**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  **p**(x)  1 para todo x



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

**F**(x) = P(**X**  x) = 

**Ejemplo**

 Experimento aleatorio: se lanza una moneda 3 veces

 **Ω** = {ccc, ccs, csc, css, scc, scs, ssc, sss }

 Sea **X** : # caras observadas

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| x | 0 | 1 | 2 | 3 |
| **p**(x) |  |  |  |  |

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

P(**X**  0) = P(X = 0) = 

P(**X**  1) = P(X = 0) + P(X = 1) =  +  = 

P(**X**  2) = P(X = 0) + P(X = 1) + P(X = 2) =  +  +  = 

P(**X** 3) = P(X= 0) + P(X= 1) + P(X= 2) + P(X= 3) =  +  +  +  = 1

 Se tiene entonces,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| x | 0 | 1 | 2 | 3 |
| **F**(x) |  |  |  | 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 , y se define en la forma siguiente:

 = 

Propiedades:

1. E(k)=k
2. E(kX)=kE(X)
3. E(X±Y)=E(X)±E(Y)
4. E(g(X))=∑g(x)p(x)

 e) Si X y Y son independientes entonces E(XY)=E(X)E(Y)=μXμY

 Para el ejemplo dado, = = 

 = 

 A veces, el interés es determinar la variabilidad de la variable aleatoria. Definimos entonces la varianza de la variable aleatoria **X**, denotada , ó σ2 mediante la siguiente ecuación:

V(X) = E[(X-E(X))2] y su forma reducida es:

 = 

donde,  = 

 Para el ejemplo dado,  = 

 = 

 Entonces,  = 

1. V(k)=0
2. V(kX)=k2V(X)
3. V(X±Y)=V(X)+V(Y) si X y Y son independientes
4. V(aX+bY)= a2V(X)+b2V(Y)+2abCov(XY)

donde Cov(XY) = E((X-μX)(Y-μY)) = E(XY)-μXμY

 La desviación estándar de la variable aleatoria **X** es la raíz cuadrada positiva de la varianza, es decir, σ = .

**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 **p**.

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

P(1)= **p** P(0)=**1-p**=**q**

Además se cumple: E(X)= p V(X)=pq

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

El experimento aleatorio se repite **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(**X** = **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 . Por otro lado, como las **n** repeticiones del experimento son independientes entre sí y calcular P(**X** = **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 ; en definitiva:

P(**X** = **x**) = 

Dado que  y , resulta que P(**X** = **x**) =  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

 = 

 Se puede demostrar que para una variable aleatoria con distribución binomial

 = **n**.**p**

 = **n**.**p**.**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:





Esto también se puede extender para más de dos grupos.

**Ejemplo:**

Si existe tres grupos el primero con *k1* elementos, el segundo grupo *k2* y el tercero con *k3* 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:



**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  se denomina distribución Poisson; donde λ es el número promedio de resultados por unidad de tiempo o región. Para una variable aleatoria con distribución Poisson se tiene  =  = λ.

# 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](http://es.wikipedia.org/wiki/Estad%C3%ADstica) la distribución uniforme es una [distribución de probabilidad](http://es.wikipedia.org/wiki/Distribuci%C3%B3n_de_probabilidad) cuyos valores tienen la misma [probabilidad](http://es.wikipedia.org/wiki/Probabilidad).

Se dice que una variable aleatoria *X* continua tiene una distribución uniforme en el intervalo [*a*,*b*] si la [función de densidad](http://es.wikipedia.org/wiki/Funci%C3%B3n_de_densidad) de probabilidad (FDP) es



La [función de distribución](http://es.wikipedia.org/wiki/Funci%C3%B3n_de_distribuci%C3%B3n) en el caso continuo entre *a* y *b* es



Su [media estadística](http://es.wikipedia.org/wiki/Media_estad%C3%ADstica) es (*a* + *b*) / 2 y su [varianza](http://es.wikipedia.org/wiki/Varianza) (*b* − *a*)2 / 12



### **Distribución exponencial**

En [estadística](http://es.wikipedia.org/wiki/Estad%C3%ADstica) la distribución exponencial es una [distribución de probabilidad](http://es.wikipedia.org/wiki/Distribuci%C3%B3n_de_probabilidad) continua con un parámetro λ > 0 cuya [función de densidad](http://es.wikipedia.org/wiki/Funci%C3%B3n_de_densidad) es



Su [función de distribución](http://es.wikipedia.org/wiki/Funci%C3%B3n_de_distribuci%C3%B3n) es



Aquí e significa el [número e](http://es.wikipedia.org/wiki/N%C3%BAmero_e).

El [valor esperado](http://es.wikipedia.org/wiki/Valor_esperado) y la [varianza](http://es.wikipedia.org/wiki/Varianza) de una [variable aleatoria](http://es.wikipedia.org/wiki/Variable_aleatoria) X con distribución exponencial son

![E[X]=\frac{1}{\lambda}]()





### **Distribución Normal**

La distribución normal, también llamada distribución de Gauss o distribución gaussiana, es la [distribución de probabilidad](http://es.wikipedia.org/wiki/Distribuci%C3%B3n_de_probabilidad) que con más frecuencia aparece en [estadística](http://es.wikipedia.org/wiki/Estad%C3%ADstica) y [teoría de probabilidades](http://es.wikipedia.org/wiki/Teor%C3%ADa_de_probabilidades). Esto se debe a dos razones fundamentalmente:

* Su [función de densidad](http://es.wikipedia.org/wiki/Funci%C3%B3n_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](http://es.wikipedia.org/wiki/Matem%C3%A1ticas).

La [función de densidad](http://es.wikipedia.org/wiki/Funci%C3%B3n_de_densidad) está dada por:



Donde μ ([Μ](http://es.wikipedia.org/wiki/%CE%9C)) es la [media](http://es.wikipedia.org/wiki/Media_aritm%C3%A9tica) y σ ([sigma](http://es.wikipedia.org/wiki/Sigma)) es la [desviación estándar](http://es.wikipedia.org/wiki/Desviaci%C3%B3n_est%C3%A1ndar) (σ2 es la [varianza](http://es.wikipedia.org/wiki/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  es la media de una muestra aleatoria de tamaño n extraída de una población que tiene media μ y varianza σ2 , entonces

****

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 (-μ)/(σ/). 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  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 ≥ 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: σ2 ≈ S2

**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 μ = 500, varianza σ2 desconocida.

Por el Teorema del Límite Central,  tiene distribución aproximadamente normal si n≥30, por lo tanto, (-μ)/(σ/) tiene distribución aproximadamente normal estándar. Adicionalmente, se puede usar la aproximación: σ ≈ S = 36



**P(480≤≤490) = P() = P()**

**= P(-3.5136≤Z≤** **-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):



Siendo la función densidad de probabilidad:



La tasa de fallos para esta distribución es:



Las ecuaciones (1), (2) y (3) sólo se aplican para valores de (t - t0) ≥ 0. Para valores de (t - t0) < 0, las funciones de densidad y la tasa de fallos valen 0. Las constantes que aparecen en las expresiones anteriores tienen una interpretación física:

* **t0** 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 - t0) = η la fiabilidad viene dada por:
R (t) = exp - (1)ß = 1/exp 1ß = 1 / 2,718 = 0,368 (36,8%)
Entonces la constante representa también el tiempo, medido a partir de t0 = 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 ß ya que como hemos visto su valor no influye en los cálculos realizados. Por esta razón también se le llama usualmente vida característica.
* **ß** 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:

1. Representa la evolución con el tiempo de la tasa de fallos, λ(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).
2. Permite fijar tiempos de reparación de componentes, siendo también en este caso el tiempo la variable independiente de la distribución.
3. 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 |
|   |  |   |  | 0 | de otra 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\*u+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 **σ2**. Se ha establecido anteriormente que la media muestral **** también tendrá **distribución normal** con media **** y varianza ****. Por lo tanto, la variable **** tendrá distribución normal estándar.

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

 ****

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: **ν = 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 ν = 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:

**tα** : valor de **t** tal que **P(T≥tα) = α,** como se 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 n1 variables aleatorias distribuidas normalmente con una varianza común  y que la otra muestra aleatoria contiene  variables aleatorias distribuidas normalmente con una varianza común  y que la otra muestra aleatoria contiene  variables aleatorias distribuidas normalmente con una varianza común . Si calculamos  de las observaciones en la muestra 1, entonces  es una estimación de . De manera similar,  calculada a partir de las observaciones de la segunda muestra es una estimación para . Así intuitivamente podríamos pensar en utilizar /  para hacer inferencias con respecto a las magnitudes relativas de  y . Si dividimos cada  por , entonces la razón siguiente



tiene una distribución con  grados de libertad. La definición general de una distribución es como sigue:

DEFINICION Sean  y  variables aleatorias ji - cuadrada con  y  grados de libertad. Respectivamente. Entonces si  y  son independientes,



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

**DISTRIBUCION DE PROBABILIDAD **

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



tienen distribuciones  independientes con



grados de libertad, respectivamente.

Así la definición implica que



tiene una distribución  con  grados de libertad del numerador y  grados de libertad del denominador.

En al figura 7.3 se muestra la gráfica de una típica función de densidad . Los valoras de  tales que  se dan en la tabla Xi Cuadrado, para los valores de 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  0.100, 0.050, 0.025, 0.010 y 0.005. Por ejemplo, si la variable  estudiada tiene 5 grados de libertad del numerador y 7 grados de libertad del denominador, 0.100= 2.88, 0.050= 3.97, 0.025  = 5.29, 0.010 = 7.46 y 0.005 =9.52. luego la probabilidad de que una variable aleatoria con una distribución  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:





En donde:



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:



donde



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.





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 <= y <= 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:





En cualquier otro punto donde



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



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



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 **y1, y2, ..., yn** de una variable **Y** denominada **variable de respuesta** las cuales corresponden a un conjunto **x1, x2, ..., xn** 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 **yi** está asociado con un valor **xi**.

Es importante reconocer que cada valor **yi** es el resultado de una medición, por lo tanto, es posible que pudiesen haber otros valores **yi** para el mismo valor dado **xi.**  Esto nos permite reconocer que **yi** proviene de una variable aleatoria **Yi** 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 **(xi, yi), i = 1, 2, ..., 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 = β0 + β1 x + ε** .

siendo **ε** el componente aleatorio de **Y**

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

 **εi ~ N(0, σ2)** (distribución normal con media **0** y varianza **σ2**)

Por lo tanto, el valor esperado de este modelo, es una recta teórica (desconocida) con los parámetros **β0** y **β1** que deben estimarse

 **E[Y] = β0 + β1 x**  .



**RECTA DE MÍNIMOS CUADRADOS**

Es un procedimiento matemático para estimar los parámetros **β0** y **β1** 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**

  .

 Es la recta de mínimos cuadrados.  son los estimadores de **β0** y **β1**



Para cada valor **** se tiene el valor observado ****  y un valor  obtenido con la recta de mínimos cuadrados: 

Sea **ei =  -** ,

Entonces, el criterio de mínimos cuadrados consiste en minimizar **** 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 = ** .

(Lea SCE: “Suma de Cuadrados del Error”)

El procedimiento matemático para realizar esta optimización es:



Con facilidad se llega al sistema de ecuaciones lineales:



De donde se obtienen finalmente los estimadores 

**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** | **xi** | **yi** | **x2i** | **xiyi** |
| 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 |
|  | **466** | **748** | **23934** | **36712** |

Sustituimos en el sistema de ecuaciones lineales:

 

De donde se obtienen: 

Ecuación de mínimos cuadrados: ** = 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:

 ** = 35.83 + 0.836 (50) = 77.63**

**REGRESIÓN LINEAL MÚLTIPLE**

Considere que una variable **Y** depende de **k** variables **x1, x2, ... , xk**

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

Modelo teórico probabilista propuesto:

 **Y = β0 + β1 x1 + β2 x2 + ... + βk xk + ∈** .

Siendo **∈** 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 **(x1,i, x2,i, ..., xk,i, yi), i = 1, 2, ..., n**

Fijados los **k** valores **x1,i, x2,i, ..., xk,i** se tiene una observación o medición **yi** la cual es uno de los posibles valores de la variable aleatoria **Yi**

Se supondrá que para cada variable aleatoria **Yi** el componente aleatorio **∈i** tiene la misma distribución de probabilidad, y que además son independientes.

 **∈i ~ N(0, σ2)** (distribución normal con media **0** y varianza **σ2**)

**MODELO DE REGRESIÓN LINEAL MÚLTIPLE DE MÍNIMOS CUADRADOS**

Para estimar los **k + 1** parámetros **β0 , β1 , β2 , ..., βk** se usará un procedimiento similar al modelo de regresión lineal simple con el método de mínimos cuadrados

  .

Sea **ei =  -** ,

en donde**:** valor observado en la muestra

 **:** valor obtenido con el modelo de mínimos cuadrados:

**Criterio de mínimos cuadrados**

 **Minimizar**

 **SCE = **.

Utilizando , **i=0, 1, 2, ..., k**, se obtienen las ecuaciones normales para encontrar los estimadores  ..., 

**Consideremos el caso específico k=2**

**Y** depende de 2 variables **x1, x2**

Modelo teórico probabilista propuesto:

 **Y = β0 + β1 x1 + β2 x2 + ∈** .

Modelo de regresión lineal múltiple de mínimos cuadrados;

 

Para encontrar  

, **i=0, 1, 2**

Se obtienen las ecuaciones normales

 

Las expresamos en notación matricial

 **A  = C ⇒ **

**REGRESIÓN LINEAL MÚLTIPLE EN NOTACIÓN MATRICIAL**

El modelo teórico probabilista puede expresarse convenientemente en notación matricial

 **Y = β0 + β1 x1 + β2 x2 + ... + βk xk + ∈**

Consideramos el caso específico **k=2** en donde **Y** depende de 2 variables **x1, x2**

Modelo teórico probabilista propuesto:

**Y = β0 + β1 x1 + β2 x2 + ∈**

Datos de la muestra:

 **(x1,i, x2,i, yi), i = 1, 2, ..., n**

Cada observación **yi** es un valor de la variable aleatoria **Yi, i = 1, 2, ..., n**

**Yi = β0 + β1 x1,i + β2 x2,i + ∈i ,i= 1, 2, ..., n**

Se puede expresar en forma desarrollada,

 **Y1 = β0 + β1 x1,1 + β2 x2,1 + ∈1**

 **Y2 = β0 + β1 x1,2 + β2 x2,2 + ∈2**

 ...

 ...

 **Yn = β0 + β1 x1,n + β2 x2,n + ∈n**

El modelo teórico expresado en notación matricial

 **Y = X β + ∈ ⇒** 

 **X** se denomina **matriz de diseño**

 **Y =** , **X =**  , **β =**, **∈ =** 

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

  ⇒ **A  = C**

 **A =  = **

 **A = XT X**

 **C =  = **

 **C = XT y**

Sea  = 

Entonces **A  = C ⇒ XT X  = XT y** ⇒ ** = (XT X)-1 (XT y)**

**:** vector con los estimadores de mínimos cuadrados

**X:** matriz de diseño (datos de la muestra)

**y:** vector de observaciones obtenidas en la muestra

**y =**

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)  

(1)Sxx = = 

(2) SCT = Syy = = 

(3) Sxy = = 

(4) SCE = 

(5) SCR = 

**Demostración de** (1)

Sxx = = =

 = == 

 = = 

ESTIMACIÓN DE LA VARIANZA

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

 ∈i ~ N(0, σ2)

**La varianza** σ2 **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** σ2**)**

 ei =  - 

 SCE = ****=

S2 = = **** **.**

**Usamos la fórmula** (4) **definida anteriormente:**

 SCE = 

**De cual se obtiene**

 ****

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

 ****

**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 el 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 = β0 + β1 x1 + β2 x2 + ... + βk xk + ∈

 Se define la varianza muestral

S2 =  **.**

**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

 R2 = , 0 ≤ R2≤ 1 **.**

**El valor de** R2 **muestra la proporción de la variabilidad de de los datos que es explicada por el modelo de mínimos cuadrados.**

Ejemplo

**Suponer que con los datos se obtuvo** R2 = 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 = β0 + β1 x1 + β2 x2 + ... + βk xk + ∈**

 **∈ ~ N(0, σ2)** .

Para estimar los **k + 1** parámetros **β0 , β1 , β2 , ..., βk** se propone el modelo de regresión lineal de mínimos cuadrados:

  .

Suposiciones

  es un estimador insesgado del parámetro **βi**

 **E[] = βi**

 El estimador  tiene distribución normal

 ** ~ N(βi, )**

Notación

 **V[] == ** (varianza de )

 **Cov[**,**] = = ** (covarianza de ,)

**Definición**

 **Matriz de varianzas-covarianzas**

 **[] = ** .

La teoría estadística demuestra la siguiente relación

 **[] = [XT X]-1 σ2 ≅ [XT X]-1 S2**

 **X** es la matriz de diseño del modelo de mínimos cuadrados

 **S2 = **  es la varianza del modelo de mínimos cuadrados .

**INTERVALO DE CONFIANZA PARA LOS PARÁMETRO βI**

 Parámetro: **βi , i = 0, 1, ..., k**

 Estimador: **, i = 0, 1, ..., k**

 El estadístico

 **t = ,** tiene distribución **t**  con **ν = n-k-1** grados de libertad

 **Definición**

 **Intervalo de confianza para βi con nivel 1 - α**

 **- tα/2  ≤ βi ≤ + tα/2 **, **i=0, 1, ..., k** .

**PRUEBA DE HIPÓTESIS PARA LOS PARÁMETROS βI**

 1) Ho: **βi = b0** (algún valor especificado para el parámetro)

 2) Ha: **βi < b0** ó **βi > b0** ó **βi ≠ b0**

 3) α nivel de significancia de la prueba

 4) Estadístico de prueba

 **t = ,** tiene distribución **t**  con **ν = n-k-1** grados de libertad

 Ha Región de rechazo de Ho en favor de Ha

 **βi < b0 t < -tα**

**βi > b0 t > tα**

 **βi ≠< b0 t<-tα/2 ∨ t > tα/2**

1. Calcule el valor del estadístico
2. 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 **x1**  y su porcentaje de asistencia a clases **x2** considerando los siguientes datos:

|  |  |  |
| --- | --- | --- |
| Calificaciónparcial**x1** | Asistencia a clases**x2** | Calificaciónfinal**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 = β0 + β1 x1 + β2 x2 + ∈** .

**b) Notación matricial**



**X =**  (matriz de diseño)

**c) Estimación de parámetros por mínimos cuadrados**



** =  = (XT X)-1 (XT y)**

 **=**

 **=**

 **=**

 **=**



**d) Pronostique la calificación final si la calificación parcial es 75 y el**

**porcentaje de asistencia clases es 80**

  **= 93.08**

**e) Análisis de varianza**

**(80 + 77 + 94 + 70 + 51 + 70) = 73.667**

 **SCT =  = 1005.33**

**SCR =  = 906.7503**

 **SCE =  = 98.583**

**f) Coeficiente de determinación**

 **R2 == 0.9019 = 90.19%**

**g) Estimación de la varianza**

 **S2 = = 32.861**

**h) Matriz de varianza-covarianza**

 **≅ **

 **=  (32.861)**

 **= **

**i) Varianza de los estimadores de mínimos cuadrados**

 **V[] == , i = 0, 1, 2**

 **σ11 = 1609.33**

 **σ22 = 0.15654**

 **σ33 = 0.17937**

**j) Intervalo de confianza para  con nivel 95%**

 **1 - α = 0.95, ν = n–k–1 = 6-2-1 = 3 ⇒ tα/2 = t0.025 = 3.182** (Tabla T)

 ** - tα/2 σ11 ≤ β0 ≤ + tα/2 σ11**

 **-134.071 – 3.188≤ β0 ≤ -134.071 + 3.188**

 **-261.72 ≤ β0 ≤ -6.4204**

**k) Pruebe con 5% de significancia que β1 > 1**

 **Ho: β1 = 1**

 **Ha: β1 > 1**

 **α = 0.05**

 **ν = n-k-1 = 3, tα = t0.05 = 2.353** (Tabla T)

 **Región de rechazo de Ho : t > 2.353**

 **t =  =  = 1.2354**

 **No se puede rechazar Ho**

**l) Pruebe la normalidad del error con 5% de significancia mediante la**

 **Prueba de Kolmogorov-Smirnov**

 **Ho: ∈ ~ N(0, σ2)** (distribución normal con media **0** y varianza **σ2**)

 **Ha: ⎤ Ho**

 **α = 0.05**

 **∈i ≅ ei = yi - , i=1, 2, .., 6**

** = **

 **σ2 ≅ S2 = 32.861**

 **Zi =  = ** (valores estandarizados)

 **F0(Zi)** (distribución normal estándar acumulada)

 **Estadístico de prueba**

 **Dn = max| Sn(xi) – F0(xi)|** (para este ejemplo **xi** son los valores **ei**)

 **Región de rechazo de Ho**

 **α = 0.05, n = 6 ⇒ D0.05 = 0.521** (Tabla K-S)

 **Dn > 0.521**

 **Datos ordenados x(i)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **i** | **xi (ordenados)**  | **Sn(xi)** | **F0(xi)** | **|Sn(xi)- F0(xi)|** |
| **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** |

 **Dn = max| Sn(xi) – F0(xi)| = 0.1437**

**Conclusión: Dn no cae en la región de rechazo, por lo tanto no se puede**

 **rechazar Ho**