Curso de R y estadística básica

[Felipe de J. Muñoz González]

fmunoz@lcg.unam.mx

Introducción
Descargar Presentación

Distribuciones de probabilidad

Distribuciones de probabilidad

Cuando una variable aleatoria toma diversos valores, la probabilidad asociada a cada uno de tales valores puede ser organizada como una **distribucion de probabilidad**.

Definiciones

Las distribuciones de probabilidad pueden representarse a traves de una una formula que se le denomina **funcion de probabilidad**. Existen dos tipos de distribuciones:

- Distribuciones discretas
- Distribuciones continuas

Distribuciones de probabilidad

Definiciones

Función	Distribución	Tipo	media	varianza
bern	Binomial	Discreta	$\mu = np$	σ^2=npq
geom	geometrica	Discreta	$\mu = 1/p$	$\sigma^2=(1-p)/p^2$
hyper	hipergeométrica	Discreta	$\mu = nk/N$	$\sigma^2=(N-n)/(n-1)\ n\ k/N(1 - k/N)$
nbinom	binomial negativa	Discreta	$\mu = [(r-1)(1-p)/p]$ si $r>1$, 0 si $r<=1$	$\sigma^2=$
pois	Poisson	Discreta	$\mu = \lambda$	$\sigma^2 = \lambda$
beta	beta	Continua	$\mu = \alpha/(\alpha + \beta)$	$\sigma^2=\alpha\beta/(\alpha + \beta)^2 /(\alpha + \beta + 1)$

Distribuciones de probabilidad

Definiciones

Función	Distribución	Tipo	media	varianza
cauchy	Cauchy–Lorentz	Continua	$\mu = x_0$	Not Defined
exp	exponencial	Continua	$\mu = 1/\lambda$	$\sigma^2 = 1/\lambda^2$
chisq	chi-cuadrado	Continua	$\mu = k$	$\sigma^2=2k$
fisher	F	Continua	$\mu = d_2/(d_2 - 2)$	$\sigma^2=$
gamma	gamma	Continua	$\mu = p/a$	$\sigma^2 = p/a^2$

Función	Distribución	Tipo	media	varianza
logis	logística	Continua	$\mu = \mu$	$\sigma^2 = \pi^2 * s^2 /3$
norm	normal	Continua	$\mu = \mu$	$\sigma^2 = \sigma^2$
t	t-Student	Continua	$\mu = 0$	$\sigma^2 = v/(v-2)$
unif	uniforme	Continua	$\mu = (a+b)/2$	$\sigma^2=(b-a)^2/12$

Distribuciones de probabilidad

Definiciones

A cada nombre de función dado por R se le agrega un prefijo ‘d’ para obtener la función de densidad, ‘p’ para la función de distribución acumulada, ‘q’ para la función cuantil o percentil y ‘r’ para generar variables pseudo-aleatorias (random). La sintaxis es la siguiente:

dxxx(x, …) pxxx(q, …) qxxx(p, …) rxxx(n, …)

Donde xxx indica el nombre de cualquiera de las distribuciones, x y q son vectores que toman valores en el soporte de la distribución, p es un vector de probabilidades y n es un valor entero.

Distribuciones de probabilidad

Distribución Binomial

La distribución binomial es una distribucón de probabilidad que toma valor 1 para la probabilidad de éxito (p) y valor q para la probabilidad de fracaso (q = 1 – p).

Se dice que la variable aleatoria X se distribuye como una Binomial de parametro p si sigue un proceso de bernulli.

- Los experimentos constan de ensayos repetidos
- Cada ensayo se clasifica como éxito o fracaso
- La probabilidad de éxito p es constante
- Los ensayos

$X \sim Be(p)$

$f(x) = p^{x} (1 - p)^{1-x}$; con $x = \{0, 1\}$

Distribuciones de probabilidad

Distribución Binomial

- dbinom(x, size, prob, log = F); Devuelve resultados de la función de densidad.
- pbinom(q, size, prob, lower.tail = T, log.p = F); Devuelve resultados de la función de distribución acumulada.
- qbinom(p, size, prob, lower.tail = T, log.p = F); Devuelve resultados de los cuantiles de la binomial.
- rbinom(n, size, prob); Devuelve un vector de valores binomiales aleatorios.

Con:

- x, q: Vector de cuantiles.
- p: Vector de probabilidades.
- n: Número de observaciones
- size: Números de ensayos(debe ser cero o máss).
- prob: Probabilidad de éxito en cada ensayo.
- log, log.p: Parámetro booleano, si es TRUE, las probabilidades p se ofrecen como log(p).
- lower.tail: Parámetro booleano, si es TRUE (por defecto), las probabilidades son P[X ≤ x], de lo contrario, P[X > x].

Distribuciones de probabilidad

Distribución Binomial

```
> dbinom(3, 10, 0.5) # Calculamos la P(X=3) de una Binomial(10
> dbinom(6, 10, 0.5) # Calculamos la P(X=6) de una Binomial(10
> dbinom(c(2, 1, 0), 10, 0.5) # Calculamos la P(X=2,1,0) de un
> sum(dbinom(c(2, 1, 0), 10, 0.5)) # Calculamos la P(X<=2) de
> sum(dbinom(c(0, 1, 2), 10, 0.5))
> # Que coincide con la función pbinom
> pbinom(2,10,0.5)
> pbinom(2,10,0.5, lower.tail=F) # Calculamos la P(X>2) de una
> pbinom(10,10,0.5)
> qbinom(c(0.90), 10, 0.5)# Los valores de X que tiene una pro
> qbinom(c(0.95), 10, 0.5) # Los valores de X que tiene una pr
> rbinom(2,10,0.5) # 2 números aleatorios de una binomial
> rbinom(9,10,0.5)  # 9 números aleatorios de una binomial
```

Ejemplo

Considere un conjunto de experimentos de Bernoulli en el que se seleccionan tres articulos al azar de un proceso de producción, luego seinspeccionan y se clasifican como defectuosos o no defectuosos. Un artículodefectuoso se designa como éxito. El número de éxitos es una variable aleatoriaX que toma valores integrales de 0 a 3. Los ocho resultadosposibles son.

Resultado	**NNN**	**NDN**	**NND**	**DNN**	**NDD**	**DND**	**DDN**	**DDD**
x	0	1	1	1	2	2	2	3

Los artículos se seleccionan de forma independiente y se asume que el proceso produce 25% de artículos defectuosos.

P(NDN) = P(N)P(D)P(N) = (3/4)(1/4)(3/4) = 9/64

Distribuciones de probabilidad

Distribución Binomial

El número x de éxitos en n experimentos de Bernoulli se denomina **Variable aleatoria binomial**. La distribución de probabilidades de esta variable aleatoria discreta se llama **distribución binomial**.

- Se denota como b(x,n,p)

Para la distribución de probabilidad de X es:

- P(X=2) = f(2) = b(2;3,1/4) = 9/64

En forma general, una probabilidad p de éxito y q=1-p el fracaso. La distribucion de la variable aleatoria binomial X, el númerode éxitos en n ensayos independientes es:

- b(x;n,p) = Conv(n,x)$p^{x}q^{n-x}$, x = 0,1,2,3,4,…,n
- $(q + p)^{n}$ = Conv(n,0) q^{n} + Conv(n,1) $p^{1}q^{n-1}$ + Conv(n,2) $p^{2}q^{n-2}$ + … + Conv(n,n) p^{n}
- p + q = 1 --> sum(b(x;n,p)) =1

Distribuciones de probabilidad

Experimentos multinominal

Un experimento bimodal se convierte en un experimento **multinominal** si cada prueba contiene más de dos resultados posibles.

En general, si un ensayo puede tener una de las k consecuencias $E_1,E_2,E_3,\ldots,E_k$ con probabilidad $P_1,P_2,P_3,\ldots,P_k$, la **distribución multinomial** dará la probabilidad de que E_i ocurra x_i veces, donde

$x_1 + x_2 + \ldots + x_k = n$

denotamos esta distribución como

$f(x_1,x_2,\ldots,x_k;p_1,p_2,\ldots,p_k)$

Distribuciones de probabilidad

Experimentos multinominal

Ejemplo La complejidad de las llegadas y salidas de los aviones en un aeropuerto es tal que a menudo se utiliza la simulación para modelar las condiciones "ideales". Para un aeropuerto A se tienen 3 pistas de aterrizaje y un avión llega con las siguientes probabilidades.

- Pista 1: $p_1 = 4/18$
- Pista 2: $p_2 = 3/18$
- Pista 3: $p_3 = 11/18$

¿Cuál es la probabilidad de que 6 aviones que llegan al azar se distribuyan de la siguiente manera?

- Pista 1: 2 aviones
- Pista 2: 1 avión
- Pista 3: 3 aviones

$f(2,1,3;4/18,3/18,11/18,6) = 6!/2!1!3!\ \mathit{2^2/9^2}\ 11^3/18^3 = 0.1127$

Distribuciones de probabilidad

Distribución Geométrica

La distribución Geométrica es una serie de ensayos de Bernoulli independientes, con probabilidad constante p de éxito, donde la variable aleatoria X denota el número de ensayos hasta el primer éxito.

- X ~ G(p)
- f(x) = p(1-p)$_{x-1}$; con x={0,1}

```
- dgeom(x, prob, log = F); Devuelve resultados de la función
- pgeom(q, prob, lower.tail = T, log.p = F); Devuelve result
- qgeom(p, prob, lower.tail = T, log.p = F); Devuelve result
- rgeom(n, prob); Devuelve un vector de valores binomiales a
```

Distribuciones de probabilidad

Distribución Geométrica

con:

- x, q: Vector de cuantiles que representa el número de fallos antes del primer éxito.
- p: Vector de probabilidades.
- n: Números de valores aleatorios a devolver.
- prob: Probabilidad de éxito en cada ensayo.
- log, log.p: Parámetro booleano, si es TRUE, las probabilidades p se ofrecen como log(p).
- lower.tail: Parámetro booleano, si es TRUE (por defecto), las probabilidades son $P[X \leq x]$, de lo contrario, $P[X > x]$.

Distribuciones de probabilidad

Distribución Geométrica

```
> dgeom(4, 0.5) # Calculamos la P(X=4) de una G(0.5); represen

> pgeom(4, 0.5, lower.tail = F) # Calculamos la P(X>4) de una

> Lo comprobamos; P(X > 4) = 1 - P(X<=3) = 1 - [P(X=0)+P(X=1)-
> 1 - (dgeom(0, 0.5) + dgeom(1, 0.5) + dgeom(2, 0.5)+ dgeom(3,
```

Distribuciones de probabilidad

Distribución Binomial Negativa

Surge de un contexto semejante al que conduce a la distribución geométrica. La distribución que asigna a la variable Y, el número de ensayo en el que ocurre el r-ésimo éxito.

Entonces X tiene una distribución Binomial Negativa.

- X ~ BN(x,r,p)

su función de probabilidad

- P(y) = conv(x-1,r-1)p^rqx-r; para x {r,r+1,r+2,...}

Distribuciones de probabilidad

Distribución Binomial Negativa

Para obtener valores que se basen en la distribución Binomial Negativa, R dispone de cuatro funciones basados en “nbinom”:

- dnbinom(x, size, prob, mu, log = F); Devuelve resultados de la función de densidad.
- pnbinom(q, size, prob, mu, lower.tail = T, log.p = F); Devuelve resultados de la función de distribución acumulada.
- qnbinom(p, size, prob, mu, lower.tail = T, log.p = F); Devuelve resultados de los cuantiles de la Binomial Negativa.

-rnbinom(n, size, prob, mu); Devuelve un vector de valores binomiales aleatorios.

Distribuciones de probabilidad

Distribución Binomial Negativa

con:

- x, q: Vector de cuantiles (Valores enteros positivos). Corresponde a número de pruebas falladas.
- q: Vector de cuantiles.
- p: Vector de probabilidades.
- n: Números de valores aleatorios a devolver.
- prob: Probabilidad de éxito en cada ensayo.
- size: Número total de ensayos. Debe ser estrictamente positivo.
- mu: Parametrización alternativa por la media.
- log, log.p: Parámetro booleano, si es TRUE, las probabilidades p se ofrecen como log(p).
- lower.tail: Parámetro booleano, si es TRUE (por defecto), las probabilidades son $P[X \leq x]$, de lo contrario, $P[X > x]$.

```
> # Suponga que el 60% de los elementos no están defectuosos.
> # Para encontrar la probabilidad de localizar el quinto elem
> # Calculamos P(X = 7), con r=5
> dnbinom(7-5, 5, 0.6)
> # si queremos calcular en el séptimo ensayo o antes
> pnbinom(7-5, 5, 0.6, lower.tail = T)
```

Distribuciones de probabilidad

Distribución De Poisson

La distribución de Poisson expresa la probabilidad de un número k de eventos ocurriendo en un tiempo fijo si estos eventos ocurren con una frecuencia media conocida y son independientes del tiempo discurrido desde el último evento.

- $X \sim P(\lambda)$

La función de densidad de la distribución de Poisson es:

- $f(k,\lambda) = e^{-\lambda}\lambda^{k}/k!$

Para obtener valores que se basen en la distribución de Poisson, R dispone de cuatro funciones basados en“pois”:

- dpois(x, lambda, log = F); Devuelve resultados de la función de densidad.
- ppois(q, lambda, lower.tail = T, log.p = F); Devuelve resultados de la función de distribución acumulada.
- qpois(p, lambda, lower.tail = T, log.p = F); Devuelve resultados de los cuantiles de la Poisson.
- rpois(n, lambda); Devuelve un vector de valores binomiales aleatorios

Distribuciones de probabilidad

Distribución normal

La distribución normal, distribución de Gauss o distribución gaussiana, a una de las distribuciones de probabilidad de variable continua que con más frecuencia aparece en fenómenos reales.

Se dice que una variable aleatoria continua X sigue una distribución normal de parámetros μ y σ y se denota X ~ N(μ, σ) si su función de densidad está dada por:

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

Para obtener valores que se basen en la distribución normal, R dispone de cuatro funciones basados en “norm”:

- dnorm(x, mean = 0, sd = 1, log = F); Resultados de la función de densidad.
- pnorm(q, mean = 0, sd = 1, lower.tail = T, log.p = F); Resultados de distribución acumulada.
- qnorm(p, mean = 0, sd = 1, lower.tail = T, log.p = F); Resultados de los cuantiles de la Normal.
- rnorm(n, mean = 0, sd = 1); Vector de valores normales aleatorios

Distribuciones de probabilidad

Distribución normal

```
> # Calcular la P(Z>1) de una N(0,1)
> pnorm(1, mean = 0, sd = 1, lower.tail = F)
> # Calcular la P(-2<Z<2) de una N(0,1)
> pnorm(c(2), mean = 0, sd = 1) - pnorm(c(-2), mean = 0, sd =
> # Calcular la P(0<Z<1.96) de una N(0,1)
> pnorm(1.96, mean = 0, sd = 1) - pnorm(0, mean = 0, sd = 1)
> # Calcular la P(Z<=z)=0,5793 de una N(0,1)
> qnorm(0.5793, mean = 0, sd = 1)
> # Calcular la P(Z>150) de una Normal de media 125 y la desvi
> pnorm(150, mean = 125, sd = 50, lower.tail = F)
```

Distribuciones de probabilidad

Distribución t-Student

La distribución t-Student es una distribución de probabilidad que surge del problema de estimar la media de una población normalmente distribuida cuando el tama˜no de la muestra es peque˜no.

La distribución t de Student es la distribución de probabilidad del cociente:

- Z/sqrt(V/v)

donde:

1. Z tiene una distribución normal de media nula y varianza 1
2. V tiene una distribución chi-cuadrado con ν grados de libertad
3. Z y V son independientess La función de densidad de t es

$$f(t) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}(1+t^2/\nu)^{-(\nu+1)/2}$$

Test-T

```
> qpois(0.985, 3)
```

Distribuciones de probabilidad

Distribución normal

La distribución normal, distribución de Gauss o distribución gaussiana, a una de las distribuciones de probabilidad de variable continua que con más frecuencia aparece en fenómenos reales.

La gráfica de su función de densidad tiene una forma acampanada y es simétrica respecto de un determinado parámetro. Esta curva se conoce como **campana de Gauss**.

Se dice que una variable aleatoria continua X sigue una distribución normal de parámetros μ y σ y se denota X ∼ N(μ, σ) si su función de densidad está dada por:

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

Distribuciones de probabilidad

Distribución normal

- dnorm(x, mean = 0, sd = 1, log = F); Devuelve resultados de la función de densidad.
- pnorm(q, mean = 0, sd = 1, lower.tail = T, log.p = F); Devuelve resultados de la función de distribución acumulada.
- qnorm(p, mean = 0, sd = 1, lower.tail = T, log.p = F); Devuelve resultados de los cuantiles de la Normal.
- rnorm(n, mean = 0, sd = 1); Devuelve un vector de valores normales aleatorios.

Con:

- x, q: Vector de cuantiles.
- p: Vector de probabilidades.
- n: Números de observaciones.
- 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, $P[X > x]$.

Distribuciones de probabilidad

Distribución normal

```
> # Calcular la P(Z>1) de una N(0,1)
> pnorm(1, mean = 0, sd = 1, lower.tail = F)
> # Calcular la P(-2<Z<2) de una N(0,1)
> pnorm(c(2), mean = 0, sd = 1) - pnorm(c(-2), mean = 0, sd =
> # Calcular la P(0<Z<1.96) de una N(0,1)
> pnorm(1.96, mean = 0, sd = 1) - pnorm(0, mean = 0, sd = 1)
> # Calcular la P(Z<=z)=0,5793 de una N(0,1)
> qnorm(0.5793, mean = 0, sd = 1)
> # Calcular la P(Z>150) de una Normal de media 125 y la desvi
> pnorm(150, mean = 125, sd = 50, lower.tail = F)
```

Distribuciones de probabilidad

Distribución t-Student

La distribución t-Student es una distribución de probabilidad que surge del problema de estimar la media de una población **normalmente distribuida** cuando el **tama˜no de la muestra es peque˜no**.

Aparece de manera natural al realizar la prueba t de Student para la determinación de las diferencias entre dos medias muestrales y para la construcción del intervalo de confianza para la diferencia entre las medias de dos poblaciones cuando se desconoce la desviación típica de una población y ésta debe ser estimada a partir de los datos de una muestra.

T = Z/sqrt(V/v)

- Z variable aleatoria distribuida segun una normal típica (media = 0, varianza=1)
- V es una variable con distribución chi^2
- Z y V son independientes

Distribuciones de probabilidad

Distribución t-Student

- dt(x, df, ncp, log = F); Devuelve resultados de la función de densidad.
- pt(q, df, ncp, lower.tail = T, log.p = F); Devuelve resultados de la función de distribución acumulada.
- qt(p, df, ncp, lower.tail = T, log.p = F); Devuelve resultados de los cuantiles de la t-Student.
- rt(n, df, ncp); Devuelve un vector de valores t-Student aleatorios.

Con:

- x, q: Vector de cuantiles.
- p: Vector de probabilidades.
- n: Números de observaciones.
- df: Grados de libertad.
- ncp: Parámetro que determina la centralidad de la gráfica t-Student. Si se omite, el estudio se realiza con la gráfica centralizada en 0.
- 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 ≤ x], de lo contrario, P[X > x].

Distribuciones de probabilidad

Distribución t-Student

```
> # Calcular P(T >= 1.3) con 7 grados de libertad.
> pt(1.3, 7, lower.tail = F)
> # Calcular P(T < 2.30) con 20 grados de libertad.
> pt(2.30,20, lower.tail = T)
> # P(T >= t) = 0.05 con 25 grados de libertad.
> qt(0.05, 25, lower.tail = F)
> # Calcular 5 números aleatorios con 25 grados de libertad.
> rt(5, 25)
```

Distribuciones de probabilidad

Distribución Chi-cuadrado

En estadística, la distribución X^2 (de Pearson) es una distribución de probabilidad continua con un parámetro k que representa los grados de libertad de la variable aleatoria:

$X = Z^2{}_1 + Z^2{}_2 + \cdots + Z^2{}_k$

donde Z_i son variables de distribución normal, de media cero y varianza uno.

$$f(x;k) = \begin{cases} \dfrac{1}{2^{k/2}\Gamma(k/2)}\, x^{(k/2)-1} e^{-x/2} & \text{para } x > 0, \\ 0 & \text{para } x \leq 0 \end{cases}$$

donde Γ es la función gamma

Distribuciones de probabilidad

Distribución Chi-cuadrado

```
- dchisq(x, df, ncp=0, log = F); Devuelve resultados de la fun
- pchisq(q, df, ncp=0, lower.tail = T, log.p = F); Devuelve re
acumulada.
- qchisq(p, df, ncp=0, lower.tail = T, log.p = F); Devuelve re
- rchisq(n, df, ncp=0); Devuelve un vector de valores chi-Cuad
```

Con:

- x, q: Vector de cuantiles.
- p: Vector de probabilidades.
- n: Números de observaciones.
- df: Grados de libertad.
- ncp: Parámetro que determina la centralidad de la gráfica chi-Cuadrados. Si se omite, el estudio se realiza con la gráfica centralizada en 0.
- log, log.p: Parámetro booleano, si es TRUE, las probabilidades p son devueltas como log (p). lower.tail: Parámetro

Distribuciones de probabilidad

Distribución Chi-cuadrado

```
> # Calcular X^2(0.52, 7) con 7 grados de libertad.
> qchisq(0.52, 7, lower.tail = F)
> # P(X^2 < x) = 0.80 con 25 grados de libertad
> qchisq(0.8, 25, lower.tail = T)
> # P(X^2 >= 18.49) con 24 grados de libertad.
> pchisq(18.49, 24, lower.tail = F)
> # Calcular 5 números aleatorios de una dist. Chi-cuadrado co
> rchisq(5, 24)
```

Distribuciones de probabilidad

Distribución F

La distribución F de Snedecor es una distribución de probabilidad continua. Una variable aleatoria de distribución F se construye como el siguiente cociente:

$F=U_1d_1 / U_2d_2$

- U1 y U2 siguen una distribución chi-cuadrada con d1 y d2 grados de libertad respectivamente, y
- U1 y U2 son estadísticamente independientes

Distribuciones de probabilidad

Distribución F

```
df(x, df1, df2, ncp, log = F); Devuelve resultados de la func
pf(q, df1, df2, ncp, lower.tail = T, log.p = F); Devuelve resu
qf(p, df1, df2, ncp, lower.tail = T, log.p = F); Devuelve resu
rf(n, df1, df2, ncp); Devuelve un vector de valores de la dist
```

Con:

- x, q: Vector de cuantiles.
- p: Vector de probabilidades.
- n: Números de observaciones.
- df1, df2: Grados de libertad, df1 corresponde al numerador y df2 al denominador.
- ncp: Parámetro que determina la centralidad de la gráfica de la distribución F. Si se omite, el estudio se realiza con la gráfica no centralizada.
- 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, $P[X > x]$.

Distribuciones de probabilidad

Distribución F

```
> # Calcular F(0.15, 3, 2). con 3 y 2 grados de libertad.
> qf(0.15, 3, 2, lower.tail=F)
> # P(F < f) = 0.025 con df1 = 20 y df2 = Infinito.
> qf(0.025, 20, Inf, lower.tail=T)
> # P(F >= 198.50) con df1 = 10 y df2 = 2.
> pf(198.50, 10, 2, lower.tail=F)
> # Calcular 5 números aleatorios de una dist. F de Snedecor c
> rf(5,24,10)
```

Distribuciones de probabilidad

Distribución beta

.right-column[En estadística la distribución beta es una distribución de probabilidad continua con dos parámetros a y b cuya función de densidad para valores 0 < x < 1 es

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

Donde Γ es la función gamma.

Distribuciones de probabilidad

Distribución beta

```
- dbeta(x, shape1, shape2, ncp = 0, log = F); Devuelve result
- pbeta(q, shape1, shape2, ncp = 0, lower.tail = T, log.p = F
distribución acumulada.
- qbeta(p, shape1, shape2, ncp = 0, lower.tail = T, log.p = F
distribución Beta.
- rbeta(n, shape1, shape2, ncp = 0); Devuelve un vector de va
```

Con:

- x, q: Vector de cuantiles.
- p: Vector de probabilidades.
- n: Números de observaciones. shape1, shape2: Parámetros de la Distribución Beta. Shape1 = α y Shape2 = β. Ambos deben ser positivos.
- ncp: Parámetro lógico que determina si la distribución es central o no.
- 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, $P[X > x]$.

Distribuciones de probabilidad

Distribución beta

```
> # Mediante una distribución beta con ALPHA = 5 y BETA = 4.
> # Calcule la probabilidad al menos del 75%
> pbeta(0.75, 5, 4, lower.tail = F)
> # Menos del 50%
> pbeta(0.5, 5, 4, lower.tail = T)
> # P(X < x)=0.25
> qbeta(0.25, 5, 4, lower.tail = T)
> # P(X > x) = 0.5
> qbeta(0.5, 5, 4, lower.tail = F)
> # Calcular 5 números aleatorios de una dist. Beta con ALPHA
> rbeta(5,5,4)
```

.right-column[

Distribuciones de probabilidad

La distribución gamma es una distribución de probabilidad continua con dos parámetros k y λ cuya función de densidad para valores x > 0 es

Distribución gamma

$$f(x) = \lambda^k e^{-\lambda x} \frac{x^{k-1}}{\Gamma(k)}$$

Aquí e es el número e y Γ es la función gamma.

Distribuciones de probabilidad

Distribución gamma

```
- dgamma(x, shape, rate, scale = 1/rate, log = F); Devuelve re
- pgamma(q, shape, rate, scale = 1/rate, lower.tail = T, log.p
acumulada.
- qgamma(p, shape, rate, scale = 1/rate, lower.tail = T, log.p
de la distribución Gamma.
- rgamma(n, shape, rate, scale = 1/rate); Devuelve un vector d
```

Con:

- x, q: Vector de cuantiles.
- p: Vector de probabilidades.
- n: Números de observaciones.
- rate: Alternativa para especificar el valor de escala (Scale). Por defecto, su valor es igual a 1. shape, scale: Parámetros de la Distribución Gamma. Shape = a y Scale = s = 1/rate. Debe ser estrictamente positivo el parámetro Scale.
- 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 ≤ x], de lo contrario, P[X > x].

Distribuciones de probabilidad

Distribución gamma

```
> # Mediante una distribución gamma con ALPHA = 3 y BETA = 0.5
> # Calcule la probabilidad de que sea mejor de 10
> pgamma(10, 3, rate = 0.5, lower.tail = F)
> # Entre 4 y 8
> pgamma(8, 3, rate = 0.5, lower.tail = T) - pgamma(4, 3, rate
> # P(X <x) = 0.7
> qgamma(0.7, 3, rate = 0.5, lower.tail = T)
> # P(X > x) = 0.5
> qgamma(0.5, 3, rate = 0.5, lower.tail = F)
> # Calcular 5 números aleatorios de una dist. gamma con ALPHA
> rgamma(5, 3, rate = 0.5)
```

Regresión

Definición

La regresión es una técnica estadística que analiza la relación de dos o mas variables que principalmente se utilizada para inferir datos a partir de otros y hallar una respuesta de lo que puede suceder. Esta nos permite conocer el cambio en una de las variables llamadas respuesta y que corresponde a otra conocida como variable explicativa.

Se pueden encontrar varios tipos de regresión, por ejemplo:

- Regresión lineal simple
- Regresión múltiple (varias variables)
- Regresión logística

Algunas ecuaciones regresión son:

- Regresión Lineal : $y = A + Bx$
- Regresión Logarítmica : $y = A + BLn(x)$
- Regresión Exponencial : $y = Ac(bx)$
- Regresión Cuadrática : $y = A + Bx + Cx2$

Regresión

Regresión lineal

La función en R para obtener modelos de regresión lineal simple, Y=aX+b, es “lm”, aunque también se puede utilizar esta función para el análisis de la varianza y análisis de covarianza.

```
> a.docencia <- c(3,1,1,2,5,6,12,7,3,10,6,11,4,4,16,4,5,3,5,2
> edad <- c(35,27,26,30,33,42,51,35,45,37,43,36,36,56,29,35,37
> # Y=a~nos de docencia y X=edad
> lm(a.docencia~edad)->r1
> r1
```

Hay que tener en cuenta que el orden en el que se escriben las variables es de gran importancia, en este caso la variable dependiente es los a˜nos de docencia.

$y=a_0+a_1x$

y = edad

x = años de docencia

Por lo tanto, la recta es: y(x) = (1,3081) − (0,1156)x

Ejercicio: Realizar el caso donde la variable dependiente es la edad.

Regresión

Regresión lineal

.right-column[Dibujar la recta de regresión

```
> # Defino los datos
> x <- c(3, 5, 2, 3, 1, 4, 6, 4)
> y <- c(150, 150, 250, 200, 350, 200, 50, 125)
> # Defino la recta de regresión
> lm(y~x)->ryx

> # Definimos el eje X
> litros <- seq(0:length(x))
> # Defino la recta
> precio <- (ryx$coefficients[1])+ (ryx$coefficients[2])*aceite# E
> #Dibujo los puntos, se~nalados con una X
> plot(x, y, pch="X", col=2, xlab="Litros", ylab="Precio")
> #Dibujo la recta
> lines(precio, col=4)
```

```
> # Creamos el diagrama de puntos
> plot(x, y)
> # Dibujamos la recta de regresión
> abline(lm(y ~ x))
> predict(ryx) # Son los valores de y cuando aplicamos la recta de regresión a los
```

Regresión

Regresión lineal

Si lo que queremos es ajustar el modelo para poder usarlo posteriormente para predecir datos utilizaremos la función “predict”. Esta función obtiene todas las posibles predicciones para la variable x según la posición en la que se encuentren sus datos.

Predecir un vector de valores y utilizar summary.

```
> x <- c(3, 5, 2, 3, 1, 4, 6, 4)
> y <- c(150, 150, 250, 200, 350, 200, 50, 125)
> # Recta de regresión sin termino independiente
> lm(y~0+x)
```

Regresión

Regresión Polinomial

Para calcular la función de regresión polinomial $Y = a_0 + a_1X + a_2X^2 + \ldots + a_pX^p$ utilizamos la función lm(y˜x + I(x^2) + I(x^3) + ... + I(x^p)).

```
> # Cuadrática
> lm(y~x+x^2)
```

Regresión

Regresión Polinomial

Para calcular la función de regresión polinomial $Y = a_0 + a_1X + a_2X^2 + \ldots + a_pX^p$ utilizamos la función lm(y˜x + I(x^2) + I(x^3) + ... + I(x^p)).

```
> # Cuadrática
> lm(y~x+x^2)
```

Regresión

Regresión Polinomial sin término independiente

```
> lm(y~0+x+x^2)
> lm(y~0+x+I(x^2)+I(x^3))
> lm(y~0+x+I(x^2)+I(x^3)+I(x^4))
```

Regresión

Regresión Potencial

Para calcular la función de regresión potencial $Y = aX^b$ utilizamos la función lm(log(y)˜ log(x)).

```
> lm(log(y)~log(x))
```

Regresión

Regresión exponencial

Para calcular la función de regresión potencial $Y = e^{a+bX}$ utilizamos la función lm(log(y)˜x).

```
> lm(log(y)~x)
```

Regresión

Regresión logarítmica

Para calcular la función de regresión logarítmica Y = a + b log(x) utilizamos la función lm(y˜ log(x)).

```
> lm(y~log(x))
```

That's all folks (for now)!

Slideshow created using remark.