

“APLICACIONES DEL ENTORNO LINEX A LA ENSEÑANZA DE LA ESTADÍSTICA EN SECUNDARIA”



**Sociedad Extremeña de Educación Matemática
"Ventura Reyes Prósper"**



JUNTA DE EXTREMADURA



Consejería de Educación

APUNTES DE ESTADÍSTICA

En este documento se explicará todo lo necesario en el campo de la Estadística para entender correctamente la base teórica de los Módulos. Dicha teoría está distribuida en los siguientes epígrafes:

- 0 Introducción al Análisis Estadístico.
 - 1 Estadística Descriptiva.
 - 2 Regresión y Correlación.
 - 3 Cálculo de Probabilidades.
 - 4 Variables Aleatorias.
 - 5 Estimación de parámetros.
 - 6 Contraste de Hipótesis
-

0 Introducción al Análisis Estadístico.

- Una POBLACIÓN es un conjunto de elementos definido por una o más características que tienen todos los elementos, y sólo ellos.
- Cada elemento de la población se llama INDIVIDUO.
- El número de individuos de una población, cuando es finito, se conoce como TAMAÑO POBLACIONAL (N).
- A veces, no todos los elementos de la población están accesibles para su estudio. Entonces se distingue entre:
 - **Población Teórica:** Conjunto de elementos a los que se quiere extrapolar los resultados del estudio.
 - **Población Estudiada:** Conjunto de elementos accesibles en el estudio.

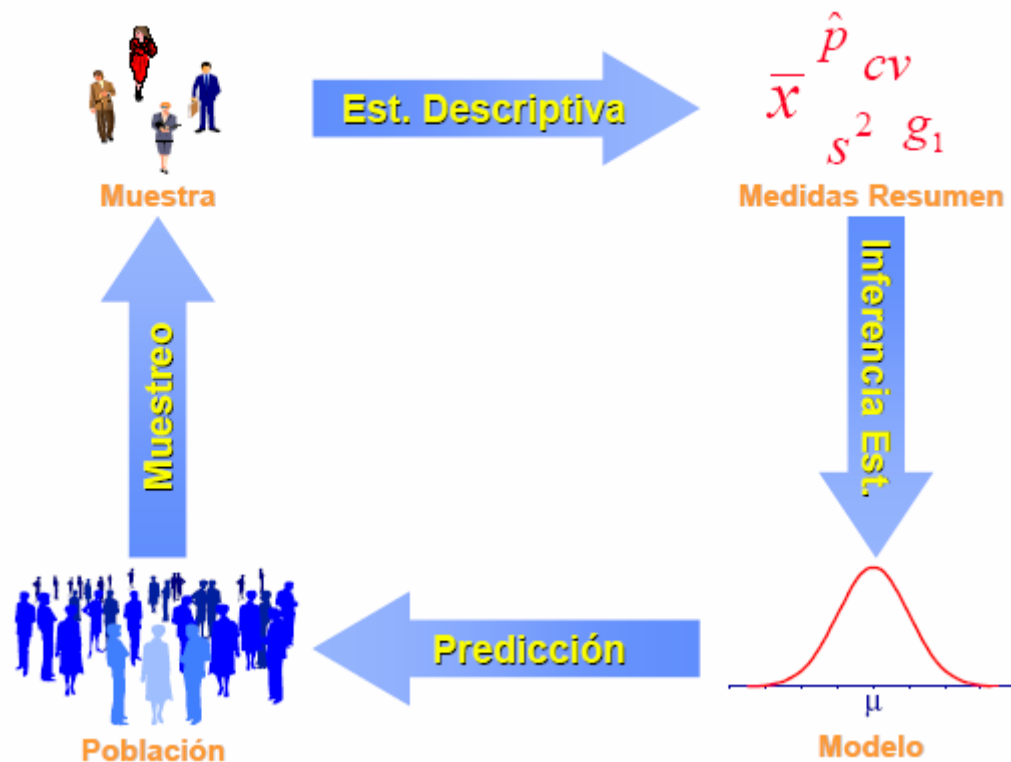
- El científico trata de estudiar el mundo que le rodea; un mundo que está lleno de variaciones que dificultan la determinación del comportamiento de las cosas.
 - Esta variabilidad en el mundo real es el origen de la ESTADÍSTICA.
 - La Estadística actúa como disciplina puente entre la concepción del mundo y los modelos matemáticos que tratan de explicar dicha concepción, proporcionando una metodología para evaluar las discrepancias entre la realidad y los modelos teóricos.
 - Esto la convierte en una herramienta indispensable en las ciencias aplicadas que requieran el análisis de datos y el diseño de experimentos.
-
- Una POBLACIÓN es un conjunto de elementos definido por una o más características que tienen todos los elementos, y sólo ellos.
 - Cada elemento de la población se llama INDIVIDUO.
 - El número de individuos de una población, cuando es finito, se conoce como TAMAÑO POBLACIONAL (N).
 - A veces, no todos los elementos de la población están accesibles para su estudio. Entonces se distingue entre:
 - **Población Teórica:** Conjunto de elementos a los que se quiere extrapolar los resultados del estudio.
 - **Población Estudiada:** Conjunto de elementos accesibles en el estudio.

- El científico estudia un determinado fenómeno en una población para comprenderlo, obtener conocimiento sobre el mismo, y así poder controlarlo.
- Para tener un conocimiento **completo** de la población es necesario estudiar **todos** los individuos de la misma.
- Sin embargo, esto no siempre es posible por distintos motivos:
 - El tamaño de la población es infinito, o bien es finito pero demasiado grande.
 - Las pruebas a que se someten los individuos son destructivas.
 - El coste tanto de dinero, como de tiempo, que supondría estudiar a todos los individuos es excesivo.

- Cuando no es posible o conveniente estudiar todos los individuos de la población, se estudia sólo una parte de la misma.
- Una MUESTRA es un subconjunto de la población.
- Al número de individuos que componen la muestra se le llama TAMAÑO DE LA MUESTRA (n).
- Habitualmente, el estudio de una población se realiza a partir de muestras extraídas de dicha población.
- Generalmente, el estudio de la muestra sólo aporta conocimiento aproximado de la población. Pero en muchos casos **es suficiente**.



- **Características de la deducción:** Si las premisas son ciertas, garantiza la certeza de las conclusiones. Sin embargo, *¡no aporta conocimiento nuevo!*
 - **Características de la inducción:** No garantiza la certeza de las conclusiones (*¡cuidado con las extrapolaciones!*), pero *¡es la única forma de generar conocimiento nuevo!*
- 1 El estudio de una población comienza por la selección de una muestra representativa de la misma. De esto se encarga el MUESTREO.
 - 2 El siguiente paso consiste en estudiar las muestras extraídas y obtener resultados numéricos que resuman la información contenida en las mismas. De esto se encarga la ESTADÍSTICA DESCRIPTIVA.
 - 3 La información obtenida es proyectada sobre un modelo matemático que intenta reflejar el comportamiento de la población. Tras construir el modelo, se realiza una crítica del mismo para validarlo. De todo esto se encarga la INFERENCIA ESTADÍSTICA.
 - 4 Finalmente, el modelo validado nos permite hacer suposiciones y predicciones sobre la población de partida con cierta confianza.



- El proceso de selección de los elementos que compondrán una muestra se conoce como MUESTREO.



- Para que una muestra refleje información fidedigna sobre la población global debe ser representativa de la misma.
- Existen muchas técnicas de muestreo. **El objetivo es obtener una muestra representativa de la población.**

Modalidades de Muestreo

- **Muestreo Aleatorio:** Elección aleatoria de los individuos de la muestra. Todos tienen la misma probabilidad de ser elegidos (*equiprobabilidad*).
- **Muestreo No Aleatorio:** Los individuos se eligen de forma no aleatoria.
- Sólo *las técnicas aleatorias* evitan el sesgo de selección, y por tanto, *garantizan la representatividad de la muestra* extraída, y en consecuencia la validez de la inferencia.
- *Las técnicas no aleatorias no sirven para hacer generalizaciones*, ya que no garantizan la representatividad de la muestra. Sin embargo, son menos costosas y pueden utilizarse en estudios exploratorios.

1 Estadística Descriptiva.

- La característica objeto de estudio puede ser de dos tipos:
 - De carácter cualitativo (ATRIBUTOS).
 - De carácter cuantitativo (VARIABLES ESTADÍSTICAS).
- A su vez, los atributos se dividen en:
 - NOMINALES: No existe un orden entre las modalidades.
 - ORDINALES: Existe un orden entre las modalidades
- Y las variables estadísticas en:
 - DISCRETAS. Reciben valores aislados.
 - CONTÍNUAS. Pueden recibir cualquier valor de un intervalo.
- Ejemplos:
 - Atributos: Color de ojos (nominal), Calificación (ordinal).
 - Variables Discretas: Número de hijos, coches, etc.

- Variables Continuas: Estatura, Peso, etc.
- El estudio de una variable estadística X comienza por extraer una muestra de tamaño n de la población, y clasificar los valores x_i de la variable obtenidos en la muestra.
- Existen dos formas de clasificar estos valores:
 - SIN AGRUPAR. Ordenar todos los valores obtenidos en la muestra de menor a mayor. Se utiliza con variables discretas y para muestras con pocos valores diferentes.
 - AGRUPADOS. Agrupar los valores en clases (intervalos) y ordenar dichas clases de menor a mayor. Se utiliza con variables discretas con muchos valores diferentes, y con variables continuas.



Recuento de Frecuencias



Frecuencias Absolutas y Relativas

- A continuación, para cada valor x_i obtenido se calcula:
 - FRECUENCIA ABSOLUTA (n_i). Es el número de individuos de la muestra que presentan dicho valor.
 - FRECUENCIA RELATIVA (f_i). Es la frecuencia absoluta dividida entre el tamaño de la muestra.
- FRECUENCIA ABSOLUTA ACUMULADA (N_i). Es la suma de las i primeras frecuencias absolutas.

$$f_i = \frac{n_i}{n}$$

$$N_i = n_1 + \dots + n_i$$

- FRECUENCIA RELATIVA ACUMULADA (F_i). Es la frecuencia absoluta acumulada dividida por el tamaño de la muestra.

$$F_i = \frac{N_i}{n}$$

Distribución de Frecuencias

- Al conjunto de valores observados en la muestra junto a sus respectivas frecuencias se le denomina DISTRIBUCIÓN DE LA MUESTRA.
- La distribución de una muestra suele reflejarse mediante una TABLA DE FRECUENCIAS.

Valores de X	Frecuencia Absoluta	Frecuencia Relativa	Frec. Abs. Acumulada	Frec. Rel. Acumulada
x_1	n_1	f_1	N_1	F_1
...
x_i	n_i	f_i	N_i	F_i
...
x_k	n_k	f_k	$N_k=n$	$F_k=1$
	n	1		

Estadísticos

- La tabla de frecuencias contiene toda la información sobre la muestra, pero no suele aclarar el comportamiento de la variable estadística estudiada.
- Para reflejar este comportamiento se calculan varios números llamados ESTADÍSTICOS que son indicadores de ciertos aspectos de la distribución objeto de estudio.
- Los estadísticos se clasifican en tres grupos:
 - Medidas de Posición
 - Medidas de Dispersión
 - Medidas de Forma

Medidas de Posición

- Pueden ser de dos tipos:
 - Medidas de Tendencia Central. Determinan valores alrededor de los cuales se agrupa la distribución. Las más importantes son:
 - › Media Aritmética
 - › Mediana
 - › Moda
 - Otras Medidas de Posición. Dividen la distribución en partes con el mismo número de observaciones. Las más importantes son:
 - › Cuantiles: Cuartiles, Deciles, Percentiles.

Media Aritmética \bar{x}

- Es la suma de todos los resultados de las observaciones, dividida por el tamaño de la muestra.

$$\bar{x} = \frac{\overset{n_1}{x_1} + \dots + \overset{n_k}{x_k}}{n}$$

- A partir de la tabla de frecuencias puede calcularse como

$$\bar{x} = \frac{x_1 n_1 + \dots + x_k n_k}{n} = \frac{\sum_{i=1}^k x_i n_i}{n}$$

$$\bar{x} = x_1 f_1 + \dots + x_k f_k = \sum_{i=1}^k x_i f_i$$

- No puede calcularse para los atributos.

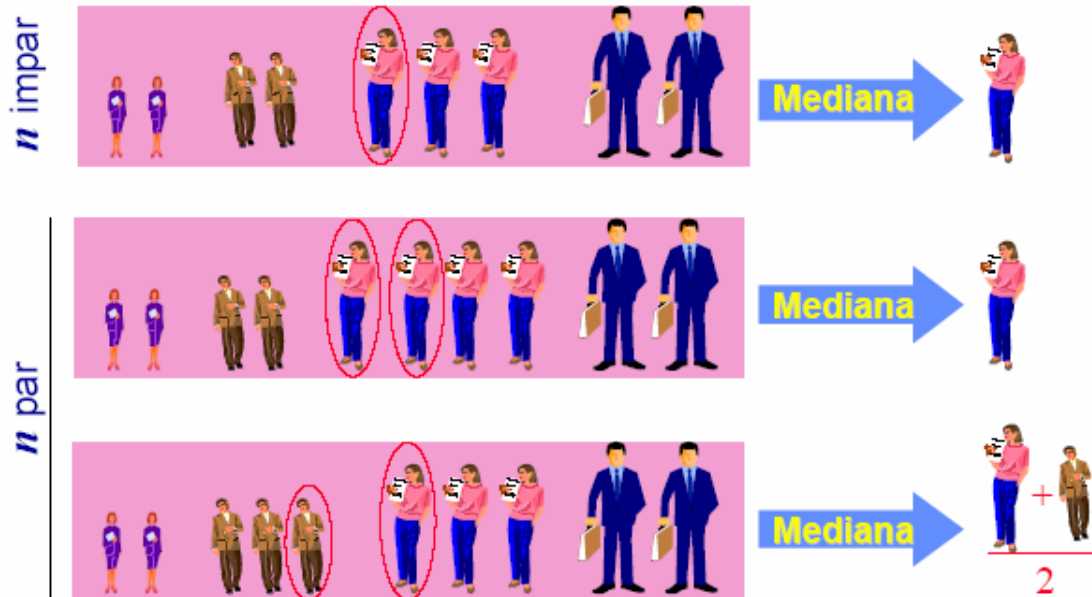
Mediana Me

- Es el valor de la variable que, suponiendo a los individuos de la muestra dispuestos en orden creciente, divide a la distribución en dos partes iguales.
- No tiene sentido en atributos nominales.
- La mediana cumple

$$N_{Me} = n/2 \quad \text{y} \quad F_{Me} = 0,5.$$

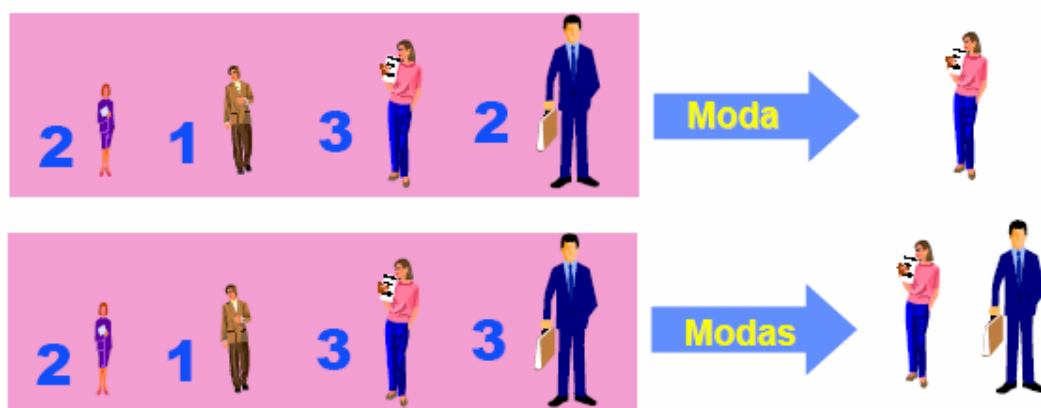
- El cálculo de la mediana se realiza de forma distinta en variables discretas y en variables continuas.

- Para datos sin agrupar pueden darse varios casos:



Moda Mo

- Es el valor de la variable más frecuente en la muestra.
- En ocasiones puede haber más de una moda.
- Con variables continuas se toma como clase modal la clase con mayor frecuencia en la muestra.



Cuantiles

- Son valores de la variable que dividen la distribución, supuesta ordenada de menor a mayor, en partes que contienen el mismo número de observaciones.
- Los más utilizados son:
 - CUARTILES: Dividen la distribución en 4 partes iguales.

$$C_1, C_2, C_3$$

- DECILES: Dividen la distribución en 10 partes iguales.

$$D_1, \dots, D_9$$

- PERCENTILES: Dividen la distribución en 100 partes iguales.

$$P_1, \dots, P_{99}$$

Medidas de Dispersión

- Recogen información respecto a la mayor o menor concentración de los valores de la variable, alrededor de algún valor central.
- Las más empleadas son:
 - Recorrido.
 - Rango Intercuartílico.
 - Varianza.
 - Desviación Típica.
 - Coeficiente de Variación.

- El **RECORRIDO** es la diferencia entre los valores máximo y mínimo que toma la variable en la muestra.

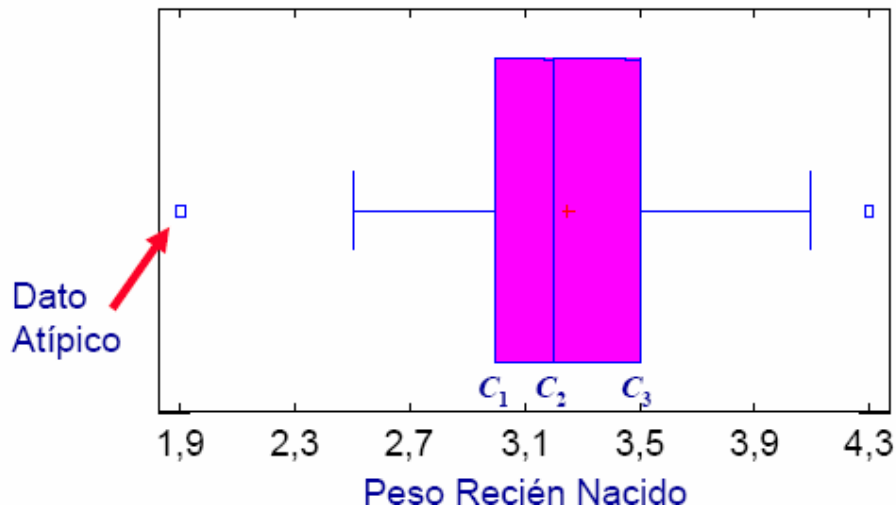
$$Re = \max\{x_i \mid i = 1, \dots, n\} - \min\{x_i \mid i = 1, \dots, n\}$$

- El **RANGO INTERCUARTÍLICO** es la diferencia entre el tercer y el primer cuartil.

$$RI = C_3 - C_1$$

Diagrama de Cajas

- Representa la distribución de los tres cuartiles, y también del rango intercuartílico.
- También se utiliza para detectar los valores atípicos (*outliers*).



Varianza y Desviación Típica

- La VARIANZA es la media de los cuadrados de las desviaciones de los valores observados respecto de la media de la muestra.
- A partir de la tabla de frecuencias puede calcularse como:

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

- Una forma equivalente de calcular la varianza es:

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

- La DESVIACIÓN TÍPICA es la raíz cuadrada positiva de la varianza.

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

- Cuanto mayor sea la dispersión entre las observaciones de la muestra, mayores serán la varianza y la desviación típica, y menos representativa la media.



Igual Media



Coeficiente de Variación C_v

- Se define como el cociente entre la desviación típica y el valor absoluto de la media de la muestra.

$$C_v = \frac{s}{|\bar{x}|}$$

- Al tratarse de una medida adimensional se utiliza para comparar la dispersión de muestras de distintas variables.
- También se utiliza para comparar lo representativa que es la media en muestras distintas. Para valores pequeños la media es más representativa que para valores grandes.

Medidas de Forma

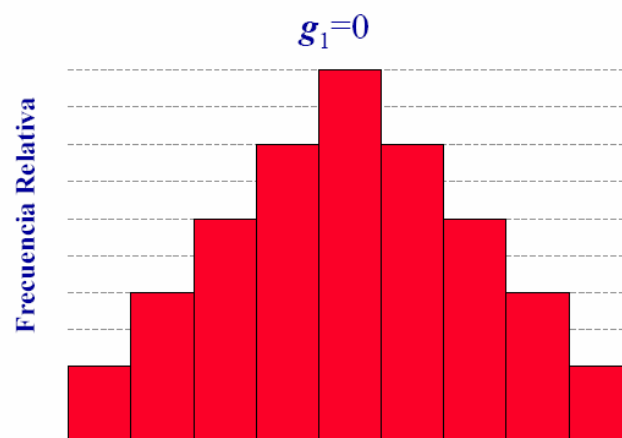
- Son medidas que tratan de caracterizar aspectos de la forma de la distribución de una muestra.
- Los aspectos más relevantes son:
 - **Simetría.** El estadístico más utilizado es el Coeficiente de Asimetría de Fisher.
 - **Apuntamiento.** El estadístico más utilizado es el Coeficiente de Apuntamiento o Curtosis.

Coefficiente de Asimetría de Fisher

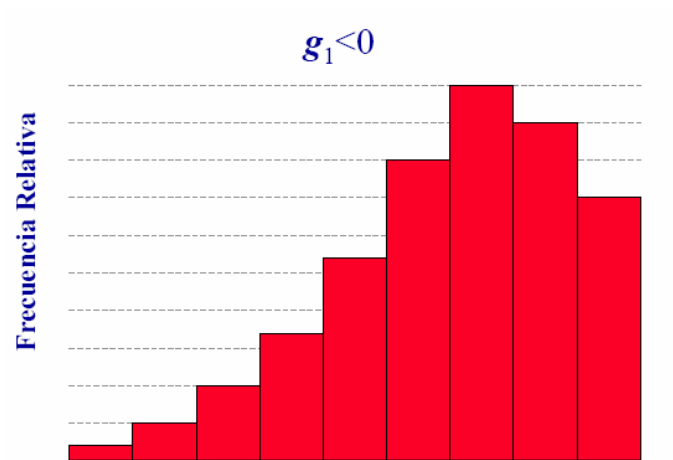
- Viene dado por la expresión siguiente:

$$g_1 = \frac{\sum_{i=1}^k (x_i - \bar{x})^3 f_i}{s^3}$$

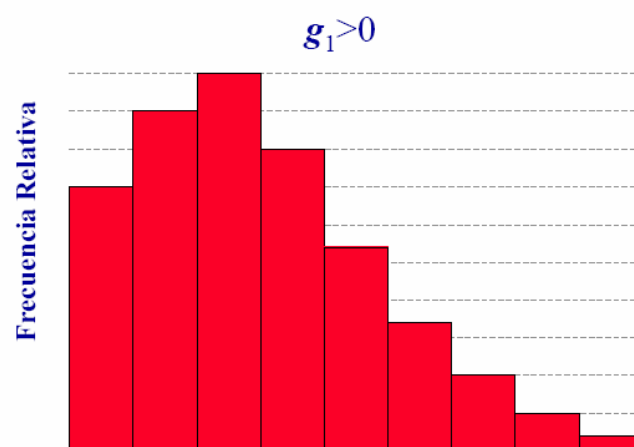
- Representa el grado de simetría de las observaciones de la muestra con respecto a la media, de manera que:
 - $g_1=0$ indica que hay el mismo número de valores a la derecha y a la izquierda de la media (**simétrica**).
 - $g_1<0$ indica que la mayoría de los valores son mayores que la media (**asimétrica a la izquierda**).
 - $g_1>0$ indica que la mayoría de los valores son menores que la media (**asimétrica a la derecha**).

Distribución Simétrica

Distribución Asimétrica a la Izquierda



Distribución Asimétrica a la Derecha



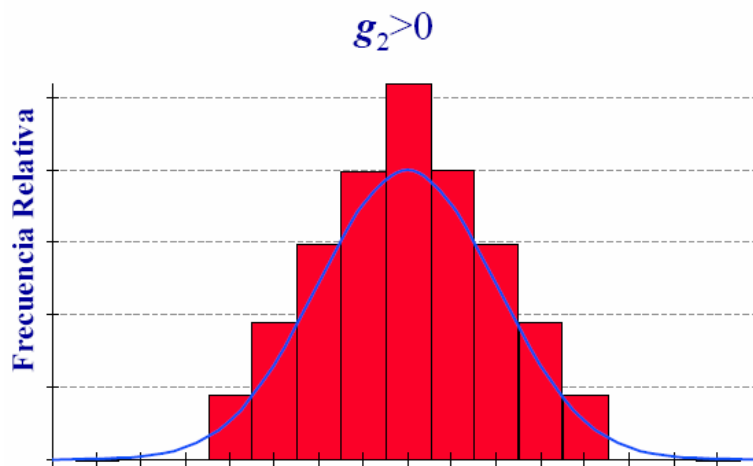
Coefficiente de Apuntamiento o Curtosis g_2

- Viene dado por la expresión siguiente:

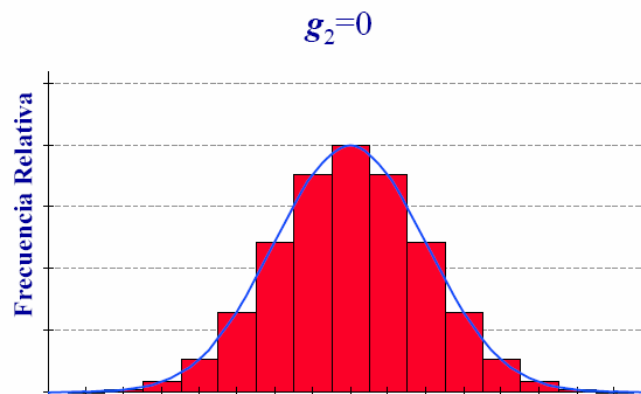
$$g_2 = \frac{\sum_{i=1}^k (x_i - \bar{x})^4 f_i}{s^4} - 3$$

- Representa el grado de apuntamiento de las observaciones de la muestra con respecto a una distribución normal, de manera que:
 - $g_2=0$ indica que la distribución tiene un apuntamiento normal (*mesocúrtica*).
 - $g_2<0$ indica que la distribución es menos apuntada que la normal (*platicúrtica*).
 - $g_2>0$ indica que la distribución es más apuntada que la normal (*leptocúrtica*).

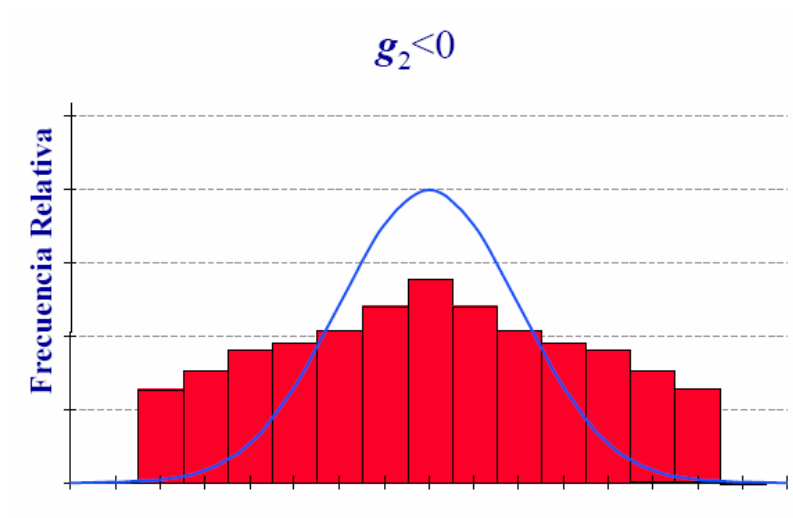
Distribución Leptocúrtica



Distribución Mesocúrtica



Distribución Platicúrtica



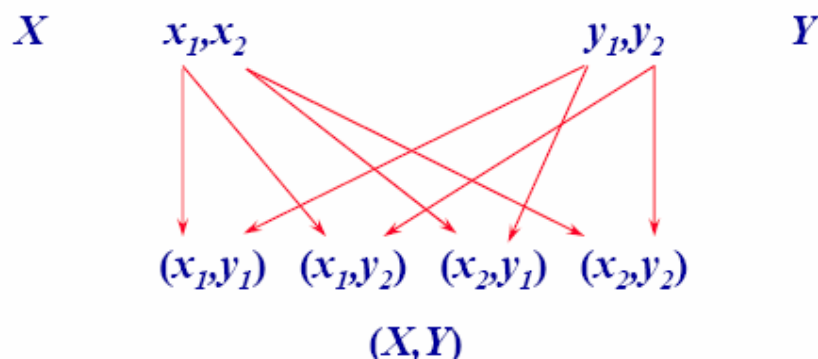
2 Regresión y Correlación.

Variables Dependientes

- A menudo se suele estudiar un conjunto de variables estadísticas que se suponen relacionadas.
- El objetivo de la estadística en este caso es dar medidas del grado y del tipo de dependencia entre dichas variables.
- Generalmente, se considera una VARIABLE DEPENDIENTE Y que se supone relacionada con otras variables X_1, \dots, X_n llamadas VARIABLES INDEPENDIENTES.
- El caso más simple es el de una sola variable independiente, y hablaremos de DEPENDENCIA SIMPLE. Para más de una variable independiente hablaremos de DEPENDENCIA MÚLTIPLE.

Variable Estadística Bidimensional

- Al estudiar la dependencia simple entre dos variables X e Y , no se pueden estudiar sus distribuciones por separado, sino que hay que estudiarlas en conjunto.
- Para ello, conviene definir una VARIABLE ESTADÍSTICA BIDIMENSIONAL (X, Y) , cuyos valores serán pares formados por los valores de las variables estadísticas unidimensionales X e Y .



Distribución Bidimensional

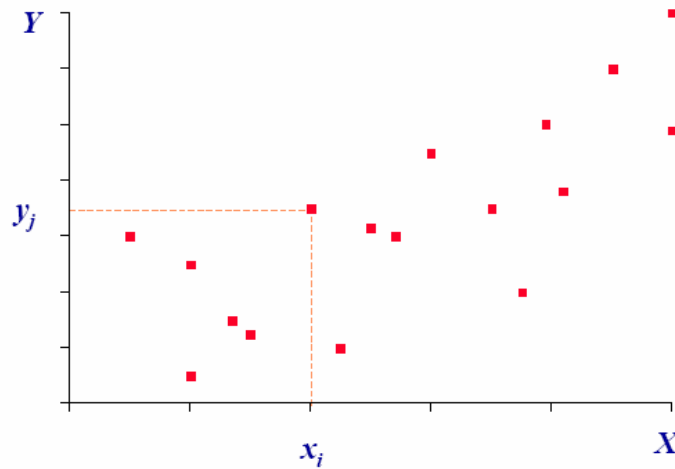
- Al conjunto de valores de la variable bidimensional y sus respectivas frecuencias se le denomina DISTRIBUCIÓN CONJUNTA.
- La distribución conjunta de una variable bidimensional se refleja mediante una TABLA DE FRECUENCIAS BIDIMENSIONAL.

$X \backslash Y$	y_1	...	y_j	...	y_q
x_1	n_{11}	...	n_{1j}	...	n_{1q}
...
x_i	n_{i1}	...	n_{ij}	...	n_{iq}
...
x_p	n_{p1}	...	n_{pj}	...	n_{pq}

Diagrama de Dispersión

- La representación gráfica que más se utiliza en el estudio de la dependencia de dos variables es el DIAGRAMA DE DISPERSIÓN.
- Consiste en representar sobre un plano coordenado los puntos que se corresponden con los valores (x_p, y_j) de la variable bidimensional (X, Y) . El conjunto de todos estos puntos recibe el nombre de NUBE DE PUNTOS.
- En un diagrama de dispersión sólo se recogen los valores observados en la muestra, no las frecuencias de los mismos. Para reflejar las frecuencias tendríamos que recurrir a otro tipo de representación.

Diagrama de Dispersión



Distribuciones Marginales

- A cada una de las distribuciones de las variables unidimensionales que conforman una variable bidimensional se les llama DISTRIBUCIONES MARGINALES.
- Las distribuciones marginales se forman sumando las frecuencias de los valores por filas y columnas.

X/Y	y_1	...	y_j	...	y_q	n_x
x_1	n_{11}	...	n_{1j}	...	n_{1q}	$n_{1\cdot}$
...
x_i	n_{i1}	...	n_{ij}	...	n_{iq}	$n_{i\cdot}$
...
x_p	n_{p1}	...	n_{pj}	...	n_{pq}	$n_{p\cdot}$
n_y	$n_{\cdot 1}$...	$n_{\cdot j}$...	$n_{\cdot q}$	n

Covarianza s_{xy}

- Uno de los estadísticos que más se utiliza en el estudio de la dependencia simple es la COVARIANZA.
- Se obtiene a partir de la tabla bidimensional como:

$$s_{xy} = \frac{\sum_{i=1}^p \sum_{j=1}^q (x_i - \bar{x})(y_j - \bar{y})n_{ij}}{n}$$

- y también como

$$s_{xy} = \frac{\sum_{i=1}^p \sum_{j=1}^q x_i y_j n_{ij}}{n} - \bar{x} \bar{y}$$

- La covarianza también puede utilizarse para medir la dispersión de la nube de puntos.

Regresión

- La regresión es una técnica para estudiar el tipo de dependencia que existe entre una variable dependiente y una o varias variables independientes.
- El objetivo es determinar una ecuación mediante la que pueda estimarse el valor de la variable dependiente en función de los valores de las independientes.
- Cuando sólo exista una variable independiente, hablaremos de REGRESIÓN SIMPLE, mientras que cuando haya más de una variable independiente, hablaremos de REGRESIÓN MÚLTIPLE.

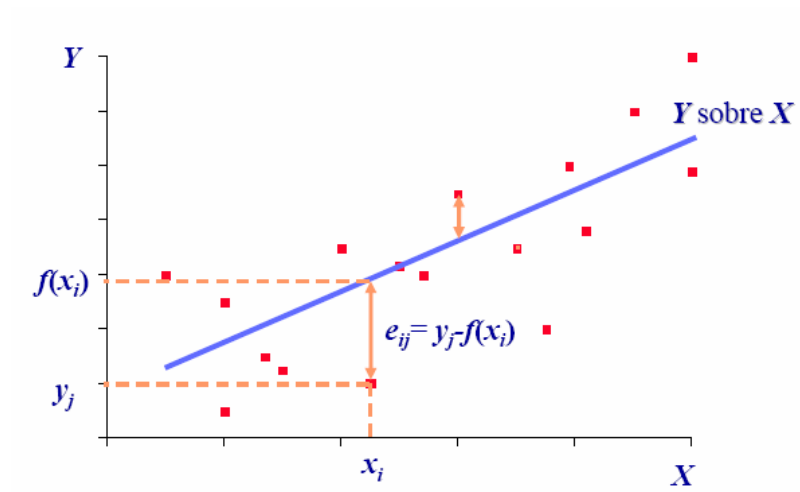
- A la vista de la nube de puntos representativa de los valores (x_i, y_j) de la distribución, se elige la familia de curvas que mejor se adapte a ella.
- Podemos hablar de regresión de Y sobre X (Y como función de X), o de regresión de X sobre Y (X como función de Y).
- La regresión toma nombres diferentes según el tipo de curva elegido:
 - Regresión Lineal: $y = a + bx$
 - Regresión Parabólica: $y = a + bx + cx^2$
 - Regresión Exponencial: $y = ae^{bx}$
 - Regresión Potencial: $y = ax^b$
 - Regresión Logarítmica: $y = a + b \log x$

Método de Mínimos Cuadrados

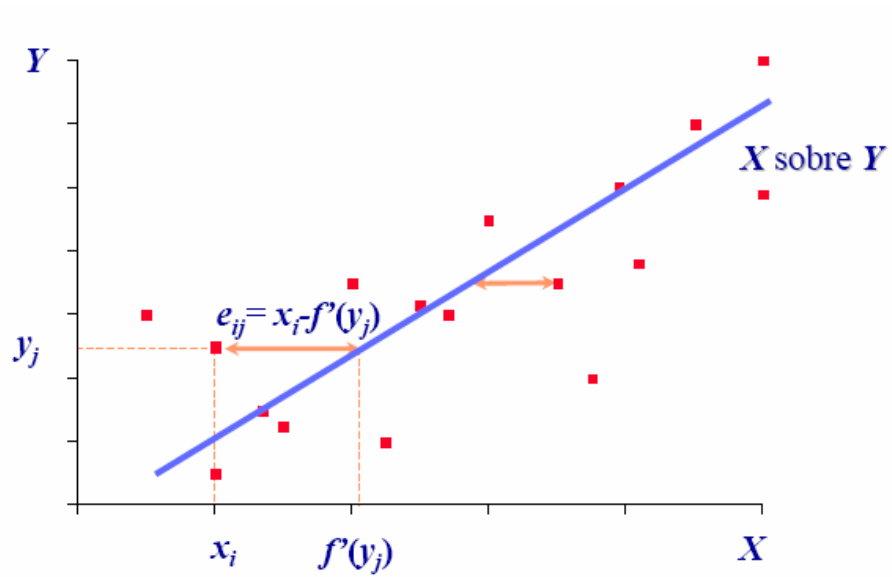
- Una vez elegida la familia de curvas que mejor se adapta a la nube de puntos, se determina, dentro de dicha familia, la curva que mejor se ajusta a la distribución.
- El criterio del mejor ajuste que utilizaremos es el método de MÍNIMOS CUADRADOS. Para ello se minimiza la función en la que se recogen las distancias (residuos) al cuadrado entre los valores de la variable y los teóricos predichos por la familia de curvas elegida.
- Para la regresión simple de Y sobre X , la función a minimizar adopta la forma:

$$\Phi = \sum_{i=1}^p \sum_{j=1}^q [y_j - f(x_i, a_1, a_2, \dots, a_n)]^2 n_{ij}$$

Residuos o Errores en Y



Residuos o Errores en X



Recta de Regresión de Y sobre X

- En el supuesto de regresión lineal simple el tipo de curva elegido para ajustar la nube de puntos es una línea recta.
- En el caso de la regresión de Y sobre X , la función que habría que minimizar es la siguiente:

$$\Phi = \sum_{i=1}^p \sum_{j=1}^q (y_j - a - bx_i)^2 n_{ij}$$

- Los valores de los parámetros, a y b , que hacen mínima esta suma son:

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

- Y la recta de regresión de Y sobre X tendrá por ecuación:

$$y - \bar{y} = \frac{s_{xy}}{s_x^2} (x - \bar{x})$$

Recta de Regresión de X sobre Y

- Para la recta de regresión de X sobre Y :

$$\Phi = \sum_{i=1}^p \sum_{j=1}^q (x_i - a' - b'y_j)^2 n_{ij}$$

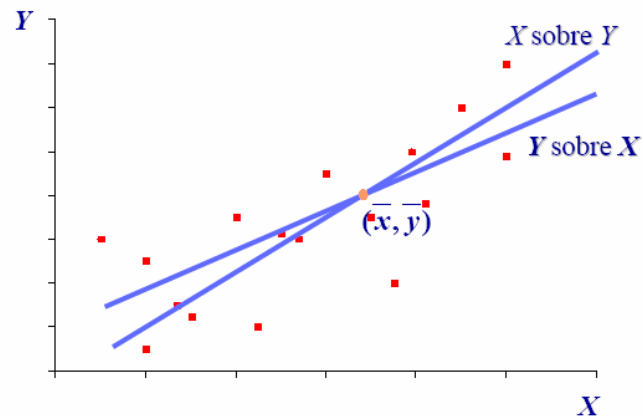
$$a' = \bar{x} - \frac{s_{xy}}{s_y^2} \bar{y} \quad b' = \frac{s_{xy}}{s_y^2}$$

- Y la ecuación de la recta de regresión de X sobre Y :

$$x - \bar{x} = \frac{s_{xy}}{s_y^2} (y - \bar{y})$$

- En función de las ecuaciones obtenidas se comprueba fácilmente que ambas rectas de regresión pasan por el punto (\bar{x}, \bar{y}) , denominado CENTRO DE GRAVEDAD de la distribución.

Rectas de Regresión



Coeficiente de Regresión

- En la ecuación de la recta de regresión de Y sobre X , la pendiente de la recta se conoce como COEFICIENTE DE REGRESIÓN DE Y SOBRE X

$$b_{yx} = \frac{s_{xy}}{s_x^2}$$

- El signo de la pendiente depende de la covarianza:
 - $s_{xy} > 0$. Dependencia lineal positiva (a medida que crece X , también crece Y).
 - $s_{xy} < 0$. Dependencia lineal negativa (a medida que crece X , decrece Y).
- El coeficiente de regresión mide el número de unidades que aumenta o disminuye la variable dependiente por cada unidad que aumenta la variable independiente.

Correlación

- Una vez fijado el tipo de dependencia entre dos o más variables, pasamos a estudiar el grado de dependencia para dicho tipo. De esto se encarga la CORRELACIÓN.
- Para cada tipo de dependencia, tendremos el correspondiente tipo de correlación.
- La correlación se basa en el estudio de los residuos de la nube de puntos. Cuanto menores sean éstos, más se ajustaría la curva de regresión a los puntos, y más intensa será la correlación.

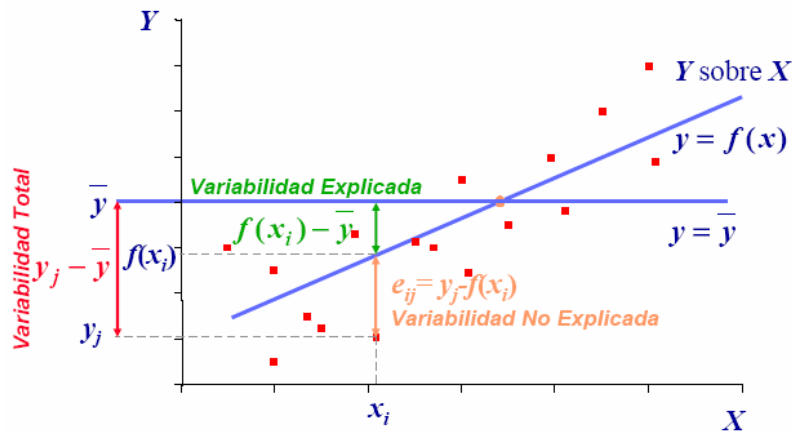
Varianza Residual

- Una primera medida de la bondad del ajuste de regresión es la VARIANZA RESIDUAL, que se calcula como la media de los cuadrados de los residuos.

$$s_{ry}^2 = \frac{\sum_{i=1}^p \sum_{j=1}^q (y_j - f(x_i))^2 n_{ij}}{n} = \frac{\sum_{i=1}^p \sum_{j=1}^q e_{ij}^2 n_{ij}}{n}$$

- Cuanto más alejados estén los puntos de la curva de regresión, mayor será la varianza residual y menor la dependencia. Esto hace de la varianza residual una medida inversamente proporcional al grado de dependencia.

Variabilidad Explicada frente a Variabilidad No Explicada



Coeficiente de Determinación R^2

- A partir de la varianza residual se define el COEFICIENTE DE DETERMINACIÓN como:

$$R^2 = 1 - \frac{s_{ry}^2}{s_y^2}$$

- El coeficiente de determinación mide la proporción de variabilidad de la variable dependiente explicada por el modelo de regresión utilizado, y por tanto, toma valores entre 0 y 1.
- Cuanto mayor sea R^2 , mayor será el grado de dependencia entre X e Y expresado por el modelo de regresión.
 - $R^2=0$. Independencia (según el modelo de regresión).
 - $R^2=1$. Dependencia funcional (correlación perfecta).

Coeficiente de Correlación R

- A la raíz del cuadrada del coeficiente de determinación se le conoce como COEFICIENTE DE CORRELACIÓN:

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

- De las dos raíces se toma siempre la del signo igual al de la covarianza, y toma valores

$$-1 \leq R \leq 1$$

- El coeficiente de correlación también es una medida del grado de dependencia entre X e Y que será mayor cuanto más próximo esté a 1 o a -1:
 - $R=0$. Independencia (según el modelo de regresión, pero, *¡puede haber relación de otro tipo!*).
 - $R=1$. Dependencia funcional positiva.
 - $R=-1$. Dependencia funcional negativa.

Coeficiente de Correlación Lineal r

- En el caso del estudio de la dependencia lineal, el coeficiente de correlación R se denomina COEFICIENTE DE CORRELACIÓN LINEAL y se nota r .
- En este caso, la fórmula general del coeficiente de correlación se convierte en

$$r = \frac{S_{xy}}{S_x S_y}$$

3 Cálculo de Probabilidades.

- El estudio de una característica en una población se realiza a través de experimentos aleatorios.
- Un EXPERIMENTO ALEATORIO es aquel en el que se conoce cuál es el conjunto de resultados posibles antes de su realización pero se desconoce cuál será el resultado concreto del mismo.
- Al conjunto E de todos los posibles resultados individuales de un experimento aleatorio se le llama ESPACIO MUESTRAL.
- Cualquier subconjunto del espacio muestral es un SUCESO ALEATORIO
- Existen distintos tipos de sucesos:
 - SUCESO IMPOSIBLE. Es el subconjunto vacío \emptyset . El suceso imposible nunca ocurre.
 - SUCESOS ELEMENTALES. Son aquellos formados por un solo elemento.
 - SUCESOS COMPUESTOS. Son aquellos formados por dos o más elementos.
 - SUCESO SEGURO o CIERTO. Es el espacio muestral E . El suceso seguro siempre ocurre.

Operaciones entre sucesos

- Dados dos sucesos A y B de un cierto experimento aleatorio, definimos:
 - UNIÓN DE SUCESOS. Es el suceso $A \cup B$ que ocurre siempre que ocurre A o B .
 - INTERSECCIÓN DE SUCESOS. Es el suceso $A \cap B$ que ocurre siempre que ocurre A y B .
 - SUCESO CONTRARIO. Es el suceso \overline{A} que ocurre siempre que no ocurre A .
 - DIFERENCIA DE SUCESOS. Es el suceso $A - B$ que ocurre siempre que ocurre A y no ocurre B .

- Diremos que dos sucesos son INCOMPATIBLES si se cumple que $A \cap B = \emptyset$.

Concepto de Probabilidad

- Para fijar el concepto de probabilidad podemos partir de un ejemplo clásico: supongamos que disponemos de una moneda no trucada y realizamos el experimento de lanzar la moneda anotando el suceso cara o cruz. Es un hecho experimental comprobado que ante 10 lanzamientos de la moneda podemos obtener 8 caras, pero si realizamos 5000 lanzamientos, el número de caras será muy próximo a 2500.
- El ejemplo anterior se concreta en el resultado experimental enunciado como LEY DE LOS GRANDES NÚMEROS: la frecuencia relativa de un suceso tiende a estabilizarse en torno a un valor a medida que el número de pruebas del experimento crece indefinidamente.
- A ese valor al que tiende la frecuencia relativa del suceso lo denominamos PROBABILIDAD.

Probabilidad “a posteriori” y “a priori”

- La definición de probabilidad que se apoya en la Ley de los Grandes Números tiene el grave inconveniente de que parte de la realización del experimento en múltiples ocasiones, mientras que a efectos prácticos buscamos una definición “a priori”, es decir, intentamos dar la probabilidad del suceso sin necesidad de realizar el experimento.
- Apoyándose en las propiedades de las frecuencias relativas se enuncia la definición axiomática de probabilidad, que pretende resolver el problema que plantea el concepto de probabilidad “a posteriori”.

Definición Axiomática de Probabilidad

- Se llama probabilidad a toda aplicación que asocia a cada suceso A del espacio de sucesos de un experimento aleatorio, un número real $P(A)$, que cumple los siguientes axiomas:

- 1 La probabilidad de un suceso cualquiera es positiva o nula:

$$P(A) \geq 0.$$

- 2 La probabilidad de la unión de dos sucesos incompatibles es igual a la suma de las probabilidades de cada uno de ellos:

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

- 3 La probabilidad del suceso cierto es igual a la unidad:

$$P(E) = 1.$$

Consecuencias de los axiomas

- Algunos de los resultados fundamentales que se pueden probar con los axiomas de probabilidad son los siguientes:

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

- $P(\emptyset) = 0.$

- Si A está contenido en B entonces: $P(A) \leq P(B).$

- $P(A) \leq 1.$

- Si A y B son sucesos compatibles, es decir, su intersección no es vacía, entonces:

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

- Si el suceso A está compuesto por los sucesos elementales e_1, e_2, \dots, e_n :

$$P(A) = \sum_{i=1}^n P(e_i)$$

- Como consecuencia del último enunciado de la diapositiva anterior se puede demostrar fácilmente la REGLA DE LAPLACE, la cual afirma que si todos los sucesos elementales de un experimento aleatorio son **equiprobables**, la probabilidad de cualquier suceso de dicho experimento será el cociente entre el número de sucesos elementales que comprende y el número total de sucesos elementales del experimento.

$$P(A) = \frac{\text{casos favorables}}{\text{casos posibles}}$$

- La Regla de Laplace tiene multitud de aplicaciones en los conocidos problemas sobre dados, cartas..., aunque su aplicación está siempre supeditada a que los sucesos elementales sean equiprobables (*¡cuidado con los dados o las cartas "trucados"!*).

Probabilidad Condicionada

- A menudo nos interesa hallar la probabilidad de un suceso, A , supeditada a que previamente ha ocurrido otro suceso, B . Por ejemplo, ¿cuál es la probabilidad de aprobar estadística perteneciendo al grupo 4?
- La PROBABILIDAD CONDICIONADA del suceso A , supuesto ha ocurrido el suceso B , se escribe como $P(A/B)$, y se define por:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad \text{si } P(B) \neq 0$$

Sucesos Independientes

- Decimos que el suceso A es INDEPENDIENTE del suceso B , si el suceso B no influye para nada en la probabilidad del A .

$$P(A/B) = P(A)$$

- Si A es independiente de B , entonces B es independiente de A , ($P(B/A)=P(B)$).
- Si dos sucesos son independientes, se cumple la importante relación:

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

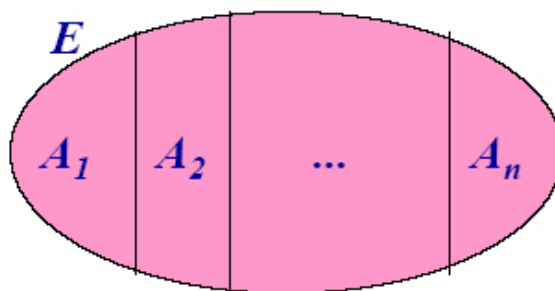
Sistema Completo de Sucesos

- Una colección de sucesos A_1, A_2, \dots, A_n constituye un SISTEMA COMPLETO DE SUCESOS para un determinado experimento aleatorio si se verifican:

1 $A_1 \cup A_2 \cup \dots \cup A_n = E$

2 A_1, A_2, \dots, A_n son incompatibles dos a dos.

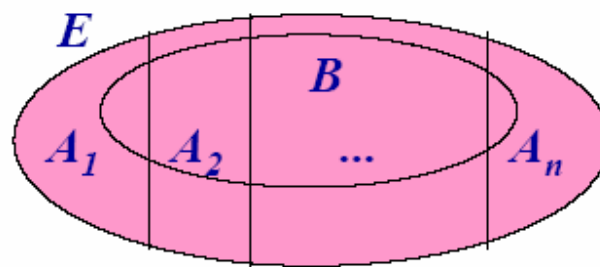
- Gráficamente:



- Conocer las probabilidades de un determinado suceso B en cada una de las partes de un sistema completo A_1, A_2, \dots, A_n , puede ser útil para calcular la probabilidad global de dicho suceso.
- En las condiciones expuestas en el punto anterior el TEOREMA DE PROBABILIDAD TOTAL afirma:

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

- Gráficamente:



Teorema de Bayes

- Los sucesos de un sistema completo de sucesos A_1, A_2, \dots, A_n , también pueden verse como las distintas hipótesis ante un determinado hecho B .
- En estas condiciones resulta útil poder calcular las probabilidades a posteriori $P(A_i/B)$ de cada una de las hipótesis. (Ej. Problemas de diagnóstico).
- El TEOREMA DE BAYES afirma que si tenemos un sistema completo de sucesos A_1, A_2, \dots, A_n , y conocemos $P(B/A_i)$ para todo $i=1, \dots, n$, entonces:

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

4 Variables Aleatorias.

- Una VARIABLE ALEATORIA X es una función real definida sobre espacio muestral E de un experimento aleatorio.

$$X: E \rightarrow \mathbb{R}$$

- El conjunto de posibles valores que puede tomar la variable aleatoria se llama RANGO O RECORRIDO.
- Definimos una variable aleatoria asignando a cada resultado del experimento un número:
 - Si el resultado del experimento es numérico, los valores de la variable coincidirán con el resultado del mismo.
 - Si el resultado del experimento es cualitativo, hacemos corresponder a cada cualidad un número arbitrariamente.

Tipos de Variables Aleatorias

- Las variables aleatorias se clasifican en dos tipos: discretas y continuas.
- Decimos que una variable aleatoria es DISCRETA (V.A.D.) cuando su recorrido es finito o infinito numerable, es decir, X sólo puede tomar los valores aislados x_1, x_2, \dots, x_n . Ejemplos: nº de hijos de una familia, nº de usuarios de una línea de autobuses, medicamentos distintos consumidos por una persona al día...
- Decimos que una variable aleatoria es CONTINUA (V.A.C.) cuando su recorrido lo forman todos los infinitos puntos posibles de un intervalo. Ejemplos: Peso y estatura de una persona, tiempo de respuesta ante un medicamento...

V.A.D.:
Función de Probabilidad $f(x)$

- Como los valores de una variable aleatoria están asociados a los sucesos elementales del correspondiente experimento aleatorio, cada valor tendrá asociada una probabilidad.
- La función $f(x)$ que asigna a cada valor x_i de la variable su probabilidad $P(x_i)$ se llama FUNCIÓN DE PROBABILIDAD:

$$f(x_i) = P(X = x_i) \quad i = 1, \dots, n$$

- Y se cumple:

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

V.A.D.:
Distribución de Probabilidad

- Los valores de una V.A.D. junto con sus respectivos valores de la función de probabilidad, dan lugar a la DISTRIBUCIÓN DE PROBABILIDAD de la V.A.D.
- Suele expresarse en forma de tabla:

X	x_1	x_2	...	x_n	Σ
$f(x)$	$f(x_1)$	$f(x_2)$...	$f(x_n)$	1

- Al igual que la distribución de frecuencias de una variable reflejaba cómo se distribuía la variable en una muestra, la distribución de probabilidad de una variable aleatoria sirve para reflejar cómo se distribuyen los valores de dicha variable en toda la población.

V.A.D.:
Función de Distribución $F(x)$

- Una forma equivalente de caracterizar la distribución de probabilidad de una V.A.D. es mediante probabilidades acumuladas.
- Llamaremos **FUNCIÓN DE DISTRIBUCIÓN** de una variable aleatoria X , y la notaremos $F(x)$, a la función que asigna a cada valor x de la variable, la probabilidad de que X tome un valor menor o igual que x .

$$F(x) = p(X \leq x)$$

Un Ejemplo

- Supongamos el experimento aleatorio de lanzar dos monedas y anotar el número de caras obtenidas:
 - $Re(X) = \{0,1,2\}$
 - $P(X=0) = 0,25$, $P(X=1) = 0,5$, $P(X=2) = 0,25$.
- La función de probabilidad y de distribución son:

X	0	1	2
$f(x)$	0,25	0,5	0,25
$F(x)$	0,25	0,75	1

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

Parámetros Descriptivos

- A partir de la distribución de probabilidad se pueden calcular diversos parámetros descriptivos de determinados aspectos de la distribución (centralización, dispersión...). Los más importantes son

- MEDIA O ESPERANZA MATEMÁTICA:

$$E[X] = \mu = \sum_{i=1}^n x_i f(x_i)$$

- VARIANZA:

$$V[X] = \sigma^2 = \sum_{i=1}^n (x_i - \mu)^2 f(x_i) = \sum_{i=1}^n x_i^2 f(x_i) - \mu^2$$

- DESVIACIÓN TÍPICA:

$$D[X] = \sigma = +\sqrt{\sigma^2}$$

Modelos de Distribución

- En teoría, para obtener la distribución de probabilidad de una variable aleatoria en una población, es necesario conocer el valor de la variable en todos los individuos de la población, lo cual muchas veces es imposible.
- Sin embargo, dependiendo de la naturaleza del experimento, a veces es posible obtener la distribución de probabilidad de una variable aleatoria sin medirla en toda la población.
- Dependiendo del tipo de experimento, existen diferentes modelos de distribuciones discretas. Los más habituales son:
 - Distribución Uniforme.
 - Distribución Binomial.
 - Distribución de Poisson.

Distribución Uniforme $U(k)$

- Cuando por la simetría del experimento, todos los valores x_1, \dots, x_k de una V.A.D. X son igualmente probables, se dice que la variable sigue un modelo de distribución UNIFORME, y se denota por $X \sim U(k)$, donde k es el número de valores distintos que puede tomar la variable.
- Recorrido: $Re(X) = \{x_1, \dots, x_k\}$.
- Función de Probabilidad:

$$f(x) = \frac{1}{k}$$

- Su media y su desviación típica son:

$$\mu = \sum_{i=1}^k x_i \frac{1}{k} \quad \sigma = +\sqrt{\sum_{i=1}^k (x_i - \mu)^2 \frac{1}{k}}$$

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

- Consideremos una prueba en la que únicamente se pueden presentar los sucesos A , que llamaremos “éxito”, y su contrario \bar{A} , que llamaremos “fracaso”, con probabilidades $P(A)=p$ y $P(\bar{A})=1-p$, constantes.
- Consideremos el experimento que consiste en repetir n veces la prueba anterior. La V.A.D. X que recoge el número de veces que hemos obtenido “éxito” como resultado de las n repeticiones de la prueba, se dice que sigue un modelo de distribución BINOMIAL, y se denota por $X \sim B(n, p)$.
- Recorrido: $Re(X) = \{0, 1, \dots, n\}$.
- Función de Probabilidad:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

- Se cumple que :

$$\mu = np$$

$$\sigma = +\sqrt{np(1-p)}$$

Distribución de Poisson $P(\lambda)$

“Ley de los Casos Raros”

- Consideremos un experimento en el que observamos la
- Recorrido: $Re(X)=\{0,1,2,3,\dots\}$. Teóricamente infinito.
- Función de Probabilidad:

$$f(x) = e^{-\lambda} \frac{\lambda^x}{x!}$$

- Se cumple que:

$$\mu = \lambda \quad \sigma = +\sqrt{\lambda}$$

- La distribución de Poisson aparece como límite de la distribución Binomial si suponemos que el número n (repeticiones del experimento) es muy grande y p (probabilidad de “éxito”) muy pequeña. Por ello la distribución de Poisson recibe también el nombre de “LEY DE LOS CASOS RAROS”.
distribución de POISSON, y se denota por $X \sim P(\lambda)$.

$$\left. \begin{array}{l} n \rightarrow \infty \\ p \rightarrow 0 \end{array} \right\} \Rightarrow B(n, p) \rightarrow P(\lambda), \quad \lambda = np$$

V.A.C.: ***Función de Densidad $f(x)$***

- El proceso de medida del valor de una V.A.C. marca todo el desarrollo de la teoría que concierne a dichas variables. De hecho, no es posible conocer el valor exacto de una variable continua y tan sólo podemos clasificarlo dentro de un intervalo. Por lo tanto, tampoco podremos asignar una probabilidad no nula a un determinado valor, y hablaremos de PROBABILIDADES ASOCIADAS A INTERVALOS.
- La función que nos permitirá asociar una probabilidad a un intervalo dado de la V.A.C. es la función de densidad.
- Llamaremos FUNCIÓN DE DENSIDAD de una V.A.C. X a una función $f(x)$, integrable en \mathcal{R} , que cumple:

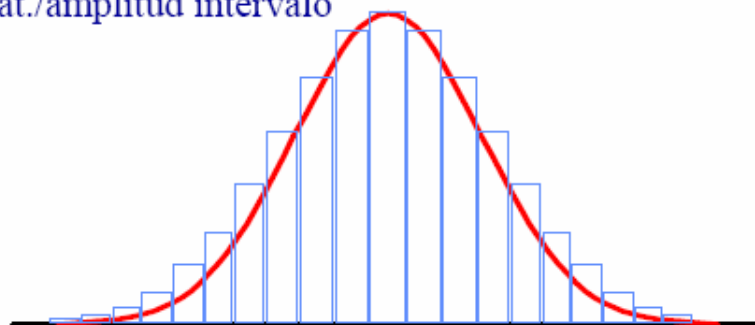
- 1 $f(x) \geq 0, \forall x \in \mathcal{R}$

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

$f(x)$ de Forma Intuitiva

- Si en el proceso de medida tomamos más y más medidas y las clasificamos en intervalos cada vez más pequeños, el polígono de frecuencias relativas tiende a una curva suave que representa la función de densidad de probabilidad de la V. A. C.

frec. relat./amplitud intervalo



Intervalos de X

V.A.C.: Función de Distribución $F(x)$

- Si $f(x)$ es una función de densidad de la V.A.C. X , la PROBABILIDAD de que el valor de la misma pertenezca al intervalo $[a, b]$, se define por:

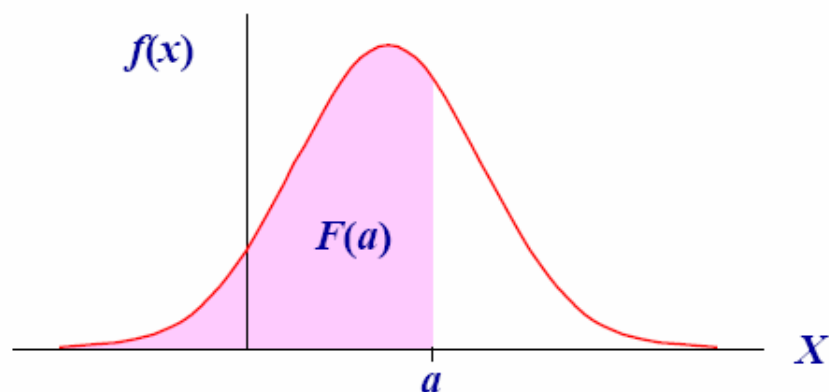
$$p(a \leq X \leq b) = \int_a^b f(x) dx$$

- Y la FUNCIÓN DE DISTRIBUCIÓN $F(x)$, asociada a $f(x)$, se define por:

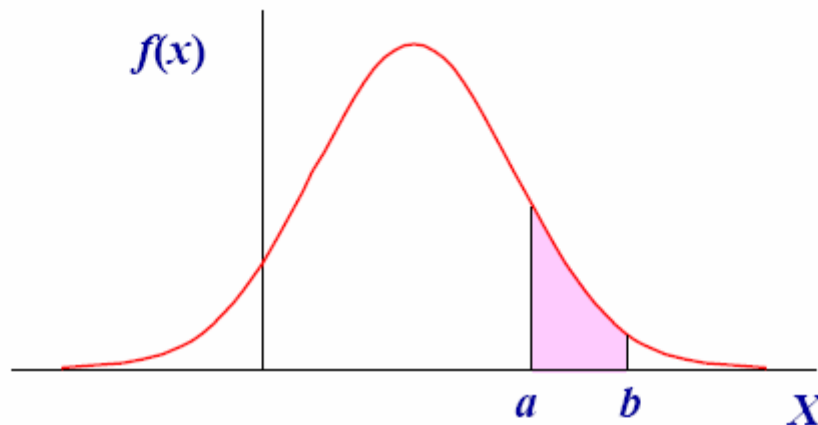
$$F(x) = p(X \leq x) = \int_{-\infty}^x f(t) dt$$

$F(x)$ de Forma Gráfica

- Supongamos un ejemplo en el que la función de densidad viene dada por la gráfica adjunta. La función de distribución en un punto $x=a$ es el área encerrada entre $f(x)$ y el eje de abscisas desde $-\infty$ hasta a :



Probabilidad como Área Debajo de la Función de Densidad



$$\begin{aligned}
 p(a \leq X \leq b) &= \int_a^b f(x) dx = \\
 &= \int_{-\infty}^b f(x) dx - \int_{-\infty}^a f(x) dx = \\
 &= F(b) - F(a)
 \end{aligned}$$

V.A.C.: Parámetros Descriptivos

- A partir de la función de densidad se pueden calcular diversos parámetros descriptivos de determinados aspectos de la distribución (centralización, dispersión...). Los más importantes son

– ESPERANZA MATEMÁTICA O MEDIA:

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

– VARIANZA:

$$V[X] = \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2$$

– DESVIACIÓN TÍPICA:

$$D[X] = \sigma = +\sqrt{\sigma^2}$$

Distribución Normal $N(\mu, \sigma)$

- Una variable aleatoria continua X cuyo recorrido es \mathfrak{R} , se dice que tiene una DISTRIBUCIÓN NORMAL O GAUSSIANA, de parámetros μ y σ , y se nota $X \sim N(\mu, \sigma)$ si su función de densidad es:

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

- Se cumple que la media de una variable aleatoria normal $N(\mu, \sigma)$ coincide con μ , y que la desviación típica coincide con σ .
- La gráfica de esta función es una especie de campana, conocida como CAMPANA DE GAUSS, centrada en μ .

Propiedades de la Normal

- La función de densidad es simétrica respecto a la media y por tanto, su coeficiente de asimetría es $g_1=0$.
- También es mesocúrtica, y por tanto, su coeficiente de apuntamiento vale $g_2=0$.
- La media, la mediana y la moda coinciden

$$\bar{x} = Me = Mo$$

- Tiende asintóticamente a 0 cuando $x \rightarrow \pm\infty$.
- Se cumple que

$$p(\mu - \sigma \leq X \leq \mu + \sigma) = 0.68$$

$$p(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0.95$$

$$p(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0.99$$

Importancia de la Normal **El Teorema Central del Límite**

- La Distribución Normal es la distribución continua más importante, ya que aproxima satisfactoriamente los resultados observados en muchos procesos de medición.
- Como justificación, el TEOREMA CENTRAL DEL LÍMITE establece que cuando los resultados de un experimento son debidos a un conjunto muy grande de causas independientes que actúan sumando sus efectos, se puede esperar que los mismos sigan una distribución normal.
- Por eso muchas variables de la naturaleza y sobre todo biológicas siguen distribuciones normales.

Tipificación

- Recordemos que:

$$p(a \leq X \leq b) = \int_{-\infty}^b f(x) dx - \int_{-\infty}^a f(x) dx = F(b) - F(a)$$

- Luego el problema del cálculo de probabilidades se reduce al conocimiento de $F(b)$ y $F(a)$. Para $X \sim N(\mu, \sigma)$:

$$F(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

- Si dispusiésemos de tablas que recogiesen el valor de la función de distribución $F(x)$, nos ahorraríamos el cálculo de esta integral. Sin embargo, esta integral depende de μ y σ , y por tanto necesitaríamos una tabla distinta para cada valor posible de estos parámetros, lo cual es inviable.
- La solución pasa por tipificar la variable aleatoria.

La Normal Estándar $N(0,1)$

- Al aplicar la transformación de tipificación a una variable normal $X \sim N(\mu, \sigma)$, se obtiene otra variable Z que se conoce como VARIABLE NORMAL TIPIFICADA o ESTÁNDAR:

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

- Por las propiedades de las transformaciones lineales, se cumple que Z también es una variable aleatoria normal, y tiene media 0 y desviación típica 1, es decir $Z \sim N(0, 1)$.
- La función de densidad de la normal estándar es

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

La Distribución Normal Tipificada

- Los valores de la función de distribución de la normal estándar suelen estar tabulados en una tabla.
- Así pues, para calcular cualquier probabilidad con una variable normal $X \sim N(\mu, \sigma)$ cualquiera, lo habitual es convertirla en una normal estándar mediante la tipificación, y recurrir a la tabla de su función de distribución para buscar las probabilidades acumuladas deseadas.

$$\begin{aligned} P(a \leq X \leq b) &= P\left(\frac{a - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right) = \\ &= P\left(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right) = F\left(\frac{b - \mu}{\sigma}\right) - F\left(\frac{a - \mu}{\sigma}\right) \end{aligned}$$

5 Estimación de parámetros.

- El primer paso en la fase de Inferencia consiste en proponer un modelo de distribución conocido X , que creemos refleja el comportamiento de la población. Dicho modelo dependerá de ciertos parámetros que debemos averiguar.
- Para determinar estos parámetros necesitaríamos conocer la característica estudiada en toda la población, lo cual no es posible en la mayoría de los casos (inviabilidad económica, física, temporal...).
- Para evitar estos inconvenientes, se recurre al estudio de una muestra, a partir de la cual podemos inferir las características de la población de que ha sido extraída.
- La parte de la Inferencia que se encarga de proponer aproximaciones para estos parámetros, a partir de la información obtenida en la muestra, es la ESTIMACIÓN DE PARÁMETROS.

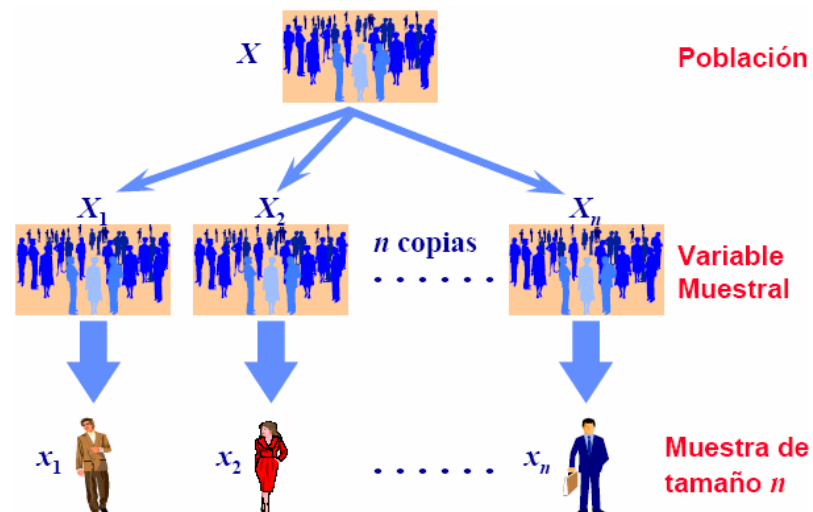
Ventajas e Inconvenientes del Muestreo

- Estudiar un número reducido de individuos de una muestra en lugar de toda la población tiene indudables ventajas:
 - Menor coste.
 - Mayor rapidez
 - Mayor facilidad.
- Pero también presenta algunos inconvenientes:
 - Necesidad de conseguir una muestra representativa.
 - Posibilidad de cometer errores en las estimaciones (**sesgos**).
- Afortunadamente, estos errores pueden ser superados. La representatividad de la muestra se consigue eligiendo la modalidad de muestreo más apropiada para el tipo de estudio. En el caso de los errores, aunque no podremos evitarlos, trataremos de reducirlos al máximo y acotarlos.

Variable Aleatoria Muestral

- Una muestra de tamaño n de una población X , puede entenderse como un valor de una variable aleatoria n -dimensional (X_1, \dots, X_n) , que se conoce como VARIABLE MUESTRAL.
- Los valores que puede tomar esta variable, serían todas las posibles muestras de tamaño n que pueden extraerse de la población X .
- En el caso del *muestreo aleatorio simple*, se cumple
 - Cada una de las variables X_i sigue la misma distribución que la población X (extracciones con reemplazamiento).
 - Cada una de las variables X_i es independiente del resto (independencia de las extracciones).

Obtención de una Muestra



Distribución de la Variable Aleatoria Muestral

- La distribución de probabilidad de los valores de la variable muestral, depende claramente de la distribución de probabilidad de los valores de la población.
- Ejemplo: Supongamos una población en la que la cuarta parte de las familias no tienen hijos, la mitad de las familias tiene 1 hijo, y el resto tiene 2 hijos.

Distribución Poblacional

X	$P(x)$
0	0.25
1	0.5
2	0.25

Muestras de tamaño 2

(X_1, X_2)	$P(x_1, x_2)$
(0,0)	0.0625
(0,1)	0.125
(0,2)	0.0625
(1,0)	0.125
(1,1)	0.25
(1,2)	0.125
(2,0)	0.0625
(2,1)	0.125
(2,2)	0.0625

Distribución Muestral

Distribución de un Estadístico Muestral

- Por ser función de una variable aleatoria, un estadístico en el muestreo es también una variable aleatoria.
- Su distribución de probabilidad, también depende de la distribución de la población estudiada y de los parámetros que determinan ésta (μ , σ , p).
- Ejemplo:

Distribución Poblacional

X	$P(x)$
0	0.25
1	0.5
2	0.25

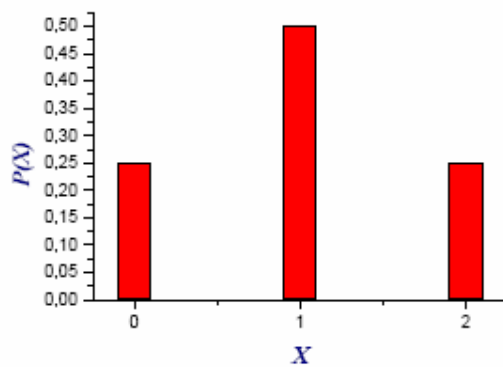
Muestras de tamaño 2

Distribución de \bar{x}

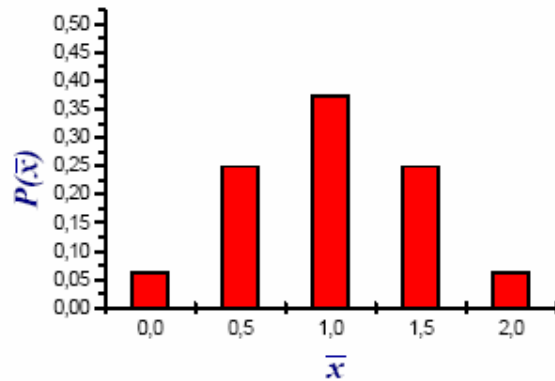
\bar{x}	$P(\bar{x})$
0	0.0625
0.5	0.25
1	0.375
1.5	0.25
2	0.0625

Distribución de un Estadístico Muestral

- Seguimos con el ejemplo:



Distribución Poblacional



Distribución de \bar{x}

¿Cuál es la probabilidad de obtener una media muestral que aproxime la media poblacional con un error máximo de 0.5?

Teorema Central del Límite

- La suma de n variables aleatorias independientes, cuando n es suficientemente grande, tiene una distribución aproximadamente **normal**, de media la suma de sus medias y varianza la suma de sus varianzas.

X_1, \dots, X_n v.a. independientes con $E[X_i] = \mu_i$ y $\text{Var}[X_i] = \sigma_i^2$

$$X = X_1 + \dots + X_n \quad \xrightarrow{n \geq 30} \quad X \sim N\left(\sum \mu_i, \sqrt{\sum \sigma_i^2}\right)$$

- La mayoría de las variables biológicas son causa de múltiples factores que suman sus efectos de manera independiente. Esto justifica que la mayoría de las variables biológicas tengan una distribución normal.

Distribución de la Media Muestral (Muestras Grandes $n \geq 30$)

- La media muestral es la suma de n variables aleatorias independientes, idénticamente distribuidas.
- Aplicando el teorema central del límite tenemos:

$$\bar{X} = \frac{X_1}{n} + \dots + \frac{X_n}{n}$$

$$E\left[\frac{X_i}{n}\right] = \frac{\mu}{n} \quad \longrightarrow \quad \text{T.C.L.} \quad \longleftarrow \quad \text{Var}\left[\frac{X_i}{n}\right] = \frac{\sigma^2}{n^2}$$

$$\bar{X} \sim N\left(\sum_{i=1}^n \frac{\mu}{n}, \sqrt{\sum_{i=1}^n \frac{\sigma^2}{n^2}}\right) = N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Distribución de la Proporción Muestral (Muestras Grandes $n \geq 30$)

- Una proporción es la media de una variable dicotómica (0,1). Esta variable se conoce como VARIABLE DE BERNOUILLI $B(p)$, que es un caso particular de la binomial para $n=1$.
- Aplicando de nuevo el teorema central del límite tenemos:

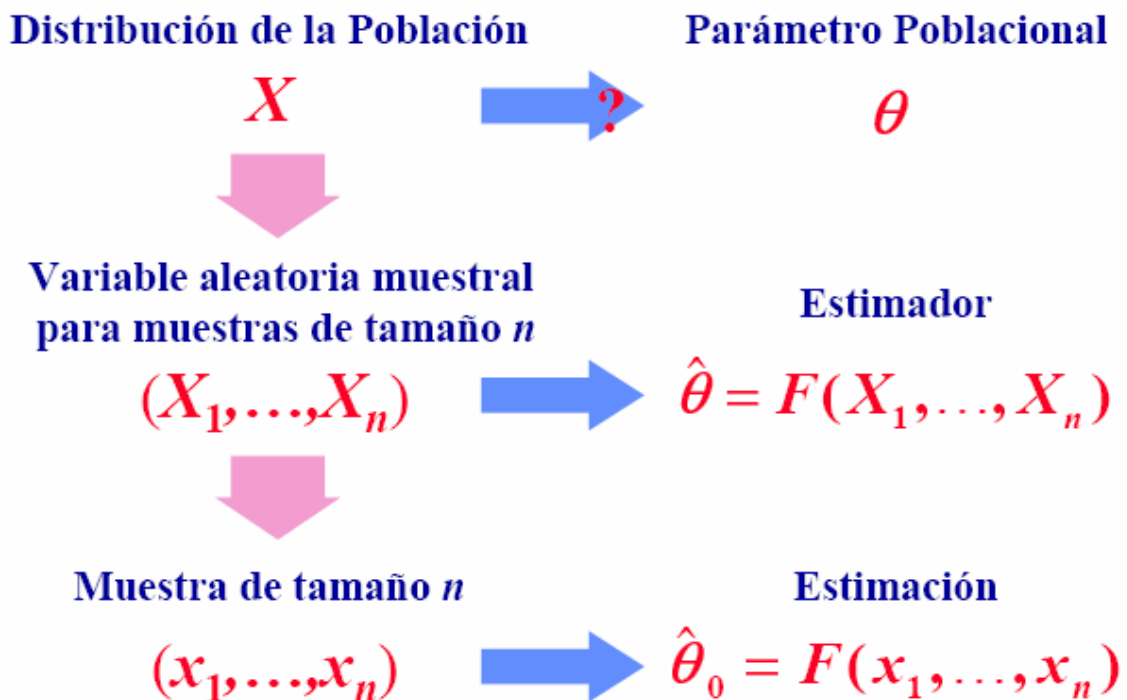
$$\hat{p} = \bar{X} = \frac{X_1}{n} + \dots + \frac{X_n}{n} \quad \text{con } X_i \sim B(p)$$

$$E\left[\frac{X_i}{n}\right] = \frac{p}{n} \quad \longrightarrow \quad \text{T.C.L.} \quad \longleftarrow \quad \text{Var}\left[\frac{X_i}{n}\right] = \frac{p(1-p)}{n^2}$$

$$\hat{p} \sim N\left(\sum_{i=1}^n \frac{p}{n}, \sqrt{\sum_{i=1}^n \frac{p(1-p)}{n^2}}\right) = N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

Dos Tipos de Estimación**Estimadores y Estimaciones**

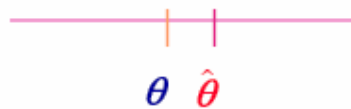
- Los estadísticos muestrales pueden utilizarse para aproximar los parámetros de la población, y cuando un estadístico se utiliza con este fin se le llama ESTIMADOR DEL PARÁMETRO.
- El valor obtenido cuando se evalúa un estimador de un parámetro para una muestra concreta se llama ESTIMACIÓN DEL PARÁMETRO.
- El estimador es una función y es único. La estimación no es única, sino que depende de la muestra tomada. Para cada muestra habrá una estimación.



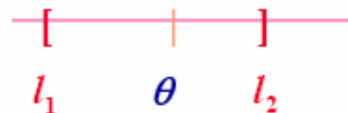
Dos Tipos de Estimación

- La estimación de parámetros puede realizarse de dos formas:
 - ESTIMACIÓN PUNTUAL. Se da un único valor aproximado del parámetro desconocido, pero no se especifica la bondad de la aproximación.
 - ESTIMACIÓN POR INTERVALOS. Se da un intervalo de valores, dentro del cual se cree que está el verdadero valor del parámetro con un cierto grado de seguridad.

Estimación puntual



Estimación por intervalos



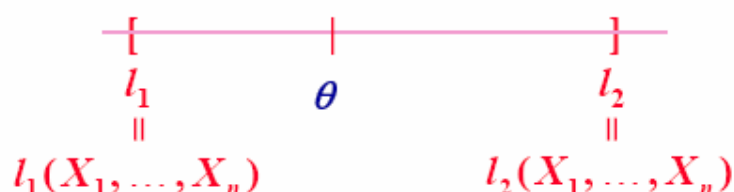
Estimador de la Media Poblacional

- Cuando el parámetro de interés es la media de la población μ , el estimador que se suele utilizar es la media muestral. Para muestras de tamaño n resulta la siguiente variable aleatoria:

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}$$

Estimación por Intervalos

- En ocasiones, más que dar un valor aproximado para un parámetro, conviene dar todo un intervalo de valores y decir que existe una confianza razonable en que el verdadero valor del parámetro se encuentre dentro de dicho intervalo.
- En este caso, en lugar de utilizar un estadístico, como ocurría con los estimadores puntuales, utilizaremos dos, uno para cada límite del intervalo.



Intervalos de Confianza

- Dados dos estadísticos muestrales L_1 y L_2 , se dice que el intervalo $I=[L_1, L_2]$ es un INTERVALO DE CONFIANZA para el parámetro poblacional θ , con un NIVEL DE CONFIANZA $1-\alpha$ (o SIGNIFICACIÓN α), si la probabilidad de que los estadísticos que determinan los límites del intervalo tomen valores tales que θ esté comprendido entre ellos es igual a $1-\alpha$, es decir

$$P(L_1(X_1, \dots, X_n) \leq \theta \leq L_2(X_1, \dots, X_n)) = 1 - \alpha$$

- La confianza $1-\alpha$ suele fijarse a niveles altos: 0,90 ($\alpha=0,1$), 0,95 ($\alpha=0,05$) o 0,99 ($\alpha=0,01$).
- La amplitud del intervalo se conoce como ERROR O IMPRECISIÓN DE LA ESTIMACIÓN. Cuanto más preciso sea un intervalo, menor será la confianza de que el valor del parámetro se encuentre allí.

I.C. Para la Media

- Supongamos que la población X se distribuye como una normal $N(\mu, \sigma)$, de media μ desconocida y desviación típica σ conocida.
- Para muestras de tamaño n , la media muestral sigue también una distribución normal con los siguientes parámetros:

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

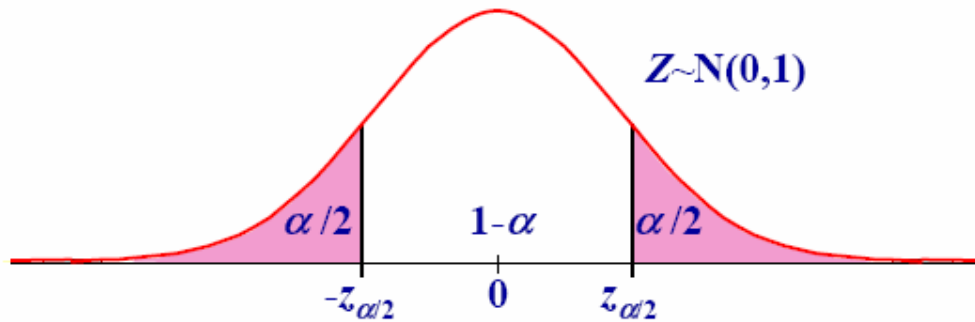
- Tipificando, llegamos a:

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

- Para conseguir un intervalo de confianza para μ con un nivel de confianza $1-\alpha$ tenemos que buscar en la tabla de la normal estándar valores l_1 y l_2 tales que:

$$P(l_1 \leq Z \leq l_2) = 1 - \alpha$$

- Como la distribución normal es simétrica, para que el intervalo sea lo más preciso posible lo mejor es tomar valores opuestos que dejen colas respectivas de probabilidad $\alpha/2$.



$$P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$$

- A partir de aquí llegamos fácilmente a los estadísticos que darán los límites del intervalo de confianza:

$$P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$$



$$P(-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}) = 1 - \alpha$$



$$P(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

- Finalmente llegamos al siguiente intervalo de confianza para μ con un nivel de confianza $1-\alpha$:

$$(\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}})$$

- Como se puede comprobar, el intervalo está centrado en la media muestral, que es precisamente el mejor estimador puntual para μ .
- La precisión del intervalo depende de dos factores:
 - **Nivel de confianza $1-\alpha$** . A mayor nivel de confianza, menor precisión.
 - **Tamaño de la muestra n** . A mayor tamaño muestral, mayor precisión.
- Generalmente, se fija el nivel de confianza y se busca el tamaño muestral para conseguir una determinada precisión.

Elección del Tamaño Muestral

- Si analizamos, por ejemplo, el caso general del intervalo de confianza para la media de una población normal

$$(\bar{x} - z_{\alpha/2} \frac{s_{n-1}}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{s_{n-1}}{\sqrt{n}})$$

la amplitud del intervalo es

$$A = 2z_{\alpha/2} \frac{s_{n-1}}{\sqrt{n}}$$

- Se observa que en la amplitud del intervalo influyen varios factores:
 - El nivel de significación α , que fija el valor de $z_{\alpha/2}$.
 - La variabilidad de la población, estimada por s_{n-1} .
 - El tamaño de la muestra n .

- El nivel de significación α suele fijarse a niveles bajos para garantizar intervalos con gran nivel de confianza.
- La variabilidad de la población suele estimarse a partir de estudios previos, es decir, utilizando la cuasivarianza obtenida para una muestra anterior.
- Por consiguiente, el único factor que aún no está fijado es el tamaño de la muestra. De esta forma, una vez fijado el nivel de significación y estimada la varianza de la población, podemos controlar la amplitud del intervalo, tan sólo con el tamaño de la muestra.

$$n = \frac{4z_{\alpha/2}^2 s_{n-1}^2}{A^2}$$

6 Contraste de Hipótesis

Hipótesis Estadística

- ▶ A parte de los métodos de estimación de parámetros vistos, el científico necesita medios para comprobar el grado de certeza de ciertas suposiciones sobre la distribución que sigue la población que se está estudiando, y que permiten la construcción de un modelo fiable.
- ▶ Muchos estudios estadísticos tienen por objeto ver si algún parámetro de una variable aleatoria se ajusta a un valor concreto o comparar dos poblaciones para ver si siguen la misma distribución.
- ▶ Cualquier afirmación o conjetura que determina, parcial o totalmente, la distribución de una población se realiza mediante una HIPÓTESIS ESTADÍSTICA.

Contraste de Hipótesis

- ▶ En general, nunca se sabe con absoluta certeza si una hipótesis es cierta o falsa, ya que, para ello tendríamos que medir a todos los individuos de la población. Por ello, al plantear una hipótesis es necesario realizar un prueba o test que nos permita tomar una decisión: aceptarla como cierta o rechazarla por falsa.
- ▶ Una hipótesis se contrasta comparando sus predicciones con la realidad que se obtiene de las muestras: si coinciden, dentro del margen de error admisible, mantendremos la hipótesis, en caso contrario la rechazaremos y buscaremos nuevas hipótesis capaces de explicar los datos observados. *Las decisiones se toman sobre una base de probabilidad.*
- ▶ Los procedimientos que conducen a la aceptación o rechazo de la hipótesis forman la parte de la Inferencia Estadística llamada CONTRASTE DE HIPÓTESIS.

Hipótesis Nula y Alternativa

- ▶ En la mayoría de los casos un contraste supone tomar una decisión entre dos hipótesis conocidas como HIPÓTESIS NULA e HIPÓTESIS ALTERNATIVA.
- ▶ La HIPÓTESIS NULA se representa como H_0 , y es la hipótesis que mantendremos a no ser que los datos observados en la muestra indiquen su falsedad. Suele ser la hipótesis más conservadora.
- ▶ La HIPÓTESIS ALTERNATIVA se representa como H_1 , y suele ser la negación de la hipótesis nula. Generalmente es lo que se quiere demostrar.
- ▶ Ambas hipótesis se eligen de acuerdo con el principio de “simplicidad científica”: *solamente debemos abandonar un modelo simple por otro más complejo cuando la evidencia a favor del último sea fuerte (Navaja de Occam).*

Un Ejemplo de Elección de Hipótesis

- ▶ En el caso de un juicio, en el que el juez debe decidir si el acusado es culpable o inocente, tendríamos
$$H_0 : \text{Inocente} \qquad H_1 : \text{Culpable}$$
ya que la inocencia se asume, mientras que la culpabilidad hay que demostrarla.
- ▶ Según esto, el juez sólo aceptaría la hipótesis alternativa cuando hubiese pruebas significativas de la culpabilidad del acusado.
- ▶ El objetivo del investigador (fiscal en este caso) consistiría en intentar demostrar dicha culpabilidad (rechazar la hipótesis nula).
- ▶ *¡Esta metodología siempre favorece a la hipótesis nula!*

Tipos de Contrastes

- ▶ **Contrastes Paramétricos.** Contrastan el valor de uno o más parámetros poblacionales. A su vez se clasifican en:
 - **Contrastes de una población:** Se contraste un valor concreto o un intervalo para un parámetro de una población. Ejemplo: Contrastar si la nota media de un grupo ha sido 6.
 - **Contrastes de comparación de poblaciones:** Se comparan los parámetros de las distribuciones de dos o más poblaciones. Ejemplo: Comparar las notas medias de dos o más grupos.
- ▶ **Contrastes No Paramétricos.** Se contrastan características estructurales de la distribución, como el modelo de distribución de los datos, la independencia de los mismos, etc. Ejemplo: Contrastar si las notas de un grupo siguen una distribución normal.

Tipos de Hipótesis en Contrastes Paramétricos de una Población

- ▶ Cuando se quiere contrastar el valor de un parámetro poblacional θ , con un valor concreto θ_0 , la hipótesis nula se suele expresar como:

$$H_0 : \theta = \theta_0.$$

- ▶ La hipótesis alternativa puede tener dos formas:
 - **Contraste Bilateral.** Si desconocemos la dirección en que H_0 puede ser falsa:

$$H_1 : \theta \neq \theta_0.$$

- **Contraste Unilateral.** Si sabemos en qué dirección puede ser H_0 falsa, hacia la izquierda ($\theta < \theta_0$) o hacia la derecha ($\theta > \theta_0$):

$$H_1 : \theta < \theta_0 \text{ (con cola a la izquierda)}$$

$$H_1 : \theta > \theta_0 \text{ (con cola a la derecha).}$$

Tipos de Hipótesis en Contrastes de Comparación de Poblaciones

- ▶ Cuando se quieren comparar dos parámetros θ_1 y θ_2 de dos poblaciones, entonces la hipótesis nula suele expresarse como:

$$H_0 : \theta_1 = \theta_2.$$

- ▶ Como antes, la hipótesis alternativa puede tener dos formas:
 - **Contraste Bilateral.** Si desconocemos la dirección en que H_0 puede ser falsa:

$$H_1 : \theta_1 \neq \theta_2.$$

- **Contraste Unilateral.** Si sabemos en qué dirección puede ser H_0 falsa, hacia la izquierda ($\theta_1 < \theta_2$) o hacia la derecha ($\theta_1 > \theta_2$):

$$H_1 : \theta_1 < \theta_2 \text{ (con cola a la izquierda)}$$

$$H_1 : \theta_1 > \theta_2 \text{ (con cola a la derecha).}$$

Errores en un Contraste

- ▶ Como dijimos antes, nunca tendremos la seguridad de conocer la certeza o falsedad de una hipótesis, de modo que al aceptarla o rechazarla es posible que estemos equivocandonos.
- ▶ Los errores que se pueden cometer en un contraste de hipótesis son de dos tipos:
 - **Error de tipo I.** Se comete cuando se rechaza la hipótesis nula siendo esta verdadera.
 - **Error de tipo II.** Se comete cuando se acepta la hipótesis nula siendo esta falsa.

Probabilidades de los Errores Nivel de Significación

- ▶ Aunque nunca sabremos si estamos cometiendo un error, si que podemos saber cual es la probabilidad de cometerlo.
- ▶ La probabilidad de cometer un error de tipo I se llama NIVEL DE SIGNIFICACIÓN y se designa por α .

$$\alpha = P(\text{Rechazar } H_0 / H_0 \text{ es cierta}).$$

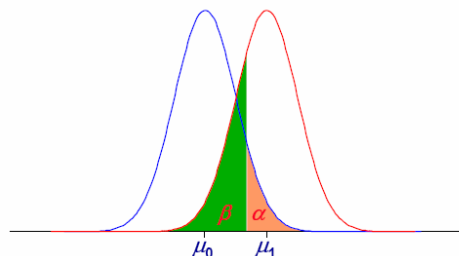
- ▶ Y la probabilidad de cometer un error de tipo II se designa por β .

$$\beta = P(\text{Aceptar } H_0 / H_0 \text{ es falsa}).$$

- ▶ Así pues, al realizar un contraste de hipótesis, pueden darse las cuatro situaciones que aparecen en el siguiente cuadro:

Decisión\Realidad	H_0 cierta	H_0 falsa
Aceptar H_0	Decisión correcta Probabilidad $1-\alpha$	Error Tipo II Probabilidad β
Rechazar H_0	Error Tipo I Probabilidad α	Decisión correcta Probabilidad $1-\beta$

Relación entre α y β



- ▶ Se observa que al disminuir el nivel de significación α aumenta β y viceversa.
- ▶ La única forma de disminuir los dos tipos de errores a la vez es aumentando el tamaño muestral.

Control de los Errores

- ▶ En principio, puesto que se pretende favorecer la hipótesis nula, el error del tipo I suele ser más grave que el error del tipo II, y por tanto, α suele fijarse a niveles bajos (0.1, 0.05 o 0.01).
- ▶ No obstante, también nos interesa controlar el error del tipo II, ya que de lo contrario será difícil detectar la falsedad de la hipótesis nula (que es lo que se persigue la mayoría de las veces), aunque haya pruebas contundentes.
- ▶ Pero la hipótesis alternativa no suele fijar el valor del parámetro o parámetros, de modo que, para poder calcular β es necesario fijar dichos valores.

Potencia de un Contraste Curva de Potencia

- ▶ Para medir la capacidad de un contraste para detectar la falsedad de la hipótesis nula se suele utilizar la POTENCIA DE UN CONTRASTE, que se define como la probabilidad de rechazar la hipótesis nula, cuando la hipótesis alternativa fijada es verdadera, es decir, $1-\beta$.
- ▶ Como la potencia depende del valor del parámetro fijado en la hipótesis alternativa, se puede definir una función para la potencia como

$$\text{Potencia}(x) = P(\text{Rechazar } H_0 / \theta = x).$$

- ▶ Esta función nos da la probabilidad de rechazar la hipótesis nula para cada valor del parámetro y se conoce como CURVA DE POTENCIA.

Ejemplo

- Supongamos que en un grupo de alumnos queremos contrastar la hipótesis de que la proporción de aprobados es de la mitad o si por el contrario es menos. Entonces tenemos que plantear el contraste unilateral

$$H_0 : p = 0.5$$

$$H_1 : p < 0.5$$

- Si tomamos una muestra de 10 alumnos de dicho grupo, bajo la hipótesis nula, la variable X que mide el número de aprobados en la muestra seguirá una distribución $B(10,0.5)$. Según este modelo, el valor esperado de X es 5 aprobados.
- Ahora bien, si en la muestra sólo observamos 2 aprobados, y en base a ello, decidimos rechazar la hipótesis nula, ¿cuál sería la probabilidad de equivocarnos en la decisión y cometer un error de tipo I?
- Si decidimos rechazar la hipótesis cuando en la muestra observemos una diferencia mayor de 2 con el número esperado de aprobados, entonces la probabilidad de equivocarnos será

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0.0547.$$

- Si la probabilidad máxima de error de tipo I que estamos dispuestos a tolerar es $\alpha=0.05$, ¿Qué valores de la muestra nos servirían para rechazar la hipótesis nula?

$$P(X \leq 1) = P(X = 0) + P(X = 1) = 0.0107.$$

- Es decir, sólo rechazaríamos la hipótesis nula cuando en la muestra obtuviésemos 0 o 1 alumnos aprobados.
- Supongamos ahora que la proporción real de aprobados en el grupo es de 0.2 (fijamos la hipótesis alternativa a $p=0.2$).
- Si utilizamos la regla de decisión anterior de rechazar la hipótesis nula cuando obtuviésemos 0 o 1 aprobados, ¿cuál será la potencia de este contraste?

- Trabajando ahora con $X \sim B(10,0.2)$, tenemos

$$\begin{aligned} \beta &= P(X \geq 2) = 1 - P(X < 2) = 1 - P(X = 0) - P(X = 1) = \\ &= 1 - 0.3758 = 0.6242. \end{aligned}$$

Luego la potencia del contraste sería $1-\beta=0.3758$.

Estadístico del Contraste

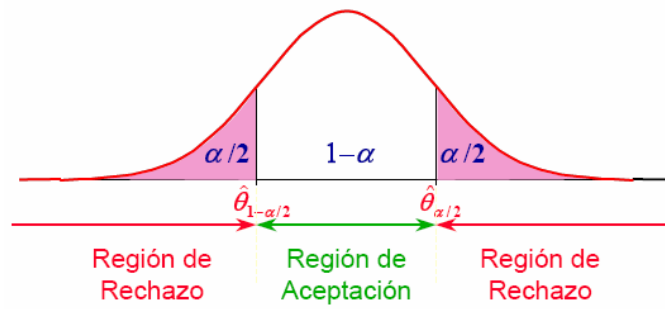
- ▶ La decisión que se plantea en el contraste, entre la aceptación o el rechazo de la hipótesis nula, se realiza en base a un estadístico en el muestreo, relacionado con el parámetro o característica que queremos contrastar, y cuya distribución debe ser conocida suponiendo cierta la hipótesis nula, y una vez fijado el tamaño de la muestra. Este estadístico recibe el nombre de **ESTADÍSTICO DEL CONTRASTE**.
- ▶ Para cada muestra que tomemos, el estadístico nos dará una estimación a partir de la cual tomaremos la decisión de aceptar o rechazar la hipótesis nula: Si la estimación difiere demasiado del valor que propone H_0 , entonces la rechazaremos, y en caso contrario la aceptaremos.

Región de Aceptación y Región de Rechazo

- ▶ Al fijar el nivel de significación α , el conjunto de valores que puede tomar el estadístico del contraste queda dividido en dos partes: las estimaciones que conducirían a la aceptación de H_0 , que se llama **REGIÓN DE ACEPTACIÓN**, y las estimaciones que conducirían a su rechazo, denominada **REGIÓN DE RECHAZO**.
- ▶ En el ejemplo anterior, para un nivel de significación $\alpha=0.05$, tomando como estadístico de contraste el número de aprobados en la muestra, vimos que los valores que llevarían al rechazo de la hipótesis nula eran 0 y 1. Así pues, la región de rechazo estaría formada por los valores $\{0,1\}$ y la de aceptación por todos los demás $\{2,3,4,5,6,7,8,9,10\}$.

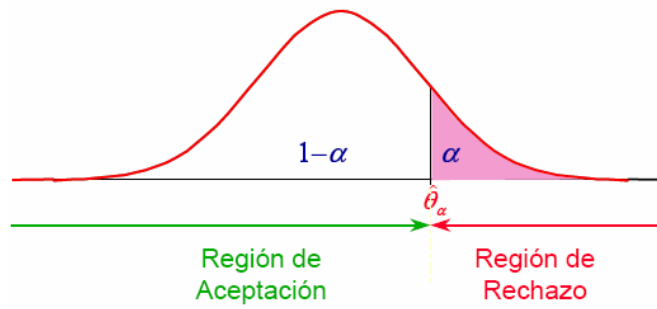
Contraste Bilateral

Distribución del estadístico del contraste $\hat{\theta}$



Contraste Unilateral

Distribución del estadístico del contraste $\hat{\theta}$



Pasos para realizar un Contraste de Hipótesis

- ▶ En resumen, para realizar el contraste de hipótesis, el investigador debe dar los siguientes pasos:
 1. Formular la hipótesis nula y la alternativa.
 2. Fijar el nivel de significación α que está dispuesto a aceptar y la potencia deseada $1-\beta$.
 3. Calcular el tamaño muestral necesario n .
 4. Construir el estadístico del contraste bajo la certeza de la hipótesis nula.
 5. Delimitar las regiones de aceptación y rechazo.
 6. Tomar una muestra de tamaño n .
 7. Aplicar el estadístico del contraste a la muestra para obtener la estimación.
 8. Aceptar o rechazar la hipótesis nula dependiendo de si la estimación cae en la región de aceptación o en la de rechazo respectivamente.

P-Valor de un Contraste

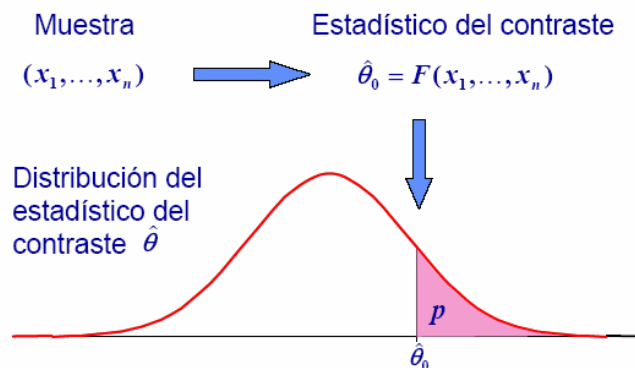
- ▶ En la práctica, además de lo anterior, cuando se toma una decisión en un contraste, conviene indicar el grado de confianza que se tiene en dicha decisión.
- ▶ En general, siempre que la estimación del estadístico caiga dentro de la región de rechazo, rechazaremos H_0 , pero evidentemente, si dicha estimación se aleja bastante de la región de aceptación tendremos más confianza en el rechazo que si la estimación está cerca del límite entre las regiones de aceptación y rechazo.
- ▶ Por este motivo, al realizar un contraste, también se calcula la probabilidad de obtener una discrepancia mayor o igual a la observada entre la estimación y el valor esperado según la hipótesis nula. Esta probabilidad se conoce como *P-VALOR* del contraste y, en cierto modo, expresa la confianza que se tiene al tomar la decisión.

Cálculo del P-Valor

- El cálculo del p -valor depende de si el contraste es unilateral o bilateral:

Contraste	p -Valor
Bilateral	$2P(\hat{\theta} > \hat{\theta}_0 / H_0 \text{ es cierta})$
Unilateral a la izquierda	$P(\hat{\theta} > \hat{\theta}_0 / H_0 \text{ es cierta})$
Unilateral a la derecha	$P(\hat{\theta} < \hat{\theta}_0 / H_0 \text{ es cierta})$

- Cuanto más próximo esté el p -valor a 1, mayor confianza existe al aceptar la hipótesis nula, y cuanto más próximo esté a 0, mayor confianza hay al rechazarla.
- Si se ha fijado de antemano el nivel de significación α , entonces se aceptará H_0 si $p > \alpha$ y se rechazará si $p < \alpha$. De este modo, el p -valor nos da información de para qué niveles de significación puede rechazarse la hipótesis nula y para cuales no.

P-Valor en un Contraste Unilateral

Contraste para la Media

- ▶ Población: $X \sim N(\mu, \sigma)$ con μ desconocida y σ conocida.
- ▶ Hipótesis:

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$
- ▶ Nivel de significación: α
- ▶ Tamaño de la muestra: n
- ▶ Estadístico del contraste: Por el teorema central del límite sabemos que

$$\bar{X} \sim N\left(\mu_0, \frac{\sigma}{\sqrt{n}}\right)$$

de este modo, utilizaremos el estadístico del contraste

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0,1)$$

- ▶ Se utiliza también para poblaciones no normales con $n \geq 30$,

Contraste de Igualdad de Medias

- ▶ Poblaciones: $X_1 \sim N(\mu_1, \sigma_1)$ y $X_2 \sim N(\mu_2, \sigma_2)$ con μ_1 y μ_2 conocidas, y σ_1 y σ_2 conocidas.
- ▶ Hipótesis:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$
- ▶ Nivel de significación: α
- ▶ Tamaños de las muestras: n_1 y n_2
- ▶ Estadístico del contraste:

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

