f_{i.} \quad \forall i = 1, 2, \dots, r \quad \forall j = 1, 2, \dots, s$$

Que también se puede expresar:

$$f_{ij} = f_{i.} \cdot f_{.j} \quad \forall i, j$$

f_{ij} : Frecuencias relativas del par (X, Y)
 $f_{i.}$: Frecuencias marginales relativas de X
 $f_{.j}$: Frecuencias marginales relativas de Y

La medida de la dependencia se puede realizar mediante diversos coeficientes de relación: de correlación (covarianza y coeficiente de correlación) para variables y de contingencia (coeficiente χ^2) para atributos.

II.2.1. Covarianza

Viene dada por la expresión:

$$S_{XY} = \sum_{i=1}^r \sum_{j=1}^s (x_i - \bar{x}) \cdot (y_j - \bar{y}) \cdot f_{ij}$$

x_i, y_j : Valores (o marcas de clase)
 \bar{x}, \bar{y} : Medias aritméticas
 f_{ij} : Frecuencias relativas del par (X, Y)

La covarianza se basa en el cálculo de las desviaciones de las observaciones de cada variable con respecto a su media, pero a diferencia de la varianza puede ser negativa. Es un indicador que detecta la relación lineal —valor absoluto alto— y si ésta es positiva o negativa, pero tiene problemas para distinguir relaciones no lineales de situaciones en las que no haya relación —en ambos casos se va a aproximar a cero—. También presenta el problema de venir acompañada de las unidades de las variables y de depender del número de observaciones.

II.2.2. Coeficiente de correlación

Debido a Pearson. Se define como sigue:

$$r = \frac{S_{XY}}{S_X \cdot S_Y}$$

S_{XY} : Covarianza del par (X, Y)
 S_X : Desviación típica de X
 S_Y : Desviación típica de Y

Supera algunas carencias de la covarianza, ya que es una medida adimensional y definida en $[-1, 1]$; su signo depende del de S_{xy} . Los extremos del intervalo implican que la relación lineal es perfecta, mientras que el cero significa que las variables están incorreladas.

II.2.3. Coeficiente chi-cuadrado

También debido a Pearson. Mide el grado de asociación entre variables cualitativas con h y k categorías, respectivamente, mediante la comparación de frecuencias esperadas y observadas:

$$\chi^2 = \sum_{i=1}^h \sum_{j=1}^k \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

o_{ij} : Frecuencias observadas o empíricas
 e_{ij} : Frecuencias esperadas o teóricas

Es una medida no acotada que siempre toma valores no negativos, y donde el cero indica independencia plena; aunque es de difícil interpretación por sí sola —de hecho, se puede utilizar como estadístico en la resolución de un contraste de hipótesis no paramétrico, como se verá más adelante—, cuanto más relacionadas estén las variables, más se alejará de cero, dado que más diferentes serán entre sí las frecuencias esperadas y las observadas. Es decir, si ambos tipos de frecuencias coinciden, es que las variables no tienen relación alguna entre sí.

III. AJUSTE Y REGRESIÓN BIDIMENSIONAL

A partir de una serie estadística procedente de una distribución bivariada se puede intentar encontrar una relación que exprese los valores de la variable dependiente Y en función de los de la independiente X . Para ello hay dos opciones:

- **Ajuste:** Prefijar una clase funcional (sea una recta, una parábola, etc.).
- **Regresión:** Estimar un valor de Y para cada valor de X .

La regresión viene determinada por un conjunto de puntos, de manera que uniendo los contiguos mediante segmentos rectilíneos se obtiene una poligonal de regresión. El proceso sólo tiene sentido cuando en la serie bidimensional hay muchos valores de Y para cada uno de X ; así, los puntos de la regresión se obtienen utilizando distribuciones condicionadas mediante el método de regresión a la media, que consiste en asociar a cada valor de X el valor medio de la distribución de Y condicionada a cada uno de dichos valores.

La poligonal de regresión puede servir a su vez de referencia de la clase funcional de ajuste, por lo que ambas técnicas se consideran complementarias. En el caso del ajuste, la cuestión es elegir la clase y determinar los parámetros específicos que identifiquen la función dentro de ella para que ésta se adapte de la mejor forma posible a la nube de puntos; para ello, se utiliza el método de los mínimos cuadrados.

III.1. Mínimos cuadrados

Fijada la clase funcional, se han de determinar los parámetros que caractericen a la función de ajuste; esto se consigue haciendo mínima la suma de las diferencias al cuadrado entre los valores observados y los valores ajustados.

III.1.1. Caso lineal

Sean $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ los valores observados y $f(x, a, b) = a + b \cdot x$ la recta de ajuste de los valores de Y en función de los de X , se obtienen a y b minimizando la función error cuadrático:

$$H(a, b) = \sum_{i=1}^n [y_i - (a + b \cdot x_i)]^2$$

Dada la siguiente notación:

$$y_i^* = f(x_i) = a + b \cdot x_i$$

$$e_i = y_i - y_i^*$$

y_i^* : Valor i -ésimo ajustado de Y

y_i : Valor i -ésimo observado de Y

Donde e_i representa el residuo o error de la observación (x_i, y_i) , tal que $\sum_{i=1}^n e_i = 0$, y admitiendo que se verifican las condiciones suficientes de mínimo, se tiene:

$$a = \bar{y} - b \cdot \bar{x}$$

$$b = \frac{S_{xy}}{S_x^2}$$

\bar{x}, \bar{y} : Medias aritméticas
 S_{XY} : Covarianza del par (X, Y)
 S_X^2 : Varianza de X

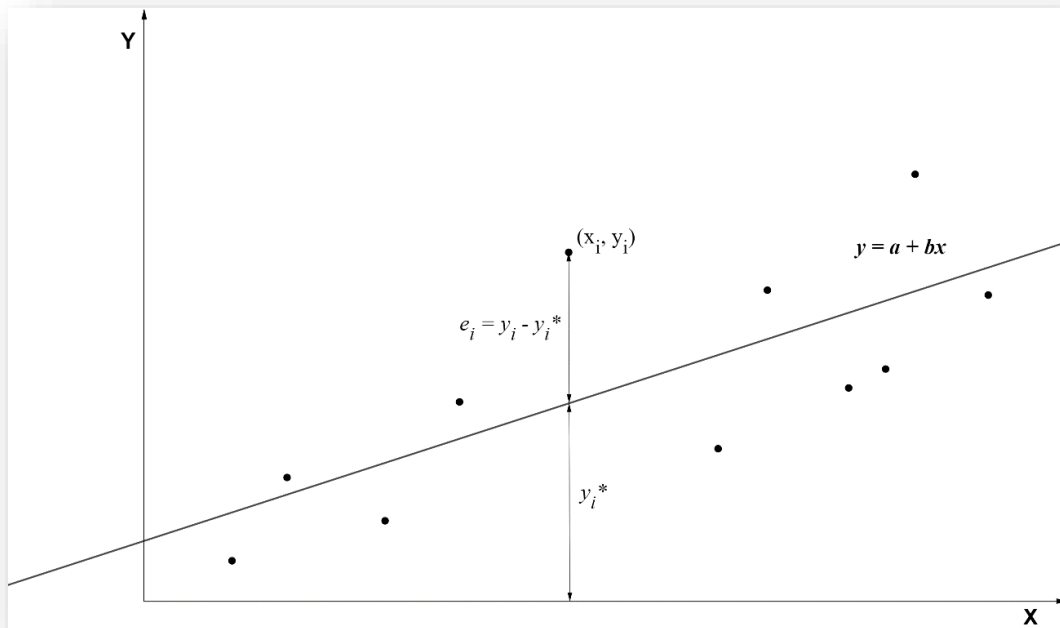


Figura III.1: Recta de ajuste

El valor b , que es la pendiente de la recta de ajuste, recibe el nombre de coeficiente de regresión y representa el incremento de Y para aumentos unitarios de X .

III.2. Bondad de ajuste

Para estimar el grado de aproximación entre los valores observados y los ajustados se recurre a la varianza residual:

$$S_e^2 = \frac{\sum_{i=1}^n e_i^2}{n} = \frac{\sum_{i=1}^n (y_i - y_i^*)^2}{n}$$

e_i : Residuo o error de la observación (x_i, y_i)

n : Número total de observaciones

y_i : Valor i -ésimo observado de Y

y_i^* : Valor i -ésimo ajustado de Y

Que coincide con el error cuadrático medio (ECM) y que es siempre positiva salvo que sea nula, supuesto en el que los valores observados coinciden con los ajustados y, por tanto, el ajuste es perfecto. En consecuencia, si se ajustan los mismos datos a dos modelos diferentes, el mejor será aquel que tenga un ECM menor.

III.2.1. Coeficiente de determinación

Dado que S_e^2 viene acompañada de la unidad de la variable no se pueden hacer comparaciones entre distribuciones con unidades diferentes, por lo que se ha de construir un coeficiente abstracto que permita expresar la bondad del ajuste en términos porcentuales: es el caso del coeficiente de determinación, R^2 .

Para el caso lineal equivale al cuadrado del coeficiente de correlación de Pearson, r , y da una medida de la relación lineal entre dos variables sin tener en cuenta el sentido de la relación, pero se puede redefinir de manera general como:

$$R^2 = \frac{S_{y^*}^2}{S_y^2} = \frac{S_y^2 - S_e^2}{S_y^2} = 1 - \frac{S_e^2}{S_y^2}$$

$S_{y^*}^2$: Varianza de los valores ajustados de Y

S_y^2 : Varianza de los valores observados de Y

S_e^2 : Varianza residual

R^2 varía entre 0 y 1 y multiplicado por 100 expresa el porcentaje de la variabilidad de Y que queda explicado por el ajuste; por tanto, un $R^2 = 1$ indica que el ajuste es perfecto.

IV. COMBINATORIA

El estudio de las diferentes formas en que se puede llevar a cabo la ordenación de objetos es muy importante en el recuento de casos favorables y posibles y, por tanto, en el cálculo de probabilidades.

IV.1. Variaciones

IV.1.1. Con repetición

Se tienen m elementos diferentes y se pretende que ocupen n lugares, de modo que cada elemento pueda ocupar más de un lugar:

$$VR_{m,n} = m^n$$

Ejemplo: El número de quinielas que es necesario hacer para acertar con seguridad el pleno al quince es $VR_{3,15}$.

IV.1.2. Sin repetición

Se tienen m elementos diferentes y se pretende que ocupen n lugares, siendo $n < m$, de modo que cada elemento sólo ocupe un lugar:

$$V_{m,n} = \frac{m!}{(m-n)!}$$

Ejemplo: En una carrera hípica con diez competidores, para asegurar el acierto de los tres primeros caballos son necesarios $V_{10,3}$ intentos.

IV.2. Permutaciones

IV.2.1. Sin repetición

Dan el número de ordenaciones distintas que se pueden hacer con n elementos; son equivalentes a las variaciones sin repetición de n elementos para que ocupen n lugares:

$$P_n = V_{n,n} = n!$$

Ejemplo: Si se quieren colocar siete libros en una estantería, se puede hacer de P_7 maneras.

IV.2.2. Con repetición

Dados n grupos de elementos a_i de distinto tipo, siendo m la suma de todos los elementos de todos los grupos, el número total de ordenaciones es:

$$PR_m^{a_1, a_2, \dots, a_n} = \frac{m!}{a_1! a_2! \dots a_n!}$$

Ejemplo: Si se desea repartir tres relojes, dos bicicletas y cuatro pelotas entre nueve niños de manera que cada uno de ellos reciba un regalo se tienen $PR_9^{3,2,4}$ maneras de hacerlo.

IV.3. Combinaciones

IV.3.1. Sin repetición

Si se pretende seleccionar n elementos de un total de m sin importar el lugar que ocupen, el número total de combinaciones sin repetición es:

$$C_{m,n} = \frac{V_{m,n}}{P_n} = \frac{m!}{(m-n)! n!}$$

Ejemplo: En una carrera de diez competidores en la que se clasifican los tres primeros, puede haber $C_{10,3}$ combinaciones de clasificados.

IV.3.2. Con repetición

Cada uno de los grupos que pueden formarse con r elementos elegidos entre n posibles sin importar que se repitan:

$$CR_{n,r} = \binom{n+r-1}{r} = \frac{(n+r-1)!}{r! (n-1)!}$$

Ejemplo: Si se quieren repartir tres bolas iguales entre cinco cajas diferentes, pudiendo estar todas dentro de la misma, se pueden hacer $CR_{5,3}$ combinaciones diferentes.

V. PROBABILIDAD

Los experimentos o fenómenos no determinísticos, en los que el conocimiento de las condiciones en que estos se desarrollan no garantiza los resultados, se denominan aleatorios y hacen imprescindible el uso de una función que asigne niveles de certidumbre, de probabilidad, a cada uno de los posibles desenlaces.

V.1. Conjuntos

Colección de elementos definida por su:

- **Extensión:** Qué elementos le pertenecen.
- **Descripción:** Qué propiedades han de cumplir dichos elementos.

Un conjunto admite subconjuntos.

V.1.1. Operaciones entre conjuntos

Para dos conjuntos A y B :

- **Unión:** Resulta en un conjunto C formado por todos los elementos de A y B . Se denota por \cup .
- **Intersección:** Resulta en un conjunto C formado por los elementos comunes de A y B . Se denota por \cap .

Son operaciones conmutativas y asociativas; además, se verifica la distributiva de cada operación respecto de la otra.

V.1.2. Conjuntos característicos

- **Conjunto vacío:** No tiene ningún elemento. Se considera subconjunto de cualquier conjunto. Se denota por \emptyset .
- **Conjunto universal:** Formado por todos los elementos del mismo tipo. Todo conjunto se puede considerar subconjunto de él. Se denota por Ω .
- **Conjuntos disjuntos:** No poseen elementos comunes. $A \cap B = \emptyset$.
- **Partición:** O sistema completo de sucesos. Agrupación de conjuntos disjuntos dos a dos y cuya unión equivale al conjunto universal.
- **Conjunto complementario:** Para A , conjunto formado por todos los elementos del conjunto universal que no están en A . Se denota por \bar{A} .
- **Diferencia de conjuntos:** Dados A y B , se puede definir mediante la intersección con el complementario del que se pretende restar, de tal forma que $A \setminus B = A \cap \bar{B}$.

V.2. Álgebra de sucesos

En un fenómeno aleatorio, bajo las mismas condiciones se obtienen resultados diferentes, pero, al igual que ocurre para uno determinista, su caracterización requiere la constatación de ciertas regularidades, que en este caso sólo aparecen al realizar un gran número de intentos. Si se prolonga indefinidamente el experimento, se llega al conjunto potencialmente infinito, denominado universo o población, de todas las pruebas asociadas a él.

La frecuencia de un suceso tiende a aproximarse a un valor fijo al aumentar el número de intentos; el valor límite ideal de estas frecuencias equivale a la probabilidad del suceso.

V.2.1. Espacio muestral

Conjunto de todos los resultados posibles de un experimento aleatorio. Se denota por Ω , dado que es un concepto análogo al del conjunto universal. Puede ser de tres tipos: finito, infinito numerable o continuo.

Se denomina suceso a todo subconjunto del espacio muestral; puede ser elemental, si consta de un solo elemento (obtener un número de 1 a 6 al lanzar un dado), o compuesto, si está formado por la unión de sucesos elementales (obtener pares o tríos con el mismo dado).

V.2.2. Relaciones entre sucesos

Son análogas a las que se establecen para los conjuntos, de manera que es posible definir la siguiente correspondencia:

Teoría de conjuntos	Cálculo de probabilidades
Conjunto universal	Suceso seguro (espacio muestral)
Punto del conjunto universal	Suceso elemental
Subconjunto	Suceso
Conjuntos disjuntos	Sucesos incompatibles
Unión de conjuntos	Unión de sucesos
Conjunto vacío	Suceso imposible
Conjunto complementario	Suceso contrario
Intersección de conjuntos	Intersección de sucesos
Partición	Sistema completo

V.3. Definición de probabilidad

Medida que cuantifica la incertidumbre de que se obtenga un determinado suceso al realizar un experimento aleatorio. Existen tres definiciones posibles:

- **Clásica:** Debida a Laplace. La probabilidad de un suceso A es el cociente entre el número de casos favorables al suceso y el número de casos posibles.
- **Frecuentista:** Debida a Bernoulli. La probabilidad de un suceso es el valor límite de su frecuencia relativa al repetir indefinidamente la experimentación.
- **Axiomática:** Debida a Kolmogórov. Dado un espacio medible (Ω, \mathcal{A}) , una función sobre \mathcal{A} tal que $P: \mathcal{A} \rightarrow \mathbb{R}$ se dice de probabilidad si cumple que $P(A) \geq 0$, $P(\Omega) = 1$ y $P(\cup_{i \in \mathbb{N}} A_i) = \sum_{i \in \mathbb{N}} P(A_i)$, siendo A_i sucesos incompatibles.

V.4. Propiedades de la función de probabilidad

Dados dos sucesos complementarios A y B (por ejemplo, que salga cara o que salga cruz al lanzar una moneda):

$$P(A \cup B) = P(\Omega) = 1$$

$$P(B) = 1 - P(A)$$

Dado que B es complementario de A , aquel se puede denotar como \bar{A} y reformular la anterior expresión de la siguiente manera:

$$P(\bar{A}) = 1 - P(A)$$

A su vez, sacar cara y sacar cruz simultáneamente es imposible, por lo que son sucesos disjuntos o incompatibles (mutuamente excluyentes):

$$P(A \cap B) = P(\emptyset) = 0$$

Por otro lado, si A y B son dos sucesos cualesquiera:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Nótese que, para sucesos incompatibles, la anterior expresión se simplifica, ya que la probabilidad de la intersección es nula:

$$P(A \cup B) = P(A) + P(B)$$

Por último, si A está incluido en B :

$$P(A) \leq P(B)$$

V.5. Probabilidad condicionada

La probabilidad de un determinado suceso puede verse condicionada por información previa; así, la probabilidad de que A ocurra si ocurre B (probabilidad de A dado B):

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

Si despejamos el numerador, y dada la conmutatividad de la intersección de conjuntos:

$$P(A \cap B) = P(A | B) \cdot P(B) = P(B | A) \cdot P(A) \quad [1]$$

Por lo que la primera expresión se puede reformular:

$$P(A | B) = \frac{P(B | A) \cdot P(A)}{P(B)} \quad [2]$$

V.5.1. Independencia

Dos sucesos son independientes si el resultado de uno no afecta al resultado de otro; es decir, la información previa existente no condiciona la probabilidad de ocurrencia:

$$P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \cdot P(B)}{P(B)} = P(A)$$

A modo de recordatorio, es interesante hacer notar que:

- Para sucesos incompatibles: $P(A \cup B) = P(A) + P(B)$
- Para sucesos independientes: $P(A \cap B) = P(A) \cdot P(B)$

V.5.2. Teorema de la probabilidad total

Para un experimento en dos etapas:

- En la primera, los posibles sucesos A_1, A_2, \dots, A_n constituyen un sistema completo —es decir, son incompatibles dos a dos y su unión equivale al espacio muestral— de tal forma que se conocen las probabilidades a priori $P(A_1), P(A_2), \dots, P(A_n)$ y son todas mayores a cero.
- En la segunda, los resultados posibles B_j tienen probabilidades desconocidas que dependen de lo ocurrido en la etapa precedente.

Si se conocen las probabilidades condicionadas con respecto a A_i de un cierto suceso B , se tiene que:

$$P(B) = \sum_{i=1}^n P(A_i \cap B) = \sum_{i=1}^n P(B | A_i) \cdot P(A_i)$$

Se observa su relación con [1].

V.5.3. Teorema de Bayes

De manera análoga, se puede determinar la probabilidad de un suceso A_k sabiendo que como resultado final del experimento se ha obtenido otro B :

$$P(A_k | B) = \frac{P(B | A_k) \cdot P(A_k)}{P(B)} = \frac{P(B | A_k) \cdot P(A_k)}{\sum_{i=1}^n P(B | A_i) \cdot P(A_i)}$$

Se observa su relación con [2].

VI. VARIABLE ALEATORIA

La abstracción cuantificada de los experimentos aleatorios se consigue asignando un número real a cada suceso del espacio muestral; esta correspondencia se realiza a través de una función llamada variable aleatoria (VA). Esta asociación no es única, por lo que la fijación de distintos valores puede dar lugar a una infinidad de funciones.

Así, para un experimento aleatorio consistente en lanzar una moneda dos veces al aire, el espacio muestral consta de cuatro posibles resultados:

$$\Omega = \{CC, CX, XC, XX\}$$

C: Cara

X: Cruz

Si definimos la VA discreta X “número de caras”, se observa que a cada suceso del experimento le corresponde uno de tres posibles valores (ya que se pueden obtener dos caras, una o ninguna), que conforman el rango de X :

$$R_X = \{0, 1, 2\}$$

Por otro lado, dado que cada suceso tiene una probabilidad de ocurrencia, existe una distribución de probabilidad asociada al conjunto de posibles valores x_i :

x_i	$p(x_i)$
0	$1/4 = 0,25$
1	$2/4 = 0,50$
2	$1/4 = 0,25$

VI.1. Caracterización de variables aleatorias

Una VA es discreta si toma valores aislados, como en el caso de la moneda, y continua si toma cualquier valor dentro de un intervalo (la estatura, por ejemplo), por lo que la probabilidad de que tome exactamente un valor es 0.

Si la VA toma muchos valores diferentes no es factible darlos todos ellos con sus probabilidades de una forma explícita y se recurre a una serie de funciones:

- *De cuantía:* para VA discretas
- *De densidad:* para VA continuas
- *De distribución:* para ambos tipos

VI.1.1. Caso discreto

La función de cuantía (o de probabilidad) es la regla que asocia a cada valor x_i que toma la VA su probabilidad de ocurrencia p_i , de tal manera que:

$$p_i = P(X = x_i)$$
$$\sum_{i=1}^{n/\infty} p_i = 1$$

Para el ejemplo de la moneda:

$$f(x) = \begin{cases} 0,25, & x = 0 \\ 0,50, & x = 1 \\ 0,25, & x = 2 \end{cases}$$

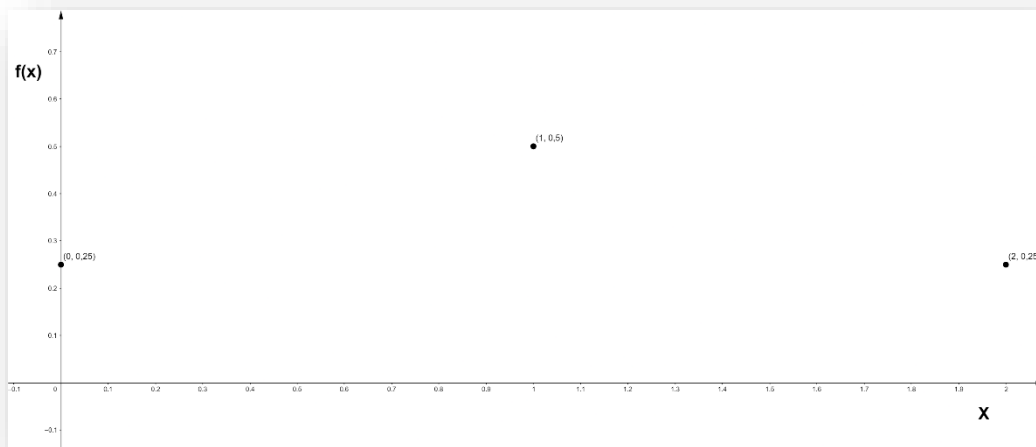


Figura VI.1: Función de cuantía de una VA discreta

Así como la función de cuantía es $f(x) = P(X = x)$, la de distribución es $F(x) = P(X \leq x)$, y asigna a cada valor de la VA la probabilidad de que X sea menor o igual que ese valor; es decir, acumula toda la probabilidad entre $-\infty$ y el punto donde está definida.

Para el ejemplo de la moneda:

$$F(x) = \begin{cases} 0, & x < 0 \\ 0,25, & 0 \leq x < 1 \\ 0,75, & 1 \leq x < 2 \\ 1, & x \geq 2 \end{cases}$$

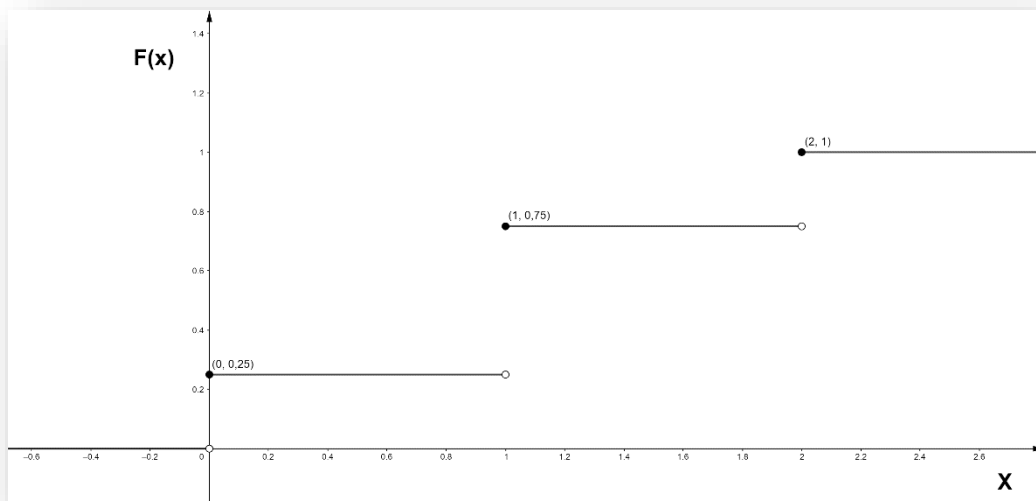


Figura VI.2: Función de distribución de una VA discreta

Sus propiedades son:

- Está definida en toda la recta real

- Es no decreciente y no negativa
- Toma el valor 0 en $-\infty$ y el 1 en $+\infty$
- Tiene discontinuidades de salto en los puntos donde la función de cuantía es distinta de 0

VI.1.2. Caso continuo

A pesar de que la probabilidad de que la VA tome un determinado valor es cero, no todos ellos tienen la misma frecuencia (la probabilidad de medir alrededor de 170 cm es mucho mayor que la de medir alrededor de 210 cm). Por tanto, se puede estimar la probabilidad de que la VA tome valores dentro de un intervalo de amplitud Δ , que será pequeña pero no nula —evidentemente, la cantidad de probabilidad que cabe en un intervalo depende de la amplitud del mismo—:

$$[x - \Delta/2, x + \Delta/2]$$

Si se levanta sobre el intervalo un rectángulo de altura $h(x)$ tal que su área sea igual a la probabilidad antes mencionada:

$$A = h \cdot \Delta = P(x - \Delta/2 \leq X \leq x + \Delta/2)$$

Se puede asociar dicha altura a una densidad de probabilidad —de manera análoga al concepto de densidad física— por ser igual al cociente entre la probabilidad y la amplitud del intervalo:

$$h = \frac{P(x - \Delta/2 \leq X \leq x + \Delta/2)}{\Delta}$$

Es importante no confundir la densidad de probabilidad, que es la masa de probabilidad por unidad de intervalo, con la probabilidad en sí misma.

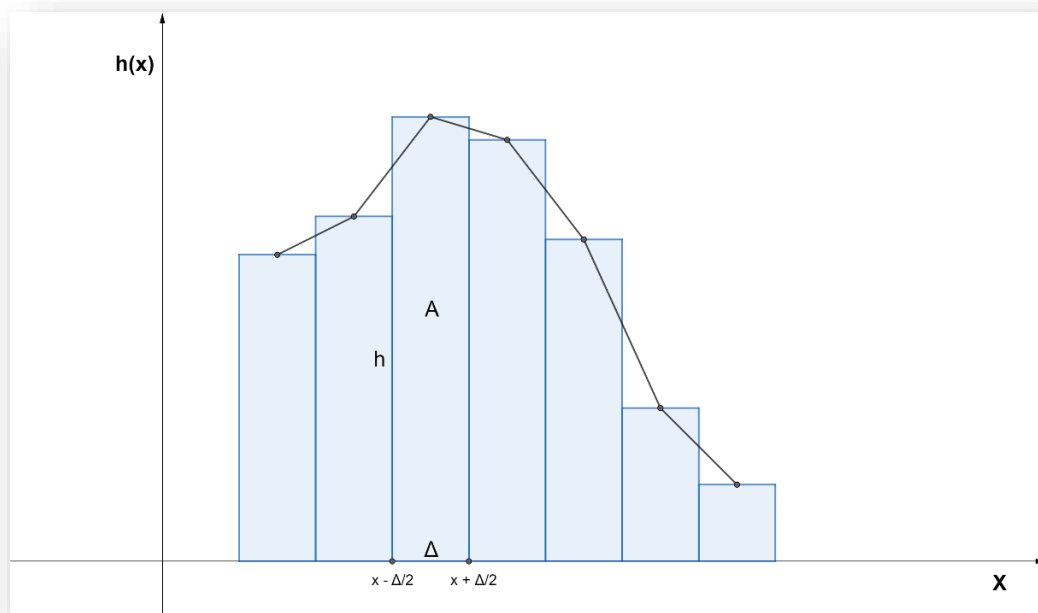


Figura VI.3: Histograma de frecuencias

Si se reproduce este proceso para varios intervalos se obtiene un histograma, y si se unen los centros de las caras superiores de los rectángulos, un polígono de

frecuencias. La suma de todas las áreas ha de ser 1, ya que equivale a la probabilidad de que X tome cualquier valor.

Si se hace tender a cero la anchura de los intervalos, el polígono de frecuencias se transforma en una curva, representación gráfica de la función de densidad de la VA:

$$\lim_{\Delta \rightarrow 0} h(x) = f(x)$$

La probabilidad de que la VA se encuentre entre a y b equivale al área bajo la curva entre a y b ; es decir:

$$P(a \leq X \leq b) = \int_a^b f(x) \cdot dx$$

$$P(-\infty \leq X \leq +\infty) = \int_{-\infty}^{+\infty} f(x) \cdot dx = 1$$

Cualquier función no negativa que verifique la anterior condición será una función de densidad.

Por otro lado, y en relación con lo que se acaba de exponer, la función de distribución de la VA, que da la probabilidad de que X tome valores menores o iguales a x , se obtiene por integración entre $-\infty$ y x de la función de densidad:

$$F(X) = P(X \leq x) = \int_{-\infty}^x f(x) \cdot dx$$

Sus propiedades son idénticas a las del caso discreto, salvo que no presenta discontinuidades de salto, sino que es continua en todo \mathbb{R} . Además, dado que para una VA continua $P(X = x) = 0$, $P(X \leq x) = P(X < x)$ —y viceversa—.

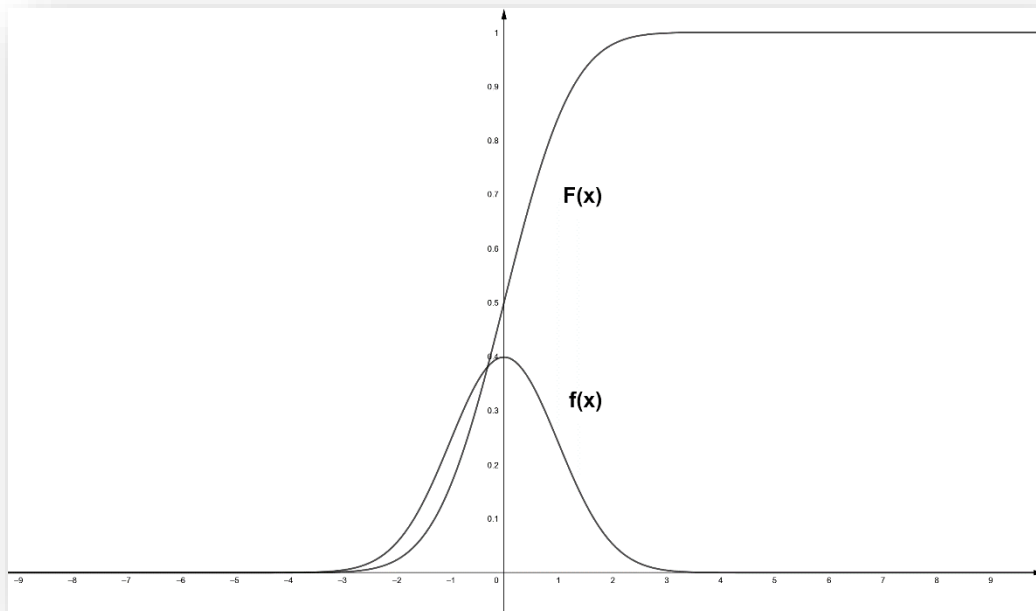


Figura VI.4: Funciones de densidad y distribución para una normal tipo

VI.2. Esperanza matemática

Se define la esperanza matemática o valor esperado de la VA discreta X como:

$$E[X] = \sum_{i \geq 1} x_i \cdot p_i$$

Mientras que para el caso continuo:

$$E[X] = \int_{-\infty}^{+\infty} x \cdot f(x) \cdot dx$$

La esperanza coincide con la media poblacional, μ , por lo que a partir de ella se puede definir la varianza poblacional como el valor esperado de la diferencia cuadrática entre la VA y su valor esperado:

$$V[X] = \sigma^2 = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

En general, a partir de la función esperanza se pueden definir los momentos de orden k de una VA respecto al origen y la media:

$$\begin{aligned}\alpha_k &= E[X^k] \\ \mu_k &= E[(X - E[X])^k]\end{aligned}$$

Se observa que el momento de orden 1 respecto al origen es μ y de orden 2 respecto a la media es σ^2 .

VII. MODELOS PROBABILÍSTICOS

La identificación del patrón probabilístico por el cual se rige un experimento aleatorio facilita enormemente su estudio y caracterización. Así, en función del tipo de variable aleatoria se tienen las siguientes distribuciones:

- **Discretas:** Binomial y de Poisson
- **Continuas:** Exponencial, gamma (caso general de la exponencial), uniforme, normal y de Weibull

VII.1. Distribución binomial

Se habla de experimento de Bernoulli cuando la realización de pruebas repetidas e independientes tiene únicamente dos posibles resultados: E (éxito) y F (fracaso), con probabilidades de ocurrencia invariantes p y q , respectivamente.

La VA discreta Y que se asocia al experimento toma los valores 1 para E y 0 para F , por lo que su rango es $R_Y = \{0, 1\}$, y su función de probabilidad:

$$P(Y = y) = \begin{cases} p, & y = 1 \\ q = 1 - p, & y = 0 \end{cases}$$

Se dice que Y sigue una distribución de Bernoulli de parámetro p :

$$Y \sim B(p)$$

$$E[Y] = \mu_Y = p$$

$$V[Y] = \sigma_Y^2 = p \cdot (1 - p)$$

Por otro lado, la distribución de probabilidad por la cual se rige el número de éxitos k al realizar n pruebas de Bernoulli independientes y con probabilidades de éxito iguales se denomina binomial.

Por tanto, se puede definir una VA discreta X de rango $R_X = \{0, 1, \dots, n\}$ asociada a dicho experimento como suma de variables aleatorias independientes de Bernoulli, de tal forma que:

$$X = \sum_{i=1}^n Y_i$$

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}$$

$$X \sim B(n, p)$$

$$E[X] = \mu_X = n \cdot p$$

$$V[X] = \sigma_X^2 = n \cdot p \cdot (1 - p)$$

Es interesante hacer notar que para $n = 1$ la distribución binomial se convierte en la de Bernoulli. Así, el lanzamiento de una moneda al aire constituye un experimento de Bernoulli, mientras que su repetición 15 veces permite definir una VA discreta “número de caras” que sigue una distribución binomial: $X \sim B(15, 0.5)$.

VII.2. Distribución de Poisson

Un proceso de Poisson implica la ocurrencia de un determinado suceso en un soporte continuo (que puede ser tanto espacial como temporal) con cierta

estabilidad para una unidad prefijada de dicho soporte y de forma independiente; por ejemplo, el número de defectos por metro cuadrado de una pieza de tela o el número de coches que pasan por un semáforo durante un intervalo de tiempo. En todos los casos, el número de ocurrencias se denota por k .

La distribución de Poisson, de carácter discreto, se obtiene como límite de la binomial cuando n tiende a infinito, p tiende a cero y el promedio de éxitos, $n \cdot p$, tiende a estabilizarse alrededor de un número, λ , que caracteriza a la distribución:

$$\begin{aligned} X &\sim P(\lambda) \\ R_X &= \{0, 1, 2, \dots\} \\ P(X = k) &= \frac{\lambda^k}{k!} \cdot e^{-\lambda} \\ E[X] = \mu_X = V[X] = \sigma_X^2 &= \lambda \end{aligned}$$

En esta distribución, la mayor parte de la masa de probabilidad queda repartida entre un número pequeño de valores, por lo que también se llama de los sucesos raros.

Lo dicho anteriormente está referido a una unidad de soporte de tamaño 1; si se generaliza para un tamaño t , la función de cuantía queda como sigue:

$$P_t(X = k) = e^{-\lambda \cdot t} \cdot \frac{(\lambda \cdot t)^k}{k!}$$

VII.3. Distribución exponencial

A partir de un proceso de Poisson en el que el soporte sea temporal, se puede definir la VA X “tiempo transcurrido entre dos eventos consecutivos”, que es continua y tiene un rango $R_X = [0, +\infty)$.

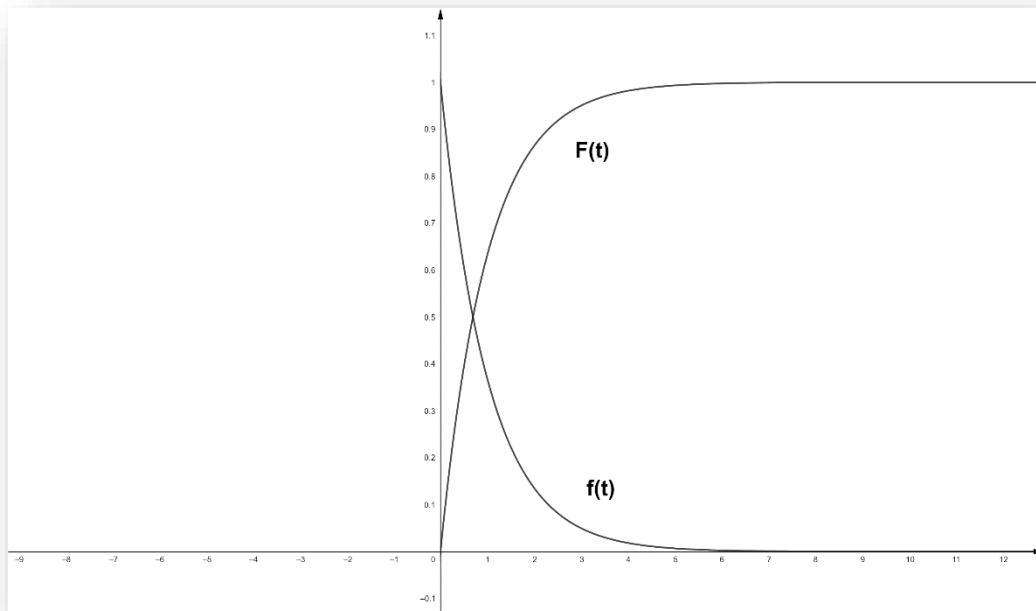


Figura VII.1: Funciones de densidad y distribución para la exponencial

La probabilidad de que X supere dicho tiempo es equivalente a la de que no suceda nada (es decir, no haya eventos: $k = 0$) entre 0 (momento de ocurrencia del primer suceso) y t (del segundo), por lo que dada la función de cuantía generalizada de Poisson:

$$P(X > t) = P(Y = 0) = e^{-\lambda \cdot t}$$

Se puede definir así una función de distribución:

$$F(t) = P(X \leq t) = 1 - e^{-\lambda \cdot t}$$

De la que por derivación se obtiene una de densidad:

$$f(t) = \lambda \cdot e^{-\lambda \cdot t}$$

Se dice, por tanto, que la VA sigue una distribución exponencial:

$$X \sim E(\lambda)$$

$$E[X] = \mu_X = \frac{1}{\lambda}$$

$$V[X] = \sigma_X^2 = \frac{1}{\lambda^2}$$

VII.4. Distribución gamma

Una VA X sigue una distribución gamma, $X \sim \Gamma(a, p)$, si su función de densidad es:

$$f(x) = \begin{cases} \frac{a^p}{\Gamma(p)} \cdot e^{-a \cdot x} \cdot x^{p-1}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

Para $p > 0$, $\Gamma(p) = \int_0^{+\infty} x^{p-1} \cdot e^{-x} \cdot dx$. Por otro lado:

$$E[X] = \mu_X = \frac{p}{a}$$

$$V[X] = \sigma_X^2 = \frac{p}{a^2}$$

El caso particular $\Gamma(\lambda, 1)$ equivale a $E(\lambda)$.

VII.5. Distribución uniforme

Una VA X de tipo continuo sigue una distribución de tipo uniforme (o rectangular) si toma valores en un intervalo acotado por a y b , de tal manera que su función de densidad, $f(x)$, muestra el aspecto de la figura VII.2.

Es decir, todos los valores dentro de ese intervalo tienen la misma probabilidad, derivada del hecho de que la frecuencia de distribución es constante.

En resumen:

$$X \sim U(a, b)$$

$$R_X = [a, b]$$

$$f(x) = \frac{1}{b - a}$$

$$E[X] = \mu_X = \frac{a + b}{2}$$

$$V[X] = \sigma_X^2 = \frac{(b-a)^2}{12}$$

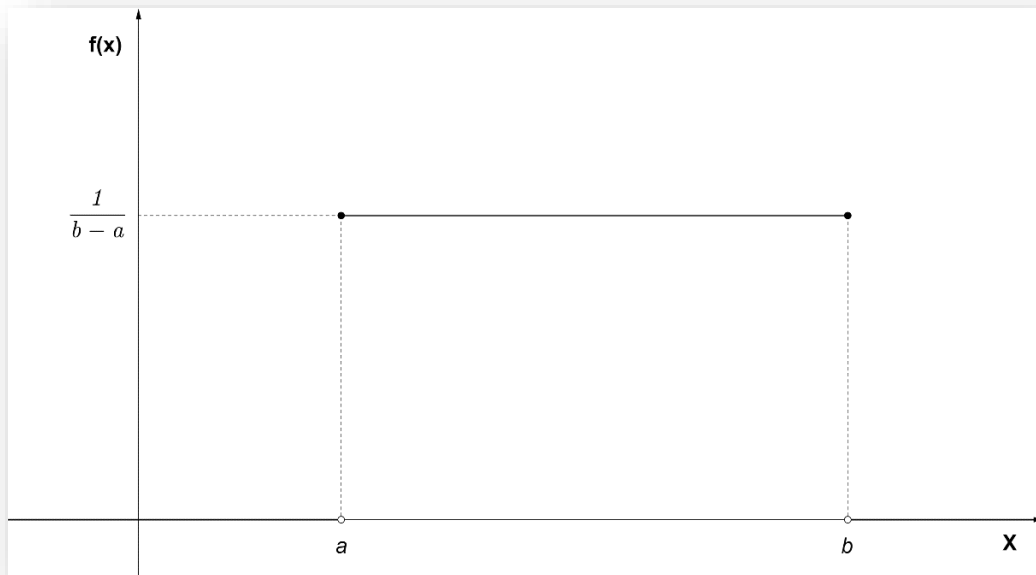


Figura VII.2: Función de densidad para una distribución uniforme

VII.6. Distribución normal

Es la distribución más importante del cálculo de probabilidades, por las siguientes razones:

- La gran mayoría de variables aleatorias de interés físico siguen una distribución normal.
- A través del teorema central del límite, otras distribuciones se pueden aproximar a una normal.
- Los errores aleatorios en los resultados de medida se suelen distribuir según una normal.
- Es la base de la inferencia estadística.

Una VA continua X sigue una distribución normal si su función de densidad es:

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot e^{-\frac{(x-\mu)^2}{2 \cdot \sigma^2}}$$

$$R_X = (-\infty, +\infty)$$

Esta distribución está caracterizada por dos parámetros: su media —que coincide con la mediana y la moda— y su desviación típica:

$$X \sim N(\mu, \sigma)$$

La función de densidad es asintótica respecto del eje de abscisas, simétrica respecto de $x = \mu$, se extiende por toda la recta real y es tanto más achatada cuanto mayor es su desviación típica —más dispersos están los datos—.

Como la mayoría de los valores se arremolina en torno a la media, se cumple que:

- El 68 % de ellos se concentran en el intervalo $[\mu - \sigma, \mu + \sigma]$
- El 95,5 %, en el $[\mu - 2 \cdot \sigma, \mu + 2 \cdot \sigma]$

- El 99,7 %, en el $[\mu - 3 \cdot \sigma, \mu + 3 \cdot \sigma]$

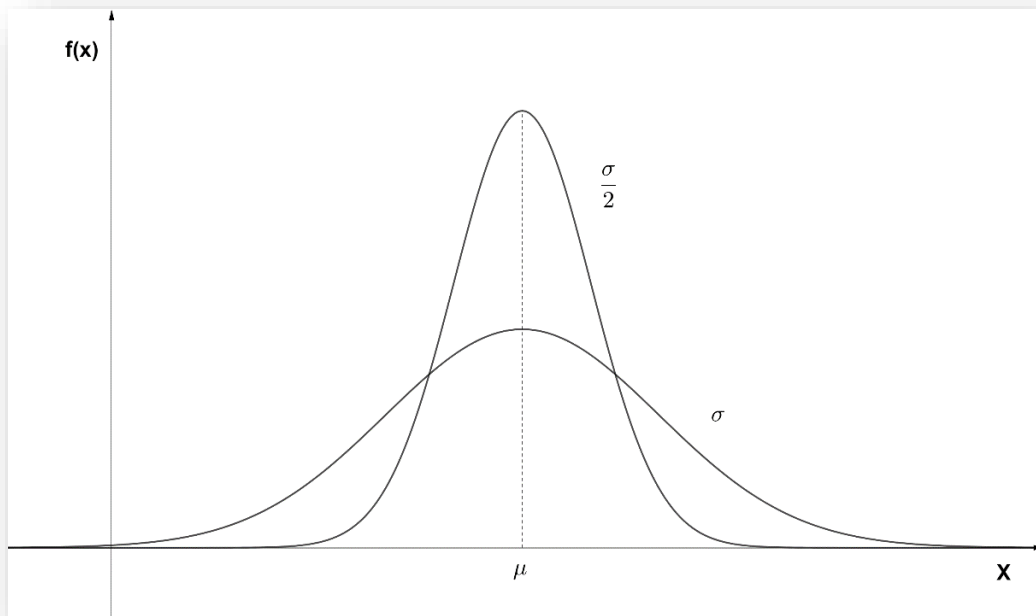


Figura VII.3: Función de densidad de la distribución normal

VII.6.1. Normal tipo

Para facilitar el cálculo se recurre a la normal tipo, $N(0, 1)$, que mediante un cambio de variable se puede generalizar a $N(\mu, \sigma)$; si $Z \sim N(0, 1)$ y $X \sim N(\mu, \sigma)$, se verifica:

$$Z = \frac{X - \mu}{\sigma}$$

$$X = \sigma \cdot Z + \mu$$

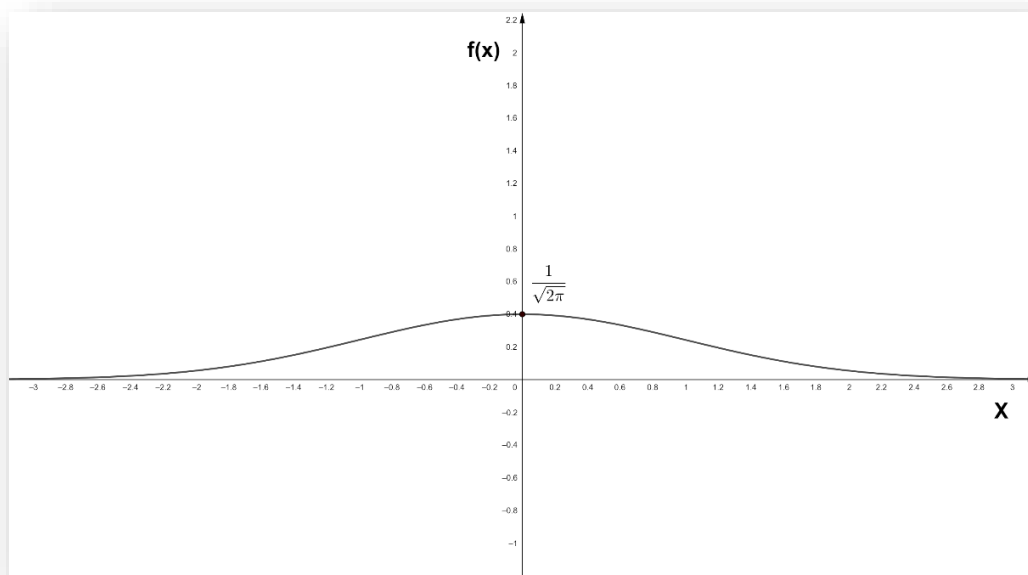


Figura VII.4: Función de densidad de la distribución normal tipo

La función de densidad de la normal tipo es:

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

$$R_X = (-\infty, +\infty)$$

Es simétrica respecto al eje OY , lo que determina que su máximo sea el único punto de corte con los ejes.

VII.6.2. Teorema central del límite

Sean X_1, X_2, \dots, X_n variables aleatorias independientes con medias μ_i , desviaciones típicas σ_i y distribuciones cualesquiera; si se define una VA $Y = X_1 + X_2 + \dots + X_n$, cuando n tiende a infinito (en la práctica, $n \geq 30$), se tiene que:

$$Y \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$$

Esto significa que cuando un experimento aleatorio es afectado por un amplio conjunto de causas independientes que actúan sumando sus efectos, es esperable que su resultado siga una distribución normal.

Este teorema permite aproximar otras distribuciones, difíciles de calcular cuando n es muy grande, a una normal; por ejemplo, una binomial, que surge del sumatorio de variables aleatorias independientes de Bernoulli, puede ser candidata si cumple ciertos requisitos, y lo mismo ocurre con una Poisson:

- Dada $X \sim B(n, p)$, $Z \sim N(0, 1)$ con $Z = (X - n \cdot p) / \sqrt{n \cdot p \cdot q}$ si $n \cdot p > 5$ y $p \leq 0,5$ o $n \cdot q > 5$ y $p > 0,5$.
- Dada $X \sim P(\lambda)$, $X \sim N(\lambda, \lambda^{1/2})$ si $\lambda > 5$

VII.6.3. Distribuciones derivadas de la normal

VII.6.3.1. Distribución chi-cuadrado

Dadas n variables aleatorias independientes Z_i distribuidas según normales tipo, se tiene que $X = \sum_{i=1}^n Z_i^2$ es una VA que sigue una distribución chi-cuadrado con n grados de libertad, denotada χ_n^2 , cuya función de densidad, de forma similar a la normal pero sesgada a la derecha, es:

$$f(x) = \frac{1}{2^{\frac{n}{2}} \cdot \Gamma\left(\frac{n}{2}\right)} \cdot x^{\frac{n}{2}-1} \cdot e^{-\frac{x}{2}}$$

$$R_X = [0, +\infty)$$

Se trata de un caso particular de la distribución gamma, $\Gamma(1/2, n/2)$, con $E[X] = \mu_X = n$ y $V[X] = \sigma_X^2 = 2 \cdot n$.

VII.6.3.2. Distribución t de Student

Dadas dos variables aleatorias independientes $X \sim N(0, 1)$ e $Y \sim \chi_n^2$, la VA $Z = X / \sqrt{Y/n}$ sigue una distribución t de Student con n grados de libertad, denotada t_n , cuya función de densidad, simétrica respecto al origen, es:

$$f(z) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n \cdot \pi} \cdot \Gamma\left(\frac{n}{2}\right)} \cdot \left(1 + \frac{z^2}{n}\right)^{-\frac{n+1}{2}}$$

$$R_Z = (-\infty, +\infty)$$

Se verifica que $E[Z] = \mu_Z = 0$ y $V[Z] = \sigma_Z^2 = n/(n-2)$ para $n > 2$.

VII.6.3.3. Distribución \mathcal{F} de Fisher

Se define como el cociente entre dos VVAA chi-cuadrado divididas por sus grados de libertad:

$$\mathcal{F}_{n,m} = \frac{\frac{\chi_n^2}{n}}{\frac{\chi_m^2}{m}}$$

Su forma es la de la chi-cuadrado y sus momentos más importantes son:

$$E[\mathcal{F}_{n,m}] = \frac{m}{m-2}$$

$$V[\mathcal{F}_{n,m}] = \frac{2 \cdot m^2 \cdot (m+n-2)}{n \cdot (m-2)^2 \cdot (m-4)}$$

La t de Student al cuadrado es un caso particular de la \mathcal{F} de Fisher: $t_n^2 \sim \mathcal{F}_{1,n}$.

VII.7. Distribución de Weibull

Se dice que una VA X sigue una distribución de Weibull con parámetro de forma $k > 0$ y de escala $\lambda > 0$, denotada por $W(k, \lambda)$, si su función de densidad es:

$$f(x) = \begin{cases} \frac{k}{\lambda} \cdot \left(\frac{x}{\lambda}\right)^{k-1} \cdot e^{-\left(\frac{x}{\lambda}\right)^k}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

$$E[X] = \mu_X = \lambda \cdot \Gamma\left(1 + \frac{1}{k}\right)$$

$$V[X] = \sigma_X^2 = \lambda^2 \cdot \left[\Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right) \right]$$

Esta función modela la distribución de fallos cuando su tasa es proporcional a una potencia del tiempo:

- $k < 1$ indica que la tasa de fallos decrece con el tiempo
- $k = 1$, que es constante
- $k > 1$, que es creciente

VIII. INFERENCIA ESTADÍSTICA

La inferencia estadística estudia grandes colectivos —poblaciones— a partir de pequeñas porciones de estos —muestras—. Para que la proyección realizada sea de calidad:

- La muestra ha de ser representativa de la población; lo contrario conlleva la aparición de sesgos y errores sistemáticos.
- A través de ella se han de alcanzar unos objetivos de precisión fijados de antemano; cuanto mayor haya de ser el nivel de precisión, mayor habrá de ser el tamaño muestral.

VIII.1. Procedimientos inferenciales

La población está representada por una variable aleatoria con una determinada distribución de probabilidad; en función del grado de conocimiento que se tenga de ella, la inferencia puede ser:

- **Paramétrica:** Se limita a determinar los parámetros de los que depende la distribución (por ejemplo, la media y la desviación típica en el caso de una normal) para caracterizar la población. Se puede seguir un enfoque:
 - *Clásico:* O frecuentista, con parámetros constantes
 - *Bayesiano:* Los parámetros se consideran variables aleatorias
- **No paramétrica:** No asume ninguna distribución sino hipótesis más generales, como la de simetría. Existen tres tipos de procedimientos:
 - Estudio de los parámetros de localización de la distribución
 - Análisis de las condiciones de estructura de la distribución
 - Comprobación de las hipótesis exigibles a los valores muestrales (independencia, ausencia de valores atípicos, etc.)

VIII.2. Obtención de la información

Todo estudio inferencial requiere, en primer lugar, un diseño muestral en el que se garanticen los niveles de representatividad y precisión deseados; éste conduce a una muestra potencial, a partir de la cual se obtienen a su vez una o varias muestras por observación, medición o encuestación.

La inferencia, y en consecuencia el muestreo, puede deberse a diversos motivos, generalmente relacionados con las limitaciones económicas o incluso físicas que supone analizar la población al completo.

VIII.2.1. Muestreo aleatorio

Desde el punto de vista estadístico, sólo un muestreo aleatorio tiene interés, puesto que, a diferencia de lo que ocurre en uno por conveniencia o comodidad, en uno probabilístico se pueden determinar los errores que afectarán al proceso inferencial.

Existen las siguientes variedades:

- **Simple con reposición:** Todas las unidades poblacionales tienen la misma probabilidad de pertenecer a la muestra y es posible medir varias veces al

mismo individuo. Las variables aleatorias derivadas son independientes e idénticamente distribuidas.

- **Simple sin reposición:** Similar al anterior, pero un mismo individuo no puede ser seleccionado varias veces. Las variables aleatorias que componen la muestra no son independientes.
- **Estratificado:** Se delimitan subpoblaciones homogéneas internamente, y la composición de la muestra se distribuye entre ellas mediante afijación:
 - *Uniforme:* Mismo número de representantes de cada estrato
 - *Proporcional:* Número de representantes de cada estrato proporcional a su tamaño
 - *Óptima:* La asignación se hace teniendo en cuenta tanto el tamaño como la variabilidad de los estratos
- **Por áreas:** Se establecen grupos de elementos físicamente próximos entre ellos; es decir, a diferencia del caso anterior, las subpoblaciones son homogéneas entre ellas pero heterogéneas internamente.

En cualquier caso, e independientemente de la metodología seguida, la última etapa de cualquier muestreo probabilístico supone siempre la realización de un muestreo simple aleatorio, lo cual implica que:

- Las variables tendrán igual distribución de probabilidad que la población de la que se ha extraído la muestra.
- Las variables serán independientes.

IX. ESTIMACIÓN PUNTUAL

En una situación paramétrica, se tiene conocimiento de la distribución de probabilidad que rige la población en estudio pero no de los parámetros que la caracterizan; es decir, se considera una población con función de distribución $F_{\underline{\theta}}(x)$ donde $\underline{\theta} \in \mathbb{R}^k$ es un vector de parámetros desconocidos. La estimación puntual consiste en determinar de la manera más precisa posible esos parámetros a través de valores numéricos.

IX.1. Estadísticos

La obtención de valores próximos a los que verdaderamente definen los parámetros de la distribución poblacional implica resumir la información muestral, lo que se puede conseguir a través de un estimador.

Un estadístico (o estimador) $T(\underline{X})$ es una función —y por tanto una VA con su propia distribución— de las variables muestrales que no depende de parámetros desconocidos; el objetivo es que el estimador resuma la muestra sin pérdida de información relevante que \underline{X} pueda contener acerca de los parámetros.

IX.2. Estimación en poblaciones normales

Dado que en inferencia estadística es usual asumir la normalidad de las poblaciones bajo estudio, a esta distribución se le da un trato preferente; así, dada $X \sim N(\mu, \sigma)$, los estimadores de los parámetros μ y σ son, respectivamente, la media y la desviación típica muestrales \bar{X} y S .

Por otro lado, muchos estadísticos siguen distribuciones derivadas de la normal, como la chi-cuadrado o la t de Student, que se obtienen como combinación de variables normales —cuyo número determina los grados de libertad—.

IX.2.1. Distribución de la media muestral

\bar{X} es una combinación lineal de VVAA normales y, por tanto, se distribuye siguiendo una normal:

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Siendo sus momentos:

$$\begin{aligned} E[\bar{X}] &= \mu \\ V[\bar{X}] &= \frac{\sigma^2}{n} \end{aligned}$$

Lo anterior también sería cierto para una VA no normal siempre que n sea suficientemente grande, según el teorema central del límite.

IX.2.2. Distribución de la varianza muestral

La relación entre media y varianza muestral viene dada por el teorema de Fisher, según el cual las VVAA \bar{X} y S^2 son independientes y el estadístico $n \cdot S^2 / \sigma^2$

sigue una distribución chi-cuadrado con $n - 1$ grados de libertad, lo que también aplica en el caso de la cuasivarianza muestral (S_c^2):

$$\frac{(n-1) \cdot S_c^2}{\sigma^2} \sim \chi_{n-1}^2$$

De los momentos de una chi-cuadrado se puede deducir:

$$E[S^2] = \frac{n-1}{n} \cdot \sigma^2$$

Esto indica que S^2 no es un estimador insesgado de σ^2 , por lo que en la mayoría de casos se utiliza la cuasivarianza muestral (o varianza muestral corregida) como estimador de la varianza poblacional, ya que:

$$E[S_c^2] = \sigma^2$$

Por otro lado:

$$V[S^2] = \frac{2 \cdot (n-1)}{n^2} \cdot \sigma^4$$

$$V[S_c^2] = \frac{2}{n-1} \cdot \sigma^4$$

IX.2.3. Distribución de la proporción muestral

Dada $X = \sum_{i=1}^n Y_i \sim B(n, p)$, siendo $Y_i \sim B(p)$, la proporción —o probabilidad según la definición clásica— de éxitos resulta ser:

$$\hat{p} = \frac{\sum_{i=1}^n Y_i}{n}$$

Es decir, el número de éxitos entre el número total de casos.

Este estimador se denomina proporción muestral y sigue una binomial, si bien para $n \geq 30$ por el teorema central del límite se cumple:

$$\hat{p} \sim N\left(p, \sqrt{\frac{p \cdot (1-p)}{n}}\right)$$

Y sus momentos son:

$$E[\hat{p}] = p$$

$$V[\hat{p}] = \frac{p \cdot (1-p)}{n}$$

IX.3. Propiedades de los estadísticos

En principio, existen infinitas formas de definir funciones a partir de elementos de la muestra para estimar un parámetro, por lo que habrá que escoger los mejores estadísticos posibles; para ello, existen cuatro propiedades que sirven de guía:

IX.3.1. Insesgadez

Dado un parámetro θ , θ^* es un estimador insesgado si:

$$E[\theta^*] = \theta$$

Si no lo es, su sesgo se puede determinar de la siguiente manera:

$$Sesgo(\theta^*) = E[\theta^*] - \theta$$

$$R_{sesgo} = (-\infty, +\infty)$$

Un estimador es asintóticamente insesgado si el sesgo tiende a cero cuando n tiende a infinito.

Como ya se reseñó anteriormente, la varianza muestral es un estimador sesgado de la varianza poblacional, y por eso se recurre a la cuasivarianza, dado que:

$$Sesgo(S^2) = \frac{-\sigma^2}{n}$$

$$Sesgo(S_c^2) = 0$$

IX.3.2. Eficiencia

Es la inversa de la varianza del estimador:

$$Eficiencia(\theta^*) = \frac{1}{V[\theta^*]}$$

Por tanto, se dice que un estimador es más eficiente que otro si su varianza es menor.

IX.3.3. Error cuadrático medio

Dados dos estimadores tales que el sesgo del primero es menor que el del segundo, pero no así sus varianzas, la elección de uno de ellos puede hacerse a través del *ECM*, que considera ambas propiedades en conjunto:

$$ECM(\theta^*) = E[(\theta^* - \theta)^2] = V[\theta^*] + (Sesgo(\theta^*))^2$$

A menor *ECM*, mejor estimador.

Como también se dijo antes, en general se prefiere la cuasivarianza a la varianza por su menor sesgo, pero no siempre, debido a que su *ECM* es mayor.

IX.3.4. Consistencia

Un estimador se dice consistente cuando a medida que se aumenta el tamaño muestral, más se aproxima al parámetro desconocido y más pequeña se hace su varianza.

X. INTERVALOS DE CONFIANZA

Al estimar de forma puntual un parámetro es necesario a su vez obtener una medida de la fiabilidad de dicha estimación; es decir, calcular una región que contenga el valor del parámetro con una cierta garantía: un intervalo de confianza.

Considérense los siguientes puntos:

- \underline{X} es una MAS procedente de una población definida por una VA X cuya distribución depende de un parámetro desconocido θ .
- $\underline{\theta}(\underline{X})$ y $\bar{\theta}(\underline{X})$ son VVAA que dependen de la muestra y acotan inferior y superiormente a θ .
- Se define el nivel de confianza γ como el grado de certeza (o probabilidad, generalmente entre 0,90 y 0,99) con el que se quiere estimar θ y el nivel de significación α como la diferencia entre la certeza (es decir, uno) y el nivel de confianza, de tal forma que $\gamma = 1 - \alpha$.

Se dice que $[\underline{\theta}(\underline{X}), \bar{\theta}(\underline{X})]$ es un intervalo de confianza de nivel $1 - \alpha$ si cumple la siguiente relación de probabilidad:

$$P[\underline{\theta}(\underline{X}) \leq \theta \leq \bar{\theta}(\underline{X})] \geq 1 - \alpha$$

Por tanto, dado que θ es, aunque desconocido, un valor constante, la probabilidad de que el intervalo aleatorio $[\underline{\theta}(\underline{X}), \bar{\theta}(\underline{X})]$ lo contenga es al menos $1 - \alpha$.

El objetivo es determinar, cumpliendo unos criterios de calidad, los extremos del intervalo, que toman valores puntuales para una realización muestral determinada. Una vez obtenida la muestra, el intervalo se vuelve fijo, por lo que no tiene sentido hablar de probabilidad: el valor pertenecerá o no a él; es decir, lo hará con probabilidad uno o cero. De ahí que se hable de confianza al considerar una muestra concreta: si γ es de 0,95, esto no significa que la probabilidad de encontrar el parámetro de la población entre los márgenes del intervalo sea del 95 %, sino que se prevé que de cada 100 realizaciones muestrales, el intervalo concreto contenga al parámetro para 95 de ellas.

Puede hablarse también de intervalos de confianza acotados sólo inferior o superiormente:

$$P[\underline{\theta}(\underline{X}) \leq \theta] \geq 1 - \alpha$$

$$P[\bar{\theta}(\underline{X}) \geq \theta] \geq 1 - \alpha$$

X.1. Longitud del intervalo

Al ampliar la longitud de un intervalo aumenta su nivel de confianza; si se considera el intervalo $(-\infty, +\infty)$, γ , obviamente, pasa a ser uno.

Por otro lado, para un nivel de confianza prefijado se comprueba que no existe un único intervalo, por lo que se puede elegir aquel que cumpla determinadas características prefijadas; desde un punto de vista práctico, la opción más interesante es el intervalo de longitud mínima o, si no es posible, aquel que minimice la expresión:

$$E[\bar{\theta}(\underline{X}) - \underline{\theta}(\underline{X})]$$

Si esto tampoco es posible, se recurre al reparto equitativo del nivel de significación entre ambas colas:

$$P[\underline{\theta}(\underline{X}) \geq \theta] = \frac{\alpha}{2}$$

$$P[\bar{\theta}(\underline{X}) \leq \theta] = \frac{\alpha}{2}$$

Ese criterio conduce a un intervalo único y, en el caso de distribución simétrica respecto de θ , de longitud mínima.

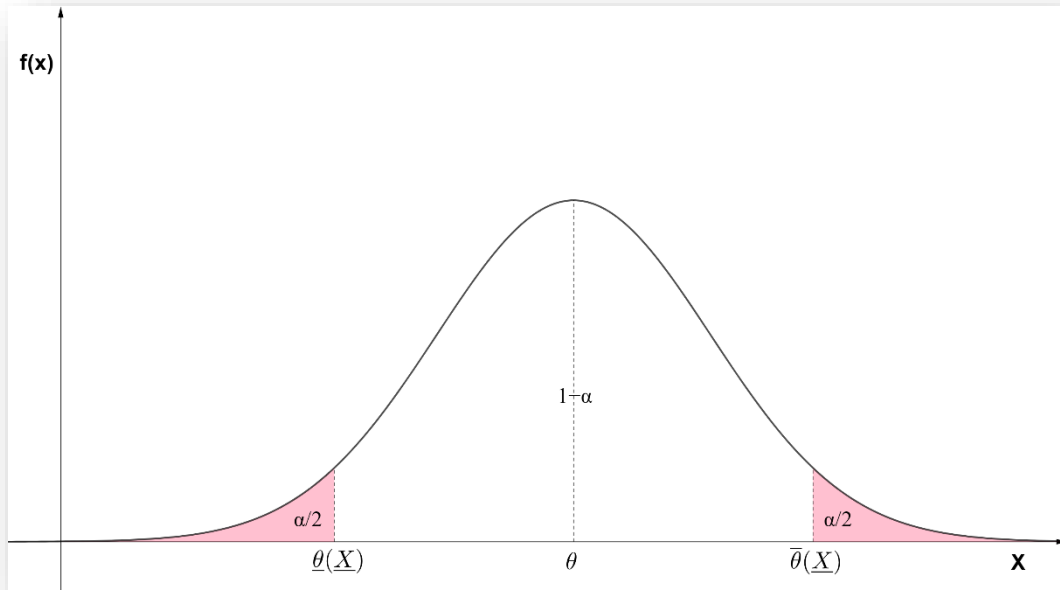


Figura X.1: Colas de una distribución simétrica respecto del parámetro

X.2. Método del pivote

Dada una MAS \underline{X} procedente de una población definida por una VA X cuya distribución depende de un parámetro desconocido θ , este método permite calcular intervalos de confianza a partir de una función de la muestra que contenga al parámetro pero cuya distribución no dependa de él.

- En primer lugar, se elige el mejor estimador puntual posible para θ , que es función de la MAS: $\hat{\theta} = T(\underline{X})$.
- A partir del estimador, se determina una función pivotal $T(\underline{X}, \theta)$, monótona para todo valor muestral \underline{x} , cuya distribución es conocida e independiente de θ : $h(\hat{\theta})$.
- Se buscan las cotas $\underline{\theta}(\underline{X})$ y $\bar{\theta}(\underline{X})$, $\lambda_1(\alpha)$ y $\lambda_2(\alpha)$, denotadas λ_1 y λ_2 por comodidad, tales que $P(\lambda_1 \leq h(\hat{\theta}) \leq \lambda_2) = 1 - \alpha$.

X.3. Intervalos en poblaciones normales

En todos los casos, se asume una población definida por una VA $X \sim N(\mu, \sigma)$ de la cual se extrae una MAS \underline{X} de tamaño n .

X.3.1. Determinación de la media conocida la varianza

Dado que se quiere determinar μ y se sabe que $\bar{X} \sim N(\mu, \sigma/\sqrt{n})$, se elige como pivote la tipificación de la media muestral:

$$T(\underline{X}, \theta) = \sqrt{n} \cdot \frac{(\bar{X} - \mu)}{\sigma} \sim N(0, 1)$$

Por tanto, para un nivel de confianza $1 - \alpha$ y una VA $Z \sim N(0, 1)$, se pretende encontrar las cotas $\lambda_1 \equiv \lambda_1(\alpha)$ y $\lambda_2 \equiv \lambda_2(\alpha)$ tales que:

$$P(\lambda_1 \leq Z \leq \lambda_2) = 1 - \alpha$$

Si α_1 y α_2 representan el reparto de α entre ambas colas, λ_1 y λ_2 se obtienen a partir de las siguientes igualdades:

$$P(Z \leq \lambda_1) = \alpha_1$$

$$P(Z \geq \lambda_2) = \alpha_2$$

Para cada elección de α_1 y α_2 tales que $\alpha_1 + \alpha_2 = \alpha$ se obtiene un intervalo diferente, por lo que siempre que sea posible se ha de elegir el de longitud mínima para este pivote; dado que la normal es simétrica respecto de su media, el intervalo más pequeño se obtiene mediante el reparto equitativo de α entre ambas colas:

$$I_{1-\alpha}(\mu) = \left[\bar{X} - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right]$$

X.3.2. Determinación de la media desconocida la varianza

Si se asume que la media y la cuasivarianza muestrales son independientes, dadas:

$$\frac{\sqrt{n} \cdot (\bar{X} - \mu)}{\sigma} \sim N(0, 1)$$
$$\frac{(n-1) \cdot S_c^2}{\sigma^2} \sim \chi_{n-1}^2$$

Se tiene:

$$\frac{\sqrt{n} \cdot (\bar{X} - \mu)}{S_c} \sim t_{n-1}$$

Usando esta función como pivote se llega al siguiente intervalo de confianza:

$$I_{1-\alpha}(\mu) = \left[\bar{X} - t_{n-1, 1-\frac{\alpha}{2}} \cdot \frac{S_c}{\sqrt{n}}, \bar{X} + t_{n-1, 1-\frac{\alpha}{2}} \cdot \frac{S_c}{\sqrt{n}} \right]$$

X.3.3. Determinación de la varianza conocida la media

Dadas n VVAA $(X_i - \mu)/\sigma \sim N(0, 1)$ independientes dos a dos, se tiene:

$$\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi_n^2$$

Usando esta función como pivote se llega al siguiente intervalo de confianza:

$$I_{1-\alpha}(\sigma^2) = \left[\frac{n \cdot S_\mu^2}{\chi_{n, 1-\frac{\alpha}{2}}^2}, \frac{n \cdot S_\mu^2}{\chi_{n, \frac{\alpha}{2}}^2} \right]$$

Siendo:

$$S_{\mu}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

X.3.4. Determinación de la varianza desconocida la media

Por el teorema de Fisher se tiene:

$$\sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma^2} \sim \chi_{n-1}^2$$

Usando esta función como pivote se llega al siguiente intervalo de confianza:

$$I_{1-\alpha}(\sigma^2) = \left[\frac{(n-1) \cdot S_c^2}{\chi_{n-1,1-\frac{\alpha}{2}}^2}, \frac{(n-1) \cdot S_c^2}{\chi_{n-1,1+\frac{\alpha}{2}}^2} \right]$$

Que para el caso de la varianza muestral queda de la siguiente manera:

$$I_{1-\alpha}(\sigma^2) = \left[\frac{n \cdot S^2}{\chi_{n-1,1-\frac{\alpha}{2}}^2}, \frac{n \cdot S^2}{\chi_{n-1,1+\frac{\alpha}{2}}^2} \right]$$

X.4. Método asintótico

En aquellos casos en que se quiera construir el intervalo de confianza para μ pero se desconozca la distribución de \bar{X} , si se dispone de una MAS de gran tamaño ($n \geq 30$), existe la posibilidad de hacerlo de manera asintótica a través del TCL en su forma de Lindeberg-Levy:

Dadas X_1, \dots, X_n VVAA independientes e idénticamente distribuidas que, por un proceso de convergencia en distribución, confluyen en una VA X tal que $E[X] = \mu$ y $V[X] = \sigma^2 < \infty$, se tiene:

$$\sqrt{n} \cdot \frac{(\bar{X} - \mu)}{\sigma} \xrightarrow{d} N(0, 1)$$

Por tanto, la metodología para obtener el intervalo será la misma que ya se ha detallado anteriormente, con la salvedad de que el nivel de confianza no será exacto sino aproximado.

X.4.1. Determinación de la proporción

Dado que para una MAS extraída de una población definida por $X \sim B(p)$ se tiene:

$$\hat{p} = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}$$

Aplicando el TCL en su forma LL se deduce que:

$$\frac{\hat{p} - p}{\sqrt{\frac{p \cdot (1-p)}{n}}} \xrightarrow{d} N(0, 1)$$

A partir de este resultado se puede construir un intervalo de confianza sustituyendo $p \cdot (1-p)$ por su estimador $\hat{p} \cdot (1-\hat{p})$:

$$I_{1-\alpha}(p) = \left[\hat{p} \pm \frac{k}{\sqrt{n}} \cdot \sqrt{\hat{p} \cdot (1 - \hat{p})} \right] = \left[\hat{p} \pm Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \right]$$

X.5. Determinación del tamaño muestral

Se puede fijar el tamaño de muestra n para que el error cometido en el proceso de estimación de los intervalos de confianza no supere un umbral ε determinado de antemano:

$$n \geq \frac{Z_{1-\frac{\alpha}{2}}^2 \cdot \hat{p} \cdot (1 - \hat{p})}{\varepsilon^2}$$

XI. CONTRASTE DE HIPÓTESIS

El contraste de hipótesis es la técnica más utilizada para tratar problemas de inferencia estadística.

Una hipótesis estadística es una conjetura sobre alguna característica o sobre la distribución de una o más VVAA; hay que diferenciar:

- **Hipótesis nula:** Es la que se desea contrastar. Se denota por H_0 .
- **Hipótesis alternativa:** Es la que se acepta cuando la evidencia muestral está claramente en contra de la hipótesis nula. Se denota por H_1 .

Un contraste de hipótesis supone la partición del espacio muestral en dos regiones, una de aceptación y otra de rechazo (o crítica), de manera que si la muestra considerada se encuentra en la segunda se rechaza H_0 , mientras que en caso contrario no, sin que ello certifique su veracidad: simplemente no hay evidencia suficiente para rechazarla.

El contraste de hipótesis es una regla de decisión-confirmación por cuanto se da más crédito a H_0 que a H_1 , por lo que sólo se ha de contrastar aquello sobre lo que se tenga una justificada sospecha de certeza.

XI.1. Clasificación de los contrastes

Si la distribución, salvo los parámetros, es conocida se trata de un contraste paramétrico, por cuanto el objetivo es obtener información sobre dichos parámetros desconocidos; si ni siquiera la distribución es conocida, se trata de un contraste no paramétrico, en el que se persigue determinar alguna característica de la población bajo estudio; por ejemplo, si se distribuye de forma normal o si los elementos que la componen son independientes.

Es preferible recurrir siempre que se pueda a los primeros, dado que al utilizar más información ofrecen mejores resultados.

XI.1.1. Contrastes paramétricos

Puede ser bilaterales, en los que se propone un valor puntual para el parámetro:

$$\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta \neq \theta_0 \end{cases}$$

En este caso, la región crítica también es bilateral, por lo que existen dos puntos críticos que la separan de la región de aceptación. Esto se debe a que, aunque el contraste se realice para un valor concreto, el rechazo no ocurre automáticamente si se obtiene cualquier otro valor distinto al esperado, sino que existe un margen de tolerancia a ambos lados; es decir, la región de aceptación no es puntual.

También pueden ser unilaterales, en los que se propone que el valor del parámetro se encuentre por encima (por la izquierda) o por debajo (por la derecha) de un determinado umbral:

$$\begin{cases} H_0: \theta \geq \theta_0 \\ H_1: \theta < \theta_0 \end{cases} \quad \begin{cases} H_0: \theta \leq \theta_0 \\ H_1: \theta > \theta_0 \end{cases}$$

En este caso, la región crítica es unilateral y está separada de la de aceptación por un único punto crítico, que debido al margen de tolerancia previamente comentado no tiene por qué coincidir con el valor propuesto.

Es importante remarcar que en todos los casos el signo igual está en la hipótesis nula.

XI.2. Errores de un contraste

De la decisión de rechazar o no la hipótesis nula se derivan varias consecuencias:

- Que se haya hecho lo correcto: rechazar H_0 siendo falsa o no rechazarla siendo cierta.
- Cometer un error de tipo 1: rechazar H_0 siendo cierta.
- Cometer un error de tipo 2: no rechazar H_0 siendo falsa.

Estos resultados, trasladados a términos probabilísticos, dan lugar a los siguientes conceptos (todos referidos al contraste):

- **Significación:** $\alpha = P(\text{rechazar } H_0 | H_0 \text{ cierta})$; es decir, la probabilidad de cometer un error de tipo 1.
- **Confianza:** $1 - \alpha = P(\text{no rechazar } H_0 | H_0 \text{ cierta})$
- **Potencia:** $1 - \beta = P(\text{rechazar } H_0 | H_0 \text{ falsa})$
- **Riesgo:** $\beta = P(\text{no rechazar } H_0 | H_0 \text{ falsa})$; es decir, la probabilidad de cometer un error de tipo 2.

Se han de minimizar las probabilidades de error, si bien éstas son complementarias —cuando disminuye la de tipo 1 aumenta la de tipo 2 y viceversa— por lo que en la práctica se fija α y se elige de entre todos los contrastes posibles aquel que haga mínimo el riesgo o —lo que es lo mismo— máxima la potencia.

XI.3. Realización de contrastes paramétricos

Una vez conocida la estructura de la población bajo estudio, Fisher propone la siguiente metodología (particularizada para un contraste unilateral a la derecha):

1. Definir la hipótesis nula H_0 a contrastar sobre el parámetro desconocido θ , lo que a su vez determinará la hipótesis alternativa H_1 .
2. Dar una medida, $d(\hat{\theta}(\underline{X}), \theta_0)$, de la discrepancia entre la evidencia muestral, representada por el estimador de θ , $\hat{\theta}(\underline{X})$, y la hipótesis nula, de tal forma que d tenga una distribución conocida cuando H_0 sea cierta.
3. Tomar una MAS, obtener $\hat{\theta}(\underline{x})$ y calcular $d_0 = d(\hat{\theta}(\underline{x}), \theta_0)$.
4. Si d_0 es muy grande, es decir, si $P(d \geq d_0) < \alpha$, se rechaza H_0 , mientras que si $P(d \geq d_0) > \alpha$, no se rechaza.

El nivel de significación es el que determina la región de aceptación y la de rechazo, para lo que se calcula un valor d_c —el punto crítico— tal que:

$$P(d > d_c | H_0) > \alpha$$

De forma que si $d_0 > d_c$ se rechaza H_0 y viceversa.

XI.3.1. Nivel crítico

Se llama nivel crítico o p-valor del contraste a la siguiente probabilidad:

$$p = P(d > d_0 | H_0)$$

Así:

- Cuando $p \geq \alpha$ no se rechaza H_0 .
- Cuando $p < \alpha$ se rechaza H_0 .

Además, el nivel crítico proporciona información adicional sobre las garantías con que se rechaza o no la hipótesis nula:

- Si p está próximo a α hay que ser prudente en la decisión adoptada y recurrir si es posible a la recolección de más información muestral.
- Si la distancia entre ambos es grande, la decisión quedará respaldada.

XII. CONTRASTES NO PARAMÉTRICOS

Se trata de una serie de técnicas inferenciales que permiten:

- Estudiar una población cuando se desconoce la distribución que sigue.
- Comprobar que se satisfacen las hipótesis exigidas para llevar a cabo un contraste paramétrico.

XII.1. Análisis de la estructura de la población

Se realiza mediante contrastes de bondad de ajuste, que permiten determinar si la suposición que se hace acerca del modo en que se distribuye una VA es correcta: si sigue una binomial, una Poisson, una exponencial o, con especial interés, de cara a poder aplicar posteriormente un contraste de hipótesis paramétrico, una normal.

XII.1.1. Contraste chi-cuadrado de Pearson

Permite contrastar, tanto para VVAA discretas como continuas, si una muestra sigue una determinada distribución de probabilidad mediante el análisis de la discrepancia entre las frecuencias empíricas u observadas y las que se deberían haber obtenido —teóricas o esperadas— si efectivamente siguiera esa distribución.

Si se desconocen m parámetros de los que caracterizan a la distribución teórica, se pueden calcular a partir de la muestra. Se recomienda esta prueba para realizaciones de al menos 25 elementos, a agrupar en k clases que han de cubrir todos los resultados posibles —en el caso continuo, toda la recta real— y cuya frecuencia esperada ha de ser al menos 5 —en caso contrario, se ha de reagrupar—.

Se calcula un estadístico que incorpora las diferencias entre las frecuencias observadas y las esperadas —a menores diferencias, mejor ajuste— y que sigue una distribución χ^2 con $k - m - 1$ grados de libertad:

$$\sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i} \sim \chi_{k-m-1}^2$$

o_i : Frecuencias observadas o empíricas
 e_i : Frecuencias esperadas o teóricas

Aunque están obviamente relacionadas, no se debe confundir esta prueba de bondad de ajuste con la prueba de independencia entre variables expuesta en el apartado 2.3. del tema II.

XII.1.2. Contraste de Kolmogorov-Smirnov

Esta prueba sólo es aplicable a distribuciones continuas —no sería de utilidad para una binomial, por ejemplo— completamente especificadas; es decir, de las que se conocen sus parámetros —por tanto, no serviría para una exponencial cualquiera, sino para una de λ concreto—. Por contra, es aplicable a muestras pequeñas y no requiere agrupar —ni reagrupar— los valores en clases, por lo que no implica pérdida de información.

El estadístico obtenido dispone de su propia tabla y, en caso de que se pretenda comprobar de manera específica el ajuste a una normal, se recomienda utilizar la corrección de Lilliefors.

Aunque la prueba Kolmogorov-Smirnov-Lilliefors es muy sensible a las diferencias entre la muestra y la distribución teórica alrededor de sus valores medios, le cuesta detectar diferencias prominentes en los extremos —confundiría una t de Student con una normal, por ejemplo—, algo que resuelve el contraste de Anderson-Darling; sin embargo, a su vez, ambos se comportan mal con muestras grandes, problema que soluciona la prueba de Shapiro-Wilk.

XII.1.3. Contraste de Shapiro-Wilk

Permite determinar de forma específica si la población se adapta a alguna estructura de probabilidad normal.

Para una realización muestral x_1, x_2, \dots, x_n :

1. Se ordenan los elementos de mayor a menor: $x_{(n)}, x_{(n-1)}, \dots, x_{(1)}$.
2. Se calculan las diferencias $x_{(n-i+1)} - x_{(i)}$ entre los valores que equidistan del centro ($x_{(n)} - x_{(1)}, x_{(n-1)} - x_{(2)}$, y así sucesivamente) y se multiplican por los coeficientes tabulados a_{n-i+1} (a_n, a_{n-1} , etc.); la suma de los productos se denota por:

$$b = \sum_{i=1}^k a_{n-i+1} \cdot ((x_{(n-i+1)} - x_{(i)}))$$

$k = n/2$ si n es par y $k = (n - 1)/2$ si n es impar.

3. Se define un estadístico experimental tal que:

$$W_{exp} = \frac{b^2}{(n-1) \cdot S_C^2} \leq 1$$

S_C^2 : Cuasivarianza muestral

El estadístico vale 1 en caso de que la muestra sea una réplica de una normal.

4. Se cae en la región crítica si:

$$W_{exp} \leq W_{n,\alpha}$$

Donde $W_{n,\alpha}$ está tabulado.

XII. 2. Contrastes de localización y escala

Si a través de una prueba de bondad de ajuste se concluye que una muestra no sigue una distribución normal, no será posible aplicarle un contraste paramétrico; sin embargo, aún se pueden realizar hipótesis sobre otras características de relevancia, como son las medidas de posición.

En particular, en este tipo de contrastes no paramétricos se analiza la situación de los elementos de la muestra con respecto a la mediana: su signo —si están por encima o por debajo de ella— y/o su distancia —el rango que ocupa el elemento en la secuencia ordenada de diferencias—.

Hay que tener en cuenta que la mediana poblacional, para una VA continua, es el único valor que divide la distribución en dos partes equiprobables:

$$P(X \geq Me) = P(X \leq Me) = \frac{1}{2}$$

$$F(Me) = 0,5$$

Haciendo uso de la mediana muestral como estimador, bajo la hipótesis nula, el número de observaciones en la muestra mayores y menores que Me_0 debe ser aproximadamente el mismo; si se confirma que la estructura de probabilidad de la población es simétrica, entonces el contraste puede ser aplicado a la media.

XII.2.1. Contraste T de Wilcoxon

También se le conoce como prueba de los rangos. Es aplicable a distribuciones continuas en las que se desea contrastar lo siguiente:

$$\begin{cases} H_0: Me = Me_0 \\ H_1: Me \neq Me_0 \end{cases}$$

El procedimiento es el siguiente:

1. Para cada elemento de la muestra, se calcula $|x_i - Me_0|$ y se consignan todas aquellas diferencias no nulas, cuyo número se denota por n' .
2. Se asignan rangos para las diferencias calculadas, desde 1 hasta n' . Las que coincidan en valor absoluto, si se da el caso, se agrupan bajo el rango promedio. Por ejemplo, si existen tres diferencias equivalentes a la unidad que suponen los rangos 1, 2 y 3, se agrupan todas bajo el rango 2,0.
3. Tras asignar rangos, se suman los correspondientes a diferencias negativas, por debajo de Me_0 , y los correspondientes a las positivas, por encima; los respectivos sumatorios, T^- y T^+ , conforman el siguiente estadístico experimental:

$$T_{exp} = \min(T^-, T^+)$$

4. Se cae en la región crítica si:

$$T_{exp} \leq T_{n', \frac{\alpha}{2}}$$

Donde $T_{n', \alpha/2}$ está tabulado.

Algunas consideraciones:

- El contraste T supone una alternativa al uso de la t de Student cuando no se puede suponer normalidad de la muestra; comprobada la simetría, también se puede realizar de manera análoga con respecto a la media, asumiendo como hipótesis nula $\mu = \mu_0$.
- Para muestras grandes ($n \geq 25$), bajo el supuesto de que H_0 es cierta, se cumple:

$$T^- \sim N\left(\frac{n \cdot (n+1)}{4}, \sqrt{\frac{n \cdot (n+1) \cdot (2 \cdot n+1)}{24}}\right)$$

- Para contrastes de tipo $H_0: Me \geq Me_0$, la región crítica viene dada por $T^+ \leq T_{n', \alpha}$, mientras que para $H_0: Me \leq Me_0$ por $T^- \leq T_{n', \alpha}$.