INICIO: ABRIL 2018

CURSO VIRTUAL

PROGRAMACIÓN ESTADÍSTICA CON R

UNIVERSIDAD PERUANA
CAYETANO HEREDIA
Escuela de Posgrado

Programación Estadística con R

UPCH

Abril 2018

UNIVERSIDAD PERUANA
CAYETANO HEREDIA
Escuela de Posgrado

Programacion en R

1. Funciones de Distribución en R

Distribucion Normal

Para obtener valores que se basen en la distribución Normal, R, dispone de cuatro funciones:

dnorm	Devuelve resultados de la función de densidad.
pnorm	Devuelve resultados de la función de distribución acumulada.
qnorm	Devuelve resultados de los cuantiles de la Normal.
rnorm	Devuelve un vector de valores de la Normal aleatorios.

Distribución Normal

uso

```
dnorm(x, mean = 0, sd = 1, log = F)
pnorm(q, mean = 0, sd = 1, lower.tail = T, log.p = F)
qnorm(p, mean = 0, sd = 1, lower.tail = T, log.p = F)
rnorm(n, mean = 0, sd = 1)
```

- `x`, `q`: vector de cuantiles.
- `p`: vector de probabilidades.
- `n`: número de observaciones.

Distribución Normal

uso

```
dnorm(x, mean = 0, sd = 1, log = F)
pnorm(q, mean = 0, sd = 1, lower.tail = T, log.p = F)
qnorm(p, mean = 0, sd = 1, lower.tail = T, log.p = F)
rnorm(n, mean = 0, sd = 1)
```

- `mean`: Vector de medias. Por defecto, su valor es 0.
- `sd`: Vector de desviación estándar. Por defecto, su valor es 1.
- `log`, `log.p`: Parámetro booleano, si es TRUE, las probabilidades p son devueltas como log (p).
- `lower.tail`: Parámetro booleano, si es TRUE (por defecto), las probabilidades son $P(X \leq x)$, de lo contrario, $P(X > x)$.

Distribución Normal

Ejemplo 1

Para comprobar el funcionamiento de estas funciones, usaremos un ejemplo de aplicación.
Imaginemos el siguiente problema: Sea Z una variable aleatoria normal con una media de 0 y una desviación estándar igual a 1.Determinar :

- $P(Z > 2)$

```
# Para resolver este apartado, necesitamos resolver:
# P( Z > 2), por lo tanto, usamos la funci n
# acumulada de distribuci n indicando que la
# probabilidad de cola es hacia la derecha:
pnorm(2, mean = 0, sd = 1, lower.tail = F)
```

UNIVERSIDAD PERUANA CAYETANO HEREDIA
Escuela de Posgrado

Distribución Normal

Ejemplo 2

Para comprobar el funcionamiento de estas funciones, usaremos un ejemplo de aplicación.

Imaginemos el siguiente problema: Sea Z una variable aleatoria normal con una media de 0 y una desviación estándar igual a 1.Determinar :

- P(-2 $\leq$ Z $\leq$ 2).

```
# Necesitamos resolver: P(-2 <= z <= 2), volvemos a
# emplear la funci n de densidad acumulada, esta
# vez, con la probabilidad de cola por defecto,
# hacia la izquierda:
pnorm(c(2), mean = 0, sd = 1) - pnorm(c(-2), mean = 0, sd
    = 1)
```

UNIVERSIDAD PERUANA
CAYETANO HEREDIA
Escuela de Posgrado

Distribución Normal

Ejemplo 3

Para comprobar el funcionamiento de estas funciones, usaremos un ejemplo de aplicación.
Imaginemos el siguiente problema: Sea Z una variable aleatoria normal con una media de 0 y una desviación estándar igual a 1.Determinar :

- $P(0 \leq Z \leq 1.73)$.

```
# Necesitamos resolver: P(0 <= z <= 1.73), este
# ejercicio se resuelve con el mismo procedimiento
# que el apartado anterior, por lo tanto, volvemos
# a emplear la funci n # de densidad acumulada:
pnorm(c(1.73), mean = 0, sd = 1) - pnorm(c(0), mean = 0,
    sd = 1)
```

UNIVERSIDAD PERUANA CAYETANO HEREDIA
Escuela de Posgrado

Distribución Normal

Ejemplo 4

Para comprobar el funcionamiento de estas funciones, usaremos un ejemplo de aplicación.

Imaginemos el siguiente problema: Sea Z una variable aleatoria normal con una media de 0 y una desviación estándar igual a 1.Determinar :

- $P(Z \leq a) = 0.5793.$

```
# En este apartado, debemos obtener el valor de
# a para que se cumpla la probabilidad, es decir:
# P(Z <= a) = 0.5793. Para ello, debemos usar
# la funci n de quantiles:
qnorm(0.5793, mean = 0, sd = 1)
```

Distribución Normal

Ejemplo 5

Para comprobar el funcionamiento de estas funciones, usaremos un ejemplo de aplicación.
Imaginemos el siguiente problema: Sea Z una variable aleatoria normal con una media de 0 y una desviación estándar igual a 1.Determinar :

- $P(Z > 200)$. Siendo la media 100 y la desviación estándar 50.

```r
# La curiosidad de este apartado es que no
# tenemos una normal est ndar , pero no hay
# problema , simplemente , debemos especificar
# los valores de la media y desviaci n est ndar
# en los argumentos de la funci n de distribuci n
# acumulada para que la tipificaci n la realice
# autom ticamente la funci n de R.
# Otra cosa importante a tener en cuenta , es que
# debemos indicar que la probabilidad de cola es
# hacia la derecha.
pnorm(c(200), mean = 100, sd = 50, lower.tail = F)
```

Distribución Normal

Importancia de la distribución normal

La distribución normal es de suma importancia en estadística por tres razones principales:

1. Numerosas variables continuas de fenómenos aleatorios tienden a comportarse probabilisticamente mediante ésta.
2. Es el límite al que convergen tanto variables aleatorias continuas como discretas.
3. Proporciona la base de la inferencia estadística clásica debido a su relación con el teorema del límite central.

Propiedades de la distribución normal

1. Su grafica tiene forma acampanada.
2. El valor esperado, la mediana y la moda tienen el mismo valor cuando la variable aleatoria se distribuye normalmente.
3. Su dispersión media es igual a 1.33 desviaciónes estándar. Es decir, el alcance intercuartil está contenido dentro de un intervalo de dos tercios de una desviación estándar por debajo de la media a dos tercios de una desviación estándar por encima de la media.

Distribución Normal

Propiedades de la distribución normal

En la práctica, algunas de las variables que observamos sólo pueden aproximar estas propiedades. Así que si el fenómeno puede mediarse aproximadamente mediante la distribución normal se tendrá:

1. Que el polígono puede verse en forma de campana y simétrico.
2. Sus mediciones de tendencia central tienen bastante parecido.
3. El valor intercuartil puede diferir ligeramente de 1.33 desviaciones estándar.
4. El dominio de la variable aleatoria normalmente distribuida generalmente caerá dentro de 3 desviaciones estándar por encima y por debajo de la media.

UNIVERSIDAD PERUANA CAYETANO HEREDIA
ESCUELA DE POSGRADO

Distribución Normal

El modelo matemático

$$NOTACIÓN: \ N(\mu, \sigma^2), \quad X \sim N(\mu, \sigma^2)$$

El modelo o expresión matemática que representa una función de densidad de probabilidad se denota mediante el símbolo $f(X)$. Para la distribución normal, se tiene la siguiente función de probabilidad.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2}$$

donde

e es la constante matemática aproximada por 2.71828

π es la constante matemática aproximada por 3.14159

UNIVERSIDAD PERUANA CAYETANO HEREDIA
Escuela de Posgrado

Distribución Normal

El modelo matemático

$$\text{Parámetros}\begin{cases} \mu_X \text{ es el valor esperado de la variable aleatoria} \\ \sigma_X \text{ es la desviación estándar de la variable aleatoria} \end{cases}$$

X es cualquier valor de la variable aleatoria continua, donde $-\infty < x < +\infty$

Así,

$$E(X) = \mu_X$$

$$Var(X) = \sigma_X^2$$

Distribución Normal

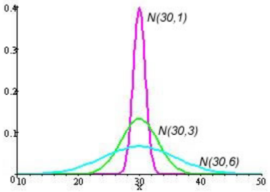
caso 1

A continuación se presentan las gráficas de las funciones de densidad Normal con el objetivo de observar cambios en la distribución de probabilidad:

Caso 1:

Cuando se mantiene la misma media, pero cambia la varianza.

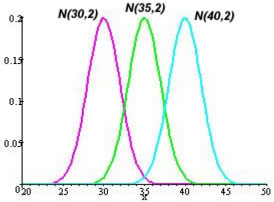
Ejemplo: $N(30, 1)$, $N(30, 3)$, $N(30, 6)$



Distribución Normal

caso 2

Cuando se mantiene la misma varianza, pero cambia la media.
Ejemplo: ($N(30,2)$, $N(35,2)$ y $N(40,2)$)



Propiedades de la Normal

Ahora, al examinar la primera y segunda derivada de $f(x)$, se pueden listar otras propiedades de la curva normal:

1. La moda, que es el punto sobre el eje horizontal donde la curva es un máximo ocurre cuando $x = \mu_X$.
2. La curva es simétrica alrededor de un eje vertical a través del valor esperado μ_X.
3. La curva tiene sus puntos de inflexión en $x = \mu_X \pm \sigma_X$, es cóncava hacia abajo si $\mu_X - \sigma_X < x < \mu_X + \sigma_X$, y es cóncava hacia arriba en cualquier otro punto.
4. La curva normal se aproxima al eje horizontal de manera asintótica conforme nos alejamos de la media en cualquier dirección.

Distribución Normal

Propiedades de la Normal

Haciendo una transformación a la variable aleatoria normal X, ésta se puede llevar a un nuevo conjunto de observaciones de una variable aleatoria normal Z con media cero y varianza 1. A dicha transformación se le conoce como estadarización de la variable aleatoria normal X:

$$Z = \frac{X - \mu_X}{\sigma_X}$$

Distribución Normal

Es bastante util aprender a hacer inferencias acerca de la población con base en información de la muestra. Varias de estas técnicas se basan en el supuesto de que la población presenta una distribución normal aproximada. Por tanto, será importante determinar si los datos de la muestra provienen de una población normal, antes de aplicar dichas técnicas.

Distribución Normal

Para determinar si los datos provienen de una distribucón aproximadamente normal, se pueden considerar tres métodos:

1. Construir en histograma de frecuencia relativa o bien un diagrama de tallos y hojas para los datos. Si los datos son aproximadamente normales, la forma de la gráfica será similar a la de la curva normal. (Con forma de joroba y simétrica alrededor de la media.)
2. Calcular el rango intercuartílico (IQR) y la desviación estándar (s), para la muestra, y luego calcular el cociente IQR/s. Si los datos son aproximadamente normales, $IQR/s \approx 1.3$.
3. Construir una gráfica de probabilidad normal para los datos Si los datos son aproximadamente normales, los puntos caerán (aproximadamente) en una línea recta.

UNIVERSIDAD PERUANA CAYETANO HEREDIA
Escuela de Posgrado

Distribución Normal

Construcción De Una Gráfica de Probabilidad Normal Para Un Conjunto De Datos

1. Haga una lista de las observaciones del conjunto de datos de muestra en orden ascendente, donde x_i representa el i-ésimo valor ordenado.
2. Para cada observación, calcule el área de cola correspondiente de la dsitribución normal estándar (z), $A_i = P\left[X \leq X_i\right]$. Empiricamente condición de continuidad.
   $$A_i = \frac{i - 0.375}{n + 0.25}$$
   donde n es el tamaño de la muestra.

Distribución Normal

Construcción De Una Gráfica de Probabilidad Normal Para Un Conjunto De Datos

1. Calcule el valor esperado estimado de x_i suponiendo normalidad, mediante la siguiente fórmula:
   $E(X_i - \overline{X}) = s \times [Z(A_i)]$
   donde s es la desviación estándar de la muestra y $Z(A_i)$ es el valor de z que recorta un área A_i de la cola inferior de la distribución normal estándar.
2. Grafique las observaciones ordenadas x_i en el eje vertical y los valores esperados estimados correspondientes, $E(x_i)$, en el eje horizontal.

Distribución Normal

Nota

Las verificaciones de normalidad dadas son sólo técnicas descriptivas. Es posible (aunque poco probable) que los datos no sean normales a pesar de que las verificaciones se satisfacen razonablemente. Por tanto, se debe tener cuidado de no asegurar que las mediciones, de hecho, se distribuyen normalmente. Sólo podemos decir que es razonable pensar que los datos provienen de una distribución normal.

Distribución Normal

Aproximación para la distribución binomial

La distribución normal frecuentemente es una buena aproximación a una distribución discreta cuando la última adquiere una forma de campana simétrica. Desde un punto de vista teórico algunas distribuciones convergen a la normal conforme sus parámetros se acercan a ciertos límites. La distribución normal es una aproximación conveniente pues la distribución acumulada se tabula más fácil. La distribución binomial se aproxima bien por la normal en problemas prácticos cuando se trabaja con la función de distribución acumulada.

Distribución Normal

Teorema. (Aplicación del Teorema del Límite Central)

Si X es una variable aleatoria binomial con media $\mu = np$ y varianza $\sigma^2 = np(1-p)$, entonces la forma limitante de la distribución de

$$Z = \frac{X - np}{\sqrt[2]{np(1-p)}}$$

cuando $n \to \infty$, es la distribución normal estándar $N(z; 0, 1)$

La distribución normal proporciona una buena aproximación de la binomial aún cuando n es pequeña y p está razonablemente cercana a 0.5.

UNIVERSIDAD PERUANA CAYETANO HEREDIA
ESCUELA DE POSGRADO

Distribución Normal

Ejemplo

Genera 100 valores aleatorios de una distribución normal de media 3 y desviación típica 2 se utiliza la semilla 111.

```
set.seed(111)
datos<-rnorm(100,3,2)
# Representamos a continuaci n el histograma.
# Si usamos el c digo
hist(datos)
# Sin embargo, si usamos el c digo
hist(datos, freq=FALSE)
# Dicha comparativa se hace ejecutando a continuaci n
# el siguiente c digo
curve(dnorm(x,3,2),add=TRUE)
# add=TRUE superpone la curva al histograma.
#aparecen representadas frecuencias relativas
# y es posible hacer una comparaci n con la
# funci n de densidad te rica.
```

Distribución Normal

Ejemplo

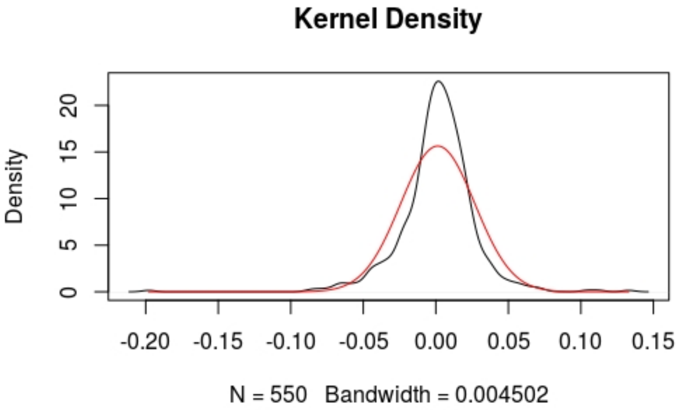
```
library(quantmod)
SPY = getSymbols("SPY",auto.assign=F)
SPY = weeklyReturn(Ad(SPY))

densitySPY = density(SPY)
plot(densitySPY,main="Kernel Density")

x = seq(min(SPY),max(SPY),length=300)
y = dnorm(x,mean=mean(SPY),sd = sd(SPY))
lines(x,y,col=2)
```

UNIVERSIDAD PERUANA CAYETANO HEREDIA
Escuela de Posgrado

Distribución Normal



Distribución Normal

```r
rm(list=ls())
library(quantmod)
getSymbols("^NDX",src="yahoo", from='1997-6-01', to='2012-6-01')
daily<- allReturns(NDX)[,c('daily')]
dailySerieTemporel<-ts(data=daily)
ss<-na.omit(dailySerieTemporel)
plot(density(ss, kernel='epanechnikov'))
set.seed(125)
lines(density(rnorm(length(ss), mean(ss), sd(ss)), kernel='epanechnikov'), col=2)
```