

VARIABLE ALEATORIA DISCRETA

Muchas veces se desea resumir con un número el resultado de un experimento aleatorio. En muchos de los ejemplos relativos a experimentos aleatorios que han sido considerados hasta ahora, el espacio muestral es sólo una descripción de los posibles resultados. En algunos casos tales descripciones son suficientes, pero en otros se hace útil asociar un número con cada resultado del espacio muestral. Es así como se llega a la definición de **variable aleatoria**.

Una **variable aleatoria X** es una función que asigna un número real a cada resultado en el espacio muestral Ω de un experimento aleatorio. El conjunto de los posibles valores de la variable aleatoria **X** se denomina rango. Diremos que la variable aleatoria es **discreta** si su rango es finito (o infinito contable).

A menudo el interés recae en la probabilidad de que una variable aleatoria **X** tome un valor particular x , esto se denota $P(X=x)$. La distribución de probabilidad de **X** será entonces la descripción del conjunto de valores posibles de **X** (rango de **X**), junto con la probabilidad asociada con cada uno de estos valores. La distribución de probabilidad de una variable aleatoria es a menudo el resumen más útil de un experimento aleatorio.

Diremos que la función $p(x)=P(X=x)$ que va del conjunto de valores posibles de la variable aleatoria **X** al intervalo $[0, 1]$ es la **función distribución de probabilidad** para **X** si y sólo si se satisfacen las siguientes propiedades:

$$0 \leq p(x) \leq 1 \text{ para todo } x$$

$$\sum_x p(x) = 1$$

Se define la distribución acumulada **F(x)** para la variable aleatoria **X** como

$$F(x) = P(X \leq x) = \sum_{t \leq x} p(t)$$

Ejemplo

Experimento aleatorio: se lanza una moneda 3 veces

$\Omega = \{ccc, ccs, csc, css, scc, scs, ssc, sss\}$

Sea **X** : # caras observadas

x	0	1	2	3
p(x)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

La distribución anterior es una distribución de probabilidades para la variable aleatoria X , en efecto $0 \leq p(x) \leq 1$ para todo x ($x = 0, 1, 2$ y 3) y además $\sum_x p(x) = 1$. Para determinar la distribución acumulada de probabilidad observe que

$$P(X \leq 0) = P(X = 0) = \frac{1}{8}$$

$$P(X \leq 1) = P(X = 0) + P(X = 1) = \frac{1}{8} + \frac{3}{8} = \frac{1}{2}$$

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8}$$

$$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1$$

Se tiene entonces,

x	0	1	2	3
$F(x)$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{7}{8}$	1

Si X es una variable aleatoria, y el experimento aleatorio que determina el valor de X se repite muchas veces, entonces se obtiene una secuencia de valores para X . A partir de esta secuencia de valores se puede identificar el valor promedio o valor esperado de la variable aleatoria X , que denotamos $E(X)$, y se define en la forma siguiente:

$$E(X) = \sum_x xp(x)$$

Propiedades:

- $E(k) = k$
- $E(kX) = kE(X)$
- $E(X \pm Y) = E(X) \pm E(Y)$
- $E(g(X)) = \sum g(x)p(x)$
- Si X y Y son independientes entonces $E(XY) = E(X)E(Y) = \mu_X \mu_Y$

Para el ejemplo dado, $E(X) = \sum_x xp(x) = 0p(0) + 1p(1) + 2p(2) + 3p(3)$

$$= 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{12}{8} = \frac{3}{2}$$

A veces, el interés es determinar la variabilidad de la variable aleatoria. Definimos entonces la varianza de la variable aleatoria \mathbf{X} , denotada $V(\mathbf{X})$, ó σ^2 mediante la siguiente ecuación:

$V(\mathbf{X}) = E[(\mathbf{X}-E(\mathbf{X}))^2]$ y su forma reducida es:

$$V(\mathbf{X}) = E(\mathbf{X}^2) - [E(\mathbf{X})]^2$$

$$\text{donde, } E(\mathbf{X}^2) = \sum_x x^2 p(x)$$

Para el ejemplo dado, $E(\mathbf{X}^2) = 0^2 p(0) + 1^2 p(1) + 2^2 p(2) + 3^2 p(3)$

$$= 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 4 \cdot \frac{3}{8} + 9 \cdot \frac{1}{8} = \frac{24}{8} = 3$$

Entonces, $V(\mathbf{X}) = 3 - \left(\frac{3}{2}\right)^2 = \frac{12-9}{4} = \frac{3}{4}$

- a) $V(k)=0$
- b) $V(kX)=k^2V(X)$
- c) $V(X\pm Y)=V(X)+V(Y)$ si X y Y son independientes
- d) $V(aX+bY)=a^2V(X)+b^2V(Y)+2abCov(XY)$
donde $Cov(XY) = E((X-\mu_x)(Y-\mu_y)) = E(XY)-\mu_x\mu_y$

La desviación estándar de la variable aleatoria \mathbf{X} es la raíz cuadrada positiva de la varianza, es decir, $\sigma = \sqrt{V(\mathbf{X})}$.

DISTRIBUCIÓN BINOMIAL

Un **ensayo Bernoulli** es un experimento aleatorio que sólo admite dos posibles resultados, denotados **éxito** y **fracaso**. La probabilidad de éxito se denota \mathbf{p} .

Por lo tanto si denotamos el éxito por 1 y el fracaso por 0 se tiene:

$$P(1)=\mathbf{p} \quad P(0)=\mathbf{1-p=q}$$

Además se cumple: $E(X)=\mathbf{p} \quad V(X)=\mathbf{pq}$

Un **proceso Bernoulli** es un proceso en el cual se verifican las siguientes condiciones:

El experimento aleatorio se repite \mathbf{n} veces en idénticas condiciones

Hay sólo dos posibles resultados en cada repetición del experimento, llamados arbitrariamente **éxito y fracaso**

La probabilidad de éxito, denotada **p**, es la misma para cada repetición (permanece constante entre repeticiones)

las **n** repeticiones del experimento aleatorio son independientes entre sí

Consideremos ahora la variable aleatoria **X**: # éxitos observados en **n** repeticiones. Suponga que se quiere determinar la probabilidad de observar **x** éxitos en **n** repeticiones; esto es, se desea determinar $P(\mathbf{X} = \mathbf{x})$. Como lo importante es observar **x** éxitos en **n** repeticiones, el orden de ocurrencia de los mismos es irrelevante; así, para contar de cuántas formas pueden observarse

x éxitos en **n** repeticiones empleamos las combinaciones $\binom{n}{x}$. Por otro lado, como las **n**

repeticiones del experimento son independientes entre sí y calcular $P(\mathbf{X} = \mathbf{x})$ equivale a calcular la probabilidad de una intersección de eventos (en las que cada evento corresponde a un éxito o a un fracaso), tenemos que la probabilidad de un punto muestral cualquiera asociado al experimento es $p^x q^{n-x}$; en definitiva:

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

Dado que $0 \leq \binom{n}{x} p^x q^{n-x} \leq 1$ y $\sum_{x=0}^n \binom{n}{x} p^x q^{n-x} = 1$, resulta que $P(\mathbf{X} = \mathbf{x}) =$

$\binom{n}{x} p^x q^{n-x}$ para $x = 0, 1, 2, \dots, n$ determina una distribución de probabilidades

denominada **distribución binomial**.

En resumen, se dice que la variable aleatoria **X** tiene distribución binomial si su función distribución de probabilidad está dada por

$$p(x) = \begin{cases} \binom{n}{x} p^x q^{n-x} & \text{si } x = 0, 1, \dots, n \\ 0 & \text{otros valores} \end{cases}$$

Se puede demostrar que para una variable aleatoria con distribución binomial

$$E(\mathbf{X}) = n \cdot p$$

$$V(\mathbf{X}) = n \cdot p \cdot q$$

DISTRIBUCIÓN HIPERGEOMETRICA:

Una variable aleatoria X tiene una distribución hipergeométrica si se toma una muestra sin reemplazo de un conjunto de N elementos, de los cuales k son considerados de una categoría en especial (aciertos) y los otros $N-k$ son considerados de otra categoría (fallas) y se desea obtener x aciertos de una muestra de n elementos ó ensayos. Se expresa de la siguiente formula:

$$P(X = x) = \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}} \quad \text{para } x = 0, 1, 2, \dots, n$$

$$E(x) = \frac{nk}{N} \quad \sigma^2 = \frac{N-n}{N-1} \cdot \frac{nk(N-k)}{N^2}$$

Esto también se puede extender para más de dos grupos.

Ejemplo:

Si existe tres grupos el primero con k_1 elementos, el segundo grupo k_2 y el tercero con k_3 . Si queremos hallar la probabilidad de escoger x elementos del primer grupo, y elementos del segundo grupo y z elementos del tercer grupo sin reemplazo; la probabilidad es la siguiente:

$$P(X = x, Y = y, Z = z) = \frac{\binom{k_1}{x} \binom{k_2}{y} \binom{k_3}{z}}{\binom{k_1 + k_2 + k_3}{x + y + z}}$$

DISTRIBUCIÓN POISSON

Los experimentos que dan valores numéricos de una variable aleatoria X que ocurre durante un intervalo de tiempo dado o en una región específica se denominan **experimentos Poisson**. El intervalo puede ser de cualquier longitud: un minuto, un día, una semana, un mes o incluso un año; y la región específica podría ser: un segmento de línea, un área o quizás una pieza de material. Un experimento Poisson se deriva de un proceso Binomial, el cual verifica las siguientes propiedades:

El número de resultados que ocurren en un intervalo o región es independiente del número de resultados que ocurren en otro intervalo o región. (Esto determina una característica que se conoce como **falta de memoria**)

La probabilidad de que ocurra un solo resultado durante un intervalo muy corto o una región pequeña es proporcional a la longitud del intervalo o al tamaño de la región y no depende del número de resultados que ocurren fuera de este intervalo o región.

la probabilidad de que ocurra más de un resultado en tal intervalo corto o que caiga en tal región pequeña es insignificante.

La variable aleatoria X : # de resultados que ocurren durante un experimento Poisson se denomina variable aleatoria Poisson y su distribución de probabilidades, dada por

$$p(x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad \text{para } x = 0, 1, \dots \text{ se denomina distribución Poisson; donde } \lambda \text{ es el número}$$

promedio de resultados por unidad de tiempo o región. Para una variable aleatoria con distribución Poisson se tiene $E(X) = V(X) = \lambda$.

DISTRIBUCIONES CONTINUAS

Las distribuciones de probabilidad son idealizaciones de los polígonos de frecuencias. En el caso de una variable estadística continua consideramos el histograma de frecuencias relativas, y se comprueba que al aumentar el número de datos y el número de clases el histograma tiende a estabilizarse llegando a convertirse su perfil en la gráfica de una función.

Distribución uniforme

En estadística la distribución uniforme es una distribución de probabilidad cuyos valores tienen la misma probabilidad.

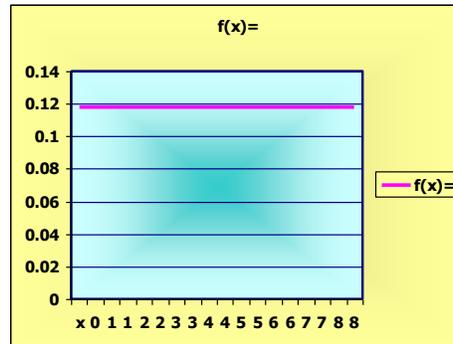
Se dice que una variable aleatoria X continua tiene una distribución uniforme en el intervalo $[a, b]$ si la función de densidad de probabilidad (FDP) es

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{para } a \leq x \leq b \\ 0 & \text{para el resto} \end{cases}$$

La función de distribución en el caso continuo entre a y b es

$$F(x) = \begin{cases} 0 & \text{para } x < a \\ \frac{x-a}{b-a} & \text{para } a \leq x < b \\ 1 & \text{para } x \geq b \end{cases}$$

Su media estadística es $(a + b) / 2$ y su varianza $(b - a)^2 / 12$



Distribución exponencial

En estadística la distribución exponencial es una distribución de probabilidad continua con un parámetro $\lambda > 0$ cuya función de densidad es

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{para } x \geq 0 \\ 0 & \text{de otro modo} \end{cases}$$

Su función de distribución es

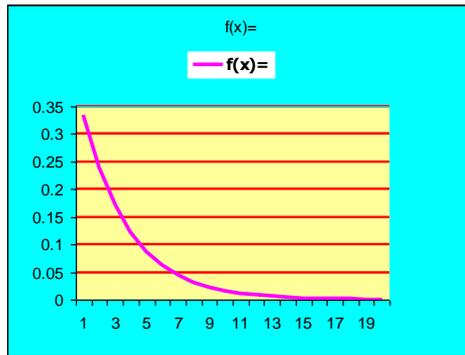
$$F(x) = P(X \leq x) \begin{cases} 0 & \text{para } x < 0 \\ 1 - e^{-\lambda x} & \text{para } x \geq 0 \end{cases}$$

Aquí e significa el número e .

El valor esperado y la varianza de una variable aleatoria X con distribución exponencial son

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

$$V(X) = \frac{1}{\lambda^2}$$



Distribución Normal

La distribución normal, también llamada distribución de Gauss o distribución gaussiana, es la distribución de probabilidad que con más frecuencia aparece en estadística y teoría de probabilidades. Esto se debe a dos razones fundamentalmente:

- Su función de densidad es simétrica y con forma de campana, lo que favorece su aplicación como modelo a gran número de variables estadísticas.
- Es, además, límite de otras distribuciones y aparece relacionada con multitud de resultados ligados a la teoría de las probabilidades gracias a sus propiedades matemáticas.

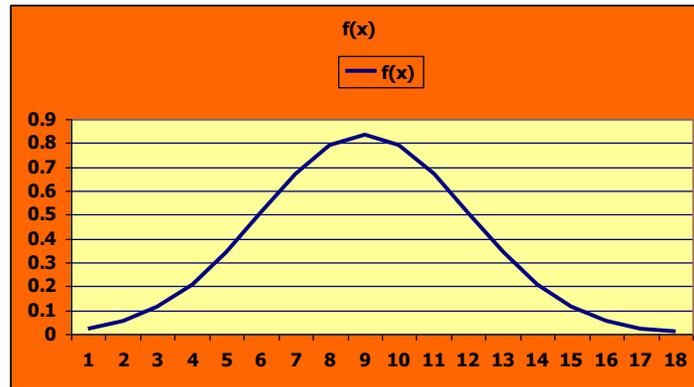
La función de densidad está dada por:

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

Donde μ (M) es la media y σ (sigma) es la desviación estándar (σ^2 es la varianza).

Muchas variables aleatorias continuas presentan una función de densidad cuya gráfica tiene forma de campana.

La importancia de la distribución normal se debe principalmente a que hay muchas variables asociadas a fenómenos naturales que siguen el modelo de la normal



TEOREMA DEL LÍMITE CENTRAL

Si \bar{X} es la media de una muestra aleatoria de tamaño n extraída de una población que tiene media μ y varianza σ^2 , entonces

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

es una variable aleatoria cuya función de distribución de probabilidad se aproxima a la de la distribución normal estándar a medida que n aumenta .

La demostración formal de este teorema requiere el manejo de límites de la función generadora de momentos de la variable $(\bar{X} - \mu)/(\sigma/\sqrt{n})$. Sin embargo, mediante simulaciones con el computador se puede verificar visualmente que sin importar la distribución de probabilidad de una variable aleatoria discreta o continua X , el límite de la variable aleatoria \bar{X} tiende a la forma tipo campana de la distribución normal cuando n crece.

Con carácter general, o al menos en los modelos de probabilidad clásicos, se admite una aproximación aceptable al modelo normal siempre que $n \geq 30$ y se dice que la muestra es "grande". Adicionalmente, en este caso, si se desconoce la varianza de la población se puede usar como aproximación la varianza muestral: $\sigma^2 \approx S^2$

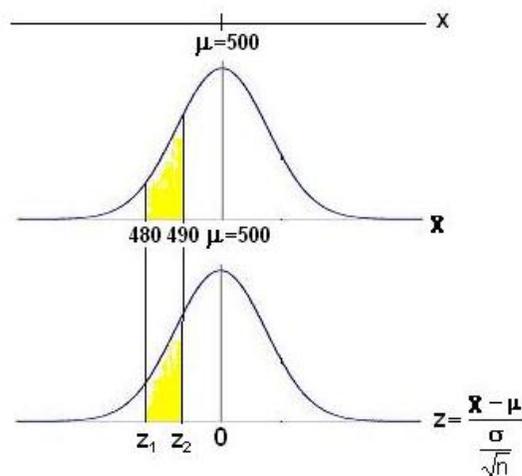
Ejemplo

Un fabricante de latas de pintura especifica que cada una cubre en promedio 500 pies cuadrados con una desviación estándar de 36 pies cuadrados. Calcule la probabilidad que la media del área cubierta por una muestra aleatoria de 40 de estas latas de pintura tenga un valor entre 480 y 490 pies cuadrados.

Solución

X: Área cubierta por una lata de pintura. Es una variable aleatoria cuya distribución de probabilidad es desconocida, con media $\mu = 500$, varianza σ^2 desconocida.

Por el Teorema del Límite Central, \bar{X} tiene distribución aproximadamente normal si $n \geq 30$, por lo tanto, $(\bar{X} - \mu) / (\sigma / \sqrt{n})$ tiene distribución aproximadamente normal estándar. Adicionalmente, se puede usar la aproximación: $\sigma \approx S = 36$



$$P(480 \leq \bar{X} \leq 490) = P\left(\frac{480 - \mu}{\frac{\sigma}{\sqrt{n}}} \leq Z \leq \frac{490 - \mu}{\frac{\sigma}{\sqrt{n}}}\right) = P\left(\frac{480 - 500}{\frac{36}{\sqrt{40}}} \leq Z \leq \frac{490 - 500}{\frac{36}{\sqrt{40}}}\right)$$

$$= P(-3.5136 \leq Z \leq -1.7568) = F(-1.7568) - F(-3.5136) = 0.0393$$

Distribución Weibull

La distribución de Weibull complementa a la distribución exponencial y a la normal, se usa cuando se sabe de antemano que una de ellas es la que mejor describe la distribución de fallos o cuando se han producido muchos fallos (al menos 10) y los tiempos correspondientes no se ajustan a una distribución más simple.

La distribución de Weibull nos permite estudiar cuál es la distribución de fallos de un componente clave de seguridad que pretendemos controlar y que a través de nuestro registro de fallos observamos que éstos varían a lo largo del tiempo y dentro de lo que se considera tiempo normal de uso.

La distribución de Weibull se representa normalmente por la función acumulativa de distribución de fallos $F(t)$:

$$F(t) = 1 - \exp\left[-\left(\frac{t-t_0}{\eta}\right)^\beta\right] \quad (1)$$

Siendo la función densidad de probabilidad:

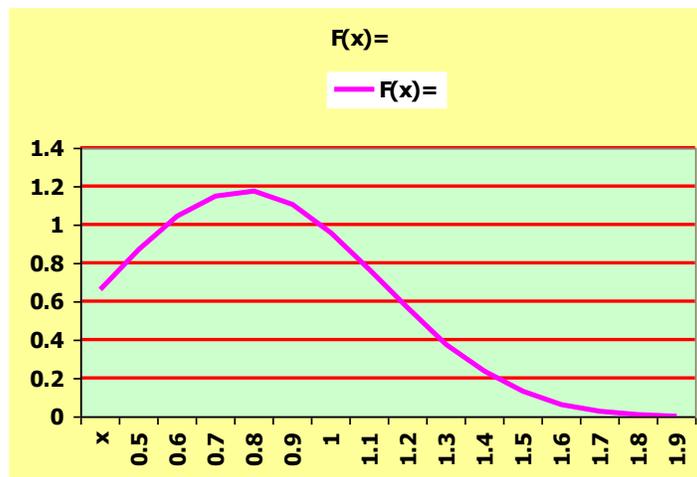
$$f(t) = \frac{\beta}{\eta} \left(\frac{t-t_0}{\eta}\right)^{\beta-1} \exp\left[-\left(\frac{t-t_0}{\eta}\right)^\beta\right] \quad (2)$$

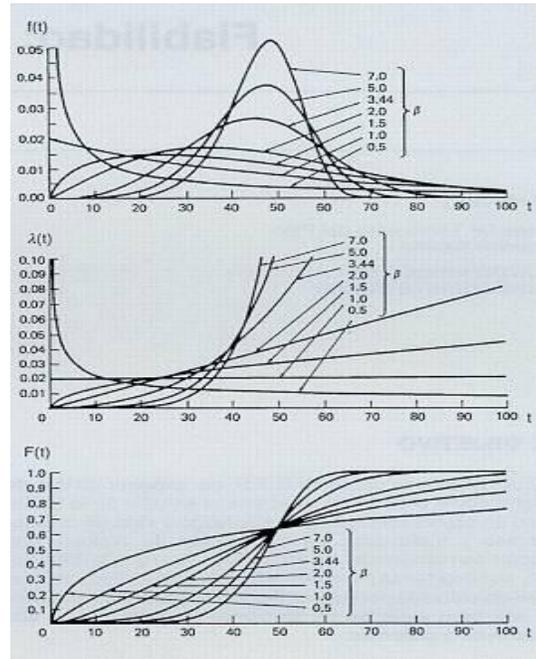
La tasa de fallos para esta distribución es:

$$\lambda(t) = \frac{\beta}{\eta} \left(\frac{t-t_0}{\eta}\right)^{\beta-1} \quad (3)$$

Las ecuaciones (1), (2) y (3) sólo se aplican para valores de $(t - t_0) \geq 0$. Para valores de $(t - t_0) < 0$, las funciones de densidad y la tasa de fallos valen 0. Las constantes que aparecen en las expresiones anteriores tienen una interpretación física:

- t_0 es el parámetro de posición (unidad de tiempos) 0 vida mínima y define el punto de partida u origen de la distribución.
- η es el parámetro de escala, extensión de la distribución a lo largo, del eje de los tiempos. Cuando $(t - t_0) = \eta$ la fiabilidad viene dada por:
 $R(t) = \exp - (1)^\beta = 1/\exp 1^\beta = 1 / 2,718 = 0,368$ (36,8%)
 Entonces la constante representa también el tiempo, medido a partir de $t_0 = 0$, según lo cual dado que $F(t) = 1 - 0,368 = 0,632$, el 63,2 % de la población se espera que falle, cualquiera que sea el valor de β ya que como hemos visto su valor no influye en los cálculos realizados. Por esta razón también se le llama usualmente vida característica.
- β es el parámetro de forma y representa la pendiente de la recta describiendo el grado de variación de la tasa de fallos.





Distribución lognormal

La distribución lognormal tiene, principalmente, las siguientes aplicaciones:

- Representa la evolución con el tiempo de la tasa de fallos, $\lambda(t)$, en la primera fase de vida de un componente, la correspondiente a los fallos infantiles en la "curva de la bañera" entendiéndose como tasa de fallos la probabilidad de que un componente que ha funcionado hasta el instante t , falle entre t y $t + dt$. En este caso la variable independiente de la distribución es el tiempo (figura 1).
- Permite fijar tiempos de reparación de componentes, siendo también en este caso el tiempo la variable independiente de la distribución.
- Describe la dispersión de las tasas de fallo de componentes, ocasionada por diferente origen de los datos, distintas condiciones de operación, entorno, bancos de datos diferentes, etc. En este caso la variable independiente de la distribución es la tasa de fallos.

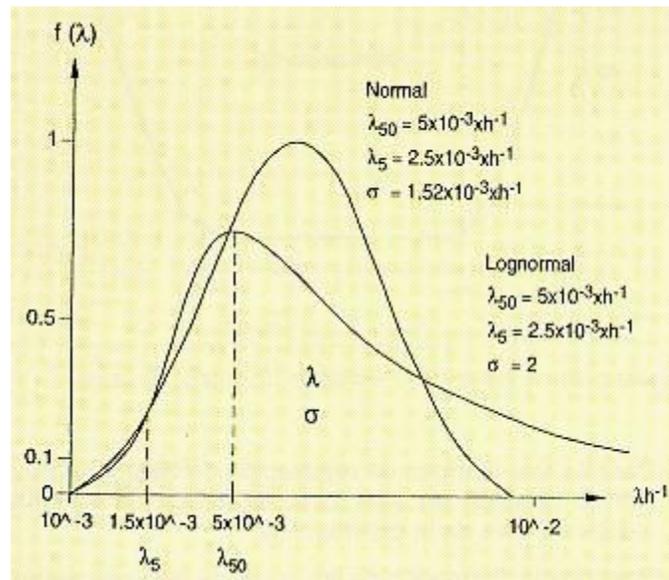
La distribución lognormal tiene dos parámetros: m^* (media aritmética del logaritmo de los datos o tasa de fallos) y σ (desviación estándar del logaritmo de los datos o tasa de fallos).

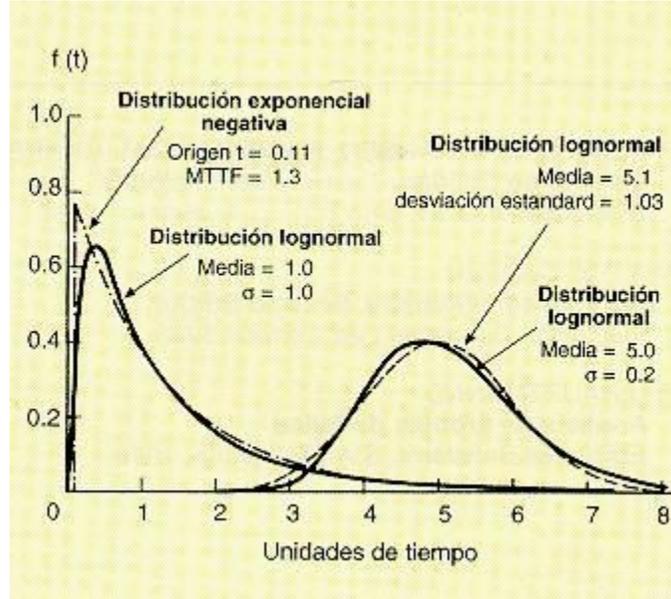
La distribución lognormal se caracteriza por las siguientes propiedades:

- Asigna a valores de la variable < 0 la probabilidad 0 y de este modo se ajusta a las tasas y probabilidades de fallo que de esta forma sólo pueden ser positivas.

- Como depende de dos parámetros, se ajusta bien a un gran número de distribuciones empíricas.
- Es idónea para parámetros que son a su vez producto de numerosas cantidades aleatorias (múltiples efectos que influyen sobre la fiabilidad de un componente).
- La esperanza matemática o media en la distribución lognormal es mayor que su mediana. De este modo da más importancia a los valores grandes de las tasas de fallo que una distribución normal con los mismos percentiles del 5% y 50% tendiendo, por tanto, a ser pesimista.

Lognormal	LOGN(m,o)
Función de densidad	$f(x)=$ si $x>0$ de otra 0 manera
Distribución acumulada	$F(x)=$ no existe ecuación
Parámetros	Parámetro de escala: m Parámetro de forma: o
Rango	[0, &]
Media	$e^{-u+o/2}$
Varianza	$e^{2*o+o^2}(e^{o^2}-1)$





LA DISTRIBUCIÓN T

Suponga que se toma una muestra aleatoria de tamaño $n < 30$ de una población con **distribución normal** con media μ y varianza σ^2 . Se ha establecido anteriormente que la media muestral \bar{X} también tendrá **distribución normal** con media $\mu_{\bar{x}} = \mu$ y varianza $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$. Por lo tanto, la

variable $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ tendrá distribución normal estándar.

Sin embargo, si la varianza de la población es desconocida, entonces la variable anterior ya no tiene distribución normal estándar y debe usarse otro estadístico denominado **estadístico T** o de "Student":

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

La distribución de este estadístico también tiene forma tipo "campana simétrica" dependiendo del valor de n . Este parámetro determina la forma particular de la distribución con la siguiente definición: $\nu = n - 1$ grados de libertad, (ν léase "nu")

Este estadístico es útil cuando por consideraciones prácticas, no se puede tomar una muestra aleatoria grande. Pero, para usar este estadístico, es necesario que la población tenga distribución normal.

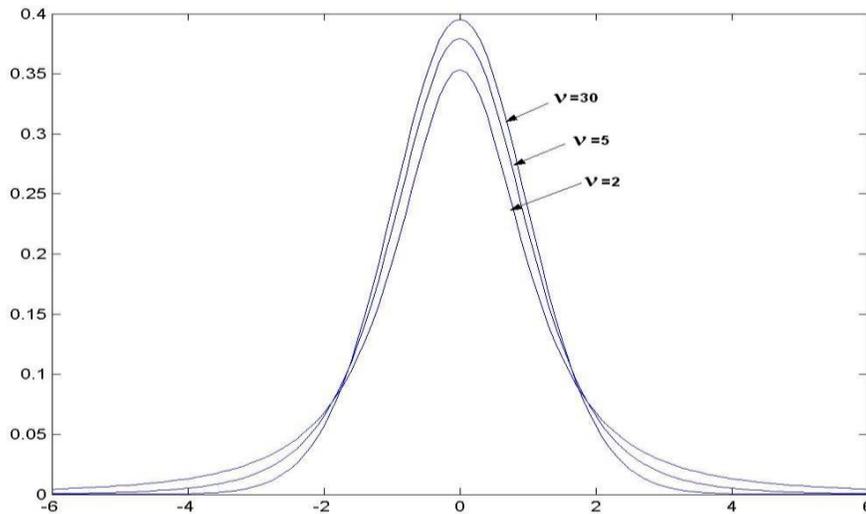
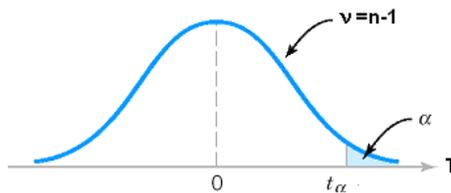


Fig. Distribución T para v = 2, 5, 30 grados de libertad.

Para usar esta distribución, si no se dispone de un utilitario informático, se usan tablas que contienen algunos valores de T para diferentes grados de libertad mediante la siguiente definición:

t_α : valor de t tal que $P(T \geq t_\alpha) = \alpha$, como se muestra en el siguiente gráfico:



DISTRIBUCION DE PROBABILIDAD F

Supóngase que deseamos comparar las varianzas de dos poblaciones normales basados en la información contenida en muestras aleatorias independiente de las dos poblaciones. Supóngase que una muestra aleatoria contiene n_1 variables aleatorias distribuidas normalmente con una varianza común σ_1^2 y que la otra muestra aleatoria contiene n_2 variables aleatorias distribuidas normalmente con una varianza común σ_1^2 y que la otra muestra aleatoria contiene n_2 variables aleatorias distribuidas normalmente con una varianza común σ_1^2 . Si calculamos S_1^2 de las observaciones en la muestra 1, entonces S_1^2 es una estimación de σ_1^2 . De manera similar, S_2^2 calculada a partir de las observaciones de la segunda muestra es una estimación para σ_2^2 . Así intuitivamente podríamos pensar en utilizar S_1^2 / S_2^2 para hacer inferencias con

respecto a las magnitudes relativas de σ_1^2 y σ_2^2 . Si dividimos cada S_i^2 por σ_i^2 , entonces la razón siguiente

$$\frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} = \left(\frac{\sigma_2^2}{\sigma_1^2} \right) \left(\frac{S_1^2}{S_2^2} \right)$$

tiene una distribución F con $(n_1 - 1)(n_2 - 1)$ grados de libertad. La definición general de una distribución F es como sigue:

DEFINICION Sean χ_1^2 y χ_2^2 variables aleatorias ji - cuadrada con ν_1 y ν_2 grados de libertad. Respectivamente. Entonces si χ_1^2 y χ_2^2 son independientes,

$$F = \frac{\chi_1^2 / \nu_1}{\chi_2^2 / \nu_2}$$

se dice que tiene una distribución F con ν_1 grados de libertad del numerador y ν_2 grados de libertad del denominador.

DISTRIBUCION DE PROBABILIDAD χ^2 - CUADRADA

Considerando nuevamente las muestras aleatorias independientes de distribuciones normales, sabemos que

$$\chi_1^2 = (n_1 - 1)S_1^2 / \sigma_1^2 \text{ y } \chi_2^2 = (n_2 - 1)S_2^2 / \sigma_2^2$$

tienen distribuciones χ^2 independientes con

$$\nu_1 = (n_1 - 1) \text{ y } \nu_2 = (n_2 - 1)$$

grados de libertad, respectivamente.

Así la definición implica que

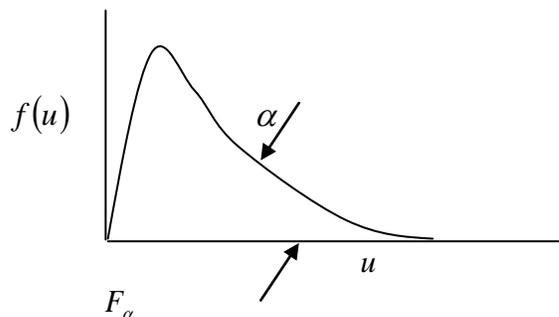
$$F = \frac{\chi_1^2 / \nu_1}{\chi_2^2 / \nu_2} = \frac{(n_1 - 1)S_1^2 / \sigma_1^2 (n_1 - 1)}{(n_2 - 1)S_2^2 / \sigma_2^2 (n_2 - 1)} = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2}$$

tiene una distribución F con $(n_1 - 1)$ grados de libertad del numerador y $(n_2 - 1)$ grados de libertad del denominador.

En la figura 7.3 se muestra la gráfica de una típica función de densidad F . Los valores de F_α tales que $P(F > F_\alpha) = \alpha$ se dan en la tabla Xi Cuadrado, para los valores de $\alpha = 0.100, 0.050, 0.025, 0.010$ y 0.005 . En la , los encabezados de las columnas corresponden a los grados de libertad del numerador, en tanto que los grados de libertad del denominador se encuentran como los encabezados principales de los renglones.

Frente a los grados de libertad del denominador (los encabezados de los renglones), se encuentran los valores de $\alpha = 0.100, 0.050, 0.025, 0.010$ y 0.005 . Por ejemplo, si la variable F estudiada tiene 5 grados de libertad del numerador y 7 grados de libertad del denominador, $F_{0.100} = 2.88$, $F_{0.050} = 3.97$, $F_{0.025} = 5.29$, $F_{0.010} = 7.46$ y $F_{0.005} = 9.52$. Luego la probabilidad de que una variable aleatoria con una distribución F con 5 grados de libertad del numerador y 7 grados de libertad del denominador exceda de 7.46 es 0.01. Lo correspondiente se afirma para los demás casos.

FIGURA
Una típica función de densidad
De probabilidad F



DISTRIBUCION PROBABILIDAD GAMA.

Los tiempos que tardan en revisar un motor de un automóvil ó avión tienen una distribución de frecuencias sesgadas. Las poblaciones asociadas a estas variables aleatorias frecuentemente tienen distribuciones que se pueden modelar adecuadamente por la función de densidad tipo gamma.

Función de densidad de probabilidad para una variable aleatoria tipo gamma:

$$\alpha, \beta > 0; 0 \leq y < \infty$$

$$f(y) = \frac{y^{\alpha-1} e^{-y/\beta}}{\beta^\alpha \Gamma(\alpha)}$$

$$0$$

En donde:

$$\tau(\alpha) = \int_0^{\alpha} y^{\alpha-1} e^{-y} dy$$

La cantidad de la de la función alfa se conoce como la función gamma. La integración directa nos da que la función uno igual a uno. La integración por partes nos da que la función de alfa menos uno alfa menos uno por la función alfa menos uno para cualquier intervalo de alfa mayor o igual a uno y que la función de n sea igual a n menos uno factorial, para un número entero n.

En el caso especial cuando alfa es un número entero, se puede expresar la función de distribución de una variable aleatoria tipo gamma como una suma de ciertas variables aleatorias de Poisson.

Si alfa no es un número entero, es imposible encontrar la antiderivada del integrando de la expresión:

$$0 < c < d < \alpha$$

donde

$$\int_c^d \frac{y^{\alpha-1} e^{-y/\beta}}{\beta^{\alpha} \tau(\alpha)} dy$$

Y por lo tanto es importante obtener las áreas bajo la función de densidad tipo gamma mediante integración directa.

Hay dos casos especiales de las variables aleatorias tipo gamma que merece consideración particular:

Una variable aleatoria tipo gamma que tiene una función de densidad con parámetros alfa igual a v entre dos y beta igual a dos se denomina variable aleatoria ji - cuadrada.

Ji - cuadrada se presenta con frecuencia en la teoría de la estadística. El parámetro v se denomina número de grados de libertad asociado a la variable aleatoria ji - cuadrada.

La función de densidad gamma para el caso especial v = 1 se denomina función de densidad exponencial.

$$\beta > 0; 0 \leq y < \infty$$

$$f(y) = \frac{1}{\beta} e^{-y/\beta}$$

En cualquier punto.

La función de densidad exponencial muchas veces es útil en los modelos de duración de componentes eléctricos.

Un fusible es un ejemplo de un componente para el cual este supuesto suele cumplirse.

LA DISTRIBUCION DE PROBABILIDAD BETA.

La distribución de probabilidad beta es una función de densidad con dos parámetros definida en el intervalo cerrado $0 \leq y \leq 1$. Se utiliza frecuentemente como modelo para fracciones, tal como la proporción de impurezas en un producto químico o la fracción de tiempo que una maquina está en reparación.

Función de densidad probabilidad:

$$\alpha, \beta > 0; 0 \leq y \leq 1$$

$$f(y) = \left\{ \frac{y^{\alpha-1} (1-y)^{\beta-1}}{B(\alpha, \beta)} \right.$$

En cualquier otro punto donde

$$B(\alpha, \beta) = \int_0^1 y^{\alpha-1} (1-y)^{\beta-1} dy = \frac{\tau(\alpha)\tau(\beta)}{\tau(\alpha + \beta)}$$

Nótese que la definición de (y) sobre el intervalo $0 \leq y \leq 1$ restringe su aplicación. Si $c \leq y \leq d$, $y = (y - c) / (d - c)$ definirá una nueva variable en el intervalo $0 \leq y \leq 1$. Así la función de densidad beta se puede aplicar a una variable aleatoria definida en el intervalo $c \leq y \leq d$ mediante una traslación y una medición en la escala.

La función de distribución acumulativa para la variable aleatoria beta se llama comúnmente función beta y esta dada por

$$F(y) = \int_0^y \frac{t^{\alpha-1} (1-t)^{\beta-1}}{B(\alpha, \beta)} dt = I_y(\alpha, \beta)$$

Para valores enteros de alfa y beta, $I_y(\alpha, \beta)$ está relacionada con la función de probabilidad binomial. Cuando $y = p$, se puede demostrar que

$$F(p) = \int \frac{y^{\alpha-1} (1-y)^{\beta-1}}{B(\alpha, \beta)} dy = \sum_{y=\alpha}^n p^y (1-p)^{n-y}$$

En donde $0 < p < 1$ y n igual a alfa más beta menos uno.

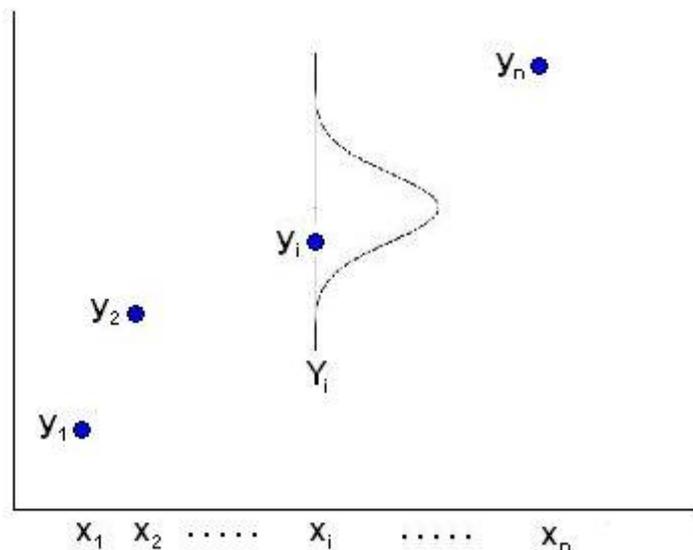
REGRESIÓN LINEAL SIMPLE

El propósito de este estudio es proporcionar los conceptos y técnicas para determinar una ecuación que describa de manera razonable a un conjunto de datos dado. Este estudio se denomina **análisis de regresión** y la ecuación empírica obtenida se denomina **ecuación de regresión** la cual sustituye a un modelo teórico no disponible

En este primer enfoque se supondrá que se tiene un conjunto de **n** mediciones u observaciones y_1, y_2, \dots, y_n de una variable **Y** denominada **variable de respuesta** las cuales corresponden a un conjunto x_1, x_2, \dots, x_n que representan los valores de una variable **X** denominada **variable de predicción**.

Se supondrá que existe una correspondencia de **X** a **Y**, de tal manera que cada valor y_i está asociado con un valor x_i .

Es importante reconocer que cada valor y_i es el resultado de una medición, por lo tanto, es posible que pudiesen haber otros valores y_i para el mismo valor dado x_i . Esto nos permite reconocer que y_i proviene de una variable aleatoria Y_i la cual debe tener alguna distribución de probabilidad. Tratemos de visualizarlo en el siguiente gráfico:



Supondremos que existe una relación lineal entre **X** y **Y**. Este hecho puede reconocerse graficando los puntos (x_i, y_i) , $i = 1, 2, \dots, n$ y observando la “tendencia lineal” de los puntos. Esta representación se denomina **gráfico de dispersión**.

Se propone un modelo lineal que tome en cuenta la aleatoriedad de **Y** y permita luego explicar los errores de medición.

Modelo probabilista propuesto:

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

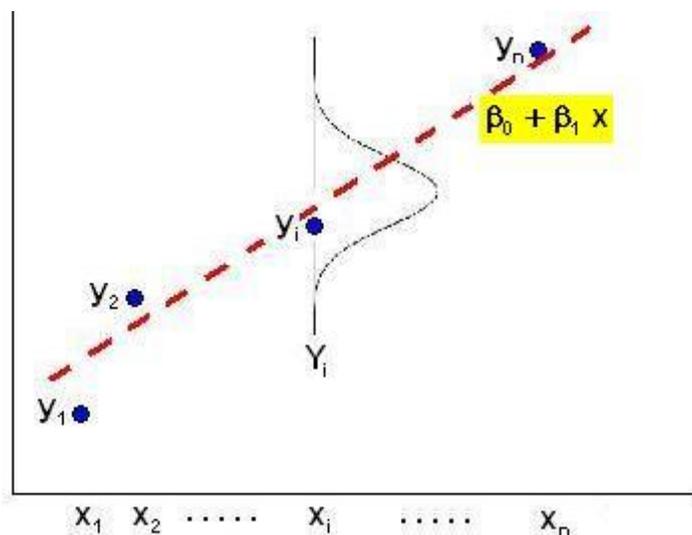
siendo ε el componente aleatorio de Y

Se supondrá que para cada variable aleatoria Y_i el componente aleatorio ε_i tiene la misma distribución de probabilidad y que además son independientes.

$$\varepsilon_i \sim N(0, \sigma^2) \quad (\text{distribución normal con media } 0 \text{ y varianza } \sigma^2)$$

Por lo tanto, el valor esperado de este modelo, es una recta teórica (desconocida) con los parámetros β_0 y β_1 que deben estimarse

$$E[Y] = \beta_0 + \beta_1 x$$



RECTA DE MÍNIMOS CUADRADOS

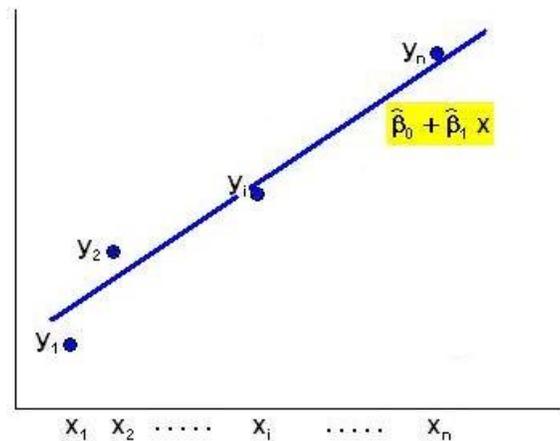
Es un procedimiento matemático para estimar los parámetros β_0 y β_1 de la recta de regresión utilizando los datos dados.

El objetivo es colocar una recta entre los puntos de tal manera la suma de las distancias de esta recta a los puntos sea la menor posible.

Definición

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

Es la recta de mínimos cuadrados. $\hat{\beta}_0$, $\hat{\beta}_1$ son los estimadores de β_0 y β_1



Para cada valor x_i se tiene el valor observado y_i y un valor \hat{y}_i obtenido con la recta de mínimos cuadrados: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

Sea $e_i = y_i - \hat{y}_i$,

Entonces, el criterio de mínimos cuadrados consiste en minimizar e_i^2 para todos los puntos. El cuadrado puede interpretarse como una manera de cuantificar las distancias. No importa si el punto está sobre o debajo de la recta

Criterio de mínimos cuadrados

$$SCE = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

(Lea SCE: "Suma de Cuadrados del Error")

El procedimiento matemático para realizar esta optimización es:

$$\frac{\partial SCE}{\partial \beta_0} = 0, \quad \frac{\partial SCE}{\partial \beta_1} = 0$$

Con facilidad se llega al sistema de ecuaciones lineales:

$$\hat{\beta}_0 n + \hat{\beta}_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i$$

$$\hat{\beta}_0 \sum_{i=1}^n x_i + \hat{\beta}_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i$$

De donde se obtienen finalmente los estimadores $\hat{\beta}_0$, $\hat{\beta}_1$

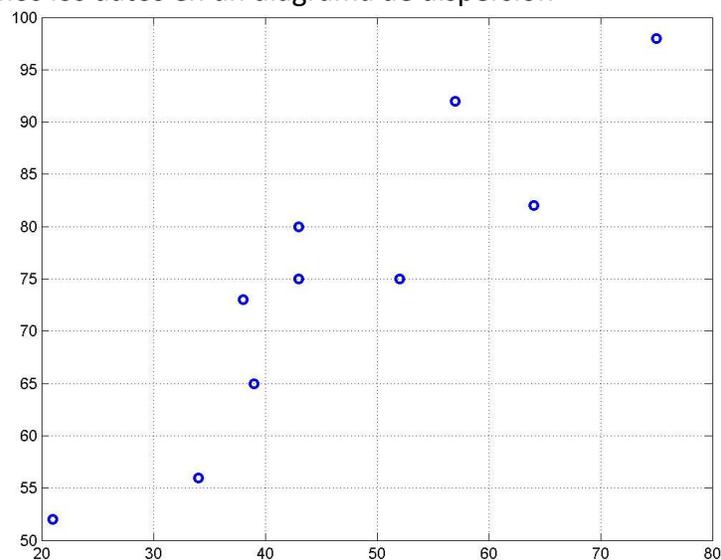
Ejemplo

Los siguientes datos corresponden a una muestra aleatoria de 10 estudiantes que han tomado cierta materia. Los datos incluyen la calificación parcial y la calificación final. Se pretende encontrar un modelo de regresión que permita predecir la calificación final que obtendría un estudiante dada su calificación parcial.

Estudiante	Nota Parcial	Nota final
1	39	65
2	43	75
3	21	52
4	64	82
5	57	92
6	43	80
7	38	73
8	75	98
9	34	56
10	52	75

Solución

Primero representamos los datos en un diagrama de dispersión



Se observa que al incrementar x (variable de predicción) también se incrementa y (variable de respuesta)

Obtención de la recta de mínimos cuadrados

Cálculos

i	x_i	y_i	x_i²	x_iy_i
1	39	65	1521	2535
2	43	75	1849	3225
3	21	52	441	1092
4	64	82	4096	5248
5	57	92	3249	5244
6	43	80	1849	3440
7	38	73	1444	2774
8	75	98	5625	7350
9	34	56	1156	1904
10	52	75	2704	3900
$\sum_{i=1}^{10}$	466	748	23934	36712

Sustituimos en el sistema de ecuaciones lineales:

$$10 \hat{\beta}_0 + 466 \hat{\beta}_1 = 748$$

$$466 \hat{\beta}_0 + 23934 \hat{\beta}_1 = 36712$$

De donde se obtienen: $\hat{\beta}_0 = 35.83$, $\hat{\beta}_1 = 0.836$

Ecuación de mínimos cuadrados: $\hat{y} = 35.83 + 0.836 x$

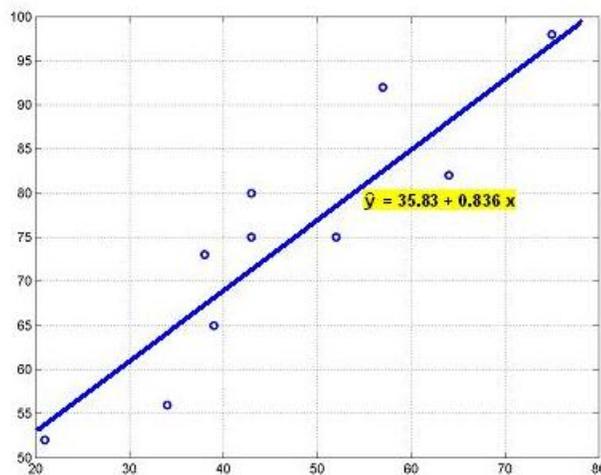


Gráfico de la recta de mínimos cuadrados

Ahora pretendamos predecir la calificación final que obtendrá un estudiante que obtuvo 50 en su calificación parcial:

$$\hat{y} = 35.83 + 0.836 (50) = 77.63$$

REGRESIÓN LINEAL MÚLTIPLE

Considere que una variable Y depende de k variables x_1, x_2, \dots, x_k

Para describir esta relación se propone un modelo de regresión lineal múltiple

Modelo teórico probabilista propuesto:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon$$

Siendo ϵ el componente aleatorio de Y

Note que cuando $k = 1$, se reduce al modelo de regresión lineal simple visto.

Suponer que se tiene una muestra aleatoria $(x_{1,i}, x_{2,i}, \dots, x_{k,i}, y_i), i = 1, 2, \dots, n$

Fijados los k valores $x_{1,i}, x_{2,i}, \dots, x_{k,i}$ se tiene una observación o medición y_i la cual es uno de los posibles valores de la variable aleatoria Y_i

Se supondrá que para cada variable aleatoria Y_i el componente aleatorio ϵ_i tiene la misma distribución de probabilidad, y que además son independientes.

$$\epsilon_i \sim N(0, \sigma^2) \text{ (distribución normal con media } 0 \text{ y varianza } \sigma^2)$$

MODELO DE REGRESIÓN LINEAL MÚLTIPLE DE MÍNIMOS CUADRADOS

Para estimar los $k + 1$ parámetros $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ se usará un procedimiento similar al modelo de regresión lineal simple con el método de mínimos cuadrados

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k$$

Sea $e_i = y_i - \hat{y}_i$,

en donde y_i : valor observado en la muestra

\hat{y}_i : valor obtenido con el modelo de mínimos cuadrados:

Criterio de mínimos cuadrados

Minimizar

$$SCE = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1,i} - \hat{\beta}_2 x_{2,i} - \dots - \hat{\beta}_k x_{k,i})^2 .$$

Utilizando $\frac{\partial SCE}{\partial \beta_i} = 0$, $i=0, 1, 2, \dots, k$, se obtienen las ecuaciones normales para encontrar los estimadores $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$

Consideremos el caso específico $k=2$

Y depende de 2 variables x_1, x_2

Modelo teórico probabilista propuesto:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon .$$

Modelo de regresión lineal múltiple de mínimos cuadrados;

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$

Para encontrar $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$

$$\frac{\partial SCE}{\partial \beta_i} = 0, \quad i=0, 1, 2$$

Se obtienen las ecuaciones normales

$$\begin{aligned} n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_{1,i} + \hat{\beta}_2 \sum_{i=1}^n x_{2,i} &= \sum_{i=1}^n y_i \\ \hat{\beta}_0 \sum_{i=1}^n x_{1,i} + \hat{\beta}_1 \sum_{i=1}^n x_{1,i}^2 + \hat{\beta}_2 \sum_{i=1}^n x_{1,i} x_{2,i} &= \sum_{i=1}^n x_{1,i} y_i \\ \hat{\beta}_0 \sum_{i=1}^n x_{2,i} + \hat{\beta}_1 \sum_{i=1}^n x_{2,i} x_{1,i} + \hat{\beta}_2 \sum_{i=1}^n x_{2,i}^2 &= \sum_{i=1}^n x_{2,i} y_i \end{aligned}$$

Las expresamos en notación matricial

$$A \hat{\beta} = C \Rightarrow \begin{bmatrix} n & \sum_{i=1}^n x_{1,i} & \sum_{i=1}^n x_{2,i} \\ \sum_{i=1}^n x_{1,i} & \sum_{i=1}^n x_{1,i}^2 & \sum_{i=1}^n x_{1,i} x_{2,i} \\ \sum_{i=1}^n x_{2,i} & \sum_{i=1}^n x_{2,i} x_{1,i} & \sum_{i=1}^n x_{2,i}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{1,i} y_i \\ \sum_{i=1}^n x_{2,i} y_i \end{bmatrix}$$

REGRESIÓN LINEAL MÚLTIPLE EN NOTACIÓN MATRICIAL

El modelo teórico probabilista puede expresarse convenientemente en notación matricial

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon$$

Consideramos el caso específico $k=2$ en donde Y depende de 2 variables x_1, x_2

Modelo teórico probabilista propuesto:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

Datos de la muestra:

$$(x_{1,i}, x_{2,i}, y_i), i = 1, 2, \dots, n$$

Cada observación y_i es un valor de la variable aleatoria Y_i , $i = 1, 2, \dots, n$

$$Y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \epsilon_i, i = 1, 2, \dots, n$$

Se puede expresar en forma desarrollada,

$$Y_1 = \beta_0 + \beta_1 x_{1,1} + \beta_2 x_{2,1} + \epsilon_1$$

$$Y_2 = \beta_0 + \beta_1 x_{1,2} + \beta_2 x_{2,2} + \epsilon_2$$

...

...

$$Y_n = \beta_0 + \beta_1 x_{1,n} + \beta_2 x_{2,n} + \epsilon_n$$

El modelo teórico expresado en notación matricial

$$Y = X \beta + \epsilon \Rightarrow \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & x_{2,1} \\ 1 & x_{1,2} & x_{2,2} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & x_{1,n} & x_{2,n} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \cdot \\ \epsilon_n \end{bmatrix}$$

X se denomina **matriz de diseño**

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{1,1} & x_{2,1} \\ 1 & x_{1,2} & x_{2,2} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & x_{1,n} & x_{2,n} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \cdot \\ \epsilon_n \end{bmatrix}$$

El sistema de ecuaciones normales del modelo de regresión lineal múltiple de mínimos cuadrados puede entonces expresarse mediante la matriz de diseño X

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \Rightarrow A \hat{\beta} = C$$

$$A = \begin{bmatrix} n & \sum_{i=1}^n x_{1,i} & \sum_{i=1}^n x_{2,i} \\ \sum_{i=1}^n x_{1,i} & \sum_{i=1}^n x_{1,i}^2 & \sum_{i=1}^n x_{1,i} x_{2,i} \\ \sum_{i=1}^n x_{2,i} & \sum_{i=1}^n x_{2,i} x_{1,i} & \sum_{i=1}^n x_{2,i}^2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdot & \cdot & 1 \\ x_{1,1} & x_{1,2} & \cdot & \cdot & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdot & \cdot & x_{2,n} \end{bmatrix} \cdot \begin{bmatrix} 1 & x_{1,1} & x_{2,1} \\ 1 & x_{1,2} & x_{2,2} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & x_{1,n} & x_{2,n} \end{bmatrix}$$

$$A = X^T X$$

$$C = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{1,i} y_i \\ \sum_{i=1}^n x_{2,i} y_i \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdot & \cdot & 1 \\ x_{1,1} & x_{1,2} & \cdot & \cdot & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdot & \cdot & x_{2,n} \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{bmatrix}$$

$$C = X^T y$$

$$\text{Sea } \hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix}$$

$$\text{Entonces } A \hat{\beta} = C \Rightarrow X^T X \hat{\beta} = X^T y \Rightarrow \hat{\beta} = (X^T X)^{-1} (X^T y)$$

- $\hat{\beta}$: vector con los estimadores de mínimos cuadrados
 X : matriz de diseño (datos de la muestra)
 y : vector de observaciones obtenidas en la muestra

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{bmatrix}$$

ANALISIS DE VARIANZA

Para simplificar la escritura de algunas expresiones de interés, se escriben las siguientes fórmulas dándoles en algunos casos una identificación.

Las demostraciones correspondientes utilizan las propiedades de las sumatorias

$$(0) \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

$$(1) \quad S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2$$

$$(2) \quad SCT = S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2$$

$$(3) \quad S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - \frac{1}{n} \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)$$

$$(4) \quad SCE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = S_{yy} - \frac{S_{xy}^2}{S_{xx}}$$

$$(5) \quad SCR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \frac{S_{xy}^2}{S_{xx}}$$

Demostración de (1)

$$\begin{aligned}
S_{xx} &= \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) = \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + n\bar{x}^2 \\
&= \sum_{i=1}^n x_i^2 - 2\bar{x}n \frac{1}{n} \sum_{i=1}^n x_i + n\bar{x}^2 = \sum_{i=1}^n x_i^2 - 2\bar{x}n\bar{x} + n\bar{x}^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2 \\
&= \sum_{i=1}^n x_i^2 - n \left(\frac{\sum_{i=1}^n x_i}{n} \right)^2 = \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2
\end{aligned}$$

ESTIMACIÓN DE LA VARIANZA

En el modelo de regresión lineal simple, el componente aleatorio es

$$\epsilon_i \sim N(0, \sigma^2)$$

La varianza σ^2 puede ser estimada con la varianza de los errores de los datos de la muestra

Definición de la varianza muestral (es un estimador insesgado de σ^2)

$$e_i = y_i - \hat{y}_i$$

$$SCE = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$s^2 = \frac{SCE}{n-2} = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Usamos la fórmula (4) definida anteriormente:

$$SCE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = S_{yy} - \frac{S_{xy}^2}{S_{xx}}$$

De cual se obtiene

$$S_{yy} = \frac{S_{xy}^2}{S_{xx}} + SCE$$

Con las fórmulas (2), (4) y (5) escritas arriba se obtiene

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

En forma simbólica

$$SCT = SCR + SCE$$

SCT: **Suma de cuadrados total (es la varianza de los datos de la muestra)**

SCR: **Suma de cuadrados de regresión (es la varianza del modelo de mínimos cuadrados propuesto)**

SCE: **Suma de cuadrados del error (es la diferencia entre los datos y el modelo de mínimos cuadrados propuesto)**

Mientras menor es el valor de SCE, mejor es la eficacia del modelo de mínimos cuadrados propuesto. Su varianza explica adecuadamente a la varianza de los datos.

Para el modelo de regresión lineal multivariado

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon$$

Se define la varianza muestral

$$s^2 = \frac{SCE}{n - k - 1}$$

Para este modelo también es válida la ecuación anterior con la misma interpretación de las fuentes de variación:

$$SCT = SCR + SCE$$

Para medir la eficacia o poder de explicación del modelo de mínimos cuadrados propuesto se define la siguiente medida:

Coeficiente de determinación del modelo de regresión lineal múltiple

$$R^2 = \frac{SCR}{SCT}, \quad 0 \leq R^2 \leq 1$$

El valor de R^2 muestra la proporción de la variabilidad de los datos que es explicada por el modelo de mínimos cuadrados.

Ejemplo

Suponer que con los datos se obtuvo $R^2 = 0.934$

Esto significa que el 93.4% de la variación de la variable de interés, es explicada por el modelo de mínimos cuadrados.

Se entiende también que el valor de SCE debe ser pequeño y esto debe interpretarse como que los datos no están muy alejados de la recta de mínimos cuadrados y que ésta los representa adecuadamente.

INFERENCIAS CON LOS PARÁMETROS DEL MODELO DE REGRESIÓN LINEAL

Modelo teórico probabilista propuesto:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon$$

$$\epsilon \sim N(0, \sigma^2)$$

Para estimar los $k + 1$ parámetros $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ se propone el modelo de regresión lineal de mínimos cuadrados:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k$$

Suposiciones

$\hat{\beta}_i$ es un estimador insesgado del parámetro β_i
 $E[\hat{\beta}_i] = \beta_i$

El estimador $\hat{\beta}_i$ tiene distribución normal
 $\hat{\beta}_i \sim N(\beta_i, \sigma^2_{\hat{\beta}_i})$

Notación

$V[\hat{\beta}_i] = \sigma^2_{\hat{\beta}_i} = \sigma_{ii}$ (varianza de $\hat{\beta}_i$)
 $Cov[\hat{\beta}_i, \hat{\beta}_j] = \sigma_{\hat{\beta}_i \hat{\beta}_j} = \sigma_{ij}$ (covarianza de $\hat{\beta}_i, \hat{\beta}_j$)

Definición

Matriz de varianzas-covarianzas

$$[\sigma_{ij}] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdot & \cdot & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \cdot & \cdot & \sigma_{2k} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{k1} & \sigma_{k2} & \cdot & \cdot & \sigma_{kk} \end{bmatrix}$$

La teoría estadística demuestra la siguiente relación

$$[\sigma_{ij}] = [X^T X]^{-1} \sigma^2 \cong [X^T X]^{-1} S^2$$

X es la matriz de diseño del modelo de mínimos cuadrados

$$S^2 = \frac{\text{SCE}}{n-k-1} \text{ es la varianza del modelo de mínimos cuadrados .}$$

INTERVALO DE CONFIANZA PARA LOS PARÁMETRO β_i

Parámetro: $\beta_i, i = 0, 1, \dots, k$

Estimador: $\hat{\beta}_i, i = 0, 1, \dots, k$

El estadístico

$$t = \frac{\hat{\beta}_i - \beta_i}{\sigma_{\hat{\beta}_i}}, \text{ tiene distribución } t \text{ con } v = n-k-1 \text{ grados de libertad}$$

Definición

Intervalo de confianza para β_i con nivel $1 - \alpha$

$$\hat{\beta}_i - t_{\alpha/2} \sigma_{\hat{\beta}_i} \leq \beta_i \leq \hat{\beta}_i + t_{\alpha/2} \sigma_{\hat{\beta}_i}, \quad i=0, 1, \dots, k$$

PRUEBA DE HIPÓTESIS PARA LOS PARÁMETROS β_i

- 1) $H_0: \beta_i = b_0$ (algún valor especificado para el parámetro)
- 2) $H_a: \beta_i < b_0$ ó $\beta_i > b_0$ ó $\beta_i \neq b_0$
- 3) α nivel de significancia de la prueba
- 4) Estadístico de prueba

$$t = \frac{\hat{\beta}_i - b_0}{\sigma_{\hat{\beta}_i}}, \text{ tiene distribución } t \text{ con } v = n-k-1 \text{ grados de libertad}$$

H_a Región de rechazo de H_0 en favor de H_a

$$\beta_i < b_0 \quad t < -t_{\alpha}$$

$$\beta_i > b_0 \quad t > t_{\alpha}$$

$$\beta_i \neq b_0 \quad t < -t_{\alpha/2} \vee t > t_{\alpha/2}$$

- 5) Calcule el valor del estadístico
- 6) Decisión

REGRESIÓN LINEAL MÚLTIPLE

Ejemplo

Se desea definir un modelo de regresión para predecir la calificación final Y que tendría un estudiante, dada su calificación parcial x_1 y su porcentaje de asistencia a clases x_2 considerando los siguientes datos:

Calificación parcial x_1	Asistencia a clases x_2	Calificación final Y
67	75	80
65	78	77
78	79	94
60	83	70
64	65	51
61	76	70

a) Modelo de regresión lineal múltiple propuesto

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

b) Notación matricial

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{bmatrix} = \begin{bmatrix} 1 & 67 & 75 \\ 1 & 65 & 78 \\ 1 & 78 & 79 \\ 1 & 60 & 83 \\ 1 & 64 & 65 \\ 1 & 61 & 76 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & 67 & 75 \\ 1 & 65 & 78 \\ 1 & 78 & 79 \\ 1 & 60 & 83 \\ 1 & 64 & 65 \\ 1 & 61 & 76 \end{bmatrix} \quad (\text{matriz de diseño})$$

c) Estimación de parámetros por mínimos cuadrados

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = (X^T X)^{-1} (X^T y)$$

$$= \begin{bmatrix} n & \sum_{i=1}^n x_{1,i} & \sum_{i=1}^n x_{2,i} \\ \sum_{i=1}^n x_{1,i} & \sum_{i=1}^n x_{1,i}^2 & \sum_{i=1}^n x_{1,i} x_{2,i} \\ \sum_{i=1}^n x_{2,i} & \sum_{i=1}^n x_{2,i} x_{1,i} & \sum_{i=1}^n x_{2,i}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{1,i} y_i \\ \sum_{i=1}^n x_{2,i} y_i \end{bmatrix}$$

$$= \begin{bmatrix} 6 & 395 & 456 \\ 395 & 26215 & 30033 \\ 456 & 30033 & 34840 \end{bmatrix}^{-1} \begin{bmatrix} 442 \\ 29431 \\ 33877 \end{bmatrix}$$

$$= \begin{bmatrix} 48.974 & -0.2880 & -0.3927 \\ -0.2880 & 4.760 \times 10^{-3} & -3.360 \times 10^{-4} \\ -0.3927 & -3.366 \times 10^{-4} & 5.458 \times 10^{-3} \end{bmatrix} \begin{bmatrix} 442 \\ 29431 \\ 33877 \end{bmatrix}$$

$$= \begin{bmatrix} -134.07 \\ 1.4888 \\ 1.4437 \end{bmatrix}$$

$$\hat{y} = -134.07 + 1.4888 x_1 + 1.4437 x_2$$

- d) Pronostique la calificación final si la calificación parcial es 75 y el porcentaje de asistencia clases es 80

$$\hat{y} = -134.07 + 1.4888 (75) + 1.4437 (80) = 93.08$$

- e) Análisis de varianza

$$\bar{y} = \frac{1}{6} \sum_{i=1}^6 y_i = \frac{1}{6} (80 + 77 + 94 + 70 + 51 + 70) = 73.667$$

$$SCT = \sum_{i=1}^6 (y_i - \bar{y})^2 = 1005.33$$

$$SCR = \sum_{i=1}^6 (\hat{y}_i - \bar{y})^2 = 906.7503$$

$$SCE = \sum_{i=1}^6 (y_i - \hat{y}_i)^2 = 98.583$$

f) Coeficiente de determinación

$$R^2 = \frac{SCR}{SCT} = \frac{906.7503}{1005.33} = 0.9019 = 90.19\%$$

g) Estimación de la varianza

$$S^2 = \frac{SCE}{n-k-1} = \frac{98.583}{6-2-1} = 32.861$$

h) Matriz de varianza-covarianza

$$\begin{aligned} [\sigma_{i,j}] &= [X^T X]^{-1} \sigma^2 \cong [X^T X]^{-1} S^2 \\ &= \begin{bmatrix} 48.974 & -0.2880 & -0.3927 \\ -0.2880 & 4.760 \times 10^{-3} & -3.360 \times 10^{-4} \\ -0.3927 & -3.366 \times 10^{-4} & 5.458 \times 10^{-3} \end{bmatrix} (32.861) \\ &= \begin{bmatrix} 1609.33 & -9.4653 & -12.904 \\ -9.4653 & 0.15654 & -0.01106 \\ -12.904 & -0.01106 & 0.17937 \end{bmatrix} \end{aligned}$$

i) Varianza de los estimadores de mínimos cuadrados

$$V[\hat{\beta}_i] = \sigma_{\beta_i}^2 = \sigma_{ii}, \quad i = 0, 1, 2$$

$$\sigma_{11} = 1609.33$$

$$\sigma_{22} = 0.15654$$

$$\sigma_{33} = 0.17937$$

j) Intervalo de confianza para $\hat{\beta}_0$ con nivel 95%

$$1 - \alpha = 0.95, \quad v = n - k - 1 = 6 - 2 - 1 = 3 \Rightarrow t_{\alpha/2} = t_{0.025} = 3.182 \quad (\text{Tabla T})$$

$$\hat{\beta}_0 - t_{\alpha/2} \sigma_{11} \leq \beta_0 \leq \hat{\beta}_0 + t_{\alpha/2} \sigma_{11}$$

$$-134.071 - 3.188 \sqrt{1609.33} \leq \beta_0 \leq -134.071 + 3.188 \sqrt{1609.33}$$

$$-261.72 \leq \beta_0 \leq -6.4204$$

k) Pruebe con 5% de significancia que $\beta_1 > 1$

$$H_0: \beta_1 = 1$$

$$H_a: \beta_1 > 1$$

$$\alpha = 0.05$$

$$v = n - k - 1 = 3, \quad t_{\alpha} = t_{0.05} = 2.353 \quad (\text{Tabla T})$$

Región de rechazo de H_0 : $t > 2.353$

$$t = \frac{\hat{\beta}_1 - 1}{\sigma_{22}} = \frac{1.4888 - 1}{\sqrt{0.15654}} = 1.2354$$

No se puede rechazar H_0

I) Pruebe la normalidad del error con 5% de significancia mediante la Prueba de Kolmogorov-Smirnov

H_0 : $\epsilon \sim N(0, \sigma^2)$ (distribución normal con media 0 y varianza σ^2)

H_a : $\neg H_0$

$$\alpha = 0.05$$

$$\epsilon_i \cong e_i = y_i - \hat{y}_i, \quad i=1, 2, \dots, 6$$

$$\begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \end{bmatrix} = \begin{bmatrix} 6.0401 \\ 1.6866 \\ -2.1121 \\ -5.0878 \\ -4.0562 \\ 3.5294 \end{bmatrix}$$

$$\sigma^2 \cong S^2 = 32.861$$

$$Z_i = \frac{e_i - 0}{S} = \frac{e_i}{S} \quad (\text{valores estandarizados})$$

$F_0(Z_i)$ (distribución normal estándar acumulada)

Estadístico de prueba

$$D_n = \max | S_n(x_i) - F_0(x_i) | \quad (\text{para este ejemplo } x_i \text{ son los valores } e_i)$$

Región de rechazo de H_0

$$\alpha = 0.05, \quad n = 6 \Rightarrow D_{0.05} = 0.521 \quad (\text{Tabla K-S})$$

$$D_n > 0.521$$

Datos ordenados $x_{(i)}$

i	x_i (ordenados)	$S_n(x_i)$	$F_0(x_i)$	$ S_n(x_i) - F_0(x_i) $
---	-------------------	------------	------------	-------------------------

1	-5.0.878	1/6=0.1667	0.1874	0.0207
2	-4.0562	2/6=0.3333	0.2396	0.0937
3	-2.1121	3/6=0.5	0.3563	0.1437
4	1.6866	4/6=0.6667	0.6157	0.0510
5	3.5294	5/6=0.8333	0.7310	0.1023
6	6.0401	6/6=1	0.8540	0.1460

$$D_n = \max | S_n(x_i) - F_0(x_i) | = 0.1437$$

Conclusión: D_n no cae en la región de rechazo, por lo tanto no se puede rechazar H_0