



Grado en Ingeniería Informática *Estadística*

Curso 2011/2012

Dpto. Estadística e I.O.
Universidad de Granada

TEMA 1. ESTADÍSTICA DESCRIPTIVA UNIDIMENSIONAL

INTRODUCCIÓN: CONCEPTOS BÁSICOS

La observación de fenómenos que acontecen en la vida real permiten establecer una clasificación de los mismos:

Fenómeno determinista: Un fenómeno es **determinista** si al repetirlo en idénticas condiciones se obtiene el mismo resultado.

Fenómeno aleatorio: Un fenómeno es **aleatorio** si al repetirlo en análogas condiciones puede presentar resultados diferentes.

La estadística se ocupa principalmente de los fenómenos aleatorios, encontrándose ante un conjunto de observaciones que presentan una variabilidad difícil de explicar y que requieren un tratamiento especial ("tratamiento estadístico") para poder efectuar conclusiones. Por lo tanto, la **estadística** es una rama de las matemáticas que trata de la recopilación, el análisis, la interpretación y la representación de una gran cantidad de datos numéricos.

Las etapas de un estudio estadístico son las siguientes:

- | | | |
|--------------------------------------|---|---------------------------|
| 1. Recogida de datos | } | → Estadística descriptiva |
| 2. Ordenación, tabulación y gráficos | | |
| 3. Descripción de características | | |
| 4. Análisis formal | } | → Inferencia estadística |

Definición

Se denomina **población** al conjunto objeto de estudio, es decir, cualquier conjunto de unidades con ciertas características comunes, sobre las que se desea información.

Definición

Cada uno de los elementos de la población se denomina **unidad estadística** o **individuo**.

La población puede ser finita o infinita, según que los elementos que la formen se presenten en número finito o infinito.

Definición

Se denomina **muestra** a un subconjunto representativo de la población.

Definición

Se llaman **caracteres** a las propiedades que se desean observar en los elementos de la población y que han de tener todos y cada uno de ellos.

En un estudio particular pueden considerarse una sola característica o varias a la vez.

Definición

Las **modalidades** son cada una de las formas en que puede presentarse un carácter.

Para estar bien definidas deben cumplir dos requisitos: exhaustividad e incompatibilidad.

Modalidades exhaustivas: Se dice que las modalidades de un carácter son **exhaustivas** si cubren todas las posibles formas en que éste se manifiesta.

Modalidades incompatibles: Se dice que las modalidades de un carácter son **incompatibles** cuando cada individuo solo puede presentar una de las modalidades.

Clasificación de caracteres según las modalidades:

Cuantitativos: Un carácter es cuantitativo cuando sus modalidades son medibles numéricamente. Los caracteres cuantitativos se denominan también **variables estadísticas**. Se subdividen en dos grupos:

- **Variables estadísticas discretas:** Son aquellas que tienen un número finito o infinito numerable de modalidades. Las modalidades son valores aislados.
- **Variables estadísticas continuas:** El número de modalidades es no numerable. Las posibles modalidades son todos los valores de un intervalo.

Cualitativos: Un carácter es cualitativo cuando sus modalidades no son medibles numéricamente. Un carácter cualitativo recibe también el nombre de **atributo**.

DISTRIBUCIÓN DE FRECUENCIAS

Definición

La *distribución de frecuencias de una variable estadística* es el conjunto de valores ordenados de la variable con sus frecuencias correspondientes.

Formalmente se representa por el conjunto de pares ordenados.

Variable cualitativa

$$\{(M_i; n_i)\}_{i=1}^k$$

o

$$\{(M_i; f_i)\}_{i=1}^k$$

Variable cuantitativa

Discreta

$$\{(x_i; n_i)\}_{i=1}^k$$

o

$$\{(x_i; f_i)\}_{i=1}^k$$

Continua

$$\{(I_i; n_i)\}_{i=1}^k$$

o

$$\{(I_i; f_i)\}_{i=1}^k$$

- M_i : cada una de las modalidades de una variable cualitativa.
- x_i : cada uno de los valores numéricos que puede tomar una variable estadística discreta.
- I_i : cada uno de los intervalos que constituyen las modalidades de una variable estadística continua, considerando que $I_i = (e_{i-1}; e_i]$, siendo e_{i-1} y e_i los extremos inferior y superior, respectivamente, del intervalo.

Se considera un carácter X con k modalidades, x_1, x_2, \dots, x_k . Las frecuencias asociadas a la modalidad x_i son:

Definición

Frecuencia absoluta (n_i): Número de individuos de la población que presentan dicha modalidad, es decir, el número de veces que se repite. Como las modalidades deben ser incompatibles y exhaustivas se verifica que

$$N = \sum_{i=1}^k n_i$$

siendo N el número total de observaciones.

Definición

Frecuencia relativa (f_i): Proporción de individuos de la población que presentan dicha modalidad. Es decir, $f_i = \frac{n_i}{N}$. Se verifica que $\sum_{i=1}^k f_i = 1$

Definición

Frecuencia absoluta acumulada (N_i): Número de individuos que presentan un valor de la variable menor o igual que el considerado, por lo tanto, es la suma de las frecuencias absolutas hasta la i -ésima modalidad,

$$N_i = n_1 + n_2 + \dots + n_i = \sum_{j=1}^i n_j \Rightarrow N_k = N = \sum_{i=1}^k n_i$$

Definición

Frecuencia relativa acumulada (F_i): Proporción de individuos de la población que presentan un valor de la variable menor o igual que el considerado, por lo tanto, es la suma de las frecuencias relativas hasta la i -ésima modalidad,

$$F_i = f_1 + f_2 + \dots + f_i = \sum_{j=1}^i f_j \Rightarrow F_k = 1 = \sum_{i=1}^k f_i$$

También puede calcularse como $F_i = \frac{N_i}{N}$

TABLA DE FRECUENCIAS DE UNA VARIABLE ESTADÍSTICA DISCRETA

Se considera una variable estadística discreta, X , que toma los valores $x_1, \dots, x_i, \dots, x_k$. La tabla estadística con los tipos de frecuencias estudiados se construye de la siguiente forma:

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
\vdots	\vdots	\vdots	\vdots	\vdots
x_i	n_i	N_i	f_i	F_i
\vdots	\vdots	\vdots	\vdots	\vdots
x_k	n_k	$N_k = N$	f_k	$F_k = 1$
Total	N		1	

TABLA DE FRECUENCIAS DE UNA VARIABLE ESTADÍSTICA CONTINUA

En las variables de tipo continuo se agrupan los valores de la variable en **intervalos** o **clases** que se denotan como $I_i = (e_{i-1}, e_i]$

Cada clase está representada por su punto medio, que recibe el nombre de **marca de clase**, y se denota por x_i , por lo tanto, se obtiene como

$$x_i = \frac{e_{i-1} + e_i}{2}$$

Se define **amplitud del intervalo** a la diferencia entre los extremos del intervalo,

$$a_i = e_i - e_{i-1}$$

Los intervalos de una población pueden elegirse de igual o distinta amplitud.

El número de intervalos, k , a utilizar no está determinado de forma fija y por tanto, se usa un k que permita trabajar cómodamente y represente bien la estructura de los datos.

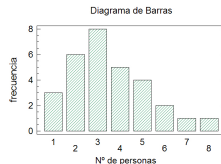
La tabla de frecuencias correspondiente a las variables estadísticas de tipo continuo con las frecuencias estudiadas es la siguiente:

$I_i = (e_{i-1}, e_i]$	x_i	n_i	N_i	f_i	F_i	a_i
$(e_0, e_1]$	x_1	n_1	N_1	f_1	F_1	a_1
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$(e_{i-1}, e_i]$	x_i	n_i	N_i	f_i	F_i	a_i
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$(e_{k-1}, e_k]$	x_k	n_k	$N_k = N$	f_k	$F_k = 1$	a_k
Total		N	1			

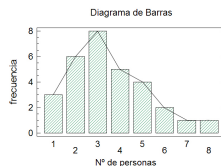
La representación gráfica tiene por objeto proporcionar una síntesis visual de la distribución de frecuencias, haciendo resaltar detalles que no resultan fácilmente perceptibles directamente en la tabla estadística.

REPRESENTACIONES GRÁFICAS DE VARIABLES ESTADÍSTICAS DISCRETAS

- **Diagrama de barras:** Sobre un sistema cartesiano se representan en el eje de abscisas los valores de la variable y sobre cada uno de estos valores se levantan barras de altura igual a su frecuencia absoluta o a su frecuencia relativa.

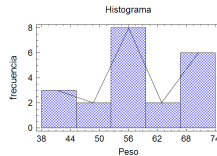
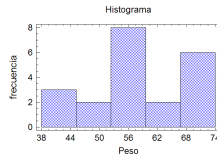


- **Polígono de frecuencias:** Es la línea que se obtiene uniendo con segmentos, en el diagrama de barras, los puntos medios de los extremos superiores de las barras recibe el nombre de **polígono de frecuencias**.



REPRESENTACIONES GRÁFICAS DE VARIABLES ESTADÍSTICAS CONTINUAS

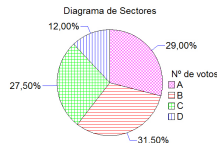
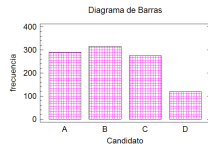
- **Histograma:** El histograma se construye representando los intervalos en el eje de abscisas y la densidad de frecuencia en el eje de ordenadas. Se dibujan rectángulos de base la amplitud a_i y de altura la densidad de frecuencia, h_i , siendo $h_i = \frac{n_i}{a_i}$ o $h_i = \frac{f_i}{a_i}$.



- **Polígono de frecuencias:** Es la línea que se obtiene uniendo con segmentos, los puntos medios de los extremos superiores de los rectángulos que forman el histograma.

REPRESENTACIONES GRÁFICAS DE VARIABLES ESTADÍSTICAS CUALITATIVAS

- **Diagrama de barras:** En unos ejes cartesianos se representan sobre el eje de abscisas las distintas modalidades del carácter y sobre el eje de ordenadas los valores de las frecuencias absolutas. A continuación, en el eje de abscisas se levantan rectángulos de base constante y de altura proporcional a la frecuencia absoluta correspondiente.
- **Gráfico de sectores:** En esta representación un círculo se divide en tantos sectores circulares como modalidades tenga el carácter, teniendo cada sector el área proporcional a la frecuencia absoluta correspondiente. Los grados de cada sector se obtienen resolviendo la proporción $\frac{n_i}{N} = \frac{\alpha_i}{360^\circ}$



OTRAS REPRESENTACIONES GRÁFICAS: DIAGRAMA DE TALLOS Y HOJAS

Se traza una línea vertical. Todas las cifras menos la última se escriben a la izquierda de la línea (forman el tallo). La última cifra se escribe a la derecha (forma la hoja). Cada tallo define una clase y se escribe sólo una vez. El número de hojas representa la frecuencia de dicha clase.

Las hojas se ordenan en cada una de las ramas. De esta manera los propios datos dan una idea visual de la zona con mayor frecuencia de observaciones. Si se gira el gráfico 90° , se obtiene el histograma clásico. Cuando se obtienen tallos con ramas muy largas, es conveniente dividir cada tallo en dos.

MEDIDAS DE POSICIÓN CENTRAL

Definición

Las **medidas de posición** tratan de resumir y sintetizar el conjunto de datos mediante un valor numérico.

Si este valor numérico se sitúa hacia el centro de la distribución se habla, entonces, de **medidas de posición central**. Las principales medidas de posición central son: la media, la mediana y la moda. Se estudiarán también otras **medidas de posición no central** llamadas cuantiles. En cada medida se distingue para su cálculo entre los casos discreto y continuo.

Definición

Media aritmética: Sea una variable X , con valores x_1, x_2, \dots, x_k y frecuencias absolutas n_1, n_2, \dots, n_k . Entonces, se define la media, y se denota por \bar{x} , como la suma ponderada de los valores de la variable por sus frecuencias.

- Caso discreto: $\bar{x} = \sum_{i=1}^k x_i f_i = \frac{1}{N} \sum_{i=1}^k x_i n_i$ siendo N el número total de observaciones.
- Caso continuo: En este caso los intervalos se representan por su marca de clase, definiéndose la media de forma análoga al caso de variable estadística discreta.

Definición

Mediana: Se define la mediana y se denota por Me , como aquel valor de la variable estadística que divide en dos conjuntos iguales a los valores de la variable supuestos ordenados de forma ascendente según el carácter.

- Caso discreto:

- ① Si no existe un valor x_i con $F_i = 0.5$, entonces la mediana es el primer valor de la variable tal que $F_i > 0.5$
- ② Si existe un valor x_i con $F_i = 0.5$, entonces la mediana será la media aritmética de los valores x_i y x_{i+1} , es decir,

$$Me = \frac{x_i + x_{i+1}}{2}$$

- Caso continuo:

- ① Si existe algún intervalo I_i , tal que $F_i = 0.5$, entonces $Me = e_i$
- ② Si no existe un intervalo I_i , tal que $F_i = 0.5$, se selecciona el primer intervalo en el que $F_i > 0.5$. A este intervalo se le denomina **intervalo mediano**, se denota por I_{Me} . El valor exacto de la mediana se obtiene aplicando al intervalo mediano la siguiente fórmula:

$$Me = e_{i-1} + \frac{0.5 - F_{i-1}}{f_i} \cdot a_i$$

Definición

Moda: Se define la moda y se nota por Mo , como el valor más frecuente de la distribución, o lo que es lo mismo, el que más se repite.

La moda puede no ser única (más de una modalidad tienen igual frecuencia máxima) o incluso no existir (cuando todas las modalidades de la variable tengan igual frecuencia).

- Caso discreto: En este caso, la moda es el valor de la variable que corresponde a la máxima frecuencia absoluta.

$$Mo = x_i \text{ tal que } n_i = \max_j n_j$$

- Caso continuo: En primer lugar, se elige el intervalo modal, $I_{Mo} = (e_{i-1}, e_i]$, que es aquel que tenga máxima altura o densidad de frecuencia $h_i = \max_j h_j$. El valor exacto de la moda se obtiene aplicando al intervalo modal la siguiente fórmula:

$$Mo = e_{i-1} + \frac{(h_i - h_{i-1})}{(h_i - h_{i-1}) + (h_i - h_{i+1})} \cdot a_i$$

OTRAS MEDIDAS DE POSICIÓN. CUANTILES

Sea X una variable estadística y sea α un número real tal que $0 < \alpha < 1$. En general, un **cuantil de orden α de la variable X** , divide a la población en dos partes, de tal manera que una proporción α de la población es menor que dicho valor y el resto mayor.

Definición

Cuartiles: Son tres valores que distribuyen la serie de datos, ordenada de forma creciente, en cuatro tramos iguales, en los que cada uno de ellos contiene el 25% de las observaciones. Se denotan por Q_1 , Q_2 y Q_3 .

Definición

Deciles: Son nueve valores que distribuyen la serie de datos, ordenada de forma creciente, en diez tramos iguales, en los que cada uno de ellos contiene el 10% de las observaciones. Se denotan por: D_1, D_2, \dots, D_9 .

Definición

Percentiles: Son noventa y nueve valores que distribuyen la serie de datos, ordenada de forma creciente, en cien tramos iguales, en los que cada uno de ellos contiene el 1% de las observaciones. Se denotan por: P_1, P_2, \dots, P_{99} .

Cálculo de un cuantil

Para calcular un cuantil $C(\alpha)$ se razona de manera análoga al cálculo de la mediana.

- Caso discreto:

- ❶ Si no existe un valor x_i con $F_i = \alpha$, entonces el cuantil de orden α es el primer valor de la variable tal que $F_i > \alpha$.
- ❷ Si existe un valor de la variable x_i que verifique $F_i = \alpha$, entonces el cuantil de orden α será

$$C(\alpha) = \frac{x_i + x_{i+1}}{2}$$

- Caso continuo:

- ❶ Si existe algún intervalo I_i , tal que $F_i = \alpha$, entonces $C(\alpha) = e_i$
- ❷ Si no existe un intervalo I_i , tal que $F_i = \alpha$, se selecciona el primer intervalo en el $F_i > \alpha$. Dicho intervalo contiene el cuantil y para determinar el valor exacto se utiliza la interpolación con la siguiente fórmula:

$$C(\alpha) = e_{i-1} + \frac{\alpha - F_{i-1}}{f_i} \cdot a_i$$

MEDIDAS DE DISPERSIÓN

Las medidas de dispersión informan de lo próximas o alejadas que están las observaciones entre sí o en relación con un valor de referencia que normalmente es una medida de centralización. De esta forma, se pueden considerar las medidas de tendencia central como muy representativas del conjunto, poco representativas, o en algunos casos, nada representativas, dependiendo de los valores adoptados por las medidas de dispersión.

Se considera la variable estadística X que toma los valores x_1, x_2, \dots, x_k (con variable estadística continua se consideran las marcas de clase de los intervalos) y frecuencias n_1, n_2, \dots, n_k .

Definición

Rango o recorrido: Es la medida de dispersión más simple y se calcula como la diferencia entre el valor máximo y el mínimo de la variable.

$$R = \max_{i=1, \dots, k} \{x_i\} - \min_{i=1, \dots, k} \{x_i\}$$

Definición

Recorrido Inter cuartílico: Es la diferencia entre el tercer y primer cuartil. Presenta la ventaja de que elimina el efecto distorsionante de los valores extremos.

$$RIQ = Q_3 - Q_1$$

Definición

Varianza: Se define la varianza y se denota por σ^2 , como la media aritmética de los cuadrados de las desviaciones entre los valores de la variable estadística y la media aritmética.

$$\sigma^2 = \sum_{i=1}^k (x_i - \bar{x})^2 f_i = \frac{1}{N} \sum_{i=1}^k (x_i - \bar{x})^2 n_i$$

La varianza siempre será mayor o igual que cero. Mientras más se aproxime a cero, más concentrados están los valores en torno a la media. Por el contrario, mientras mayor sea la varianza, más dispersos están. El inconveniente que presenta es que no está acotada superiormente, por lo que cuando los valores son grandes no se tiene una clara interpretación.

Cálculo simplificado de la varianza (Teorema de König)

Se obtiene una expresión más simple y sencilla para calcular la varianza

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

Definición

Desviación Típica: La varianza es una medida de dispersión que viene dada en unidades al cuadrado. Para mantener la misma unidad de medida de las observaciones, se define la desviación típica, y se denota por σ , como la raíz cuadrada positiva de la varianza,

$$\sigma = \sqrt{\sum_{i=1}^k (x_i - \bar{x})^2 f_i} = \sqrt{\frac{1}{N} \sum_{i=1}^k (x_i - \bar{x})^2 n_i} = \sqrt{\frac{1}{N} \sum_{i=1}^k x_i^2 n_i - \bar{x}^2}$$

Definición

Coeficiente de variación de Pearson: Se define el coeficiente de variación de Pearson de una variable estadística X , y se denota por CV_x , como el cociente entre la desviación típica y la media aritmética,

$$CV_x = \frac{\sigma_x}{\bar{x}}$$

Se utiliza para comparar la dispersión de dos o más distribuciones en las que las variables vienen expresadas en unidades distintas ya que es una medida de dispersión relativa sin dimensión, invariante respecto a cambios de escala. Presenta la ventaja de utilizar toda la información que suministra la distribución.

El coeficiente de variación representa el número de veces que la desviación típica contiene a la media aritmética, por tanto, cuanto mayor sea el coeficiente de variación significa que mayor número de veces contiene la desviación típica a la media aritmética y entonces la media aritmética es menos representativa.

MEDIDAS DE FORMA: MEDIDAS DE SIMETRÍA

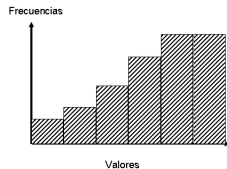
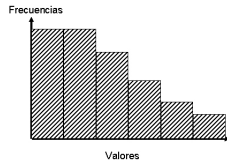
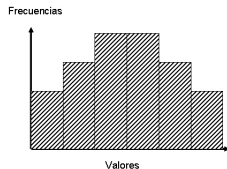
Para medir la asimetría de una distribución se utiliza el **coeficiente de asimetría de Fisher**, que viene dado por

$$\gamma_1 = \frac{\mu_3}{\sigma^3}$$

donde $\mu_3 = \frac{1}{N} \sum_{i=1}^n (x_i - \bar{x})^3 n_i = m_3 - 3m_1m_2 + 2m_1^3$. Es invariante frente a cambios de origen y escala.

Interpretación (medida de simetría):

- Cuando la distribución es **simétrica** existe la misma concentración de valores a la derecha y a la izquierda de la media. En este caso la suma de los valores menos la media al cubo será nula por lo que $\gamma_1 = 0$. Además $\bar{x} = Me = Mo$.
- Si $\gamma_1 > 0$, se tiene una distribución **asimétrica a la derecha o positiva** y se caracteriza porque su representación gráfica presenta cola a la derecha. En este caso $Mo \leq Me \leq \bar{x}$.
- Si $\gamma_1 < 0$, se tiene una distribución **asimétrica a la izquierda o negativa** y se caracteriza porque su representación gráfica presenta cola a la izquierda. En este caso $\bar{x} \leq Me \leq Mo$.



MEDIDAS DE FORMA: MEDIDAS DE APUNTAMIENTO O CURTOSIS

Las medidas de **apuntamiento o curtosis** permiten analizar el grado de concentración que presentan los valores alrededor de la zona central de la distribución.

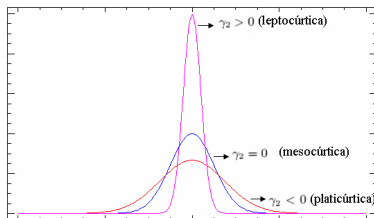
Para medir la curtosis de una distribución se utiliza el **coeficiente de curtosis de Fisher**, que viene dado por

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3$$

donde $\mu_4 = \frac{1}{N} \sum_{i=1}^n (x_i - \bar{x})^4 n_i = m_4 - 4m_1m_3 + 6m_1^2m_2 - 3m_1^4$. Es invariante frente a cambios de origen y escala.

Interpretación:

- Si $\gamma_2 = 0$, la distribución se denomina **mesocúrtica**. Presenta un grado de concentración medio alrededor de los valores centrales de la variable.
- Si $\gamma_2 > 0$, la distribución es más apuntada que la distribución normal, y se denomina **leptocúrtica**. Presenta un elevado grado de concentración alrededor de los valores centrales de la variable.
- Si $\gamma_2 < 0$, la distribución es menos apuntada que la distribución normal, y se denomina **platicúrtica**. Presenta un reducido grado de concentración alrededor de los valores centrales de la variable.



TEMA 2. ESTADÍSTICA DESCRIPTIVA BIDIMENSIONAL

INTRODUCCIÓN

Se considera que se realiza el estudio de dos caracteres simultáneamente, X e Y , sobre cada uno de los N individuos que integran la población. La variable que representa el estudio conjunto de esos dos caracteres se denota por (X, Y) y recibe el nombre de **variable estadística bidimensional**.

Se denotan $x_1, \dots, x_i, \dots, x_k$, las k modalidades del carácter X e $y_1, \dots, y_j, \dots, y_p$, las p modalidades del carácter Y . Por lo tanto, a cada individuo de la población le corresponden un par de valores (x_i, y_j) .

DISTRIBUCIÓN DE FRECUENCIAS BIDIMENSIONAL

Definición

Se llama **frecuencia absoluta** del par (x_i, y_j) , y se denota por n_{ij} , al número de individuos de la población que presentan simultáneamente el valor x_i de X y el valor y_j de Y .

La distribución de frecuencias absoluta bidimensional verifica $\sum_{i=1}^k \sum_{j=1}^p n_{ij} = N$ ya que las modalidades de cada uno de los caracteres deben ser incompatibles y exhaustivas, siendo N el número total de observaciones.

Definición

Se llama **frecuencia relativa** del par (x_i, y_j) , y se denota por f_{ij} , a la proporción de individuos de la población que presentan simultáneamente el valor x_i de X y el valor y_j de Y ,

$$f_{ij} = \frac{n_{ij}}{N}$$

Se verifica que $\sum_{i=1}^k \sum_{j=1}^p f_{ij} = 1$

Definición

Se define la **distribución de frecuencias bidimensional** como el conjunto de valores de la variable bidimensional con sus respectivas frecuencias absolutas o relativas. A cada valor de la variable, que será de la forma (x_i, y_j) , se le hace corresponder su frecuencia absoluta (n_{ij}) o relativa (f_{ij}) que se representa por:

$$\{(x_i, y_j; n_{ij})\} \text{ o } \{(x_i, y_j; f_{ij})\} \quad i = 1, \dots, k; j = 1, \dots, p$$

Los datos se pueden disponer en una tabla de doble entrada, de manera que en la primera columna aparezcan las k modalidades correspondientes a la variable X y en la primera fila las p modalidades correspondientes a la variable Y . En cada intersección se coloca la frecuencia correspondiente al cruce de las dos modalidades.

$X \backslash Y$	$(e'_0, e'_1]$ y_1	\dots	$(e'_{j-1}, e'_j]$ y_j	\dots	$(e'_{p-1}, e'_p]$ y_p	$n_{i.}$
$(e_0, e_1]$ x_1	n_{11}	\dots	n_{1j}	\dots	n_{1p}	$n_{1.}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$(e_{i-1}, e_i]$ x_i	n_{i1}	\dots	n_{ij}	\dots	n_{ip}	$n_{i.}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$(e_{k-1}, e_k]$ x_k	n_{k1}	\dots	n_{kj}	\dots	n_{kp}	$n_{k.}$
$n_{.j}$	$n_{.1}$	\dots	$n_{.j}$	\dots	$n_{.p}$	N

El primer subíndice (i), representa a la fila, y el segundo subíndice (j) a la columna.

La frecuencia absoluta $n_{i\cdot}$ representa el número de veces que se ha observado el valor x_i de X independientemente del valor presentado por Y y, por tanto, se obtiene como la suma de todas las frecuencias absolutas correspondientes a la fila i -ésima, es decir,

$$n_{i\cdot} = n_{i1} + n_{i2} + \dots + n_{ij} + \dots + n_{ip} = \sum_{j=1}^p n_{ij} \quad (i \text{ fijo})$$

La frecuencia absoluta $n_{\cdot j}$ representa el número de veces que se ha observado el valor y_j de Y independientemente del valor presentado por X y, por tanto, se obtiene como la suma de todas las frecuencias absolutas correspondientes a la columna j -ésima, es decir,

$$n_{\cdot j} = n_{1j} + n_{2j} + \dots + n_{ij} + \dots + n_{kj} = \sum_{i=1}^k n_{ij} \quad (j \text{ fijo})$$

Si se divide cada casilla de la tabla anterior por el total de los elementos de la población N , se obtiene la tabla que contiene las frecuencias relativas f_{ij} , correspondientes a los distintos pares (x_i, y_j) .

$X \setminus Y$	$(e'_0, e'_1]$ y_1	\dots	$(e'_{j-1}, e'_j]$ y_j	\dots	$(e'_{p-1}, e'_p]$ y_p	$f_{i\cdot}$
$(e_0, e_1]$ x_1	f_{11}	\dots	f_{1j}	\dots	f_{1p}	$f_{1\cdot}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$(e_{i-1}, e_i]$ x_i	f_{i1}	\dots	f_{ij}	\dots	f_{ip}	$f_{i\cdot}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$(e_{k-1}, e_k]$ x_k	f_{k1}	\dots	f_{kj}	\dots	f_{kp}	$f_{k\cdot}$
$f_{\cdot j}$	$f_{\cdot 1}$	\dots	$f_{\cdot j}$	\dots	$f_{\cdot p}$	1

La última columna está formada por las frecuencias relativas $f_{i\cdot}$, donde $f_{i\cdot} = \sum_{j=1}^p f_{ij}$. La última fila está formada por las frecuencias relativas $f_{\cdot j}$ donde $f_{\cdot j} = \sum_{i=1}^k f_{ij}$ y en la que la suma de todas las frecuencias relativas vale la unidad $\sum_{i=1}^p \sum_{j=1}^q f_{ij} = \sum_{i=1}^p f_{i\cdot} = \sum_{j=1}^q f_{\cdot j} = 1$. Cuando la variable estadística es de tipo continuo, se suelen incluir además las marcas de clase, amplitud y densidad de frecuencia.

DISTRIBUCIÓN DE FRECUENCIAS MARGINAL

La **distribución marginal de X** , que expresa cómo se distribuye la variable X independientemente de los valores presentados por la variable Y , se obtiene tomando en la tabla de doble entrada, la primera y última columnas.

X	$n_{i\cdot}$
x_1	$n_{1\cdot}$
\vdots	\vdots
x_i	$n_{i\cdot}$
\vdots	\vdots
x_k	$n_{k\cdot}$
Total	N

La **distribución marginal de Y** , que expresa cómo se distribuye la variable Y independientemente de los valores presentados por la variable X , se obtiene tomando en la tabla de doble entrada, la primera y última filas.

Y	$n_{\cdot j}$
y_1	$n_{\cdot 1}$
\vdots	\vdots
y_j	$n_{\cdot j}$
\vdots	\vdots
y_p	$n_{\cdot p}$
Total	N

Las frecuencias relativas de las distribuciones marginales se obtienen dividiendo las frecuencias absolutas entre el número total de observaciones. Es decir, la frecuencia relativa de x_i será $f_{i\cdot} = \frac{n_{i\cdot}}{N}$ y la frecuencia relativa de y_j será $f_{\cdot j} = \frac{n_{\cdot j}}{N}$.

Características de las distribuciones marginales

Las distribuciones marginales son distribuciones unidimensionales, y por tanto, se pueden calcular para ellas todas las medidas descriptivas. Por ejemplo, las medias y varianzas son:

- Media marginal de X :

$$\bar{x} = \frac{1}{N} \sum_{i=1}^k x_i n_{i.} = \sum_{i=1}^k x_i f_{i.}$$

- Varianza marginal de X :

$$\sigma_x^2 = \frac{1}{N} \sum_{i=1}^k (x_i - \bar{x})^2 n_{i.} = \sum_{i=1}^k (x_i - \bar{x})^2 f_{i.} = \frac{1}{N} \sum_{i=1}^k x_i^2 n_{i.} - \bar{x}^2$$

- Media marginal de Y :

$$\bar{y} = \frac{1}{N} \sum_{j=1}^p y_j n_{.j} = \sum_{j=1}^p y_j f_{.j}$$

- Varianza marginal de Y :

$$\sigma_y^2 = \frac{1}{N} \sum_{j=1}^p (y_j - \bar{y})^2 n_{.j} = \sum_{j=1}^p (y_j - \bar{y})^2 f_{.j} = \frac{1}{N} \sum_{j=1}^p y_j^2 n_{.j} - \bar{y}^2$$

Covarianza

Se considera como una medida de la variabilidad conjunta de las variables X e Y . Su expresión es la siguiente

$$\sigma_{xy} = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^p (x_i - \bar{x})(y_j - \bar{y})n_{ij} = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^p x_i y_j n_{ij} - \bar{x}\bar{y}$$

El signo de la covarianza indica el sentido en el que varían conjuntamente las dos variables

- Si $\sigma_{xy} > 0$, entonces las variables varían en el mismo sentido.
- Si $\sigma_{xy} < 0$, entonces las variables varían en sentidos opuestos.
- Si $\sigma_{xy} = 0$, entonces no existe relación lineal entre las variables, lo que no significa que sean independientes, pueden tener otro tipo de relación.

DISTRIBUCIÓN DE FRECUENCIAS CONDICIONADA

Las **distribuciones condicionadas** son distribuciones unidimensionales obtenidas a partir de las bidimensionales, manteniendo fijo un valor o varios en una de las variables y considerando los valores que toma la otra con sus respectivas frecuencias.

La **distribución condicionada de X respecto de $Y = y_j$** ($X|Y = y_j$) se obtiene a partir de la tabla bidimensional, tomando la primera columna de los valores de la variable X y la columna j -ésima (correspondiente al valor y_j) de frecuencias absolutas.

$X Y = y_j$	$n_{i j}$
x_1	n_{1j}
\vdots	\vdots
x_i	n_{ij}
\vdots	\vdots
x_k	n_{kj}
Total	$n_{\cdot j}$

La **distribución condicionada de Y respecto de $X = x_i$** ($Y|X = x_i$) se obtiene a partir de la tabla bidimensional, tomando la primera fila de los valores de la variable Y y la fila i -ésima (correspondiente al valor x_i) de frecuencias absolutas.

$Y X = x_i$	$n_{j i}$
y_1	n_{i1}
\vdots	\vdots
y_j	n_{ij}
\vdots	\vdots
y_p	n_{ip}
Total	$n_{i\cdot}$

Las frecuencias relativas de las distribuciones condicionadas se obtienen dividiendo las frecuencias absolutas entre el número total de observaciones que cumplen la condición requerida, que en los casos anteriores son, respectivamente, $n_{.j}$ y $n_{i.}$. Es decir, la frecuencia relativa de $x_i|Y = y_j$ será $f_{i|j} = \frac{n_{ij}}{n_{.j}} = \frac{f_{ij}}{f_{.j}}$, y la frecuencia relativa de $y_j|X = x_i$ será $f_{j|i} = \frac{n_{ij}}{n_{i.}} = \frac{f_{ij}}{f_{i.}}$.

Características de las distribuciones condicionadas

- Media condicionada de $X|Y = y_j$:

$$\bar{x}_{|Y=y_j} = \frac{1}{n_{.j}} \sum_{i=1}^k x_i n_{ij} = \sum_{i=1}^k x_i f_{i|j}$$

- Varianza condicionada de $X|Y = y_j$:

$$\sigma_{X|Y=y_j}^2 = \frac{1}{n_{.j}} \sum_{i=1}^k (x_i - \bar{x}_{|Y=y_j})^2 n_{ij} = \sum_{i=1}^k (x_i - \bar{x}_{|Y=y_j})^2 f_{i|j}$$

- Media condicionada de $Y|X = x_i$:

$$\bar{y}_{|X=x_i} = \frac{1}{n_{i.}} \sum_{j=1}^p y_j n_{ij} = \sum_{j=1}^p y_j f_{j|i}$$

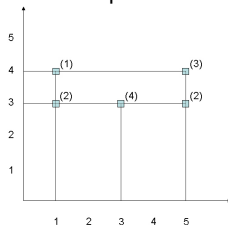
- Varianza condicionada de $Y|X = x_i$:

$$\sigma_{Y|X=x_i}^2 = \frac{1}{n_{i.}} \sum_{j=1}^p (y_j - \bar{y}_{|X=x_i})^2 n_{ij} = \sum_{j=1}^p (y_j - \bar{y}_{|X=x_i})^2 f_{j|i}$$

Diagrama de dispersión o nube de puntos

Cuando los caracteres que forman la distribución bidimensional son, ambos, cuantitativos sus valores son pares de números reales de la forma (x_i, y_j) y se pueden representar ordenados sobre un sistema de ejes cartesianos, con lo que se obtiene un conjunto de puntos sobre el plano. A este conjunto de puntos se le denomina **diagrama de dispersión o nube de puntos** y es la representación gráfica más importante asociada a una distribución bidimensional.

Hay que tener en cuenta que como el par (x_i, y_j) tiene una frecuencia n_{ij} , que en muchos casos será mayor que 1, entonces se representa el par con un punto en el que se indica la frecuencia que corresponde a ese par. En ocasiones también se puede representar utilizando distintos tamaños de puntos.



Independencia. Condición necesaria y suficiente.

La condición necesaria y suficiente para que X e Y sean independientes es que la frecuencia relativa conjunta sea igual al producto de las frecuencias marginales.

$$f_{ij} = f_{i\cdot} \cdot f_{\cdot j} \quad \forall i = 1, \dots, k, \quad j = 1, \dots, p$$

Si no se verifica esa igualdad para algún (i, j) , se dice que las variables son dependientes estadísticamente. También se puede comprobar la condición anterior en términos de las frecuencias absolutas

$$n_{ij} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{N}$$

Se concluye diciendo que si los dos caracteres son independientes entre sí se verifica que tanto las filas como las columnas de la tabla de doble entrada son proporcionales entre sí.

La independencia y la dependencia funcional son dos casos extremos de la relación entre las variables, cuando ésta no existe o es total. Generalmente cuando se estudian conjuntamente dos variables surgen los casos intermedios.

Cuando existe una dependencia estadística entre variables, el objetivo es encontrar una medida de la relación entre ambas. Se trata de buscar un modelo o función matemática que recoja esta relación y además una medida de la aproximación de dicha función a los datos reales. Por tanto, en el estudio de la dependencia estadística de dos variables hay que resolver dos problemas:

- 1 Determinar el grado de relación o dependencia entre las variables.
- 2 Encontrar un modelo aproximado de la dependencia.

La **correlación** se encarga de solucionar el primer problema cuantificando esta dependencia mediante el cálculo del coeficiente de correlación.

La **regresión** estudia desde el punto de vista estadístico la relación entre variables proporcionando un modelo de dicha relación. El modelo consiste en una función matemática cuya forma se aproxima a los datos observados. La función encontrada permitirá obtener los valores aproximados de una de las variables a partir de los valores prefijados de las otras.

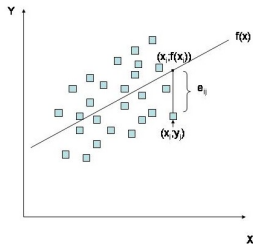
Concepto de regresión

Definición

Dadas dos variables estadísticas X e Y , se busca una función f que permita expresar los valores de Y en función de los valores de los de X (o viceversa)

$$Y = f(X).$$

Gráficamente esto equivale a encontrar una curva que aunque no pase por todos los puntos del diagrama de dispersión esté lo más próxima posible de dichos puntos. Por tanto, interesará medir el grado de ajuste entre la función teórica y la nube de puntos. El principal objetivo de este ajuste es el de realizar predicciones.



Definición

Se llama **variable dependiente** (Y), a la variable que se pretende predecir, y como su nombre indica, depende de los valores que toma otra variable.

Definición

Se denomina **variable independiente** (X), a la variable que se usa para predecir los valores de la variable dependiente.

Cuando solo se utiliza una variable independiente, la regresión se denomina **simple**. Cuando intervienen varias variables independientes la regresión se denomina **múltiple**.

El criterio que se adoptará para elegir el mejor ajuste es el "**criterio de mínimos cuadrados**", mediante el cual se pretende hacer mínima la media de la diferencia al cuadrado, entre los valores observados y los valores que da la función ajustada, que se denominan **valores ajustados** y se denotan por \hat{y}_i , obteniéndose como

$$\hat{y}_i = f(x_i)$$

Las diferencias entre los valores observados y los ajustados se denominan **residuos**, y se denotan por e_{ij} , por lo tanto,

$$e_{ij} = y_j - \hat{y}_i$$

En definitiva, se trata de minimizar la media de los residuos al cuadrado, por lo que la función a minimizar es

$$\Phi = \sum_{i=1}^k \sum_{j=1}^p e_{ij}^2 f_{ij} = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^p e_{ij}^2 n_{ij} = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^p (y_j - \hat{y}_i)^2 n_{ij}$$

Esta función será mínima cuando las derivadas parciales respecto a cada parámetro sean nulas. Calculando las derivadas parciales e igualándolas a cero se obtiene un sistema de ecuaciones denominado **sistema de ecuaciones normales**.

Dependiendo del tipo de función que mejor ajuste a los datos, la regresión se denomina de distinta forma. En el caso concreto en el que la función a utilizar sea la ecuación de una recta, se denomina **regresión lineal simple**.

Se distingue entre dos rectas de regresión:

- La recta de regresión de Y sobre X , se denota por Y/X , cuando la variable independiente es X .
- La recta de regresión de X sobre Y , se denota por X/Y , cuando la variable independiente es Y .

Recta de regresión Y/X

Se considera que la función que mejor se ajusta a la nube de puntos es una recta, por lo tanto,

$$f(X) = Y = a + bX$$

Se aplica el criterio de mínimos cuadrados para determinar el valor de a y de b , parámetros de la función que son desconocidos. En este caso,

- La variable Y es la variable dependiente.
- La variable X es la variable independiente.
- El coeficiente o parámetro b representa la pendiente de la recta, con la siguiente interpretación:
 - Si $b > 0$, representa el incremento que se produce en Y al aumentar en una unidad X .
 - Si $b < 0$, representa la disminución que se produce en Y al aumentar en una unidad X .
 - Si $b = 0$, Y no depende linealmente de X .
- El coeficiente o parámetro a representa el valor que toma Y cuando $X = 0$.

La solución que se obtiene de aplicar el criterio de mínimos cuadrados es:

$$b = \frac{\sigma_{xy}}{\sigma_x^2}$$

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

Una de las aplicaciones de la recta de regresión de Y sobre X es predecir el valor que tendrá la variable Y conocido un valor de X , diferente a los utilizados para construir la recta de regresión.

Recta de regresión X/Y

Se considera que la función que mejor se ajusta a la nube de puntos es una recta, por lo tanto,

$$f(Y) = X = a' + b'Y$$

Se aplica el criterio de mínimos cuadrados para determinar el valor de a' y de b' .

- La variable X es la variable dependiente.
- La variable Y es la variable independiente.
- El coeficiente o parámetro b' representa la pendiente de la recta, con la siguiente interpretación
 - Si $b' > 0$, representa el incremento que se produce en X al aumentar en una unidad Y .
 - Si $b' < 0$, representa la disminución que se produce en X al aumentar en una unidad Y .
 - Si $b' = 0$, X no depende linealmente de Y .
- El coeficiente o parámetro a' representa el valor que toma X cuando $Y = 0$.

El estudio de la regresión lineal es simétrico al anterior, solamente hay que cambiar el papel de las variables. El criterio de mínimos cuadrados da como resultado las siguientes soluciones

$$b' = \frac{\sigma_{xy}}{\sigma_y^2}$$

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

Una de las aplicaciones de la recta de regresión de X sobre Y es predecir el valor que tendrá la variable X conocido un valor de la variable Y , diferente a los utilizados para construir la recta de regresión.

Coeficiente de correlación lineal de Pearson

Cuando los datos tienden a agruparse en torno a una línea recta, se puede afirmar que existe **correlación lineal** entre las variables. Se distinguen dos casos:

- Si la recta tiene pendiente positiva, la correlación es **directa**, es decir, incrementos positivos de una variable implican aumentos en la otra.
- Si la recta tiene pendiente negativa, la correlación es **inversa** o **indirecta**, es decir, al aumentar una variable disminuye la otra.

El coeficiente de correlación lineal de Pearson mide el grado de asociación lineal (correlación lineal) entre las variables. Se denota por r y se define como

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}, \quad -1 \leq r \leq 1$$

El signo de r coincide con el de la covarianza, ya que las desviaciones típicas siempre son positivas.

Interpretación:

- $r = 1$: Existe una relación lineal directa perfecta, por lo que todos los puntos de la variable están sobre la recta de regresión.
- $r = -1$: Existe una relación lineal inversa perfecta, por lo que todos los puntos de la variable están sobre la recta de regresión.
- $0 < r < 1 \Rightarrow$ La relación lineal será más intensa (y directa) cuanto más se aproxime a 1, y más débil a medida que se aproxime a 0.
- $-1 < r < 0 \Rightarrow$ La relación lineal será más intensa (e inversa) cuanto más se aproxime a -1, y más débil a medida que se aproxime a 0.
- $r = 0 \Rightarrow$ No existe relación lineal entre las variables, lo que no implica que sean independientes, pueden estar relacionadas pero no linealmente. Si X e Y son independientes, entonces $r = 0$. El recíproco no es cierto.

Puede comprobarse que

$$r^2 = b \cdot b'$$

ya que

$$r^2 = \frac{\sigma_{xy}^2}{\sigma_x^2 \sigma_y^2} = \frac{\sigma_{xy}}{\sigma_x} \cdot \frac{\sigma_{xy}}{\sigma_y} = b \cdot b'$$

Coeficiente de determinación

Este término mide el grado de ajuste entre la nube de puntos y la función ajustada. Esto lleva a usar los residuos para medir la dependencia, haciendo notar que la dependencia estará íntimamente relacionada con la función que se ajuste, ya que cada una dará lugar a unos residuos distintos.

Considerando la regresión Y/X , el objetivo es explicar una variable Y a partir de una variable X y esto conlleva explicar la variabilidad de Y . Si X e Y están relacionadas, parte de esta variabilidad vendrá expresada en función de X y el resto dependerá de Y o del azar, por lo que se distingue entre la variabilidad que explica la regresión (variabilidad explicada por el modelo) y la variabilidad que no explica la regresión (también llamada residual).

$$\text{Variabilidad } Y = \text{Var.explicada por el modelo} + \text{Var.no explicada por el modelo}$$

De esta descomposición de la variabilidad se obtiene el **coeficiente de determinación**, que se denota por R^2

$$R^2 = \frac{\sigma_{exp}^2}{\sigma_y^2} = \frac{\sigma_y^2 - \sigma_{res}^2}{\sigma_y^2} = 1 - \frac{\sigma_{res}^2}{\sigma_y^2}, \quad 0 \leq R^2 \leq 1$$

Este coeficiente representa la proporción de la varianza de Y (variabilidad de la variable) que es explicada por la regresión. De esta forma, R^2 da una medida de la **bondad de ajuste**, es decir, del grado de ajuste de la regresión a los datos.

Interpretación:

- $R^2 = 0 \Rightarrow \sigma_{exp}^2 = 0$. El modelo no explica nada de la variable dependiente, es un mal ajuste.
- $R^2 = 1 \Rightarrow \sigma_{exp}^2 = \sigma_y^2$. Existe dependencia funcional exacta, todos los puntos observados están sobre la función de regresión obtenida, es decir, el modelo explica completamente el comportamiento de la variable dependiente, el ajuste de los datos al modelo es perfecto.
- $0 < R^2 < 1$. Valores cercanos a 1 indican un buen ajuste de los datos al modelo. Valores cercanos a 0 indican un mal ajuste de los datos al modelo.

El objetivo de la regresión es la predicción de valores de la variable dependiente a partir de valores conocidos de la variable independiente. Esta predicción será fiable siempre y cuando se tenga un buen ajuste.

Cuando el modelo que se utiliza es el lineal, se verifica que

$$R^2 = r^2$$

ya que en este caso se tiene que $\sigma_{res}^2 = \sigma_y^2(1 - r^2)$

El coeficiente de correlación de Pearson se utiliza como medida de bondad de ajuste de la nube de puntos a la recta de regresión. El valor obtenido permite determinar la **posición relativa de la rectas de regresión**:

- 1) Si $r^2 \neq 1$, las rectas se cortan en un solo punto, que tiene de coordenadas (\bar{x}, \bar{y}) .
- 2) Si $r^2 = 1$, las rectas coinciden. En este caso, el ajuste será perfecto.
- 3) Si $r^2 = 0$, entonces no existe relación lineal entre las variables.

TEMA 3. TEORÍA DE LA PROBABILIDAD

INTRODUCCIÓN

El objetivo de la Estadística Descriptiva es hacer una descripción sencilla de los datos correspondientes a la muestra obtenida. Pero el objetivo fundamental de la Estadística es inferir las propiedades de la población a partir de las propiedades de la muestra. Para esto es necesario un puente de unión entre la población y la muestra (modelos de probabilidad).

Se dice que un **experimento** o **prueba** es una acción que se realiza con el propósito de recoger algún tipo de observación sobre los resultados.

Definición

*Un experimento se denomina **determinístico** si, al repetir el experimento en idénticas condiciones, siempre presenta el mismo resultado. En estos fenómenos es posible saber el resultado final si se conocen el estado inicial y las condiciones de realización.*

Definición

*Un experimento es **aleatorio** cuando su resultado es impredecible, es decir, aunque el experimento se repita de la misma forma y bajo idénticas condiciones puede dar lugar a diferentes resultados.*

CONCEPTOS BÁSICOS

- **Suceso elemental**

Es cada uno de los posibles resultados que se obtienen al realizar el experimento, y estos se consideran indivisibles.

- **Espacio muestral**

Se denota por Ω y se define como el conjunto formado por todos los posibles resultados o sucesos elementales de un experimento aleatorio.

El espacio muestral puede ser:

- Finito: Se sabe cuantos sucesos elementales lo componen.
- Infinito numerable: Está compuesto por infinitos sucesos elementales que corresponden con los números naturales.
- Infinito no numerable: Está compuesto por infinitos sucesos elementales que no se corresponden con los números naturales.

- **Suceso**

Es cualquier resultado del experimento. Está compuesto por uno o más sucesos elementales.

- Suceso seguro (Ω)

Es el formado por todos los resultados posibles del experimento, es decir, el propio espacio muestral Ω .

- Suceso imposible (\emptyset)

Es aquel que no contiene ningún resultado del experimento, es decir, el conjunto vacío \emptyset .

- Suceso diferencia

Se denota por $A - B$, y es otro suceso que contiene todos los sucesos elementales de A que no son sucesos de B . Se tiene, por tanto, que $A - B = A - A \cap B$.

- Suceso complementario de un suceso

Se denota por \bar{A} , es el suceso que se realiza cuando no se verifica A .

- Se dice que el suceso A **implica** el suceso B , y se denota $A \subseteq B$, si siempre que se verifica A , se verifica B .

Operaciones con sucesos

Las operaciones con sucesos son las habituales entre conjuntos (unión, intersección, complementario, inclusión, etc).

1 Unión ($A \cup B$)

Si A y B son dos sucesos, el suceso unión $A \cup B$ es el conjunto de todos los sucesos elementales de A y los de B .

2 Intersección ($A \cap B$)

Si A y B son dos sucesos, el suceso intersección $A \cap B$ es el formado por todos los sucesos elementales que pertenecen simultáneamente a A y B .

3 Diferencia ($A - B$)

Suceso en el que se verifica que se cumple A pero no B .

4 Complementación (\bar{A})

Dado un suceso A , se define el complementario de dicho suceso, como el formado por todos los sucesos elementales del espacio muestral que no están en A .

Dos sucesos A y B se dicen **incompatibles** si al ocurrir uno no puede ocurrir el otro, es decir $A \cap B = \emptyset$.

PROPIEDADES

Con las operaciones anteriores, puede demostrarse que se verifican las siguientes propiedades.

	$\bar{\emptyset} = \Omega$	$\bar{\Omega} = \emptyset$
	Unión	Intersección
Idempotente	$A \cup \bar{A} = \Omega$ $A \cup A = A$	$A \cap \bar{A} = \emptyset$ $A \cap A = A$
Conmutativa	$A \cup B = B \cup A$	$A \cap B = B \cap A$
Asociativa	$(A \cup B) \cup C = A \cup (B \cup C)$	$(A \cap B) \cap C = A \cap (B \cap C)$
Distributiva	$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$	$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$
Leyes de Morgan	$\bar{A} \cup \bar{B} = \overline{A \cap B}$	$\bar{A} \cap \bar{B} = \overline{A \cup B}$

Álgebra de sucesos

Tanto si el fenómeno o experimento aleatorio lleva asociado un espacio muestral discreto o continuo, en la práctica se estará interesado en medir de alguna forma el grado de ocurrencia de ciertos sucesos. El objetivo por consiguiente es asignar valores a los sucesos, de modo que representen sus posibilidades de realizarse. Para poder definir o asignar estos números a los sucesos es preciso dotar al conjunto de todos los sucesos de una estructura de σ -álgebra.

Dado un experimento aleatorio y su espacio muestral asociado Ω , un álgebra de sucesos que se denota por \mathcal{A} , es un conjunto formada por todos los sucesos de Ω , que verifica:

- i) $\Omega \in \mathcal{A}$.
- ii) Si $A \in \mathcal{A}$ entonces $\bar{A} \in \mathcal{A}$.
- iii) Si $A_1, A_2, \dots \in \mathcal{A}$ entonces $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

Definición axiomática de probabilidad

La definición axiomática de probabilidad, debida a Kolmogorov, es general y se adapta a un conjunto amplio de situaciones. Para poder dar dicha definición se necesita disponer de un espacio muestral Ω sobre el que esté definida una σ -álgebra de sucesos, es decir, se impone una cierta estructura al conjunto de sucesos.

Definición

Sea Ω un espacio muestral y \mathcal{A} un σ -álgebra de sucesos definida sobre Ω . Se define una probabilidad como una función $P : \mathcal{A} \rightarrow \mathbb{R}$ que verifica los siguientes axiomas

- 1 Si $A \in \mathcal{A}$ entonces $P(A) \geq 0$.
- 2 $P(\Omega) = 1$.
- 3 Si $A_1, A_2, \dots \in \mathcal{A}$, es cualquier sucesión infinita numerable de sucesos

incompatibles, $(A_i \cap A_j = \emptyset, \forall i \neq j)$ entonces $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$.

Consecuencias de los axiomas

- 1 $P(\emptyset)=0$. El recíproco de la propiedad anterior no es cierto, es decir, si un suceso tiene probabilidad nula no tiene que ser el suceso vacío.
- 2 El tercer axioma puede particularizarse para el caso de una colección de sucesos incompatibles finitos, es decir,

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$$

- 3 Si $A \in \mathcal{A}$ entonces $P(\bar{A}) = 1 - P(A)$.
- 4 Si $A, B \in \mathcal{A}$ y $A \subset B$ entonces $P(A) \leq P(B)$
- 5 Si $A \in \mathcal{A}$ entonces $P(A) \leq 1$
- 6 Si $A, B \in \mathcal{A}$ entonces $P(B - A) = P(B) - P(A \cap B)$
- 7 Si $A, B \in \mathcal{A}$ y $A \subset B$, entonces $P(B - A) = P(B) - P(A)$
- 8 Si $A, B \in \mathcal{A}$ entonces $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- 9 Si $A, B \in \mathcal{A}$ entonces $P(A \cup B) \leq P(A) + P(B)$
- 10 Para n sucesos la desigualdad anterior se expresa como

$$P(A_1 \cup \dots \cup A_n) \leq P(A_1) + \dots + P(A_n)$$

PROBABILIDAD CONDICIONADA

Con este concepto se trata de registrar el cambio que experimenta la probabilidad de un suceso si aumenta la información de que se dispone sobre el resultado del experimento.

Sea B un suceso con probabilidad no nula, dado otro suceso A , se define la **probabilidad de A condicionada a B** , denotada por $P(A|B)$, a la probabilidad de que ocurra A supuesto que ha ocurrido B . Esta definición implica que la probabilidad condicionada se puede calcular como el cociente,

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

Del mismo modo, se tiene que

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

De esta definición se puede deducir que

$$P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$$

Se cumple que

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

INDEPENDENCIA DE SUCESOS

Dados dos sucesos A y B , se dice que son **independientes** si la ocurrencia de uno de ellos no afecta a la ocurrencia del otro, es decir, si se verifica que

$$P(A|B) = P(A) \quad \text{o} \quad P(B|A) = P(B)$$

de donde se deduce que

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

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

En otro caso se dice que los sucesos A y B son **dependientes**.

Una condición necesaria y suficiente de independencia de sucesos es considerar que dos sucesos son independientes cuando verifiquen que

$$P(A \cap B) = P(A)P(B).$$

Independencia de más de dos sucesos

La probabilidad de la intersección de más de dos sucesos viene dada por la expresión

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

Se dice que los sucesos A_1, A_2, \dots, A_n son independientes si para todo subconjunto $A_{i_1}, A_{i_2}, \dots, A_{i_l}$ de A_1, A_2, \dots, A_n , se verifica:

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_l}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_l})$$

El concepto de independencia es bastante importante ya que muchos resultados de la teoría de la probabilidad se obtienen bajo dicha hipótesis.

TEOREMA DE LA PROBABILIDAD TOTAL

Dados n sucesos A_1, \dots, A_n que forman una partición del espacio muestral, es decir,

- La unión de los n sucesos forman el espacio muestral: $\bigcup_{i=1}^n A_i = \Omega$
- Los n sucesos son incompatibles dos a dos: $A_i \cap A_j = \emptyset \quad \forall i \neq j$
- La probabilidad de cada uno de los n sucesos es positiva: $P(A_i) > 0 \quad \forall i$

y dado un suceso cualquiera $B \in \Omega$, se verifica que

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

TEOREMA DE BAYES

Dados n sucesos A_1, \dots, A_n que forman una partición del espacio muestral, y dado un suceso cualquiera $B \in \Omega$, se verifica que

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^n P(A_j)P(B|A_j)}$$

El teorema de probabilidad total y el de Bayes van a ser especialmente útiles cuando se den las siguientes circunstancias:

- a) El experimento aleatorio se puede separar en dos etapas.
- b) Es sencillo dar una partición de todo el espacio muestral Ω mediante sucesos A_1, \dots, A_n correspondientes a resultados de la primera etapa.
- c) Son conocidas o fácilmente calculables las probabilidades $P(A_1), \dots, P(A_n)$.
- d) Son conocidas o fácilmente calculables las probabilidades $P(B|A_1), \dots, P(B|A_n)$, donde B es un suceso correspondiente a resultados de la segunda etapa.

Cuando se den estas circunstancias el teorema de probabilidad total será muy útil para calcular $P(B)$, y el teorema de Bayes será muy conveniente para obtener $P(A_j|B)$.

TEMA 4. CONCEPTOS BÁSICOS DE VARIABLES ALEATORIAS

INTRODUCCIÓN

A cada uno de los posibles resultados de un experimento se le atribuye una posibilidad de ocurrir que, siguiendo los axiomas de Kolmogorov, será un número comprendido entre 0 y 1.

Asociado al concepto de probabilidad está el concepto de variable aleatoria. Una **variable aleatoria** (v.a.) representa el conjunto de valores que pueden observarse en un fenómeno aleatorio, valores que dependen del azar y sobre los cuales es posible establecer una medida de probabilidad.

Definición

Sea (Ω, \mathcal{A}, P) un espacio probabilístico donde Ω es el espacio muestral, \mathcal{A} la σ -álgebra definida sobre Ω y P una función de probabilidad. Una variable aleatoria $X : \Omega \rightarrow \mathbb{R}$ es una función que asigna a cada resultado del espacio muestral un número real.

$$\begin{aligned} X: \quad \Omega &\longrightarrow \mathbb{R} \\ s &\longrightarrow X(s) \end{aligned}$$

TIPOS DE VARIABLES ALEATORIAS

Las variables aleatorias se clasifican usualmente de acuerdo con el número de valores que pueden asumir distinguiendo entre v.a. discreta o continua:

Definición

Una v.a. X es **discreta** si el conjunto de valores que puede tomar la variable es contable. Sus posibles valores constituyen un conjunto finito o infinito numerable.

Definición

Una v.a. X es **continua** si el conjunto de valores que puede tomar la variable es un subconjunto de los números reales, es decir, sus valores consisten en uno o más intervalos de la recta de números reales.

Definición

Se llama **recorrido**, y se denota por R , al espacio formado por todos los posibles valores de la variable.

VARIABLE ALEATORIA DISCRETA

Una v.a. discreta está perfectamente definida cuando se conocen sus valores, es decir, el recorrido, y las probabilidades asociadas a los valores.

Función masa de probabilidad

Sea una v.a. discreta X que toma un número finito de valores x_1, x_2, \dots, x_k . Se define la **función masa de probabilidad** de la v.a. X como una función que asigna probabilidades a los valores de la variable, es decir,

$$\begin{aligned} f: \quad \mathbb{R} &\longrightarrow [0, 1] \\ x_i &\longmapsto P[X = x_i] = p_i, \quad i = 1, 2, \dots, k \end{aligned}$$

verificando las siguientes propiedades:

- $\sum_{i=1}^k p_i = 1$
- $p_i \geq 0, \quad \forall i = 1, \dots, k.$

Función de distribución

Se define la **función de distribución** de la v.a. X , y se denota por $F(x)$, como la probabilidad de que la v.a. X tome un valor menor o igual que x .

$$\begin{aligned} F: \quad \mathbb{R} &\longrightarrow [0, 1] \\ x &\longmapsto F(x) = P[X \leq x] = \sum_{x_i \leq x} p_i \end{aligned}$$

Propiedades

- 1 $F(+\infty) = 1$ y $F(-\infty) = 0$
- 2 $F(x)$ es una función no decreciente, si $x_i < x_j \Rightarrow F(x_i) \leq F(x_j)$
- 3 F es continua a la derecha.
- 4 $P[x_i < X \leq x_j] = F(x_j) - F(x_i)$
- 5 $P[X > x] = 1 - P[X \leq x] = 1 - F(x)$

Esperanza matemática

Se considera una v.a. discreta, X , que toma los valores $\{x_1, x_2, \dots, x_k\}$ con probabilidades $\{p_1, p_2, \dots, p_k\}$. Se define la **esperanza matemática** o media de la v.a. X , y se representa por $E[X]$, como

$$E[X] = \sum_{i=1}^k x_i p_i$$

Si se considera una función real, $g(X)$, de la v.a. X , entonces se define la esperanza matemática de $g(X)$ como

$$E[g(X)] = \sum_{i=1}^k g(x_i) p_i$$

Sean a, b y c , constantes, y X e Y variables aleatorias discretas, entonces:

- 1 $E[c] = c$
- 2 $E[aX] = aE[X]$
- 3 $E[X + a] = E[X] + a$
- 4 $E[aX + b] = aE[X] + b$
- 5 $E[X + Y] = E[X] + E[Y]$
- 6 Si $g(X)$ y $h(X)$ son dos funciones de la v.a. X , entonces

$$E[g(X) + h(X)] = E[g(X)] + E[h(X)]$$

Varianza

Se define la **varianza** de una v.a. X discreta y se representa por $Var[X]$ o σ_X^2 como

$$\sigma_X^2 = Var[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2 = \sum_{i=1}^k x_i^2 p_i - \left(\sum_{i=1}^k p_i x_i \right)^2$$

Se define la **desviación típica** de la v.a. X , y se representa por σ_X , como la raíz cuadrada positiva de la varianza.

Propiedades

- Si a y b son constantes, $Var[aX + b] = a^2 Var[X]$
- Para cualquier $k \in \mathbb{R}$, $E[(X - k)^2] = \sigma^2 + (E[X] - k)^2$
- En la propiedad anterior, si se toma $k = 0$

$$Var[X] = E[X^2] - (E[X])^2$$

VARIABLE ALEATORIA CONTINUA

Sea una v.a. continua X , que toma valores en el intervalo $[a, b] \in \mathbb{R}$. Se dice que $f(x)$ es la **función de densidad** de una v.a. continua si satisface las siguientes condiciones:

$$① \quad f(x) \geq 0, \quad \forall x \in R_X$$

$$② \quad \int_{-\infty}^{+\infty} f(x)dx = 1.$$

A partir de la función de densidad, se calcula la probabilidad de un suceso relativo a una v.a. de la siguiente forma:

$$P[x_i < X < x_j] = \int_{x_i}^{x_j} f(x)dx$$

Conviene resaltar que el valor de la función de densidad, $f(x)$, en un punto x , no es la probabilidad de aparición de ese valor x , ya que:

$$P[X = x] = \int_x^x f(t)dt = 0.$$

Función de distribución

Dada una v.a. continua X , recibe el nombre de **función de distribución** y se denota por $F(x)$, la función definida por

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

es decir, el área que encierra la función de densidad en el intervalo $(-\infty, x]$.

La función de densidad y la función de distribución de una v.a. continua están relacionadas mediante la siguiente expresión

$$f(x) = \frac{dF(x)}{dx}.$$

Propiedades

- ❶ $F(+\infty) = 1$ y $F(-\infty) = 0$
- ❷ La función de distribución es una función no decreciente y continua a la derecha.
- ❸ $P[x_i < X < x_j] = P[x_i \leq X < x_j] = P[x_i < X \leq x_j] = P[x_i \leq X \leq x_j] =$

$$= \int_{x_i}^{x_j} f(x)dx = F(x_j) - F(x_i)$$
- ❹ $P[X > x] = 1 - F(x)$

Esperanza matemática

Sea X una v.a. continua con función de densidad $f(x)$. Se define la **esperanza matemática** o media de la v.a. X , y se representa por $E[X]$ o μ_X como

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

Si se considera una función real, $g(X)$, de la v.a. X , se define la esperanza matemática de $g(X)$ como

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

Propiedades

Sean a, b , constantes y X e Y variables aleatorias continuas:

- 1 $E[aX] = aE[X]$
- 2 $E[X + a] = E[X] + a$
- 3 $E[aX + b] = aE[X] + b$
- 4 $E[X + Y] = E[X] + E[Y]$

Varianza

Se define la **varianza** de una v.a. X continua, y se representa por $Var[X]$ o σ_X^2 , como

$$Var[X] = E[(X - E[X])^2] = \int_{-\infty}^{+\infty} (x - E[X])^2 f(x) dx$$

que también se puede expresar de la siguiente forma

$$Var[X] = E[X^2] - E[X]^2 = \int_{-\infty}^{+\infty} x^2 f(x) dx - \left(\int_{-\infty}^{+\infty} x f(x) dx \right)^2$$

Se define la **desviación típica** de la v.a. X , y se representa por σ_X , como la raíz cuadrada positiva de la varianza.

Propiedades

Sean a, b , constantes y X una v.a.:

- ① $Var[aX] = a^2 Var[X]$
- ② $Var[X + a] = Var[X]$
- ③ $Var[aX + b] = a^2 Var[X]$

TEMA 5. MODELOS DE DISTRIBUCIONES DISCRETAS Y CONTINUAS

DISTRIBUCIÓN DE BERNOULLI

Se aplica este modelo a cualquier situación en la que se tiene un experimento que produce sólo dos resultados posibles incompatibles, a los que se denominan **éxito** y **fracaso**.

Se define la variable aleatoria discreta X que representa el resultado del experimento, es decir, asigna un 1 al suceso éxito con probabilidad p y un 0 al suceso fracaso con probabilidad $q = 1 - p$.

$$X : \begin{cases} 0, & \text{si ocurre el fracaso, con probabilidad } q \\ 1, & \text{si ocurre el éxito, con probabilidad } p \end{cases}$$

Su función masa de probabilidad es:

x_i	p_i
0	1-p
1	p

Definición

Se dice entonces que la variable aleatoria $X = \text{resultado del experimento}$, sigue una distribución **Bernoulli** de parámetro p y se denota como $X \rightsquigarrow B(p)$.

Función de distribución

$$F(x) = P[X \leq x] = \begin{cases} 0, & \text{si } x < 0 \\ q, & \text{si } 0 \leq x < 1 \\ 1, & \text{si } x \geq 1 \end{cases}$$

Características

- Esperanza

$$E[X] = \sum_{i=1}^k x_i p_i = 0 \cdot (1 - p) + 1 \cdot p = p$$

- Varianza

$$Var[X] = E[X^2] - (E[X])^2 = p - p^2 = p(1 - p)$$

$$\text{donde } E[X^2] = \sum_{i=1}^k x_i^2 p_i = 0^2 \cdot (1 - p) + 1^2 \cdot p = p$$

DISTRIBUCIÓN BINOMIAL

Se considera un experimento que se repite n veces de forma idéntica e independiente. Los resultados de cada realización del experimento se clasifican en dos categorías, éxito y fracaso, como en el caso de la distribución de Bernoulli.

Definición

*Se dice que la variable aleatoria discreta que cuenta el número de éxitos en n realizaciones del experimento, tiene una distribución **Binomial** de parámetros n y p , y se denota como $X \rightsquigarrow B(n, p)$.*

El recorrido de esta variable aleatoria es $R_X = \{0, 1, 2, 3, \dots, n\}$. La función masa de probabilidad viene dada por

$$P[X = k] = \binom{n}{k} p^k (1 - p)^{n-k} \quad \forall k = 0, \dots, n$$

Función de distribución

$$F(x) = P[X \leq x] = \begin{cases} 0, & x < 0 \\ \sum_{k=0}^{[x]} \binom{n}{k} p^k (1-p)^{n-k}, & 0 \leq x < n \\ 1, & x \geq n \end{cases}$$

La función masa de probabilidad y la función de distribución están tabuladas para distintos valores de los parámetros.

Características

- Esperanza: $E[X] = np$
- Varianza: $Var[X] = np(1-p)$
- Propiedad reproductiva: Sean X_1, X_2, \dots, X_m un conjunto de variables aleatorias independientes con distribución Binomial, $X_i \rightsquigarrow B(n_i, p)$. Entonces,

$$\sum_{i=1}^m X_i \rightsquigarrow B\left(\sum_{i=1}^m n_i, p\right)$$

DISTRIBUCIÓN DE POISSON

Sea X una variable aleatoria que modeliza el fenómeno aleatorio, N° de veces que se presenta un determinado hecho en un intervalo de tiempo, longitud o espacio fijado. También puede modelizar el número de ocurrencias de un suceso con poca probabilidad de ocurrir. Dicha variable se denomina de **Poisson**.

Definición

Sea X una variable aleatoria discreta que toma los valores $0, 1, 2, \dots$. Se dice que X tiene una distribución de Poisson de parámetro λ , y se denota como $X \rightsquigarrow \mathcal{P}(\lambda)$, ($\lambda > 0$) si

$$P[X = k] = e^{-\lambda} \frac{\lambda^k}{k!} \quad \forall k = 0, 1, 2, \dots$$

El parámetro de la distribución (λ) indica la esperanza de la variable aleatoria es decir, el número medio de ocurrencias del suceso considerado en el intervalo de tiempo, longitud o espacio fijado.

Función de distribución

$$F(x) = P[X \leq x] = \begin{cases} 0, & x < 0 \\ \sum_{k=0}^{[x]} \frac{\lambda^k}{k!} e^{-\lambda}, & x \geq 0 \end{cases}$$

donde $[x]$ es la parte entera de x .

La función masa de probabilidad y la función de distribución están tabuladas para distintos valores de los parámetros.

Características

- Esperanza: $E[X] = \lambda$
- Varianza: $Var[X] = \lambda$
- Propiedad reproductiva: Sean X_1, X_2, \dots, X_m , un conjunto de variables aleatorias independientes con distribución de Poisson, $X_i \rightsquigarrow \mathcal{P}(\lambda_i)$. Entonces,

$$\sum_{i=1}^m X_i \rightsquigarrow \mathcal{P}\left(\sum_{i=1}^m \lambda_i\right)$$

DISTRIBUCIÓN NORMAL

Definición

Se dice que una variable aleatoria continua X sigue una distribución **Normal** de parámetros μ y σ^2 ($\mu \in \mathbb{R}$ y $\sigma^2 > 0$), si su función de densidad es

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{1}{2\sigma^2} (x - \mu)^2 \right], \quad -\infty < x < +\infty$$

donde μ es la media y σ^2 es la varianza. Se denota como $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$.

Función de distribución

$$F(x) = P[X \leq x] = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt, \quad -\infty < x < +\infty$$

Características

- Esperanza: $E[X] = \mu$
- Varianza: $Var[X] = \sigma^2$
- Propiedad reproductiva: Sean X_1, X_2, \dots, X_m , un conjunto de variables aleatorias independientes con distribución Normal, $X_i \rightsquigarrow \mathcal{N}(\mu_i, \sigma_i^2)$. Entonces,

$$\sum_{i=1}^m X_i \rightsquigarrow \mathcal{N}\left(\sum_{i=1}^m \mu_i, \sum_{i=1}^m \sigma_i^2\right)$$

- Transformación lineal: Si se considera una transformación lineal de una v.a. Normal, la nueva variable tiene también distribución Normal

$$\left. \begin{array}{l} X \rightsquigarrow \mathcal{N}(\mu, \sigma^2) \\ Y = aX + b \end{array} \right\} \quad a, b \in \mathbb{R}, \quad a \neq 0 \quad \Rightarrow Y \rightsquigarrow \mathcal{N}(a\mu + b, a^2\sigma^2)$$

- Combinación lineal: La combinación lineal de variables aleatorias Normales e independientes tiene distribución Normal:

$$\left. \begin{array}{l} X_1, X_2, \dots, X_n \text{ v.a.i.} \\ X_i \rightsquigarrow \mathcal{N}(\mu_i, \sigma_i^2) \\ a_i \in \mathbb{R}, \end{array} \right\} \Rightarrow Y = \sum_{i=1}^n a_i X_i \rightsquigarrow \mathcal{N}\left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right)$$

Tipificación

La variable normal, Z , con $\mu = 0$ y $\sigma^2 = 1$, se denomina **distribución Normal Tipificada o Estándar**, $\mathcal{N}(0, 1)$, y su función de distribución está tabulada.

Para calcular probabilidades en el caso general para $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$, se transforma la v.a. X en la variable tipificada Z . Por tanto, utilizando la propiedad de la transformación lineal, se pueden conocer las probabilidades correspondientes a cualquier distribución Normal realizando la siguiente transformación

$$Z = \frac{X - \mu}{\sigma} \Rightarrow \left. \begin{array}{l} E[Z] = \frac{E[X] - \mu}{\sigma} = 0 \\ Var[Z] = \frac{Var[X]}{\sigma^2} = 1 \end{array} \right\} \Rightarrow Z \rightsquigarrow \mathcal{N}(0, 1)$$

De esta forma, para calcular probabilidades de una distribución normal previamente se tipifica:

$$P[a \leq X \leq b] = P\left[\frac{a - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right] = P\left[\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right]$$

Nota Para calcular probabilidades de una v.a. con distribución $\mathcal{N}(0, 1)$ se tiene en cuenta que:

$$① \quad P[X \geq a] = 1 - P[X \leq a]$$

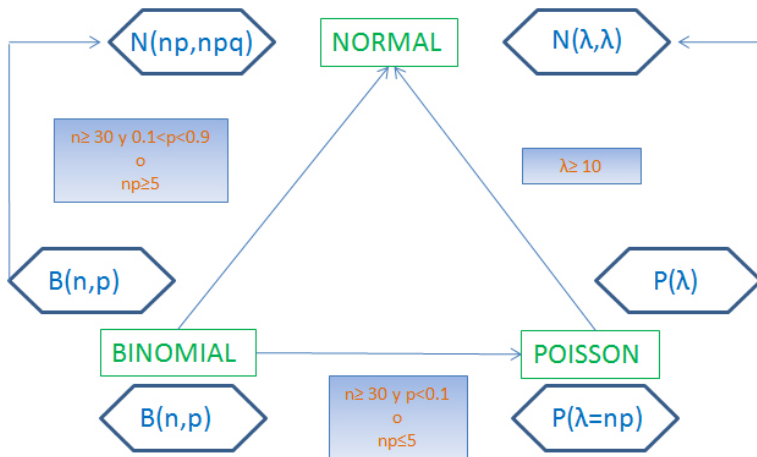
$$② \quad P[a \leq X \leq b] = P[X \leq b] - P[X \leq a]$$

$$③ \quad P[X \leq -a] = P[X \geq a] = 1 - P[X \leq a]$$

$$④ \quad P[-a \leq X \leq a] = P[X \leq a] - P[X \leq -a] = 2P[X \leq a] - 1$$

$$⑤ \quad P[|X| \leq a] = P[-a \leq X \leq a] = 2P[X \leq a] - 1$$

APROXIMACIONES ENTRE LAS DISTRIBUCIONES



Aproximación de una Binomial por una Poisson

Sea $X \rightsquigarrow B(n, p)$, de forma que $n \rightsquigarrow \infty$ y $p \rightsquigarrow 0$. Más concretamente, en el caso en el que $n \geq 30$ y $p < 0.1$ o $np < 5$, entonces se puede aproximar esta distribución a una distribución de Poisson, es decir,

$$X \rightsquigarrow B(n, p) \implies X \approx \mathcal{P}(\lambda = n \cdot p)$$

Aproximación de una Binomial por una Normal

Sea $X \rightsquigarrow B(n, p)$, con el parámetro n muy grande, concretamente $n > 30$ y además $0.1 < p < 0.9$, entonces la distribución Binomial puede aproximarse a una distribución Normal con $\mu = np$ y $\sigma^2 = npq$, es decir,

$$X \rightsquigarrow B(n, p) \approx \mathcal{N}(np, npq) \implies Z = \frac{X - np}{\sqrt{npq}} \approx \mathcal{N}(0, 1)$$

Aproximación de una Poisson por una Normal

Sea $X \rightsquigarrow \mathcal{P}(\lambda)$ con $\lambda \geq 10$, entonces la distribución de Poisson puede aproximarse a una distribución con $\mu = \lambda$ y $\sigma^2 = \lambda$, es decir,

$$X \rightsquigarrow \mathcal{P}(\lambda) \approx \mathcal{N}(\lambda, \lambda) \implies Z = \frac{X - \lambda}{\sqrt{\lambda}} \approx \mathcal{N}(0, 1)$$

Corrección por continuidad

Cuando se aproxima la distribución Binomial o Poisson a una distribución Normal, se está aproximando de una distribución discreta por una continua. Para que la aproximación sea la más adecuada se utiliza la **corrección por continuidad**, que consiste en sumar y restar 0.5 a cada valor discreto de la distribución, es decir,

- Calcular $P[X = k]$ en una distribución Binomial o Poisson, es equivalente a calcular $P[k - 0.5 \leq X \leq k + 0.5]$ en una distribución Normal.
- Calcular $P[X \leq k]$ en una distribución Binomial o Poisson, es equivalente a calcular $P[X \leq k + 0.5]$ en una distribución Normal.
- Calcular $P[X \geq k]$ en una distribución Binomial o Poisson, es equivalente a calcular $P[X \geq k - 0.5]$ en una distribución Normal.

Distribución Chi-cuadrado de Pearson

Definición

Sean Z_1, Z_2, \dots, Z_n variables aleatorias independientes e idénticamente distribuidas según una $\mathcal{N}(0, 1)$. Se considera la variable aleatoria

$$X = Z_1^2 + Z_2^2 + \dots + Z_n^2,$$

entonces se verifica que X tiene distribución **Chi-cuadrado con n grados de libertad** y se denota como,

$$X \rightsquigarrow \chi_n^2$$

La función de distribución asociada está tabulada para distintos valores de n . Para el uso de las tablas se considera que un punto (o valor crítico) $\chi_{n;p}^2$ representa el valor de la abscisa que tiene a la izquierda un área igual a p en una distribución χ_n^2 , es decir

$$P[\chi_n^2 \leq \chi_{n;p}^2] = p$$

Distribución t de Student

Definición

Sean Y y Z dos variables aleatorias independientes con $Y \rightsquigarrow \mathcal{N}(0, 1)$ y $Z \rightsquigarrow \chi_n^2$. Se define la v.a. $T = \frac{Y}{\sqrt{Z/n}}$, que tiene distribución t de **Student** con n grados de libertad. Se denota como $T \rightsquigarrow t_n$.

La función de distribución asociada está tabulada para distintos valores de n . Para el uso de las tablas se considera que un punto (o valor crítico) $t_{n;p}$ representa el valor de la abscisa que tiene a la izquierda un área o probabilidad igual a p en una distribución t_n , es decir,

$$P[t_n \leq t_{n;p}] = p$$

En la tabla sólo se encuentran valores positivos, por lo que es necesario utilizar las relaciones que surgen de la simetría de la distribución, es decir,

$$t_{n;p} = -t_{n;1-p}$$

Distribución F de Snedecor

Definición

Se consideran dos variables aleatorias independientes Y y Z tales que $Y \rightsquigarrow \chi_n^2$ y $Z \rightsquigarrow \chi_m^2$. Sea F una variable aleatoria definida como $F = \frac{Y/n}{Z/m}$, entonces se dice que la variable aleatoria F sigue una distribución F de **Snedecor** con n y m grados de libertad, y se denota por $F_{n,m}$.

La función de distribución asociada está tabulada para distintos valores de n y m . Para el uso de las tablas se considera que un punto (o valor crítico) $F_{n,m;p}$ representa el valor de la abscisa que tiene a la izquierda un área igual a p en una distribución $F_{n,m}$, es decir,

$$P[F_{n,m} \leq F_{n,m;p}] = p$$

En la tabla de probabilidades sólo se encuentran algunos valores de p . Para calcular probabilidades en el caso de que no se encuentre el valor p se utiliza la siguiente relación de inversión:

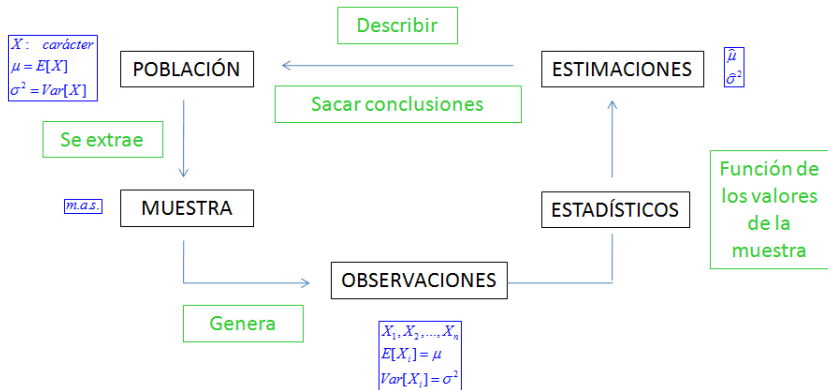
$$F_{n,m;p} = \frac{1}{F_{m,n;1-p}} \Rightarrow P[F_{n,m} \leq x] = 1 - P[F_{m,n} \leq 1/x]$$

TEMA 6. INTRODUCCIÓN A LA INFERENCIA ESTADÍSTICA

CONCEPTOS GENERALES

En la mayoría de las situaciones el investigador se ve obligado a efectuar un estudio de la población, a través de la información que se recoge en una parte de la misma, que se supone representativa del total y que se denomina muestra. Esto se produce por el exceso de coste que supondría o bien por la imposibilidad de acceder a todos ellos o incluso porque el individuo se destruya al observarlo.

El objetivo que se persigue es obtener información de la muestra y a partir de ésta poder dar resultados para la población, tratando de medir el riesgo que se asume al atribuir propiedades de la población a partir de lo que sólo se conoce en una muestra. Será, por tanto, imprescindible que la muestra que se examina sea representativa del total, por lo que su selección requerirá unos determinados requisitos que garanticen dicha representatividad.



Distribución del estadístico media muestral

Sea P una población y X una variable aleatoria asociada a dicha población cuya distribución es Normal de media μ y varianza σ^2 .

Sean X_1, \dots, X_n variables aleatorias independientes e idénticamente distribuidas, que representan n selecciones de una población normal. Se define el estadístico

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \rightsquigarrow t_{n-1}$$

donde S es la raíz cuadrada de la cuasivarianza muestral y se denomina **cuasidesviación típica muestral**.

Distribución del estadístico varianza muestral

Sean X_1, \dots, X_n variables aleatorias independientes e idénticamente distribuidas, que representan n selecciones de una población normal. Entonces el estadístico

$$Y = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} \rightsquigarrow \chi_n^2$$

Distribución del estadístico cuasivarianza muestral

Sean X_1, \dots, X_n variables aleatorias independientes e idénticamente distribuidas, que representan n selecciones de una población normal. Entonces

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

Distribución del estadístico proporción muestral

Se considera una v.a. X ="nº de éxitos en n repeticiones de un experimento" con distribución $B(n, p)$ siendo p la proporción de éxitos en la población. La proporción de éxitos en la muestra se obtiene a partir del estadístico **proporción muestral**:

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

Cuando n es grande ($n > 30$), la distribución Binomial se aproxima a una distribución Normal, por tanto,

$$X \rightsquigarrow B(n, p) \approx \mathcal{N}(np, npq)$$

Utilizando la propiedad de transformación lineal de la distribución Normal se tiene que

$$\hat{p} \rightsquigarrow \mathcal{N}\left(p, \frac{pq}{n}\right)$$

Distribución del estadístico diferencia de medias muestrales de dos poblaciones normales independientes

Sean X e Y variables aleatorias independientes tales que:

$$X \rightsquigarrow \mathcal{N}(\mu_X, \sigma_X^2), \quad Y \rightsquigarrow \mathcal{N}(\mu_Y, \sigma_Y^2)$$

Se considera:

- Una muestra aleatoria simple de tamaño n_X de X , con \bar{X} la media muestral y S_X^2 la cuasivarianza muestral.
- Una muestra aleatoria simple de tamaño n_Y de Y , con \bar{Y} la media muestral y S_Y^2 la cuasivarianza muestral.

En estas condiciones, se distinguen dos casos:

Caso A Varianzas poblacionales desconocidas pero supuestas iguales ($\sigma_X^2 = \sigma_Y^2$)

El estadístico T , definido como:

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{S_P \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \rightsquigarrow t_{n_X + n_Y - 2}$$

Caso B Varianzas poblacionales desconocidas con $n_X, n_Y > 30$

El estadístico Z , definido como:

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}} \rightsquigarrow \mathcal{N}(0, 1)$$

Distribución del estadístico cociente de varianzas muestrales de dos poblaciones Normales independientes

Sean las variables aleatorias X e Y independientes tales que:

$$X \rightsquigarrow \mathcal{N}(\mu_X, \sigma_X^2), \quad Y \rightsquigarrow \mathcal{N}(\mu_Y, \sigma_Y^2)$$

El estadístico definido como:

$$F = \frac{S_X^2 / \sigma_X^2}{S_Y^2 / \sigma_Y^2} \rightsquigarrow F_{n_X - 1, n_Y - 1}$$

Distribución del estadístico diferencia de proporciones muestrales

Sean las variables aleatorias X e Y independientes tales que:

$$X \rightsquigarrow B(n_X, p_X) \quad Y \rightsquigarrow B(n_Y, p_Y).$$

Para n_X y n_Y grandes se verifica:

$$X \rightsquigarrow \mathcal{N}(n_X p_X, n_X p_X q_X) \quad Y \rightsquigarrow \mathcal{N}(n_Y p_Y, n_Y p_Y q_Y)$$

con $q_X = 1 - p_X$ y $q_Y = 1 - p_Y$.

Se definen las proporciones muestrales como:

$$\hat{p}_X = \frac{X}{n_X} \quad \hat{p}_Y = \frac{Y}{n_Y}$$

Se define el estadístico diferencia de proporciones muestrales como:

$$\hat{p}_X - \hat{p}_Y \rightsquigarrow \mathcal{N}\left(p_X - p_Y, \frac{p_X q_X}{n_X} + \frac{p_Y q_Y}{n_Y}\right)$$

y se verifica que:

$$Z = \frac{(\hat{p}_X - \hat{p}_Y) - (p_X - p_Y)}{\sqrt{\frac{p_X q_X}{n_X} + \frac{p_Y q_Y}{n_Y}}} \rightsquigarrow \mathcal{N}(0, 1)$$

TEMA 7. ESTIMACIÓN DE PARÁMETROS

INTRODUCCIÓN

En este tema se aborda el problema de la estimación estadística, es decir, se está interesado en conocer una determinada característica y para ello se obtiene una aproximación de ésta, o lo que es lo mismo, se obtiene una estimación.

Hay diversas formas de aproximar una característica desconocida, según el objetivo que se desee conseguir:

- Si lo que se necesita es obtener un valor puntual de la característica considerada, se está ante un problema de **estimación puntual**.
- Si el objetivo es obtener conjuntos a los que pertenezca la característica estudiada con un determinado grado de credibilidad, se está ante un problema de **estimación por intervalos de confianza**.

ESTIMACIÓN PUNTUAL

Estimadores puntuales de los parámetros de una distribución normal

Sea una población P de la que se extrae una muestra de tamaño n y X una v.a. asociada a dicha población cuya distribución es $\mathcal{N}(\mu, \sigma^2)$.

- Estimador puntual de la media poblacional, μ :

Media muestral $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

- Estimadores puntuales de la varianza poblacional, σ^2 :

Varianza muestral $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$

Cuasivarianza muestral $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

siendo \bar{X} la media muestral antes definida.

Estimador puntual de la proporción poblacional

Sea una población P de N individuos de la que se extrae una muestra de tamaño n , se estudia sobre ella un carácter cualitativo.

- Estimador puntual de la proporción poblacional, p :

Proporción muestral $\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i$

ESTIMACIÓN POR INTERVALOS DE CONFIANZA

Un intervalo de confianza está formado por dos valores dentro de los cuales se afirma que se encuentra el verdadero parámetro con cierta confianza. Son unos límites o margen de variabilidad que se le da al valor estimado, para poder afirmar que el verdadero valor del parámetro está dentro de estos límites.

Sea θ el parámetro de interés. El **nivel de confianza**, $1 - \alpha$, indica la proporción de veces que se acertaría al afirmar que el parámetro está dentro de esos límites, al seleccionar un conjunto amplio de muestras. El nivel α , es la probabilidad de error, suele ser fijada de antemano y se denomina **nivel de significación**.

Para la construcción de intervalos de confianza, es necesario buscar una función de la muestra y del parámetro, $T(X_1, \dots, X_n, \theta)$, cuya distribución no dependa del parámetro. Se trata, por tanto, de calcular dos valores $\hat{\theta}_I$ y $\hat{\theta}_S$ que verifiquen

$$P[\hat{\theta}_I < T(X_1, \dots, X_n, \theta) < \hat{\theta}_S] = 1 - \alpha,$$

obteniendo un intervalo para $T(X_1, \dots, X_n, \theta)$. Despejando θ se obtiene el intervalo buscado.

Sea P una población en la que se mide una variable aleatoria X con distribución $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$.

Intervalo de confianza para la media, μ , de una población Normal

$$IC(\mu) = \left[\bar{X} - t_{n-1;1-\alpha/2} \frac{S}{\sqrt{n}}; \bar{X} + t_{n-1;1-\alpha/2} \frac{S}{\sqrt{n}} \right] = \left[\bar{X} \mp t_{n-1;1-\alpha/2} \frac{S}{\sqrt{n}} \right]$$

Intervalo de confianza para la varianza poblacional, σ^2 , de una población Normal

$$IC(\sigma^2) = \left[\frac{(n-1)S^2}{\chi_{n-1;1-\alpha/2}^2}; \frac{(n-1)S^2}{\chi_{n-1;\alpha/2}^2} \right]$$

Sean $X \rightsquigarrow \mathcal{N}(\mu_X, \sigma_X^2)$ e $Y \rightsquigarrow \mathcal{N}(\mu_Y, \sigma_Y^2)$ dos v.a. independientes. Se extraen dos muestras aleatorias de tamaños n_X y n_Y , respectivamente, ambas de forma independiente.

Intervalo de confianza para la diferencia de medias poblacionales, $\mu_X - \mu_Y$, de dos poblaciones normales con varianzas desconocidas supuestas iguales

$$IC(\mu_X - \mu_Y) = \left[(\bar{X} - \bar{Y}) \mp t_{n_X+n_Y-2; 1-\alpha/2} S_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}} \right]$$

siendo $S_p = \sqrt{\frac{(n_X-1)S_X^2 + (n_Y-1)S_Y^2}{n_X+n_Y-2}}$

Intervalo de confianza para la diferencia de medias poblacionales, $\mu_X - \mu_Y$, de dos poblaciones normales con varianzas desconocidas y $n_X, n_Y > 30$

$$IC(\mu_X - \mu_Y) = \left[(\bar{X} - \bar{Y}) \mp z_{1-\alpha/2} \sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}} \right]$$

Intervalo de confianza para el cociente de varianzas, σ_X^2/σ_Y^2 , de dos poblaciones normales

$$IC\left(\frac{\sigma_X^2}{\sigma_Y^2}\right) = \left[\frac{1}{F_{n_X-1, n_Y-1; 1-\alpha/2}} \frac{S_X^2}{S_Y^2}; \frac{1}{F_{n_X-1, n_Y-1; \alpha/2}} \frac{S_X^2}{S_Y^2} \right]$$

Intervalo de confianza para el parámetro p de una distribución binomial

Se considera una variable aleatoria X con distribución $B(n, p)$.

$$IC(p) = \left[\hat{p} \mp z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}} \right]$$

Intervalo de confianza para la diferencia de proporciones de dos distribuciones binomiales con muestras grandes

consideran $X \rightsquigarrow B(n_x, p_X)$ e $Y \rightsquigarrow B(n_y, p_Y)$. Como n_X y n_Y grandes.

$$IC(p_X - p_Y) = \left[(\hat{p}_X - \hat{p}_Y) \mp z_{1-\alpha/2} \sqrt{\frac{\hat{p}_X \hat{q}_X}{n_X} + \frac{\hat{p}_Y \hat{q}_Y}{n_Y}} \right]$$

TEMA 8. CONTRASTES DE HIPÓTESIS

CONCEPTOS BÁSICOS

El contraste de hipótesis constituye el tercer gran bloque de técnicas dentro de la Inferencia, junto con la estimación puntual y la estimación por intervalos de confianza.

El objetivo del contraste de hipótesis es diferente del de la estimación: ahora no deseamos estimar razonablemente bien el valor de un parámetro desconocido, sino que tratamos de decidir si es sensato rechazar (o aceptar) la hipótesis de que el valor de ese parámetro se sitúa en una determinada región del espacio paramétrico Θ .

En un proceso de contraste de hipótesis se pueden considerar las siguientes etapas:

- 1 Se formula un hipótesis sobre una determinada característica de interés de la población bajo estudio (generalmente un parámetro poblacional).
- 2 Se selecciona una muestra de la población.
- 3 Se comprueba si los datos están o no de acuerdo con la hipótesis que se plantea; si no es así, se puede formular una nueva hipótesis.

DEFINICIÓN DE CONTRASTE

Tipos de contrastes

Contrastes paramétricos: Se conoce la distribución de la v.a. bajo estudio y se establecen hipótesis sobre los parámetros de dicha distribución.

Contrastes no paramétricos: Se desconoce la distribución de la v.a. considerada. Se establecen hipótesis acerca de alguna propiedad de la distribución.

La primera hipótesis que se plantea en un problema de contraste recibe el nombre de **hipótesis nula**: H_0 . La hipótesis complementaria a la hipótesis nula recibe el nombre de **hipótesis alternativa**: H_1 .

Definición

Un contraste de hipótesis es una regla de decisión que lleva a aceptar o rechazar H_0 . Se especifica un conjunto C tal que si

$$(X_1, X_2, \dots, X_n) \in C \Rightarrow \text{se rechaza } H_0$$

$$(X_1, X_2, \dots, X_n) \notin C \Rightarrow \text{se acepta } H_0$$

Al conjunto C se le denomina **región crítica** o **región de rechazo**. La región crítica C se especifica mediante un estadístico que se llama **estadístico de contraste**.

Según lo anterior, la región crítica está formada por todos aquellos valores que toma el estadístico de contraste que llevan a rechazar H_0 (puntos críticos). La región de aceptación estará entonces formada por aquellos valores que toma el estadístico de contraste que llevan a aceptar H_0 .

Pasos para la realización de un contraste de hipótesis

- 1 Fijar la hipótesis nula y la alternativa, que pueden ser:

$$\begin{array}{lll} H_0 : \theta = \theta_0 & H_0 : \theta \geq \theta_0 & H_0 : \theta \leq \theta_0 \\ H_0 : \theta \neq \theta_0 & H_0 : \theta < \theta_0 & H_0 : \theta > \theta_0 \end{array}$$

- 2 Buscar el estadístico adecuado para realizar el contraste.
- 3 Obtener la región crítica.
- 4 Obtener el valor observado (valor experimental) del estadístico para la muestra seleccionada.
- 5 Decidir entre aceptar o rechazar H_0 , ya sea con el criterio de la Región Crítica.

Contrastes de hipótesis sobre la media de una población normal

Fijado un nivel de significación, α , se calcula el valor del estadístico

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \rightsquigarrow t_{n-1} \Rightarrow t_{exp} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

Contraste	Región crítica Se rechaza H_0 a nivel α si
$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	$t_{exp} < -t_{n-1;1-\alpha/2}$ o $t_{exp} > t_{n-1;1-\alpha/2}$
$H_0 : \mu \geq \mu_0$ $H_1 : \mu < \mu_0$	$t_{exp} < -t_{n-1;1-\alpha}$
$H_0 : \mu \leq \mu_0$ $H_1 : \mu > \mu_0$	$t_{exp} > t_{n-1;1-\alpha}$

Contrastes de hipótesis sobre la varianza de una población normal

Fijado un nivel de significación, α , se calcula el valor del estadístico

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \rightsquigarrow \chi_{n-1}^2 \Rightarrow \chi_{exp}^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

Contraste	Región crítica Se rechaza H_0 a nivel α si
$H_0 : \sigma^2 = \sigma_0^2$ $H_1 : \sigma^2 \neq \sigma_0^2$	$\chi_{exp}^2 < \chi_{n-1;\alpha/2}^2$ o $\chi_{exp}^2 > \chi_{n-1;1-\alpha/2}^2$
$H_0 : \sigma^2 \geq \sigma_0^2$ $H_1 : \sigma^2 < \sigma_0^2$	$\chi_{exp}^2 < \chi_{n-1;\alpha}^2$
$H_0 : \sigma^2 \leq \sigma_0^2$ $H_1 : \sigma^2 > \sigma_0^2$	$\chi_{exp}^2 > \chi_{n-1;1-\alpha}^2$

Contrastes de hipótesis sobre la proporción de una población binomial

Fijado un nivel de significación, α , se calcula el valor del estadístico

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \rightsquigarrow N(0, 1) \Rightarrow z_{exp} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Contraste	Región crítica Se rechaza H_0 a nivel α si
$H_0 : p = p_0$ $H_1 : p \neq p_0$	$z_{exp} < -z_{1-\alpha/2} \text{ o } z_{exp} > z_{1-\alpha/2}$
$H_0 : p \geq p_0$ $H_1 : p < p_0$	$z_{exp} < -z_{1-\alpha}$
$H_0 : p \leq p_0$ $H_1 : p > p_0$	$z_{exp} > z_{1-\alpha}$

Contrastes de hipótesis sobre la diferencia de medias de dos poblaciones normales, con varianzas desconocidas pero iguales

Fijado un nivel de significación, α , se calcula el valor del estadístico

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{S_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \rightsquigarrow t_{n_x+n_y-2} \Rightarrow t_{exp} = \frac{(\bar{X} - \bar{Y}) - \mu_0}{S_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$$

Contraste	Región crítica Se rechaza H_0 a nivel α si
$H_0 : \mu_x - \mu_y = \mu_0$ $H_1 : \mu_x - \mu_y \neq \mu_0$	$t_{exp} < -t_{n_x+n_y-2; 1-\alpha/2}$ o $t_{exp} > t_{n_x+n_y-2; 1-\alpha/2}$
$H_0 : \mu_x - \mu_y \geq \mu_0$ $H_1 : \mu_x - \mu_y < \mu_0$	$t_{exp} < -t_{n_x+n_y-2; 1-\alpha}$
$H_0 : \mu_x - \mu_y \leq \mu_0$ $H_1 : \mu_x - \mu_y > \mu_0$	$t_{exp} > t_{n_x+n_y-2; 1-\alpha}$

Contrastes de hipótesis sobre la diferencia de medias de dos poblaciones normales, con varianzas desconocidas y tamaños de muestras grandes

Fijado un nivel de significación, α , se calcula el valor del estadístico

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}} \rightsquigarrow N(0, 1) \Rightarrow z_{exp} = \frac{(\bar{X} - \bar{Y}) - \mu_0}{\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}}$$

Contraste	Región crítica Se rechaza H_0 a nivel α si
$H_0 : \mu_x - \mu_y = \mu_0$ $H_1 : \mu_x - \mu_y \neq \mu_0$	$z_{exp} < -z_{1-\alpha/2}$ o $z_{exp} > z_{1-\alpha/2}$
$H_0 : \mu_x - \mu_y \geq \mu_0$ $H_1 : \mu_x - \mu_y < \mu_0$	$z_{exp} < -z_{1-\alpha}$
$H_0 : \mu_x - \mu_y \leq \mu_0$ $H_1 : \mu_x - \mu_y > \mu_0$	$z_{exp} > z_{1-\alpha}$

Contrastes de hipótesis sobre el cociente de las varianzas de dos poblaciones normales

Fijado un nivel de significación, α , se calcula el valor del estadístico

$$F = \frac{S_x^2/\sigma_x^2}{S_y^2/\sigma_y^2} \rightsquigarrow F_{n_x-1; n_y-1} \Rightarrow F_{exp} = \frac{S_x^2}{S_y^2}$$

Contraste	Región crítica Se rechaza H_0 a nivel α si
$H_0 : \frac{\sigma_x^2}{\sigma_y^2} = 1$ $H_1 : \frac{\sigma_x^2}{\sigma_y^2} \neq 1$	$F_{exp} < F_{n_y-1, n_x-1; 1-\alpha/2} \text{ o } F_{exp} > F_{n_x-1, n_y-1; 1-\alpha/2}$
$H_0 : \frac{\sigma_x^2}{\sigma_y^2} \geq 1$ $H_1 : \frac{\sigma_x^2}{\sigma_y^2} < 1$	$F_{exp} < F_{n_y-1, n_x-1; 1-\alpha}$
$H_0 : \frac{\sigma_x^2}{\sigma_y^2} \leq 1$ $H_1 : \frac{\sigma_x^2}{\sigma_y^2} > 1$	$F_{exp} > F_{n_x-1, n_y-1; 1-\alpha}$

Contrastes de hipótesis sobre la diferencia de proporciones de dos poblaciones binomiales

Fijado un nivel de significación, α , se calcula el valor del estadístico

$$Z = \frac{(\hat{p}_x - \hat{p}_y) - (p_x - p_y)}{\sqrt{\frac{p_x(1-p_x)}{n_x} + \frac{p_y(1-p_y)}{n_y}}} \rightsquigarrow N(0, 1) \Rightarrow z_{exp} = \frac{(\hat{p}_x - \hat{p}_y) - p_0}{\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n_x} + \frac{\hat{p}_y(1-\hat{p}_y)}{n_y}}}$$

Contraste	Región crítica Se rechaza H_0 a nivel α si
$H_0 : p_x - p_y = p_0$ $H_1 : p_x - p_y \neq p_0$	$z_{exp} < -z_{1-\alpha/2} \text{ o } z_{exp} > z_{1-\alpha/2}$
$H_0 : p_x - p_y \geq p_0$ $H_1 : p_x - p_y < p_0$	$z_{exp} < -z_{1-\alpha}$
$H_0 : p_x - p_y \leq p_0$ $H_1 : p_x - p_y > p_0$	$z_{exp} > z_{1-\alpha}$

Contraste de la bondad de ajuste (primer caso)

Vamos a observar una muestra aleatoria (X_1, \dots, X_n) de una población X con distribución desconocida y queremos ver si, a la vista de la muestra, es razonable admitir que la distribución de X viene dada por P (un determinado modelo de probabilidad); es decir, queremos ver si los datos se ajustan bien a P . Por tanto, tenemos:

$$\begin{aligned}H_0: & \text{El modelo de probabilidad de } X \text{ es } P \\H_1: & \text{El modelo de probabilidad de } X \text{ no es } P\end{aligned}$$

Para contrastar H_0 frente a H_1 hacemos una partición (arbitraria) del espacio muestral de la población (posibles valores de X) en k clases A_1, \dots, A_k . Después, para cada A_i ($i = 1, \dots, k$) consideramos las siguientes frecuencias (absolutas):

O_i = frecuencia observada en A_i = número de elementos de la muestra (x_1, \dots, x_n) que se han situado en la clase A_i .

e_i = frecuencia esperada en la clase A_i , si la hipótesis nula es cierta = $nP(A_i)$.

El estadístico que utilizaremos para llevar a cabo este contraste es:

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

que tiene, aproximadamente una distribución χ_{k-1}^2 , si H_0 es cierta.

Si la muestra procede de P , es de esperar que haya valores parecidos para O_i y e_i y, por tanto, este estadístico debería tomar valores próximos a cero; en consecuencia, rechazaremos la hipótesis nula cuando los valores de este estadístico sean grandes y la aceptaremos cuando sean pequeños; la separación entre valores grandes y pequeños viene dada por la elección de un nivel de significación α ; en definitiva, tenemos:

Rechazamos la hipótesis nula H_0 : El modelo de probabilidad de X es P (al nivel de significación α) si:

$$\sum_{i=1}^k \frac{(O_i - e_i)^2}{e_i} > \chi_{k-1; 1-\alpha}^2$$

Contraste de la bondad del ajuste (segundo caso)

El contraste de la bondad del ajuste se puede plantear también en una situación algo más general:

Observamos una muestra aleatoria (X_1, \dots, X_n) de una población X con distribución desconocida y queremos ver si, a la vista de la muestra, es razonable admitir que la distribución de X viene dada por algún modelo de la familia $\{P_\theta : \theta \in \Theta\}$ donde $\theta = (\theta_1, \dots, \theta_r)$. Es decir, queremos ver si los datos se ajustan bien a un modelo de probabilidad de la familia $\{P_\theta : \theta \in \Theta\}$. Por tanto, tenemos:

H_0 : El modelo de probabilidad de X es algún P_θ de la familia indicada.

H_1 : El modelo de probabilidad de X no es ningún P_θ de la familia indicada.

Para contrastar H_0 frente a H_1 hacemos nuevamente una partición (arbitraria) del espacio muestral de la población (posibles valores de X) en clases A_1, \dots, A_k , y consideramos:

O_i = frecuencia observada en A_i .

e_i = frecuencia esperada en la clase A_i , si la hipótesis nula es cierta =
 $nP_\theta(A_i) \approx nP_{\hat{\theta}}(A_i)$.

(donde $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_r)$ son estimaciones de máxima verosimilitud).

El estadístico que utilizaremos es:

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

que tiene, aproximadamente una distribución χ_{k-1-r}^2 , si H_0 es cierta.

Razonando de manera análoga a como se hizo en el primer caso, llegamos a la siguiente regla para efectuar el contraste:

Rechazamos la hipótesis nula H_0 : El modelo de probabilidad de X es algún P_θ de la familia indicada (al nivel de significación α) si:

$$\sum_{i=1}^k \frac{(O_i - e_i)^2}{e_i} > \chi_{k-1-r; 1-\alpha}^2$$

Contraste de homogeneidad de poblaciones

Supongamos que disponemos de p muestras aleatorias tomadas independientemente en p poblaciones:

$$\left. \begin{array}{c} (X_{11}, \dots, X_{1n_1}) \\ \dots\dots\dots \\ (X_{p1}, \dots, X_{pn_p}) \end{array} \right\} n_1 + \dots + n_p = n$$

sobre una característica X común a todas ellas.

Queremos ver si, a la vista de las muestras obtenidas, es razonable admitir que todas las poblaciones tienen una distribución común; es decir, queremos ver si son poblaciones homogéneas. Por tanto, tenemos:

H_0 : Las p poblaciones tienen una distribución común.

H_1 : Las p poblaciones no tienen una distribución común.

Para contrastar H_0 frente a H_1 hacemos una partición (arbitraria) del espacio muestral común a las p poblaciones en k clases A_1, \dots, A_k . Después, definimos para la clase A_i ($i = 1, \dots, k$) y para la muestra de la población j -ésima ($j = 1, \dots, p$):

O_{ij} = frecuencia observada en la clase A_i con la muestra j -ésima.

e_{ij} = frecuencia esperada en la clase A_i con la muestra j -ésima, si todas las poblaciones tienen la distribución común $P = n_j P(A_i)$.

El estadístico utilizado es:

$$\sum_{j=1}^p \sum_{i=1}^k \frac{(O_{ij} - e_{ij})^2}{e_{ij}}$$

que tiene, aproximadamente una distribución $\chi^2_{(k-1)(p-1)}$ cuando H_0 es cierta.

De nuevo, podemos razonar diciendo que, si la hipótesis nula es cierta, las frecuencias observadas y esperadas serán parecidas y, por tanto, el estadístico anterior tomará valores pequeños (próximos a cero); en definitiva, tenemos:

Rechazamos la hipótesis nula H_0 : Las p poblaciones tienen una distribución común (al nivel de significación α) si:

$$\sum_{j=1}^p \sum_{i=1}^k \frac{(O_{ij} - e_{ij})^2}{e_{ij}} > \chi^2_{(k-1)(p-1); 1-\alpha}$$

Contraste de independencia

Supongamos que queremos estudiar si dos características X e Y de una población están relacionadas o no. Para hacer este estudio, obtenemos una muestra aleatoria de n pares de valores de estas características:

$$((X_1, Y_1), \dots, (X_n, Y_n))$$

Queremos ver si, a la vista de la muestra, tiene sentido admitir que X e Y son independientes. Por tanto, tenemos:

H_0 : X e Y son independientes.

H_1 : X e Y no son independientes.

Tomamos una partición (arbitraria) del espacio muestral (correspondiente a los posibles valores de X e Y) en kp clases $A_1 \times B_1, \dots, A_i \times B_j, \dots, A_k \times B_p$. Estas kp clases corresponden a tomar las clases A_1, \dots, A_k para la característica X , y las clases B_1, \dots, B_p para la característica Y .

Llamamos:

O_{ij} = frecuencia observada en la clase $A_i \times B_j$.

e_{ij} = frecuencia esperada en la clase $A_i \times B_j$, si la hipótesis nula es cierta = $nP(A_i)P(B_j)$.

El estadístico que utilizaremos para el contraste de independencia es:

$$\sum_{j=1}^p \sum_{i=1}^k \frac{(O_{ij} - e_{ij})^2}{e_{ij}}$$

que tiene, aproximadamente una distribución $\chi^2_{(k-1)(p-1)}$ cuando H_0 es cierta.

Como se puede observar, el estadístico anterior coincide con el que utilizábamos para el contraste de homogeneidad, aunque tiene un origen diferente. En definitiva, tenemos:

Rechazamos la hipótesis nula H_0 : X y Y son independientes (al nivel de significación α) si:

$$\sum_{j=1}^p \sum_{i=1}^k \frac{(O_{ij} - e_{ij})^2}{e_{ij}} > \chi^2_{(k-1)(p-1); 1-\alpha}$$

TEMA 9. OPTIMIZACIÓN SIN RESTRICCIONES

INTRODUCCIÓN: CONCEPTOS PREVIOS

Estudiaremos cómo obtener una solución óptima (si existe) o un extremo local para el siguiente problema:

$$\begin{array}{ll} \text{Max. (Min.)} & z = f(x_1, x_2, \dots, x_n) \\ \text{s.a.} & \bar{x} = (x_1, x_2, \dots, x_n), \bar{x} \in \mathbb{R}^n \end{array}$$

Las funciones cóncavas y convexas representan un papel fundamental en la Teoría de la Optimización ya que pueden garantizarnos la globalidad de los óptimos locales. Por ello vamos a iniciar este tema introduciendo el concepto de función cóncava y convexa para, posteriormente, introducir condiciones que nos permitan reconocer si una función es cóncava o convexa dependiendo de sus propiedades de diferenciabilidad.

Definición

La **matriz hessiana** asociada a una función $f(\bar{x}) = (x_1, x_2, \dots, x_n)$ es una matriz cuadrada, $H_{n \times n}$, tal que sus elementos son de la forma:

$$h_{ij} = \frac{\partial^2 z}{\partial x_i \partial x_j}$$

Definición

Denominamos **hessiano** al determinante asociado a la matriz hessiana. En \mathbb{R}^2 , el hessiano es: $H : \mathbb{R}^2 \rightarrow \mathbb{R}$

$$H = \begin{vmatrix} \frac{\partial^2 z}{\partial x_1^2} & \frac{\partial^2 z}{\partial x_1 \partial x_2} \\ \frac{\partial^2 z}{\partial x_2 \partial x_1} & \frac{\partial^2 z}{\partial x_2^2} \end{vmatrix}$$

Definición

El **menor principal de orden i** de una matriz $H_{n \times n}$ es el determinante de cualquier matriz $i \times i$ que se obtiene al suprimir las $n - i$ filas y las $n - i$ columnas correspondientes de la matriz.

Teorema: Sea la función $f(\bar{x})$ con derivadas parciales de segundo orden continuas para cada punto de $\bar{x} \in S$ (conjunto convexo de soluciones factibles). Entonces $f(\bar{x})$ es convexa sobre S si y sólo si, para cada $\bar{x} \in S$, todos los menores principales, H_i , son no negativos.

Teorema: Sea la función $f(\bar{x})$ con derivadas parciales de segundo orden continuas para cada punto de $\bar{x} \in S$ (conjunto convexo de soluciones factibles). Entonces $f(\bar{x})$ es cóncava sobre S si y sólo si, para cada $\bar{x} \in S$, los menores principales, H_i , no nulos tienen el signo que $(-1)^i$.

CONDICIONES NECESARIAS DE ÓPTIMO LOCAL

Supongamos que existen las primeras y las segundas derivadas parciales de $f(\bar{x})$ y que son continuas en todos los puntos.

Una condición necesaria para que un punto sea un extremo local para el problema nos la proporciona el teorema siguiente:

Teorema: Si $\bar{x} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n)$ es un extremo local para el problema sin restricciones, entonces

$$\nabla f(\bar{x}) = \frac{\partial f(\bar{x})}{\partial x_i} = 0, \quad \forall i$$

Para funciones diferenciables, la condición $\nabla f(\bar{x}) = 0$ es una condición necesaria para que f tenga un óptimo local en el punto \bar{x} , que sin embargo no es una condición suficiente, es decir, pueden existir puntos que anulando el gradiente no sean óptimos locales de f .

Definición

Un punto \bar{x} que satisfaga $\frac{\partial f(\bar{x})}{\partial x_i} = 0$ es un **punto estacionario** o **crítico** de la función $f(\bar{x})$.

Teorema: Si $H_i(\bar{x}) > 0, (i = 1, 2, \dots, n)$, entonces un punto estacionario será un **mínimo local** para el problema sin restricciones.

Teorema: Si $H_i(\bar{x}) \neq 0, (i = 1, 2, \dots, n)$, y tiene el signo $(-1)^i$, entonces un punto estacionario será un **máximo local** para el problema sin restricciones.

Teorema: Si $H_i(\bar{x}) \neq 0, (i = 1, 2, \dots, n)$, y no se dan ninguno de los casos anteriores, $f(\bar{x})$ presenta un **punto de inflexión** o **punto de silla** en ese punto \bar{x} .

CONDICIÓN SUFICIENTE DE ÓPTIMO LOCAL

Definición

Se denota por Q la forma cuadrática definida como sigue:

$$Q(h) = \sum_{i,j=1}^n h_i h_j H_{ij} f(\bar{x})$$

Proposición: Sean $f \in \mathcal{C}^2$ y \bar{x} un punto crítico de f , se verifica que:

- a) Si \bar{x} es un mínimo local de f , Q es semidefinida positiva o definida positiva.
- b) Si \bar{x} es un máximo local de f , Q es semidefinida negativa o definida negativa.

Proposición: Sean $f \in \mathcal{C}^2$ y \bar{x} un punto crítico de f , se verifica que:

- a) Si Q es definida positiva, \bar{x} es un mínimo local de f .
- b) Si Q es definida negativa, \bar{x} es un máximo local de f .
- c) Si Q es indefinida, \bar{x} no es ni máximo ni mínimo local de f .

Los puntos críticos de f que no son ni máximos ni mínimos locales, se denominan **puntos de silla**.

Ejemplo: Obtener el extremo de $f(\bar{x}) = x + 2y + yz - x^2 - y^2 - z^2$.

La condición necesaria para que exista extremo es: $\frac{\partial f}{\partial k} = 0$, para $k = x, y, z$.

$$\frac{\partial f}{\partial x} = 1 - 2x = 0, \quad \frac{\partial f}{\partial y} = z - 2y = 0, \quad \frac{\partial f}{\partial z} = 2 + y - 2z = 0$$

Resolviendo el sistema anterior, obtenemos, para el extremo, el punto objeto de estudio $\bar{x}^* = (1/2, 2/3, 4/3)$.

Condición suficiente:

$$H = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial x \partial z} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} & \frac{\partial^2 f}{\partial y \partial z} \\ \frac{\partial^2 f}{\partial z \partial x} & \frac{\partial^2 f}{\partial z \partial y} & \frac{\partial^2 f}{\partial z^2} \end{pmatrix} = \begin{pmatrix} -2 & 0 & 0 \\ 0 & -2 & 1 \\ 0 & 1 & -2 \end{pmatrix}$$

Estudio de los menores principales:

Orden 1:

$$|H_{1 \times 1}| = -2 < 0 \Rightarrow \text{signo } (-1)^1$$

Orden 2:

$$|H_{2 \times 2}| = \begin{vmatrix} -2 & 0 \\ 0 & -2 \end{vmatrix} = 4 > 0 \Rightarrow \text{signo } (-1)^2$$

Orden 3:

$$|H_{3 \times 3}| = \begin{vmatrix} -2 & 0 & 0 \\ 0 & -2 & 1 \\ 0 & 1 & -2 \end{vmatrix} = -8 + 2 = -6 < 0 \Rightarrow \text{signo } (-1)^3$$

Por tanto, tenemos el signo de $(-1)^i$. Así pues, en el punto \bar{x}^* , tenemos un posible máximo. Como el valor de los menores principales no depende del punto, en ese punto tenemos un máximo global.

Ejemplo: Dada $f(x, y) = x^2 + xy + y^3 - 3x - 2y + 1$, por la condición necesaria de extremo relativo, se tiene que:

$$\left. \begin{array}{l} 2x + y - 3 = 0 \\ x + 3y^2 - 2 = 0 \end{array} \right\} \quad a = (5/4, 1/2); \quad b = (5/3, -1/3)$$

Los puntos a y b verifican la condición necesaria de extremo. Veamos la condición suficiente, para lo que se obtienen, a continuación, las derivadas parciales segundas de f en cada uno de los puntos que cumplen la condición necesaria de extremo,

$$H = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 6y \end{pmatrix}$$

las matrices hessianas que permiten utilizar la condición suficiente de extremo relativo en los puntos a y b son:

$$H(a) = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}; \quad H(b) = \begin{pmatrix} 2 & 1 \\ 1 & -2 \end{pmatrix}$$

a es un mínimo local relativo ya que $H_1 = 2 > 0$, $H_2 = 5 > 0$ con lo que la forma cuadrática asociada a este punto es definida positiva y b es un punto de silla al ser la forma cuadrática asociada indefinida.

TEMA 10. OPTIMIZACIÓN CON RESTRICCIONES

INTRODUCCIÓN

El objetivo principal de la Programación Matemática es la resolución de problemas del tipo:

$$\min\{f(\bar{x}) : \bar{x} \in S\}, \quad (1)$$

donde $S \subseteq \mathbb{R}^n$ y $f : S \rightarrow \mathbb{R}$, y que debe leerse como encontrar (si existe) un elemento de S donde la función f alcance su mínimo valor. A cada elemento de S se le llama **solución** (o también **solución factible**), a S se le llama **región factible** y a f se le llama **función objetivo**. En tal sentido también afronta problemas de maximización, ya que éstos pueden reformularse como problemas de minimización:

$$\max\{f(\bar{x}) : \bar{x} \in S\} = -\min\{-f(\bar{x}) : \bar{x} \in S\}$$

Los problemas del tipo (1) se llaman **problemas de optimización**, y plantean la búsqueda de una solución factible \bar{x}^* de manera que $f(\bar{x}^*)$ sea lo menor posible. Tal búsqueda puede concluir con uno de los resultados siguientes:

- **problema no factible**: cuando no existe ninguna solución factible, es decir, $S = \emptyset$ y en tal caso escribiremos $\min\{f(\bar{x}) : \bar{x} \in S\} = +\infty$;
- **problema no acotado**: cuando existan soluciones que hagan descender infinitamente a la función objetivo, es decir, cuando para cualquier valor real M siempre existe un $\bar{x} \in S$ tal que $f(\bar{x}) < M$, y en tal caso escribiremos $\min\{f(\bar{x}) : \bar{x} \in S\} = -\infty$;
- **problema con óptimo**: cuando exista un $\bar{x}^* \in S$ tal que $f(\bar{x}^*) \leq f(\bar{x})$ para todo $\bar{x} \in S$. En este caso, $f(\bar{x}^*) = \min\{f(\bar{x}) : \bar{x} \in S\}$, y \bar{x}^* recibe el nombre de **solución óptima (global)** siendo no necesariamente única en S con esta propiedad. Un punto \bar{x}' se dice que es **solución óptima local** cuando existe un $\epsilon > 0$ tal que $f(\bar{x}') \leq f(\bar{x})$ para todo $\bar{x} \in S$ con $\|\bar{x} - \bar{x}'\| < \epsilon$, para alguna norma predefinida en el espacio \mathbb{R}^n .

Para determinar un modelo de programación matemática hemos de especificar tres conjuntos básicos de elementos:

- Las **variables principales** del modelo, que son las variables para las que queremos encontrar el mejor valor posible. Cuando trabajemos en general las variables principales serán x_1, \dots, x_n , es decir, salvo que convengamos otra cosa en un momento dado, la letra n representará siempre el número de variables principales de un modelo. En casos concretos, en lugar de x_1, x_2, x_3 podremos escribir también x, y, z , o usar las letras que consideremos más apropiadas por cuestiones de claridad.
- Las **restricciones**, que son las condiciones que hemos de imponer a las variables para que una solución sea admisible como tal. En general, consideraremos tres tipos de restricciones: restricciones de menor o igual: $g(\bar{x}) \leq b$, restricciones de mayor o igual: $g(\bar{x}) \geq b$ y restricciones de igualdad: $g(\bar{x}) = b$, donde $g : \mathbb{R}^n \rightarrow \mathbb{R}$ y $b \in \mathbb{R}$. Cuando trabajemos en general, las restricciones serán de la forma $g_i(\bar{x}) \leq b_i$ (o bien \geq o bien $=$), para $i = 1, \dots, m$, es decir, las funciones que determinan las restricciones se llamarán siempre g_i , los términos independientes se llamarán siempre b_i y el número de restricciones será siempre m , salvo por la siguiente excepción: las restricciones $x_i \geq 0$ se llaman **condiciones de no negatividad**, y en algunos contextos conviene tratarlas separadamente.

- La **función objetivo**, que es la función $f : \mathbb{R}^n \rightarrow \mathbb{R}$ que queremos maximizar o minimizar. Notemos que para determinar un modelo no sólo hemos de especificar una función objetivo, sino también si hay que maximizarla o minimizarla. Cuando no queremos especificar si buscamos el máximo o el mínimo, hablamos de **optimizar** la función objetivo. En un problema de maximizar, los óptimos de la función objetivo son sus máximos, mientras que en un problema de minimizar son sus mínimos.

Ejemplo

$$\begin{aligned} \text{Max.} \quad & x_1 + x_2 \\ \text{s.a.} \quad & 3x_1 + 2x_2 \leq 12 \\ & x_1 + 2x_2 \leq 10 \\ & x_1, x_2 \geq 0 \end{aligned}$$

Las variables principales son x_1 y x_2 , tiene cuatro restricciones:

$$3x_1 + 2x_2 \leq 12, \quad x_1 + 2x_2 \leq 10, \quad x_1 \geq 0, \quad x_2 \geq 0.$$

y la función objetivo es $f(x_1, x_2) = x_1 + x_2$.

Transformaciones de problemas

A veces es conveniente transformar un problema en otro equivalente en el sentido de que a partir de la solución óptima de uno puede calcularse fácilmente la solución óptima del otro.

Cambio del objetivo. Un problema con objetivo Max. $f(\bar{x})$ es equivalente al problema que tiene las mismas restricciones pero su objetivo es Min. $-f(\bar{x})$. Del mismo modo podemos transformar un problema de minimizar en otro de maximizar.

Eliminación de una constante en la función objetivo. Si la función objetivo es de la forma $f(\bar{x}) + k$ y eliminamos la constante k , el problema que obtenemos tiene el mismo óptimo.

Cambio de una desigualdad. Una restricción de \leq se transforma en una de \geq multiplicando sus dos miembros por -1 , y viceversa.

Igualdad por desigualdades. Una restricción de igualdad puede sustituirse por las dos restricciones que resultan de cambiar el $=$ por un \leq y un \geq .

Desigualdades por igualdades. También es posible transformar desigualdades en igualdades introduciendo las llamadas **variables de holgura**. Es costumbre hacer la transformación de tal modo que las variables de holgura sean siempre no negativas. Así, para transformar una restricción de \leq en una igualdad se le suma una variable de holgura, mientras que si es de \geq se le resta una variable de holgura.

Condiciones de signo. Una variable sometida a una condición $x \leq 0$ puede convertirse en una variable no negativa mediante el cambio $x = -x_1$, de modo que $x \leq 0$ se sustituye por $x_1 \geq 0$.

Una variable libre x se puede sustituir por variables no negativas mediante el cambio $x = x_1 - x_2$, $x_1 \geq 0$, $x_2 \geq 0$.

En vista de estas transformaciones, cualquier problema de programación matemática puede transformarse en uno de la forma

$$\begin{array}{ll} \text{Max.} & f(\bar{x}) \\ \text{s.a.} & g(\bar{x}) \leq \bar{b} \end{array}$$

o también

$$\begin{array}{ll} \text{Min.} & f(\bar{x}) \\ \text{s.a.} & g(\bar{x}) \geq \bar{b} \end{array}$$

Hechos básicos de la programación lineal

Hay dos formas especialmente importantes de presentar un problema lineal:

Forma canónica de un problema lineal. Se llama así a la presentación de un problema en la que todas las variables son no negativas y las restricciones son de \leq cuando el objetivo es maximizar o de \geq cuando el objetivo es minimizar.

$$\begin{array}{ll}\text{Max.} & c_1x_1 + \cdots + c_nx_n \\ \text{s.a.} & a_{11}x_1 + \cdots + a_{1n}x_n \leq b_1 \\ & \dots\dots\dots \\ & a_{m1}x_1 + \cdots + a_{mn}x_n \leq b_m \\ & x_1, \dots, x_n \geq 0\end{array}$$

Es conveniente escribir el problema en forma matricial, para lo cual llamamos $A = (a_{ij})$, $\bar{b} = (b_i)$, $\bar{c} = (c_j)$, con lo que la expresión anterior se reduce a

$$\begin{array}{ll}\text{Max.} & \bar{c}^t \bar{x} \\ \text{s.a.} & A\bar{x} \leq \bar{b} \\ & \bar{x} \geq 0\end{array}$$

La matriz A se llama **matriz técnica** del problema. El vector \bar{c} es el **vector de coeficientes de la función objetivo** y \bar{b} es el **vector de términos independientes**.

Forma estándar de un problema lineal. Un problema lineal está en forma estándar si todas sus variables son no negativas y todas sus restricciones son de igualdad. Matricialmente, la expresión de un problema en forma estándar con objetivo de maximizar es

$$\begin{array}{ll}\text{Max.} & \bar{c}^t \bar{x} \\ \text{s.a.} & A\bar{x} = \bar{b} \\ & \bar{x} \geq 0\end{array}$$

Podemos pasar de un problema en forma canónica a un problema en forma estándar sin más que introducir variables de holgura.

Clases de problema lineales. El esquema siguiente contiene todas las posibilidades con que nos podemos encontrar sobre existencia de óptimos en un problema lineal:

$$\text{Problema lineal} \left\{ \begin{array}{l} \text{infactible} \\ \text{factible} \left\{ \begin{array}{l} \text{acotado} \left\{ \begin{array}{l} \text{solución única} \\ \text{infinitas soluciones} \end{array} \right. \\ \text{no acotado} \end{array} \right. \end{array} \right.$$

RESOLUCIÓN GRÁFICA

Si el problema de optimización contiene dos o tres variables se puede abordar su resolución por medio de la técnica gráfica o geométrica que se describe a continuación para el caso de dos variables y que se basa en el estudio de las curvas de nivel de la función. Como es natural esta técnica tiene el inconveniente de que sólo es aplicable al caso de problemas con dos o tres variables.

La resolución geométrica de un programa matemático con dos o tres variables consiste en determinar por medio de una representación gráfica el o los puntos del conjunto factible situados sobre la curva de menor y de mayor nivel, que darán lugar, respectivamente, al mínimo y al máximo globales de la función objetivo.

Las fases del proceso de solución gráfica son:

- 1 Dibujar un sistema de coordenadas cartesianas en que las variables de decisión estén representadas por los ejes.
- 2 Dibujar en el sistema coordenado las restricciones del problema (incluyendo las de no negatividad). La intersección de todas las regiones determina la **región factible** o **espacio de soluciones**. Si esta región es no vacía, ir a la fase siguiente. En otro caso, no existe solución que satisfaga (simultáneamente) todas las restricciones y el problema se dice **infactible**.
- 3 Determinar los puntos extremos (puntos que no están situados en segmentos de línea que unen otros dos puntos del conjunto convexo) del espacio de soluciones. Evaluar la función objetivo en esos puntos y aquél o aquéllos que maximicen (o minimicen) el objetivo, corresponden a las soluciones óptimas del problema.

El proceso para resolverlo gráficamente si la función objetivo es lineal es el siguiente:

- 1 Comprobamos que la función objetivo es lineal.
- 2 Representamos el conjunto de oportunidades.
- 3 Calculamos el gradiente de la función objetivo y lo representamos gráficamente.
- 4 Representamos la curva de nivel $f = 0$ de la función objetivo. Si la función objetivo es lineal será siempre la recta perpendicular al vector gradiente.
- 5 Si la curva de nivel no pasara por el conjunto de oportunidades la movemos paralelamente a sí misma hasta que pase por S . Así obtenemos otra curva de nivel válida para alguna solución factible.
- 6 Ahora recordamos que el gradiente de la función objetivo indica hacia dónde aumenta la función, mientras que la dirección contraria indica hacia dónde disminuye. Por lo tanto, si estamos maximizando desplazaremos la curva de nivel paralelamente a sí misma en la dirección del gradiente. El último punto de S que toquemos será la solución óptima. Si el problema es de minimizar hemos de desplazar la curva de nivel en la dirección opuesta al gradiente para llegar al mínimo de f .
- 7 Si el óptimo encontrado satura al menos dos restricciones, podemos calcular sus coordenadas resolviendo el sistema de ecuaciones formado por ellas.

Como ejemplo consideramos una vez más el problema

$$\begin{aligned} \text{Max.} \quad & x_1 + x_2 \\ \text{s.a.} \quad & 3x_1 + 2x_2 \leq 12 \\ & x_1 + 2x_2 \leq 10 \\ & x_1, x_2 \geq 0 \end{aligned}$$

El proceso para resolverlo gráficamente es el siguiente:

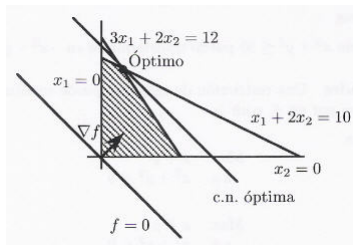


Figure: Conjunto de oportunidades.

- 1 Comprobamos que la función objetivo $f(x, y) = x_1 + x_2$ es lineal.
- 2 Representamos el conjunto de oportunidades.
- 3 Calculamos el gradiente de la función objetivo $\nabla f = (1, 1)$ y lo representamos gráficamente.
- 4 Representamos la curva de nivel $f = 0$ de la función objetivo.
- 5 La curva de nivel pasa por el conjunto de oportunidades.
- 6 El gradiente de la función objetivo indica hacia dónde aumenta la función. Por lo tanto, como estamos maximizando desplazaremos la curva de nivel paralelamente a sí misma en la dirección del gradiente. El último punto de S que toquemos será la solución óptima.
- 7 El óptimo encontrado satura al menos dos restricciones, podemos calcular sus coordenadas resolviendo el sistema de ecuaciones formado por ellas.

Para determinar las coordenadas de la solución óptima que hemos encontrado basta observar en la figura que satura las restricciones $x_1 + 2x_2 = 10$ y $3x_1 + 2x_2 = 12$. Al resolver el sistema de ecuaciones obtenemos que $(x_1, x_2) = (1, 9/2)$ y la función objetivo toma el valor $z = 5.5$.

MÉTODO SÍMPLEX

El método símplex es un procedimiento de resolución de programas lineales que consiste en obtener sucesivas soluciones factibles básicas que vayan suponiendo una mejora en el valor de la función objetivo hasta llegar al óptimo. Simplifica enormemente el proceso, a la vez que nos permite determinar si un problema es o no acotado, y si tiene una o infinitas soluciones.

Dado un problema de programación lineal, el método símplex consiste en partir de un vértice de su conjunto de oportunidades, es decir, de una solución factible básica, e ir saltando sucesivamente de una a otra adyacente, de modo que la función objetivo mejore siempre (o, al menos, no empeore nunca). Cuando llegamos a una solución desde la cual no podemos saltar a otra contigua mejor, el proceso termina, ya sea porque hemos encontrado el óptimo, ya sea porque hemos llegado a un extremo de una arista infinita a través de la cual la función objetivo mejora indefinidamente (y el problema es no acotado).

La tabla del símplex

La aplicación del algoritmo del símplex consiste en ir encontrando en forma sucesiva soluciones factibles básicas del programa, de tal forma que en cada repetición o iteración del proceso es necesario realizar un número elevado de cálculos que se simplifican considerablemente utilizando la tabla del símplex que se describe a continuación. Se considera un programa de máximo dado en la forma estándar

$$\begin{aligned} \text{Max.} \quad & \bar{c}^t \bar{x} \\ \text{s.a.} \quad & A\bar{x} = \bar{b} \\ & \bar{x} \geq 0 \end{aligned}$$

con m restricciones de igualdad, n variables y tal que el rango de A es m . Se obtiene, en primer lugar, una solución factible básica o solución inicial $\bar{x} = (h_1, \dots, h_m, 0, \dots, 0)^t$ con matrices asociadas $B = (A_1, \dots, A_m)$ y $N = (A_{m+1}, \dots, A_n)$ y se diseña la tabla del símplex.

	x_1	...	x_s	...	x_m	x_{m+1}	...	x_e	...	x_n	T.I.
x_1	1	...	0	...	0	α_{1m+1}	...	α_{1e}	...	α_{1n}	h_1
...
x_s	0	...	1	...	0	α_{sm+1}	...	α_{se}	...	α_{sn}	h_s
...
x_m	0	...	0	...	1	α_{mm+1}	...	α_{me}	...	α_{mn}	h_m
z	0	...	0	...	0	$z_{m+1} - c_{m+1}$...	$z_e - c_e$...	$z_n - c_n$	z_0

Obsérvese que si la matriz B de partida es la identidad, los cálculos se simplifican mucho ya que entonces se tiene que

$$(h_1, \dots, h_m)^t = B^{-1}\bar{b} = \bar{b},$$

es decir, que la última columna de la tabla está formada por los términos independientes del sistema de restricciones. En este caso, las α_{ij} que aparecen en la tabla serían simplemente las columnas de coeficientes de las variables no básicas que se tienen en el sistema de restricciones. Recordemos que:

$$B_j = \alpha_{1j}B_1 + \dots + \alpha_{mj}B_m, \quad \forall j = m+1, \dots, n.$$

Además, si $\bar{c}_B = \bar{0}$, como

$$z_j - c_j = \bar{c}_B B^{-1} B_j - c_j = \bar{c}_B B_j - c_j = -c_j,$$

los valores no nulos que aparecen en la fila objetivo de la tabla son los coeficientes de las variables no básicas en la función objetivo cambiados de signo.

A partir de la tabla anterior, teniendo en cuenta los resultados obtenidos, se aplica el siguiente test de optimalidad.

- Si $z_j - c_j \geq 0$ para todo $j = m + 1, \dots, n$, la solución factible básica que corresponde a esa tabla es la solución del programa, pudiendo darse el caso de soluciones múltiples si para alguna j se tiene que $z_j - c_j = 0$.
- Si alguna diferencia $z_j - c_j$ es negativa y todas las α_{ij} correspondientes a alguna de las diferencias positivas son negativas o nulas, entonces el programa no tiene solución máxima factible.
- Si no se da ninguno de los casos anteriores, se pasa a otra tabla que proporciona otra solución factible básica que mejora el valor de la función objetivo. Para ello se selecciona un elemento de la tabla llamado pivote como se indica a continuación.

Se selecciona la columna que corresponda a la mayor de las diferencias anteriores de entre las negativas (columna e de la variable que entra en la nueva solución factible básica) y la fila s (indica la variable que sale de la solución) tal que

$$\frac{h_s}{\alpha_{se}} = \min \left\{ \frac{h_i}{\alpha_{ie}} : \alpha_{ie} > 0 \right\}.$$

El elemento pivote es el de la fila s y la columna e . La nueva tabla se construye dividiendo por α_{se} los elementos de la fila pivote, con lo que el pivote pasa a valer uno, y haciendo ceros en todos los elementos de la columna pivote (incluyendo la fila objetivo) por aplicación del método de Gauss. Por último en esta nueva tabla se cambia la variable x_s por la x_e y se repite el test de optimalidad. La iteración de este proceso llevará a la solución del programa o a la conclusión de que el programa no tiene solución.

Tipos de soluciones

Según esto, las posibilidades de que el símplex termine son:

- ① *Una variable no básica x_j puede entrar pero ninguna puede salir.* El problema es no acotado.
- ② *Ninguna variable puede entrar.*

Esto significa que la función objetivo empeora (o se mantiene constante) nos movamos hacia donde nos movamos, luego estamos en una solución óptima. A su vez, pueden darse los casos siguientes:

- ① *Todas las variables no básicas cumplen $z_j - c_j \neq 0$.*

La solución es única (solución de vértice).

- ② *Hay alguna variable no básica x_j con $z_j - c_j = 0$.*

Esto significa que si hiciéramos entrar a esta variable pasaríamos a otras soluciones con el mismo valor de la función objetivo, es decir, otras soluciones óptimas. A su vez hay dos posibilidades:

- ① *Alguna variable básica x_i puede salir.*

Estamos en una solución de arista.

- ② *Ninguna variable básica puede salir.*

Estamos en una solución de arista infinita.

Ejemplo: *Se considera el siguiente programa en forma canónica:*

$$\begin{array}{ll}\text{Max.} & z = 3x_1 + 5x_2 \\ \text{s.a.} & x_1 + 2x_2 \leq 36 \\ & 2x_1 + x_2 \leq 36 \\ & x_1, x_2 \geq 0\end{array}$$

Se introducen las variables de holgura x_3 y x_4 , con lo que se tiene el programa en la forma estándar:

$$\begin{array}{ll}\text{Max.} & z = 3x_1 + 5x_2 + 0x_3 + 0x_4 \\ \text{s.a.} & x_1 + 2x_2 + x_3 = 36 \\ & 2x_1 + x_2 + x_4 = 36 \\ & x_1, x_2, x_3, x_4 \geq 0\end{array}$$

Las variables básicas son x_3 y x_4 para las que la matriz B es la identidad. La tabla inicial del símplex es la siguiente:

	x_1	x_2	x_3	x_4	T.I.
x_3	1	2	1	0	36
x_4	2	1	0	1	36
$z_j - c_j$	-3	-5	0	0	0

La solución factible básica inicial es $x_1 = 0$, $x_2 = 0$, $x_3 = 36$ y $x_4 = 36$ con valor cero en la función objetivo. Las variables básicas son x_3 y x_4 . Como hay valores negativos en la fila objetivo y los elementos de las columnas correspondientes a estos valores son positivos hay que buscar otra solución factible básica, para lo que se selecciona el elemento pivote. La columna pivote es la segunda que corresponde al mayor valor negativo de los de la fila objetivo. La fila pivote es la primera ya que

$$\frac{36}{2} = \min \left\{ \frac{36}{2}, \frac{36}{1} \right\}.$$

El pivote es el elemento $(1, 2)$, la variable x_2 entra como variable básica y sale la x_3 . Las operaciones para obtener la nueva tabla son $F_p = F_1/2$, $F_2 - F_p$ y $z + 5F_p$, en donde se denota por F las diferentes filas.

Se obtiene así la siguiente tabla:

	x_1	x_2	x_3	x_4	T.I.
x_2	1/2	1	1/2	0	18
x_4	3/2	0	-1/2	1	18
$z_j - c_j$	-1/2	0	5/2	0	90

La solución factible básica que corresponde a esta tabla es $x_1 = 0$, $x_2 = 18$, $x_3 = 0$ y $x_4 = 18$. El valor correspondiente en la función objetivo es 90 que mejora al anterior. Al haber un valor negativo en la fila objetivo con la columna correspondiente a este valor formada por números positivos se ha de buscar otra solución factible básica y, por lo tanto, otro pivote. La columna pivote es la primera que corresponde al único valor negativo en la fila objetivo. La fila pivote es la segunda, y el pivote es el elemento (2, 1). La variable x_1 entra como variable básica y sale la x_4 . Las operaciones para obtener la nueva tabla son $F_p = 2F_2/3$, $F_1 - F_p/2$ y $z + F_p/2$.

Se tiene así la siguiente tabla:

	x_1	x_2	x_3	x_4	T.I.
x_2	0	1	$2/3$	$-1/3$	12
x_1	1	0	$-1/3$	$2/3$	12
$z_j - c_j$	0	0	$7/3$	$1/3$	96

La solución factible básica correspondiente a esta tabla es $x_1 = 12$, $x_2 = 12$, $x_3 = 0$ y $x_4 = 0$. El valor correspondiente en la función objetivo es 96 y esta es la solución del programa, ya que todos los elementos de la fila objetivo son mayores o iguales que cero. Además esta solución es única pues no hay elementos nulos en la fila objetivo en los lugares de las variables no básicas de la tabla.