Tema 4. Distribuciones de Probabilidad con R

Estadística

Ángel Serrano Sánchez de León

Índice

Distribuciones de probabilidad en R

Prefijo	Tipo de función
r...	Generador de números aleatorios
d...	Densidad de probabilidad (continua) o función de masa (discreta)
p...	Función de distribución (probabilidad acumulada)
q...	Función cuantil sobre la distribución

Distribuciones de probabilidad en R

Nombre de distribución	Generador de números aleatorios	Función de Densidad/ de Masa	Función de Distribución	Función cuantil
Uniforme Discreta	-	-	-	-
Binomial	`rbinom`	`dbinom`	`pbinom`	`qbinom`
Poisson	`rpois`	`dpois`	`ppois`	`qpois`
Uniforme Continua	`runif`	`dunif`	`punif`	`qunif`
Normal	`rnorm`	`dnorm`	`pnorm`	`qnorm`
χ^2	`rchisq`	`dchisq`	`pchisq`	`qchisq`
t de Student	`rt`	`dt`	`pt`	`qt`
F	`rf`	`df`	`pf`	`qf`

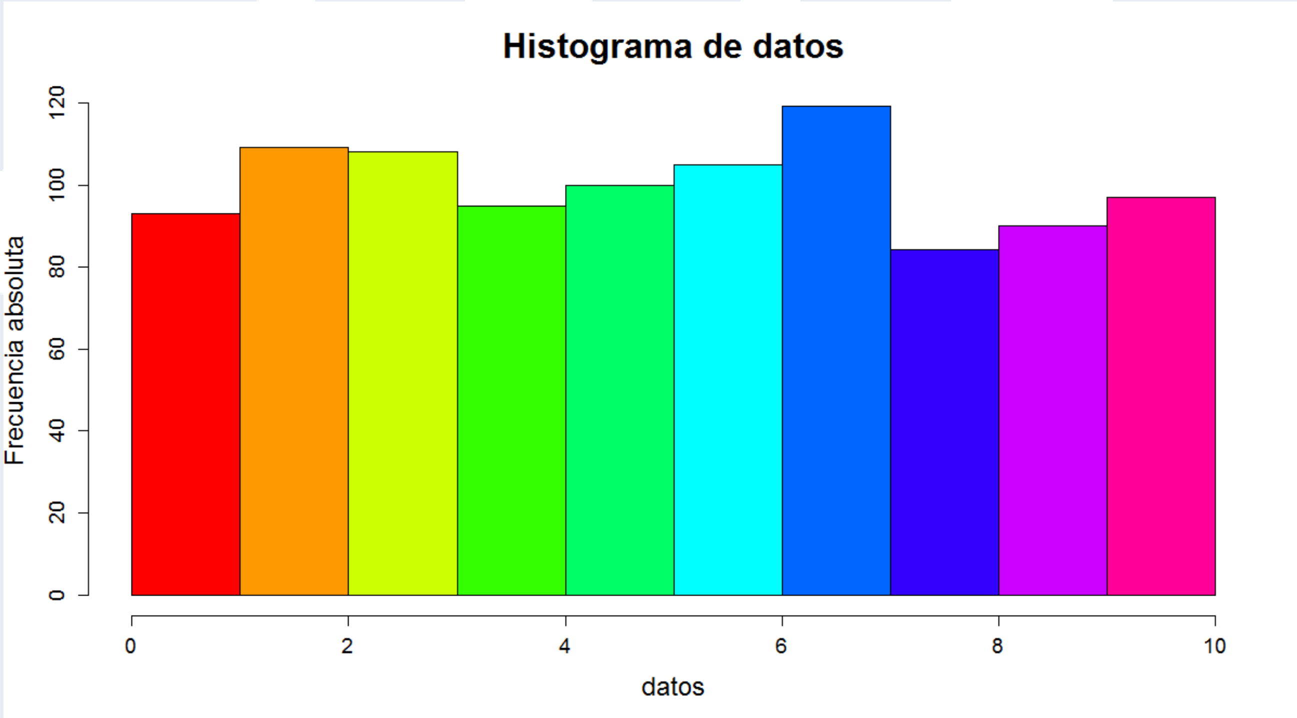
Distribución uniforme continua

Distribución uniforme continua

- Generaremos un vector de 1000 números aleatorios reales entre $a = 0$ y $b = 10$.

```
> datos <- runif(1000,0,10)
> head(datos) # Primeros valores
[1] 0.2426938 4.2880470 0.4842197 9.1586329
  8.7479441 2.7027880
> hist(datos, col=rainbow(10),
  ylab="Frecuencia absoluta",
  main="Histograma de datos",cex.axis=1.2,
  cex.lab=1.5,cex.main=2)
```

Distribución uniforme continua



Distribución uniforme continua

```
> mean(datos)
[1] 4.92457
> (0+10)/2 # Valor teórico de la media
[1] 5
```

$$E(X) = \mu = \frac{a+b}{2}$$

```
> var(datos)
[1] 8.1444
> (10-0)^2/12 # Valor teórico de la
  varianza
[1] 8.333333
```

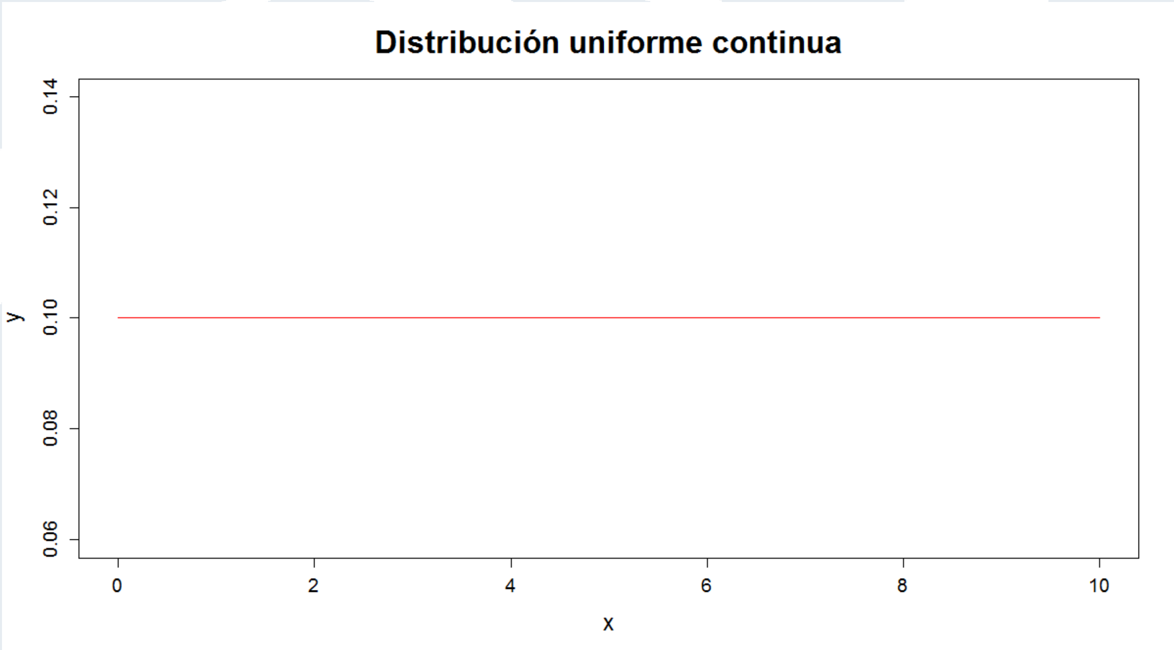
$$\mathrm{Var}(X) = \sigma^2 = \frac{(b-a)^2}{12}$$

Distribución uniforme continua

- Representemos ahora la densidad de probabilidad de la distribución uniforme continua de una variable que toma valores entre $a = 0$ y $b = 10$.

```
> x <- 0:10
> y <- dunif(x,0,10)
> plot(x,y,type="l",col="red",
  main="Distribución uniforme continua",
  ylab="y",cex.axis=1.2,cex.lab=1.5,
  cex.main=2)
```

Distribución uniforme continua



$$f(x) = \frac{1}{b-a}$$

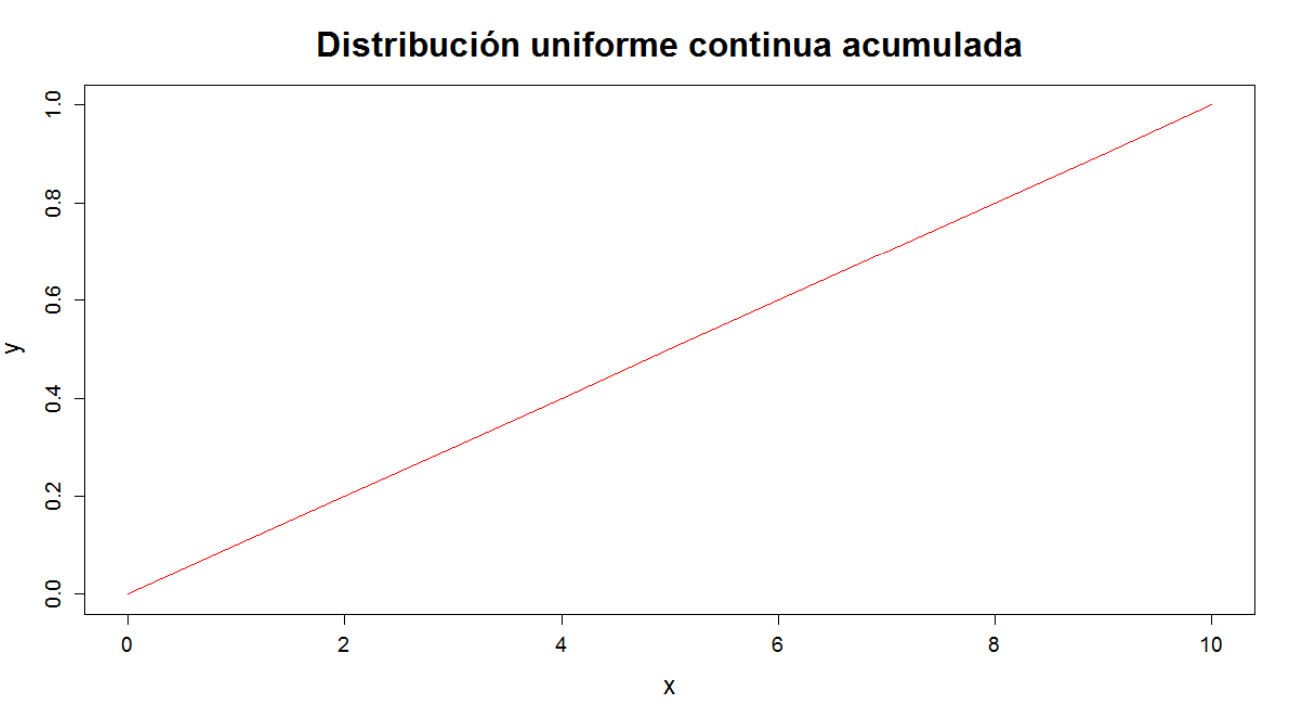
10

Distribución uniforme continua

- Veamos la función de distribución (probabilidad acumulada) para una variable que toma valores entre $a = 0$ y $b = 10$.

```
> x <- 0:10
> y <- punif(x,0,10)
> plot(x,y,type="l", col="red",
  main="Distribución uniforme continua
  acumulada",ylab="y",cex.axis=1.2,
  cex.lab=1.5,cex.main=2)
```

Distribución uniforme continua

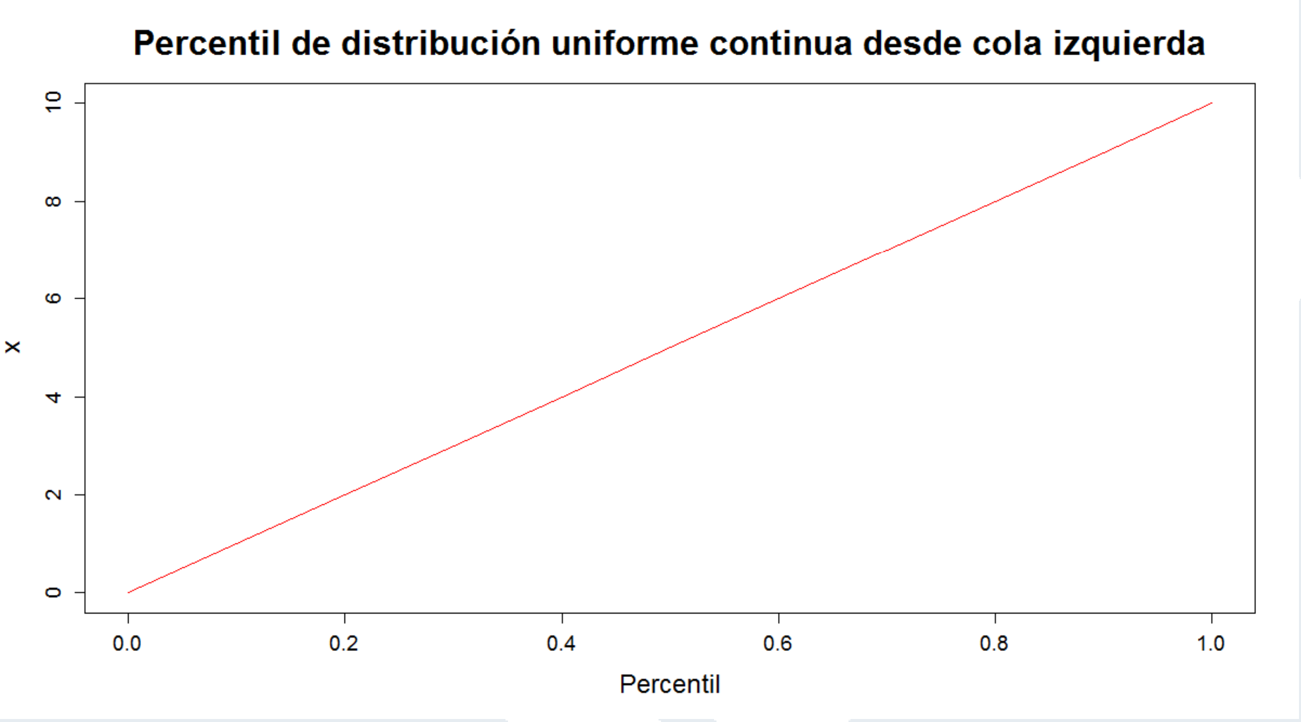


Distribución uniforme continua

- Veamos para qué valores de la variable se obtienen los diversos percentiles correspondientes al área bajo la función densidad de probabilidad medida desde la cola izquierda.

```
> x <- seq(0,1,0.1)
> y <- qunif(x,0,10)
> plot(x,y,type="l", col="red",
  main="Percentil de distribución uniforme
  continua desde cola izquierda",
  ylab="x",cex.axis=1.2,cex.lab=1.5,
  cex.main=2, xlab="Percentil")
```

Distribución uniforme continua



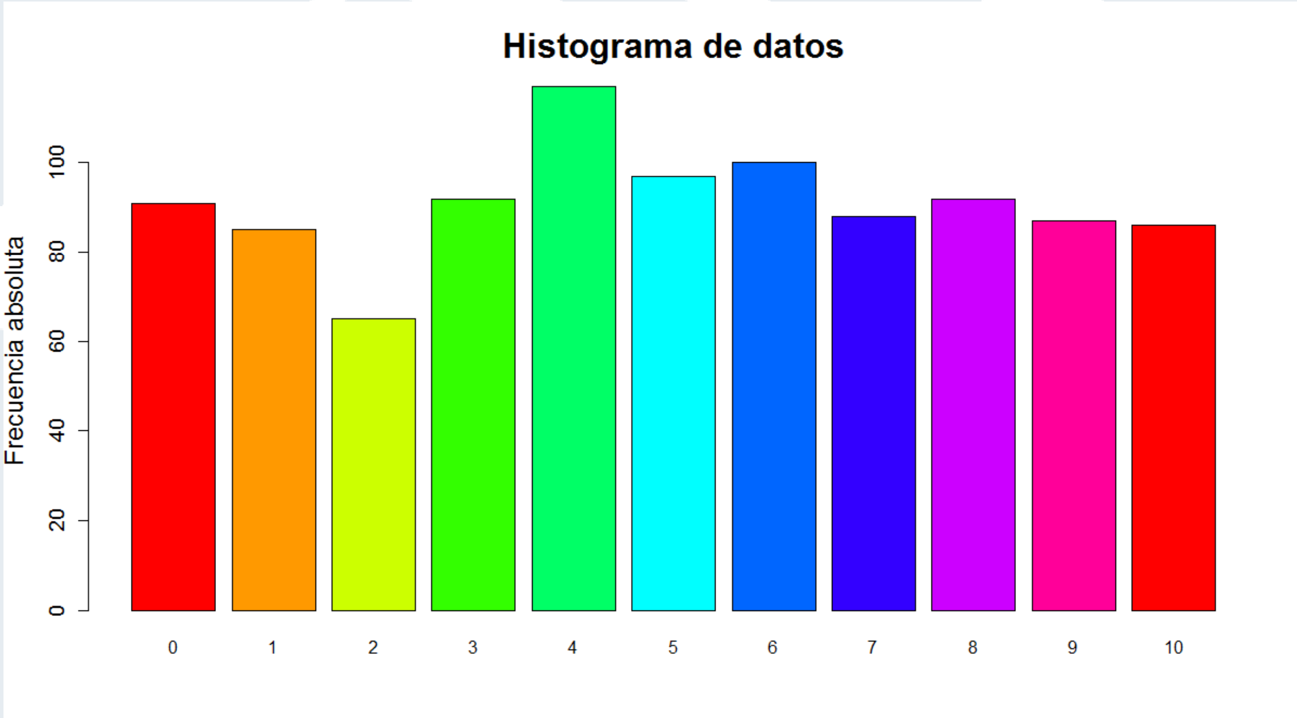
Distribución uniforme discreta

Distribución uniforme discreta

- Igual que antes, pero convertimos los números a enteros por truncamiento (no existe función específica en R).
- Generaremos un vector de 1000 números aleatorios enteros entre $a = 0$ y $b = 10$.

```
> datos <- as.integer(runif(1000,0,11))
> head(datos) # Primeros valores
[1] 4 8 2 1 3 8
> barplot(table(datos),col=rainbow(10),
  ylab="Frecuencia absoluta",
  main="Histograma de datos",cex.axis=1.2,
  cex.lab=1.5,cex.main=2)
```

Distribución uniforme discreta



Distribución uniforme discreta

```
> mean(datos)
[1] 5.039
> (0+10)/2 # Valor teórico de la media
[1] 5
```

$$E(X) = \mu = \frac{a+b}{2}$$

```
> var(datos)
[1] 9.535014
> ((10-0+1)^2 - 1)/12 # Valor teórico de
  la varianza
[1] 10
```

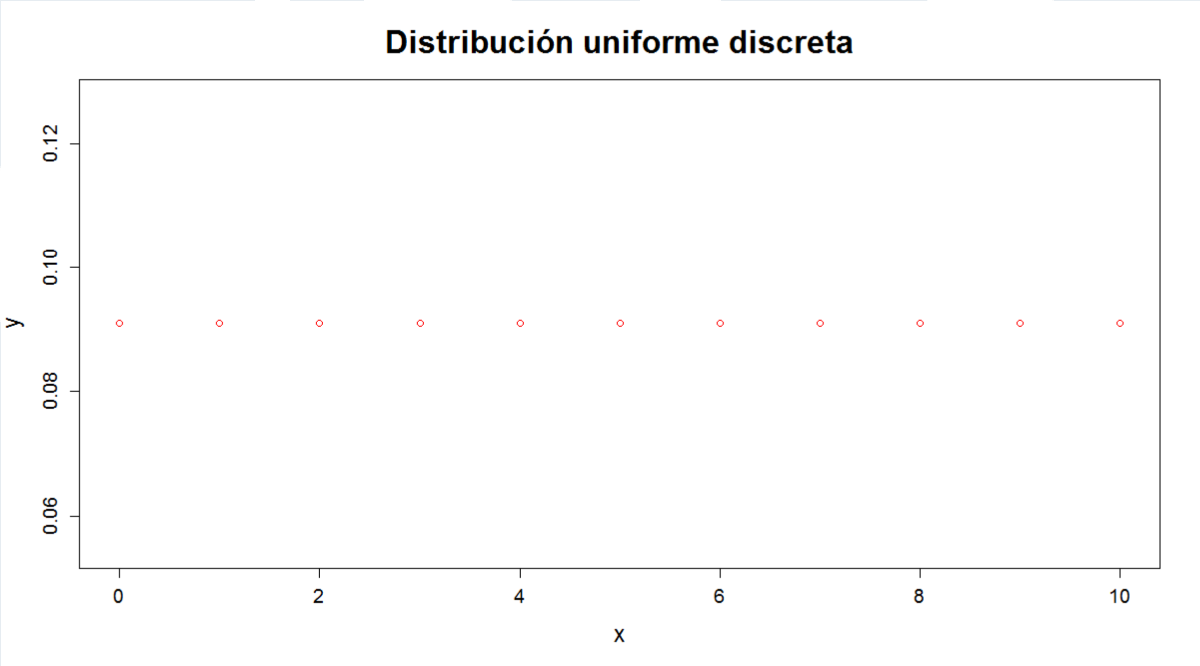
$$\operatorname{Var}(X) = \sigma^2 = \frac{(b-a+1)^2 - 1}{12}$$

Distribución uniforme discreta

- La distribución uniforme discreta teórica la calculamos mediante la uniforme continua, prestando cuidado a que los valores de *x* sean enteros.
- Para el caso $a = 0$ y $b = 10$.

```
> x <- 0:10
> y <- dunif(x,0,11) # ¡¡OJO!!
> plot(x,y,type="p",col="red",
  main="Distribución uniforme discreta",
  ylab="y",cex.axis=1.2,cex.lab=1.5,
  cex.main=2)
```

Distribución uniforme discreta



$$f(x) = \frac{1}{b-a+1}$$

20

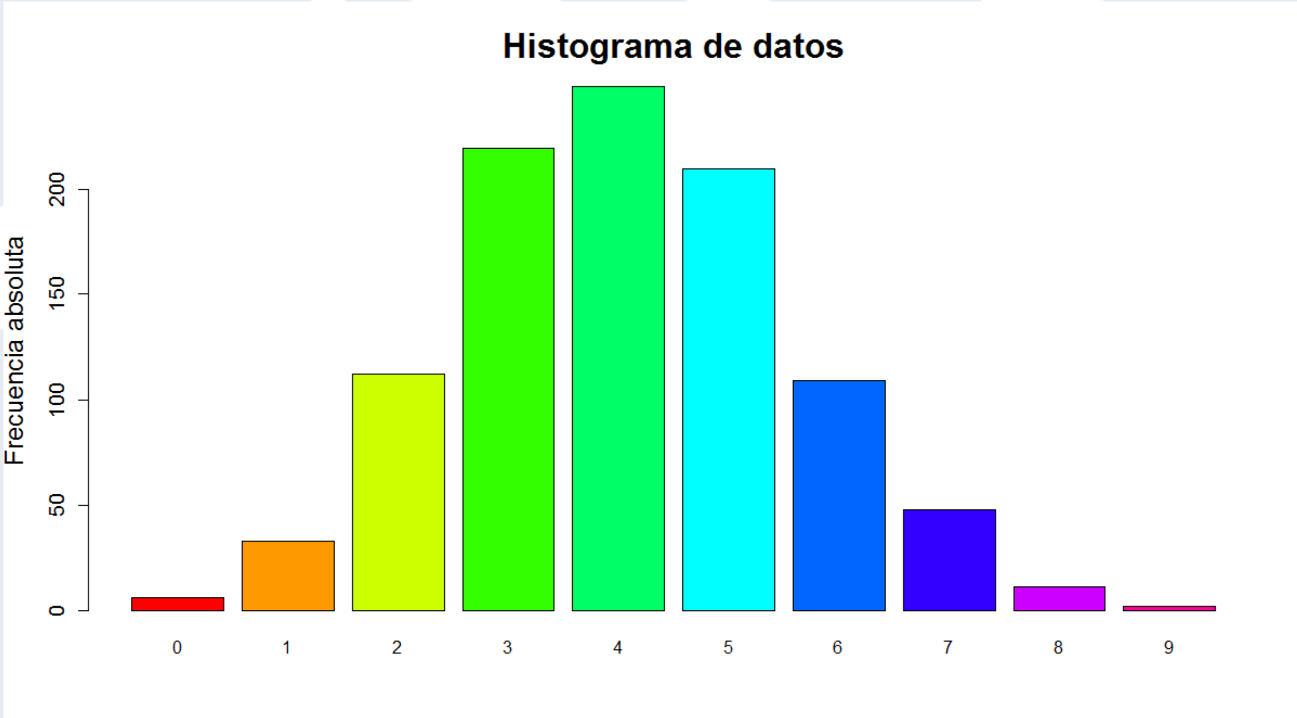
Distribución binomial

Distribución binomial

- Calculemos ahora 1000 números aleatorios modelados según la distribución binomial que en $n = 10$ intentos tiene una probabilidad de éxito $p = 0{,}4$.

```
> datos <- rbinom(1000,10,0.4)
> head(datos)
[1] 2 2 4 3 4 6
> barplot(table(datos),col=rainbow(10),
  ylab="Frecuencia absoluta",
  main="Histograma de datos",cex.axis=1.2,
  cex.lab=1.5,cex.main=2)
```

Distribución binomial



Distribución binomial

```
> mean(datos)
[1] 4.059
> 10*0.4 # Valor teórico de la media
[1] 4
```

$$\mathrm{E}(X) = np$$

```
> var(datos)
[1] 2.363883
> 10*0.4*(1-0.4) # Valor teórico de la
  varianza
[1] 2.4
```

$$\mathrm{Var}(X) = \sigma^2 = npq$$

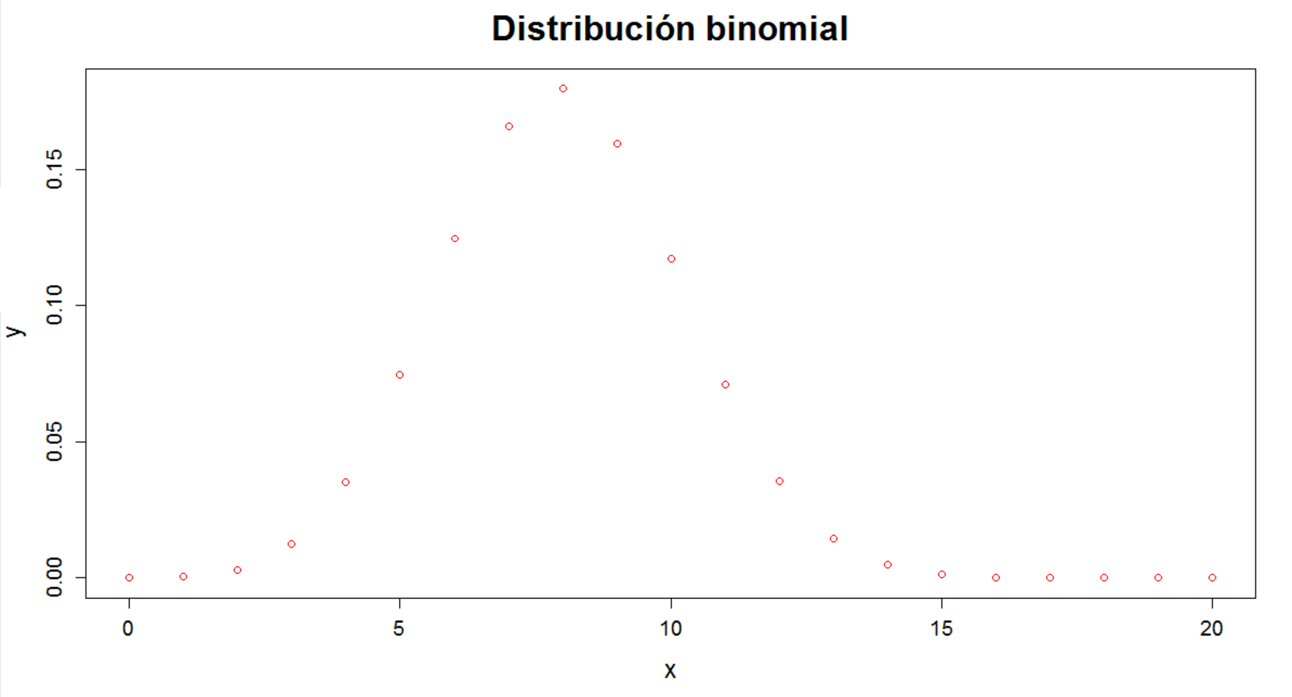
24

Distribución binomial

- Dibujemos ahora la función de masa de la distribución binomial que en $n = 20$ intentos tiene una probabilidad de éxito $p = 0{,}4$.

```
> x <- 0:20
> y <- dbinom(x,20,0.4)
> plot(x,y,type="p",col="red",
  main="Distribución binomial",ylab="y",
  cex.axis=1.2,cex.lab=1.5,cex.main=2)
```

Distribución binomial

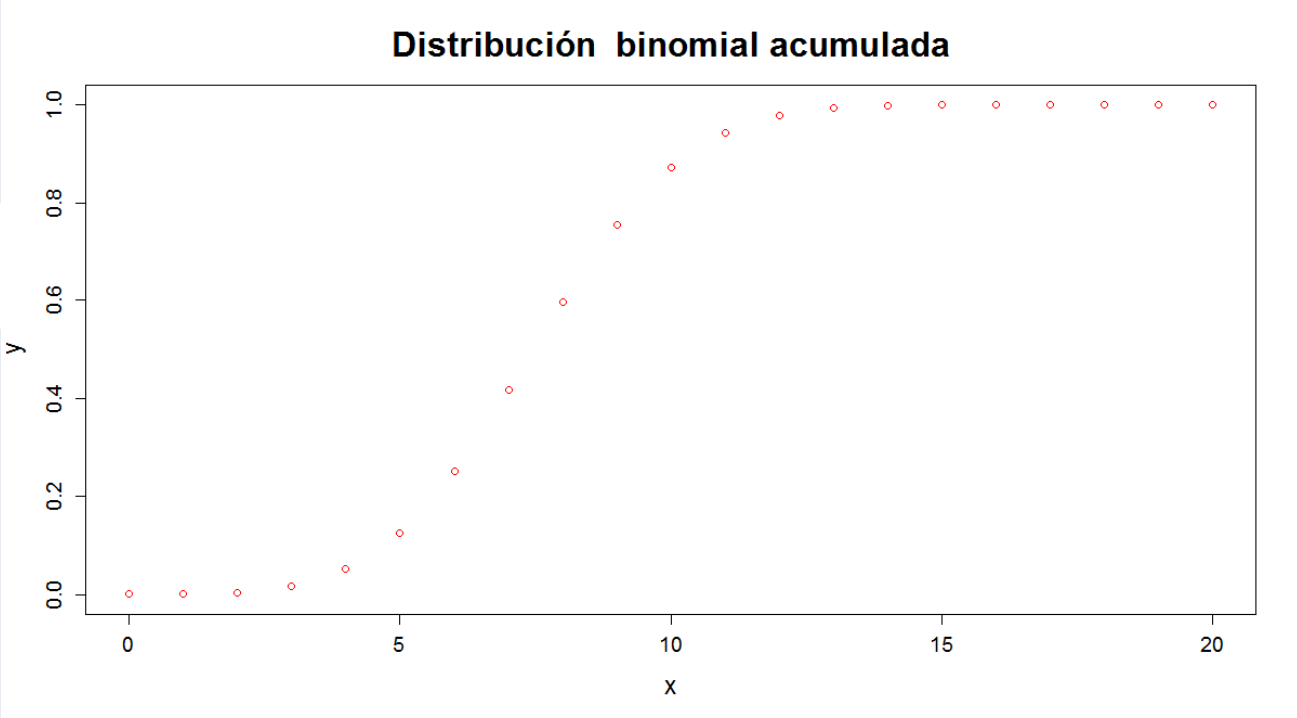


Distribución binomial

- La función de distribución de una binomial que en $n = 20$ intentos tiene una probabilidad de éxito $p = 0{,}4$ nos dice la probabilidad de tener x éxitos o menos.

```
> x <- 0:20
> y <- pbinom(x,20,0.4)
> plot(x,y,type="p",col="red",
  main="Distribución binomial acumulada",
  ylab="y",cex.axis=1.2,cex.lab=1.5,
  cex.main=2)
```

Distribución binomial

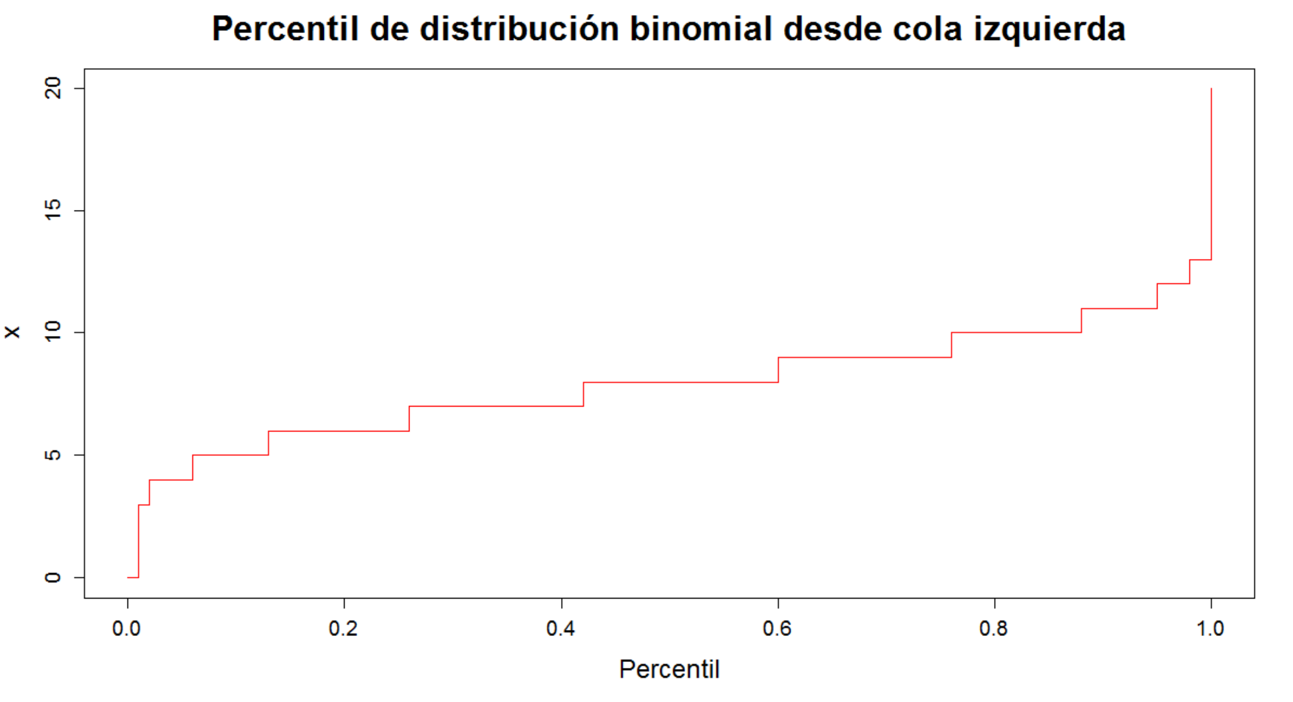


Distribución binomial

- Encontremos el valor de x que corresponde a diversos percentiles para una distribución binomial que en $n = 20$ intentos tiene una probabilidad de éxito $p = 0{,}4$.

```
> x <- seq(0,1,0.01)
> y <- qbinom(x,20,0.4)
> plot(x,y,type="s", col="red",
  main="Percentil de distribución
  binomial desde cola izquierda",
  ylab="x",cex.axis=1.2,cex.lab=1.5,
  cex.main=2,xlab="Percentil")
```

Distribución binomial



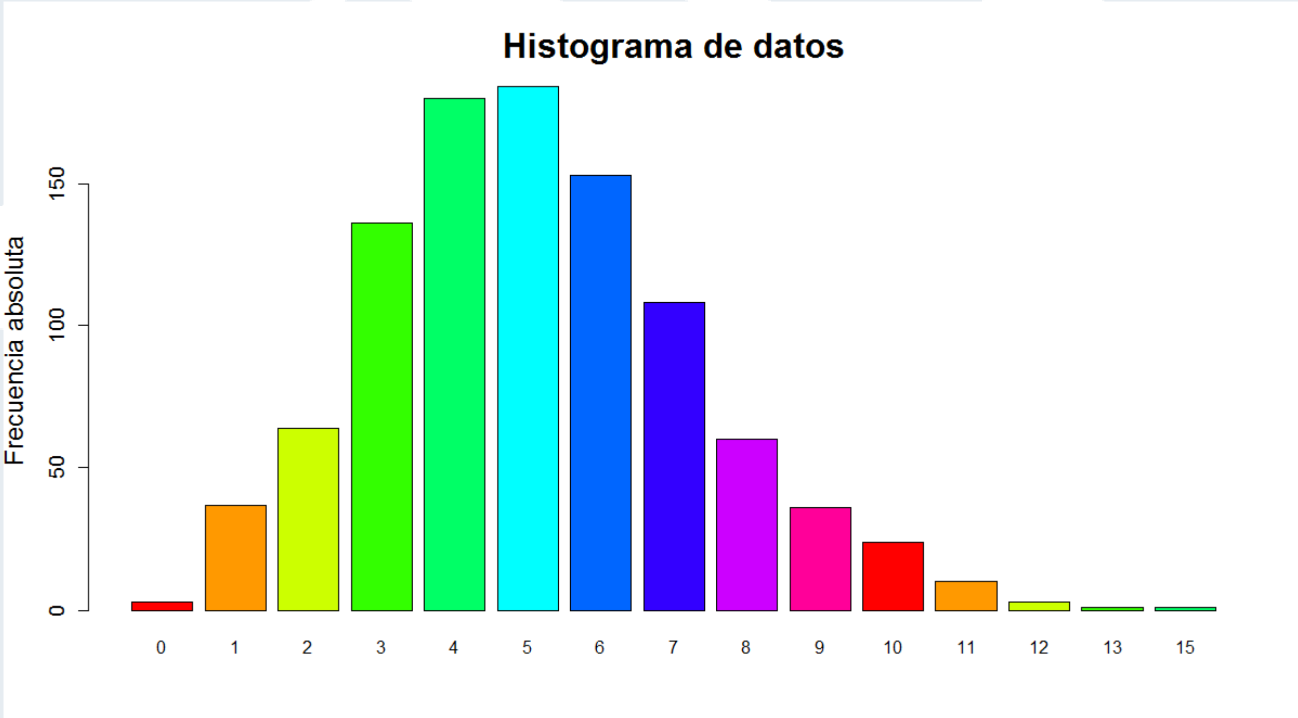
Distribución de Poisson

Distribución de Poisson

- Calculemos 1000 números aleatorios según la distribución de Poisson con un parámetro $\lambda = 5$.

```
> datos <- rpois(1000,5)
> head(datos)
[1] 5 9 6 2 4 5
> barplot(table(datos),col=rainbow(10),
  ylab="Frecuencia absoluta",
  main="Histograma de datos",
  cex.axis=1.2,cex.lab=1.5,cex.main=2)
```

Distribución de Poisson



Distribución de Poisson

```
> mean(datos) # El valor teórico es λ=5
[1] 5.105
```

$$\mathrm{E}(X) = \mu = \lambda$$

```
> var(datos) # El valor teórico es λ=5
[1] 4.932908
```

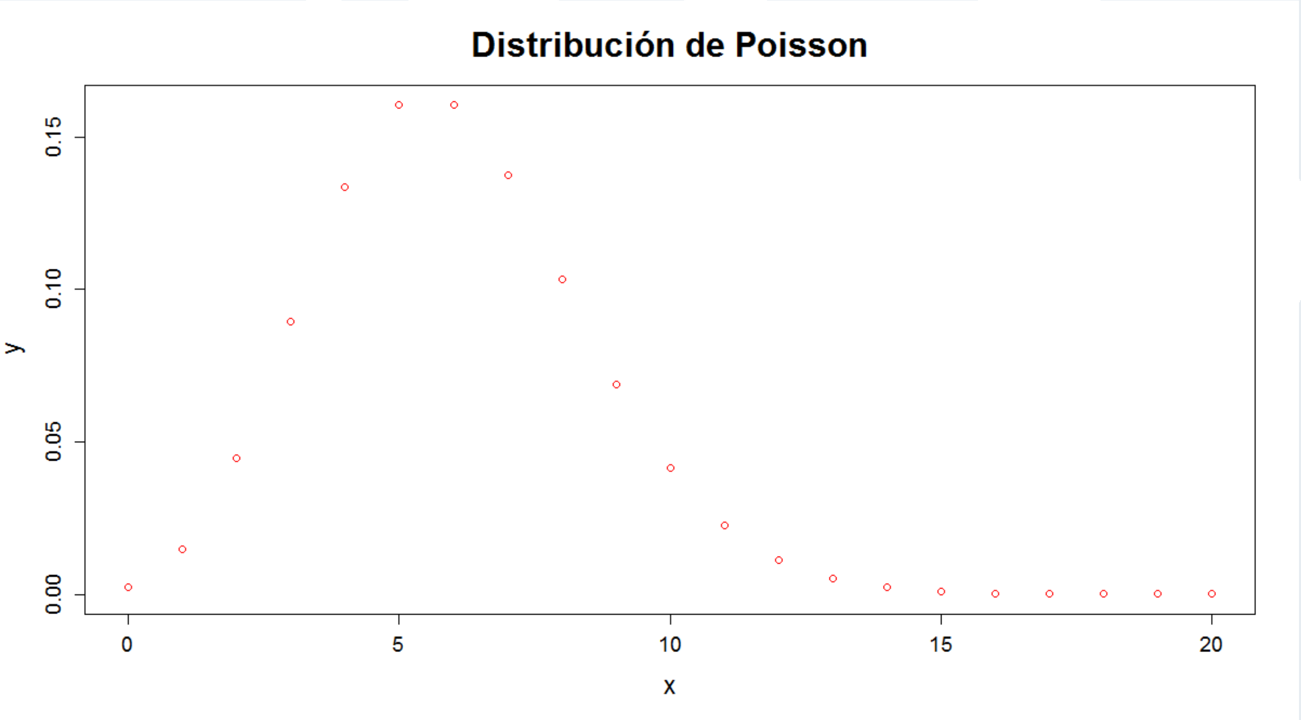
$$\mathrm{Var}(X) = \sigma^2 = \lambda$$

Distribución de Poisson

- Dibujemos la función de masa de una distribución de Poisson para un parámetro $\lambda = 6$ entre los valores 0 y 20.

```
> x <- 0:20
> y <- dpois(x,6)
> plot(x,y,type="p",col="red",
  main="Distribución de Poisson",
  ylab="y",cex.axis=1.2,cex.lab=1.5,
  cex.main=2)
```

Distribución de Poisson

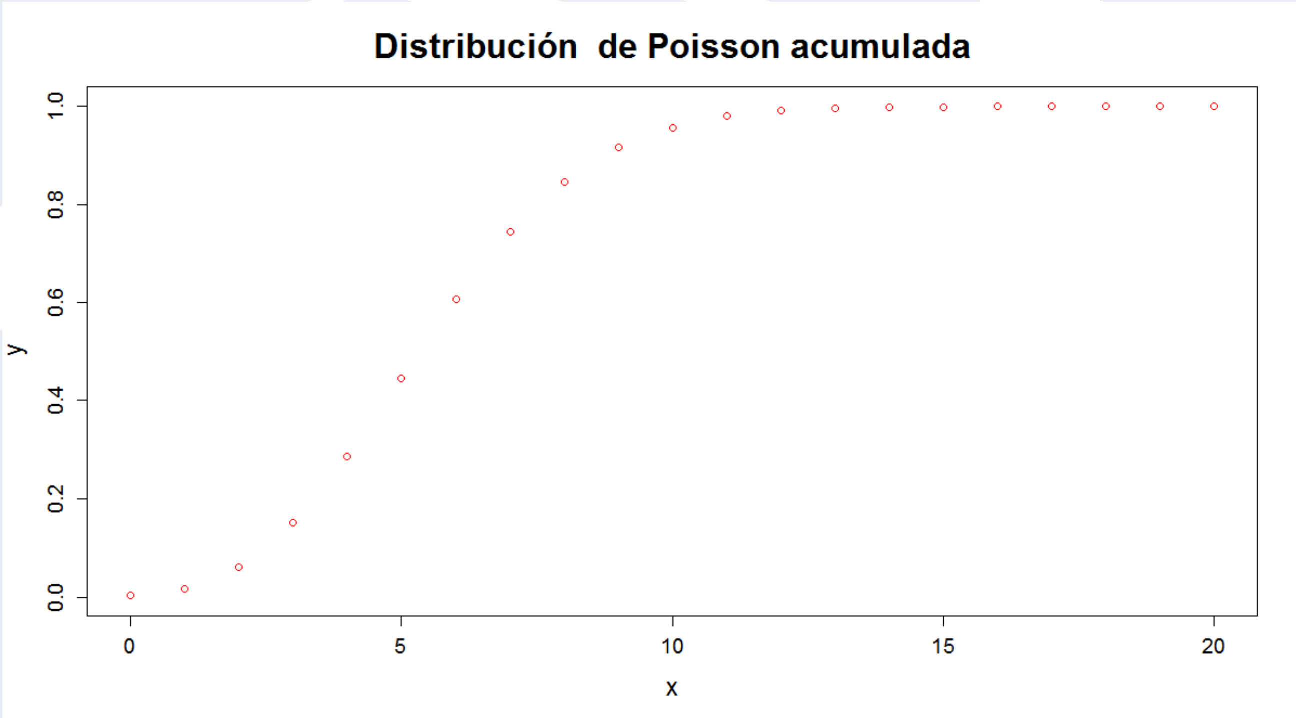


Distribución de Poisson

- Ahora la función de distribución de una distribución de Poisson para un parámetro $\lambda = 6$ entre los valores 0 y 20 sería la siguiente.

```
> x <- 0:20
> y <- ppois(x,6)
> plot(x,y,type="p", col="red",
  main="Distribución de Poisson
  acumulada",ylab="y",cex.axis=1.2,
  cex.lab=1.5,cex.main=2)
```

Distribución de Poisson

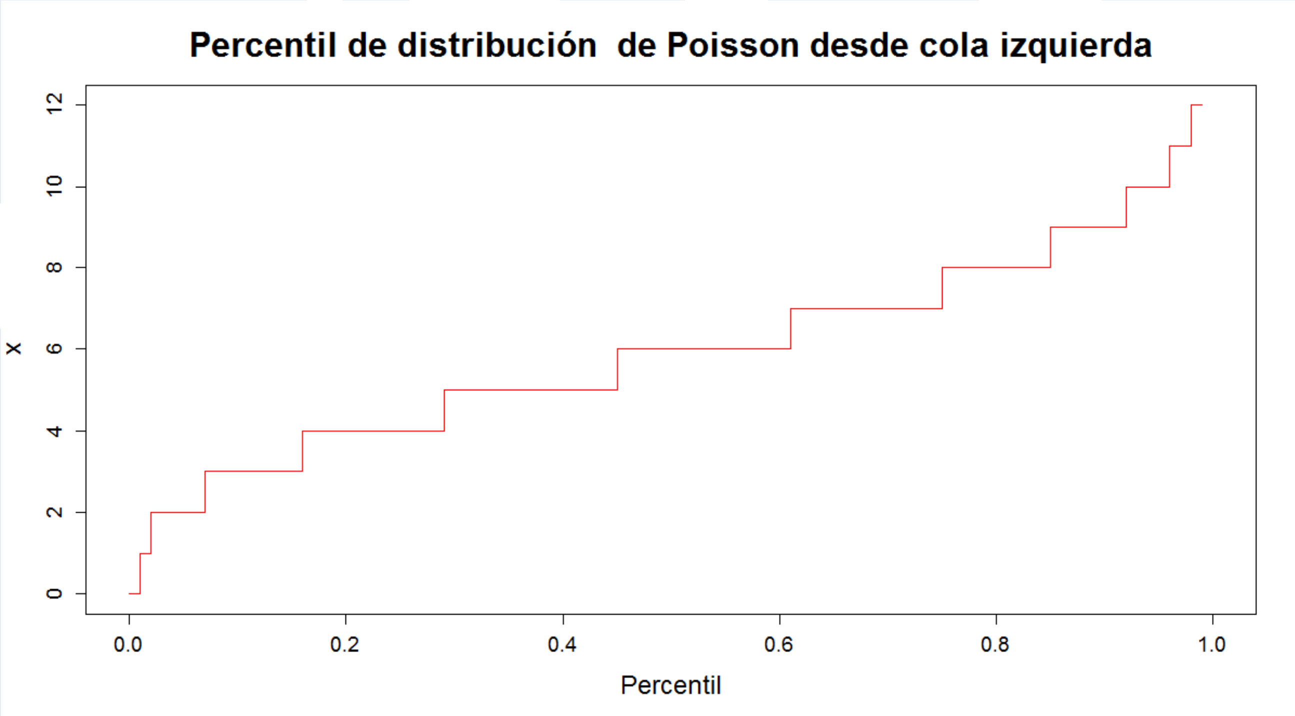


Distribución de Poisson

- Encontremos el valor de *x* que corresponde a diversos percentiles para una distribución de Poisson para un parámetro $\lambda = 6$.

```
> x <- seq(0,1,0.01)
> y <- qpois(x,6)
> plot(x,y,type="s", col="red",
  main="Percentil de distribución de
  Poisson desde cola izquierda",
  ylab="x",cex.axis=1.2,cex.lab=1.5,
  cex.main=2,xlab="Percentil")
```

Distribución de Poisson



40

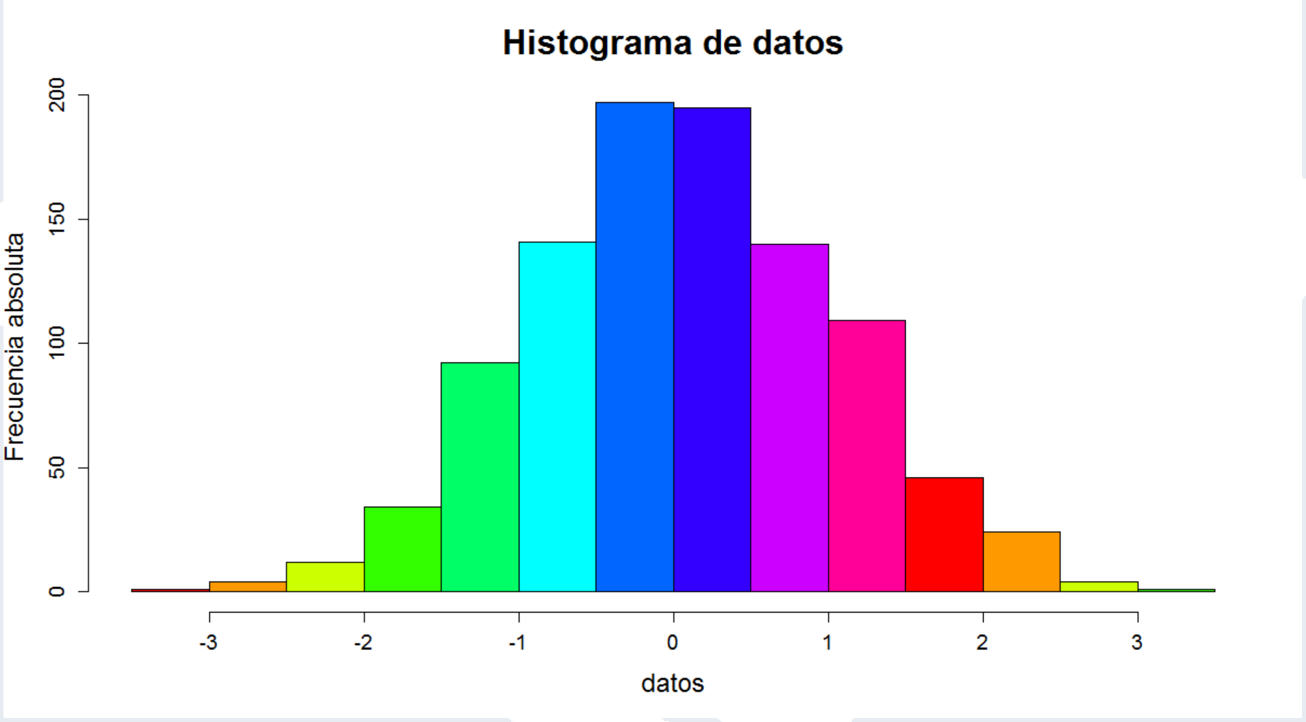
Distribución Normal

Distribución Normal

- Generaremos 1000 números aleatorios reales con media $\mu = 0$ y desviación típica $\sigma = 1$.

```
> datos <- rnorm(1000,0,1)
> head(datos)
[1] 1.4110755 -0.1783396 0.3035531 0.9312146
  -0.4680443 -0.2545553
> hist(datos,col=rainbow(10),
  ylab="Frecuencia absoluta",
  main="Histograma de datos",cex.axis=1.2,
  cex.lab=1.5,cex.main=2)
```

Distribución Normal



Distribución Normal

```
> mean(datos) # El valor teórico es 0
[1] 0.05477506
```

$$E(X) = \mu$$

```
> var(datos) # El valor teórico es 1
[1] 1.037411
```

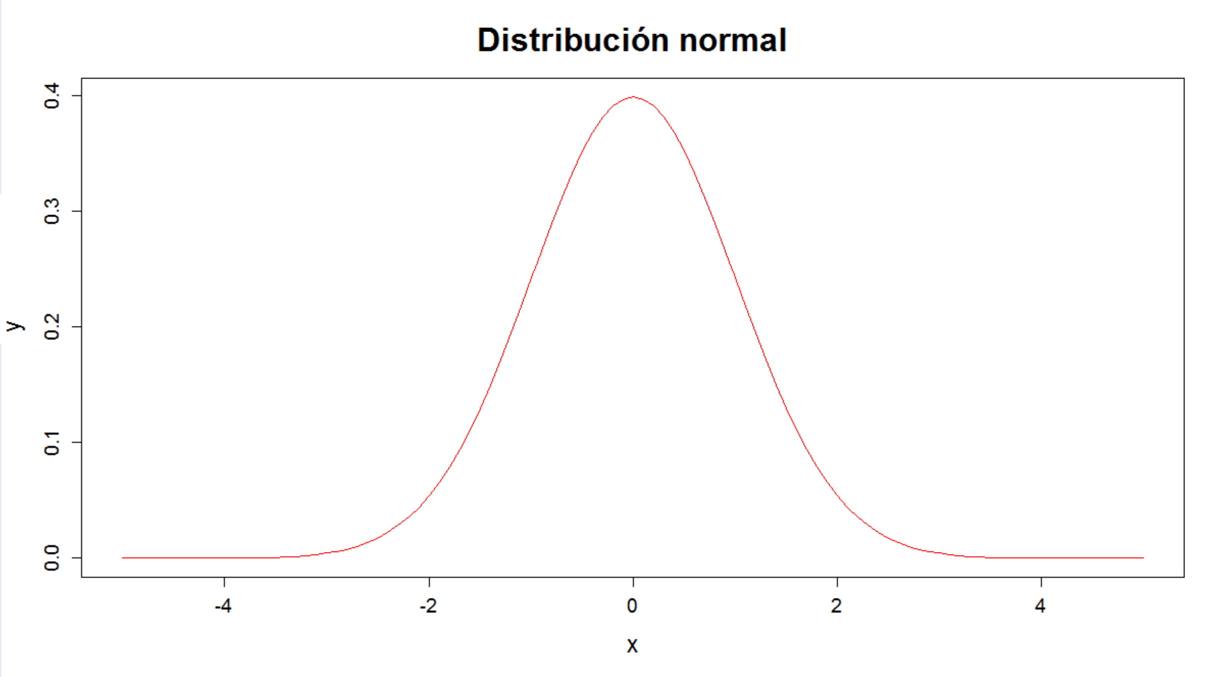
$$\mathrm{Var}(X) = \sigma^2$$

Distribución Normal

- Dibujemos la función densidad de probabilidad de una distribución Normal con media $\mu = 0$ y desviación típica $\sigma = 1$ entre los valores -5 y 5.

```
> x <- seq(-5,5,0.1)
> y <- dnorm(x,0,1) # Ídem dnorm(x)
> plot(x,y,type="l",col="red",
  main="Distribución normal",ylab="y",
  cex.axis=1.2,cex.lab=1.5,cex.main=2)
```

Distribución Normal



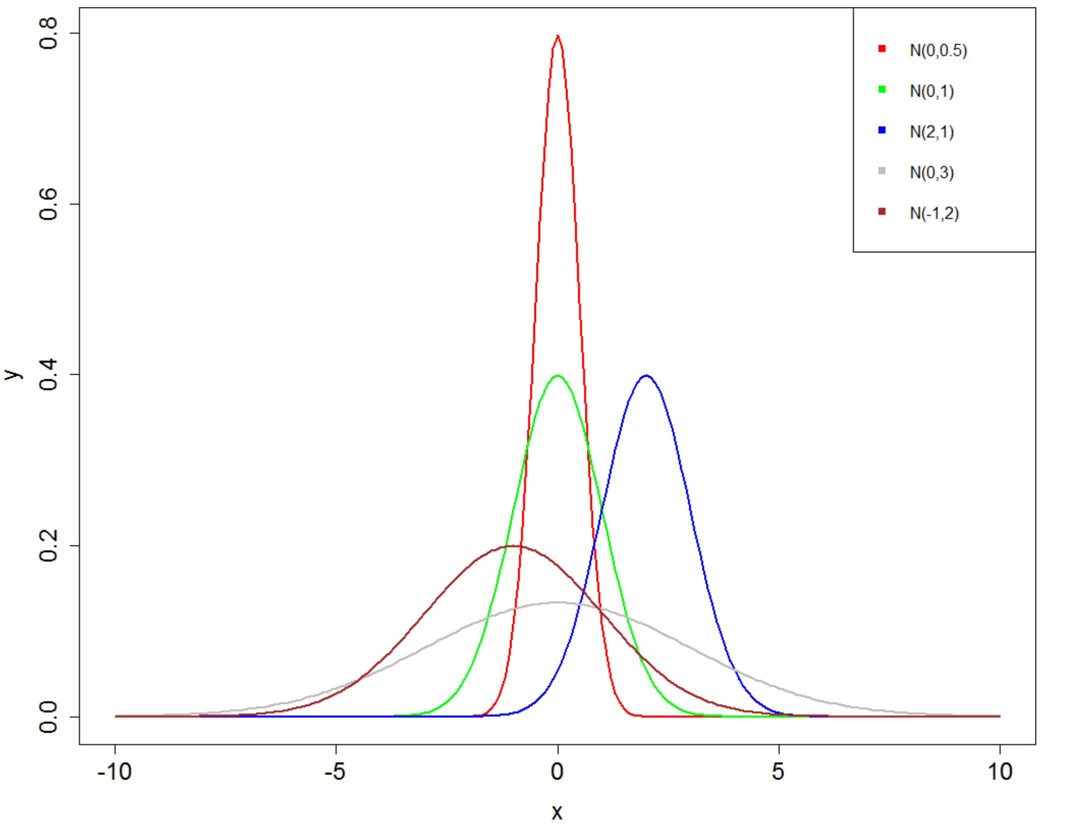
$$f(x) \equiv N(\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Distribución Normal

- Comparemos curvas normales con diferentes parámetros:

```
> x <- seq(-10,10,0.1)
> y1 <- dnorm(x,0,0.5) # N(0,0.5)
> y2 <- dnorm(x,0,1)   # N(0,1)
> y3 <- dnorm(x,2,1)   # N(2,1)
> y4 <- dnorm(x,0,3)   # N(0,3)
> y5 <- dnorm(x,-1,2)  # N(-1,2)
> plot(x,y1,type="l",col="red", xlab="x",ylab="y",cex.main=2,
  cex.axis=1.5,cex.lab=1.5,lwd=2)
> lines(x,y2,type="l",col="green",lwd=2)
> lines(x,y3,type="l",col="blue",lwd=2)
> lines(x,y4,type="l",col="grey",lwd=2)
> lines(x,y5,type="l",col="brown",lwd=2)
> legend("topright",legend=c("N(0,0.5)","N(0,1)","N(2,1)",
  "N(0,3)","N(-1,2)"),col=c("red","green","blue","grey",
  "brown"),pch=15)
```

Distribución Normal



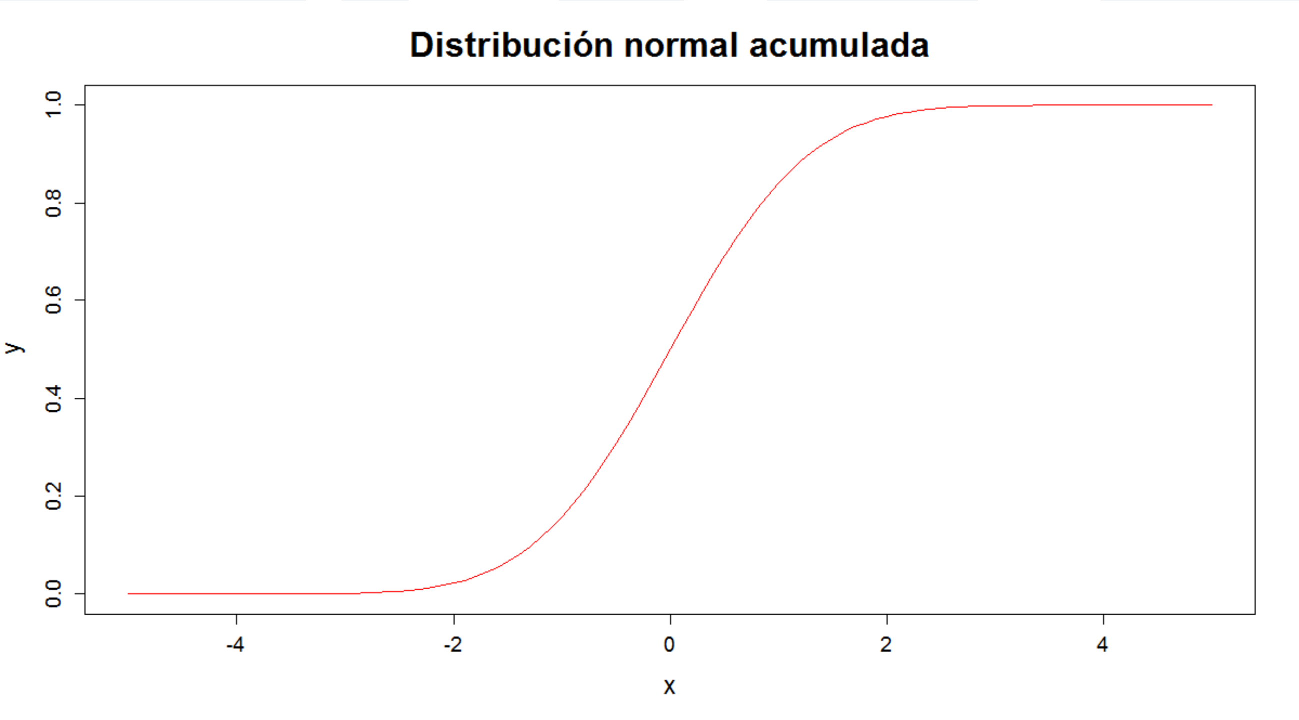
48

Distribución Normal

- La función de distribución de una Distribución Normal con media $\mu = 0$ y desviación típica $\sigma = 1$ entre los valores -5 y 5.

```
> x <- seq(-5,5,.1)
> y <- pnorm(x,0,1) # Ídem pnorm(x)
> plot(x,y,type="l",col="red",
  main="Distribución normal acumulada",
  ylab="y",cex.axis=1.2,cex.lab=1.5,
  cex.main=2)
```

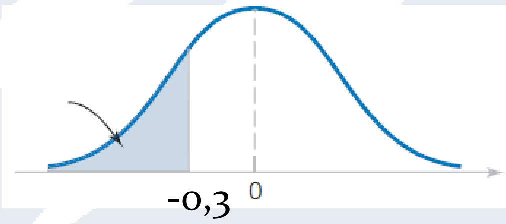
Distribución Normal



Distribución Normal

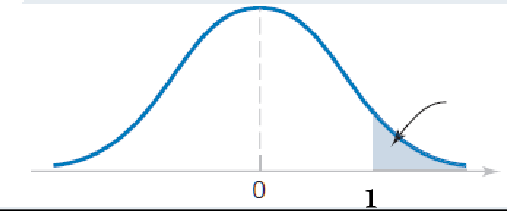
- Función de distribución (área bajo la cola izquierda de la curva normal de media $\mu = 0$ y desviación típica $\sigma = 1$ hasta el punto $x = -0{,}3$:

```
> pnorm(-0.3,0,1) # P(X<-0.3) si X ~ N(0,1)
  = Área bajo la cola izquierda
[1] 0.3820886
```



- Bajo la cola derecha desde $x = 1$:

```
> pnorm(1,0,1,lower.tail=FALSE) # P(X>1) si
  X ~ N(0,1) = Área bajo cola derecha
[1] 0.1586553
```



Distribución Normal

- Recordemos que el valor pnorm suele venir tabulado en los libros de Estadística:

```
> pnorm(-3.53)
[1] 0.0002077798
```

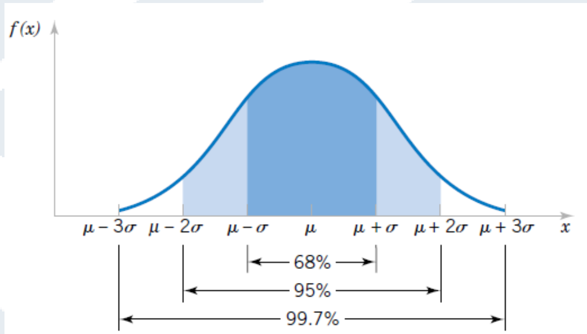
Función de distribución para la distribución normal tipificada

z	−0.09	−0.08	−0.07	−0.06	−0.05	−0.04	−0.03	−0.02	−0.01	−0.00
−3.9	0.000033	0.000034	0.000036	0.000037	0.000039	0.000041	0.000042	0.000044	0.000046	0.000048
−3.8	0.000050	0.000052	0.000054	0.000057	0.000059	0.000062	0.000064	0.000067	0.000069	0.000072
−3.7	0.000075	0.000078	0.000082	0.000085	0.000088	0.000092	0.000096	0.000100	0.000104	0.000108
−3.6	0.000112	0.000117	0.000121	0.000126	0.000131	0.000136	0.000142	0.000147	0.000153	0.000159
−3.5	0.000165	0.000172	0.000179	0.000185	0.000193	0.000200	0.000208	0.000216	0.000224	0.000233
−3.4	0.000242	0.000251	0.000260	0.000270	0.000280	0.000291	0.000302	0.000313	0.000325	0.000337
−3.3	0.000350	0.000362	0.000376	0.000390	0.000404	0.000419	0.000434	0.000450	0.000467	0.000483
−3.2	0.000501	0.000519	0.000538	0.000557	0.000577	0.000598	0.000619	0.000641	0.000664	0.000687
−3.1	0.000711	0.000736	0.000762	0.000789	0.000816	0.000845	0.000874	0.000904	0.000935	0.000968

Distribución Normal

- Es interesante conocer el área bajo la curva normal de media $\mu = 0$ y desviación típica $\sigma = 1$ alrededor del valor de la media y limitada a ambos lados por múltiplos de σ.

```
> pnorm(1)-pnorm(-1)
[1] 0.6826895
> pnorm(2)-pnorm(-2)
[1] 0.9544997
> pnorm(3)-pnorm(-3)
[1] 0.9973002
```

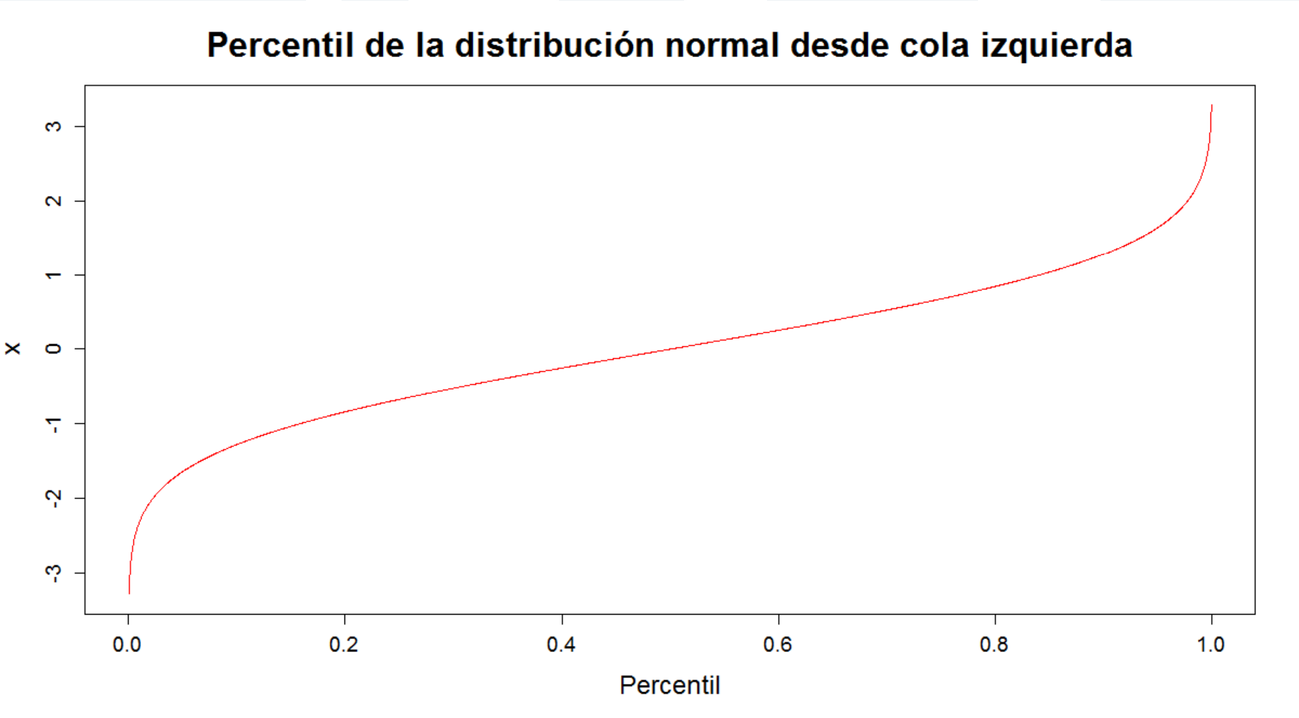


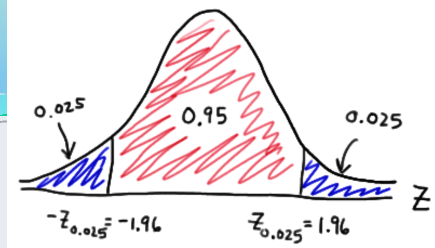
Distribución Normal

- Para encontrar el valor de x para el cual el área bajo la cola izquierda de la curva normal de media $\mu = 0$ y desviación típica $\sigma = 1$ corresponde a los diversos percentiles.

```
> x <- seq(0,1,0.0005)
> y <- qnorm(x,0,1) # Ídem qnorm(x)
> plot(x,y,type="l",col="red",
  main="Percentil de la distribución normal
  desde cola izquierda",ylab="x",
  cex.axis=1.2,cex.lab=1.5,cex.main=2,
  xlab="Percentil")
```

Distribución Normal





Distribución Normal

- Supongamos que medimos el área α bajo las dos colas simétricas de la curva normal de media $\mu = 0$ y desviación típica $\sigma = 1$.
- Tienen de especial relevancia para temas posteriores los valores de x correspondientes a unas áreas α bajo la curva normal estándar del:
  - El 0,5 % bajo la cola izquierda.
  - El 2,5 % bajo la cola izquierda.
  - El 97,5 % bajo la cola izquierda = 2,5 % bajo la cola derecha.
  - El 99,5 % bajo la cola izquierda = 0,5 % bajo la cola derecha.

```
> alpha <- c(0.005,0.025,0.975,0.995)
> qnorm(alpha)
[1] -2.575829 -1.959964 1.959964 2.575829
```

Aproximación Normal a la Binomial

- Probabilidad de que una variable binomial tome un valor entre 40 y 60, ambos inclusive, en n = 100 intentos, con probabilidad de éxito p = 0,5 (np>>5).

```
> pbinom(60,100,0.5) - pbinom(39,100,0.5)
[1] 0.9647998
```

- Ídem con la aproximación Normal a la Binomial (con corrección de continuidad).

```
> mu <- 100*0.5 # n*p
> sigma <- sqrt(100*0.5*(1-0.5)) # sqrt(n*p*q)
> pnorm(60.5,mu,sigma) - pnorm(39.5,mu,sigma)
[1] 0.9642712
```

Aproximación Normal a Poisson

- Probabilidad de que una variable de Poisson con parámetro $\lambda = 20$ (>5) tome un valor entre 9 y 15, ambos inclusive.

```
> ppois(15,20) - ppois(8,20)
[1] 0.1544259
```

- Ídem con la aproximación Normal a Poisson (con corrección de continuidad).

```
> mu <- 20 # lambda
> sigma <- sqrt(20) # sqrt(lambda)
> pnorm(15.5,mu,sigma) - pnorm(8.5,mu,sigma)
[1] 0.1520891
```

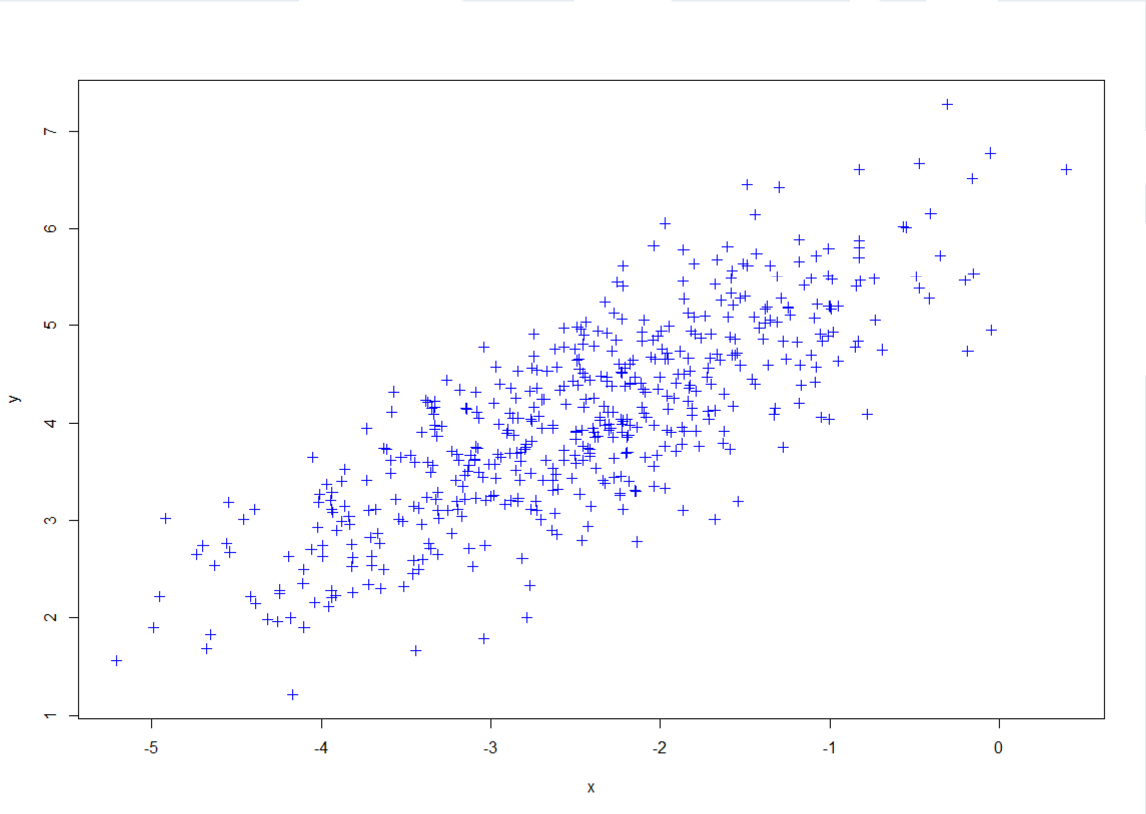
Distribución Normal 2D

- Vamos a generar una serie de 500 números aleatorios según una Distribución Normal Bivariante.

$$f_n(x_1,\ldots,x_n)=\frac{1}{\sqrt{(2\pi)^n|\Sigma|}}\exp\left[-\frac{1}{2}(\vec{x}-\vec{\mu})^T\Sigma^{-1}(\vec{x}-\vec{\mu})\right] \qquad \vec{x}=\begin{pmatrix}x\\y\end{pmatrix},\ \vec{\mu}=\begin{pmatrix}-2,5\\4\end{pmatrix},\ \Sigma=\begin{bmatrix}1 & 0,8\\0,8 & 1\end{bmatrix}$$

```
> install.packages("mvtnorm")
> require("mvtnorm")
> mu <- c(-2.5,4)
> sigma <- matrix(c(1,0.8,0.8,1),nrow=2)
> datos <- rmvnorm(500, mu, sigma)
> plot(datos,pch=3,xlab="x",ylab="y",col="blue")
```

Distribución Normal 2D



Distribución Normal 2D

- Para evaluar la distribución evaluada en un punto $p = (x,y)$ se usa la función dmvnorm:

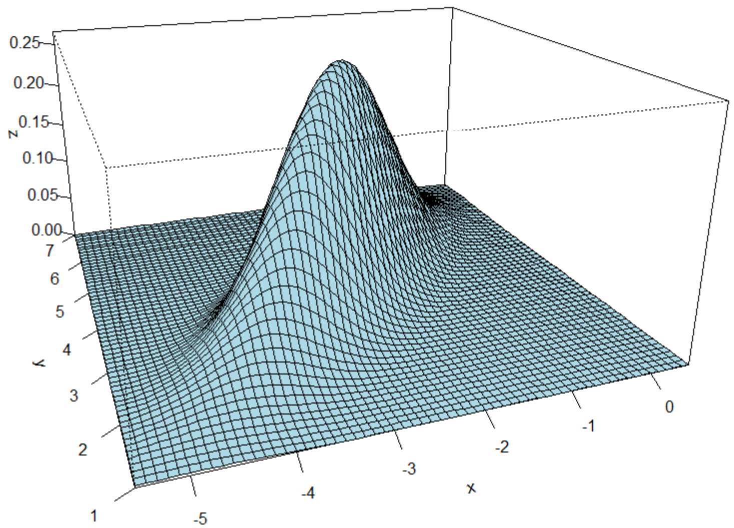
```
> p <- c(-2.5,4) # Punto (x,y)
> dmvnorm(p,mu,sigma)
[1] 0.2652582
> p <- c(-2,5)
> dmvnorm(p,mu,sigma)
[1] 0.1419825
> p <- c(0,2)
> dmvnorm(p,mu,sigma)
[1] 2.603216e-12
```

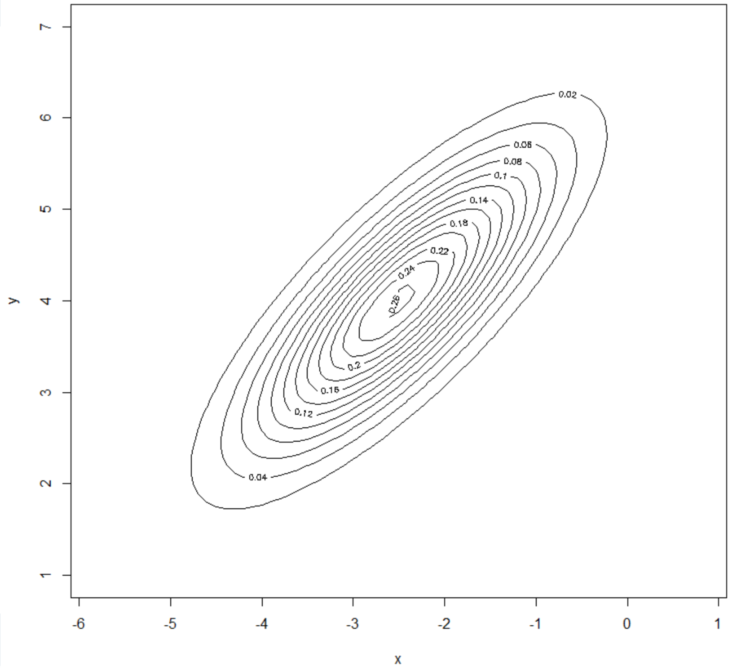
Distribución Normal 2D

- Dibujemos ahora la función de densidad conjunta en la Distribución Normal Bivariante, primero como una gráfica 3D (persp) y luego como un diagrama de contornos de nivel constante (contour):

```
> x <- seq(-5.5,0.5,0.1)
> y <- seq(1,7,0.1)
> f <- function(x,y) { dmvnorm(c(x,y),
  mean=cbind(-2.5,4),sigma=matrix(c(1,0.8,0.8,1),
  nrow=2))}
> z <- outer(x,y,f)
> persp(x, y, z, theta=-20, phi=20, expand=0.5,
  col="lightblue", ticktype="detailed")
> contour(x, y, z, xlab="x", ylab="y", asp=1)
```

Distribución Normal 2D





63

Distribución Normal 2D

- La función de distribución (=volumen debajo de la superficie de la gaussiana 2D) se calcula con pmvnorm:

```
> limiteInferior <- c(-2.5,4)  # Límites inferiores de integral
> limiteSuperior <- c(Inf,Inf) # Límites superiores de integral
> mu
[1] -2.5 4.0
> sigma
     [,1] [,2]
[1,] 1.0  0.8
[2,] 0.8  1.0
> pmvnorm(lower=limiteInferior,upper=limiteSuperior,
  mean=mu,sigma)
[1] 0.3975836
attr(,"error")
[1] 1e-15
attr(,"msg")
[1] "Normal Completion"
```

$$P(X > -2{,}5; Y > 4) = \int_{-2{,}5}^{\infty} \int_{4}^{\infty} f(x, y, \vec{\mu}, \Sigma) dxdy = 0{,}3975836$$

64

Distribución χ^2

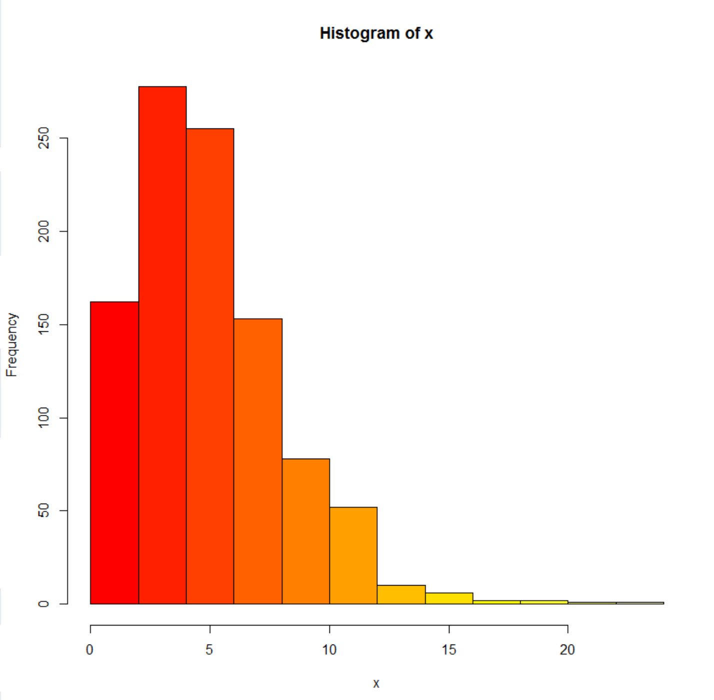
Distribución χ^2

- Generación de números aleatorios según la distribución χ^2 con k = 5 grados de libertad:

```
> x <- rchisq(1000,5)
> hist(x,col=heat.colors(12))
> mean(x)
[1] 4.967784
> var(x)
[1] 9.593942
```

Los valores teóricos del valor esperado y la varianza son:

$$\mathrm{E}(X) = k \qquad \mathrm{Var}(X) = \sigma^2 = 2k$$

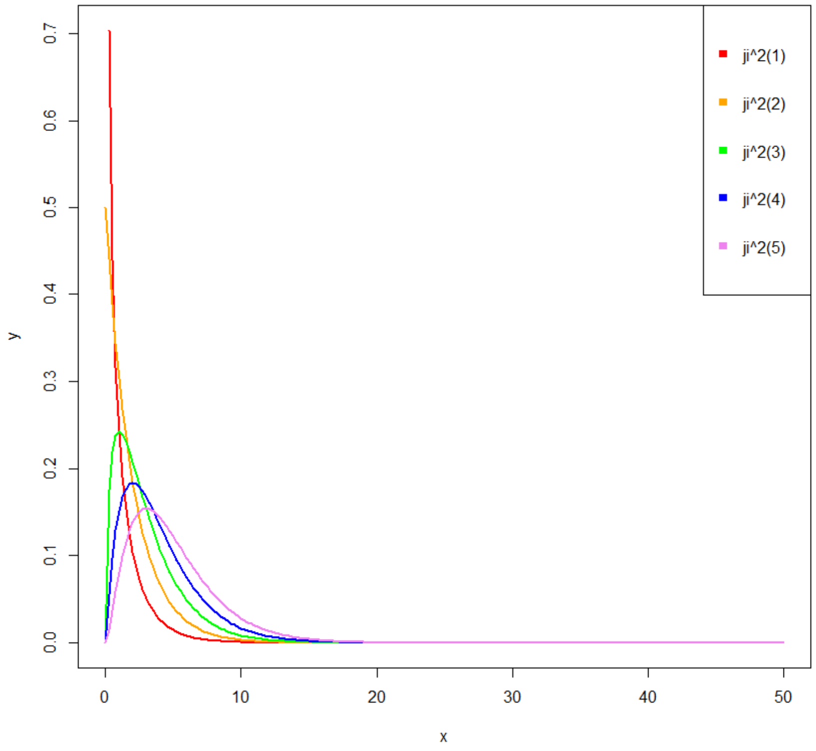


Distribución χ^2

- La distribución χ^2 para un valor *x* y un cierto número *k* de grados de libertad se calcula con dchisq.

```
> x <- seq(0,50,.25)
> y1 <- dchisq(x,1)  # ji^2(1)
> y2 <- dchisq(x,2)  # ji^2(2)
> y3 <- dchisq(x,3)  # ji^2(3)
> y4 <- dchisq(x,4)  # ji^2(4)
> y5 <- dchisq(x,5)  # ji^2(5)
> plot(x,y1,col="red",type="l",ylab="y",lwd=2)
> lines(x,y2,col="orange",type="l",ylab="y",lwd=2)
> lines(x,y3,col="green",type="l",ylab="y",lwd=2)
> lines(x,y4,col="blue",type="l",ylab="y",lwd=2)
> lines(x,y5,col="violet",type="l",ylab="y",lwd=2)
> legend("topright",legend=c("ji^2(1)","ji^2(2)","ji^2(3)",
  "ji^2(4)","ji^2(5)"),col=c("red","orange","green","blue",
  "violet"),pch=15)
```

Distribución χ^2



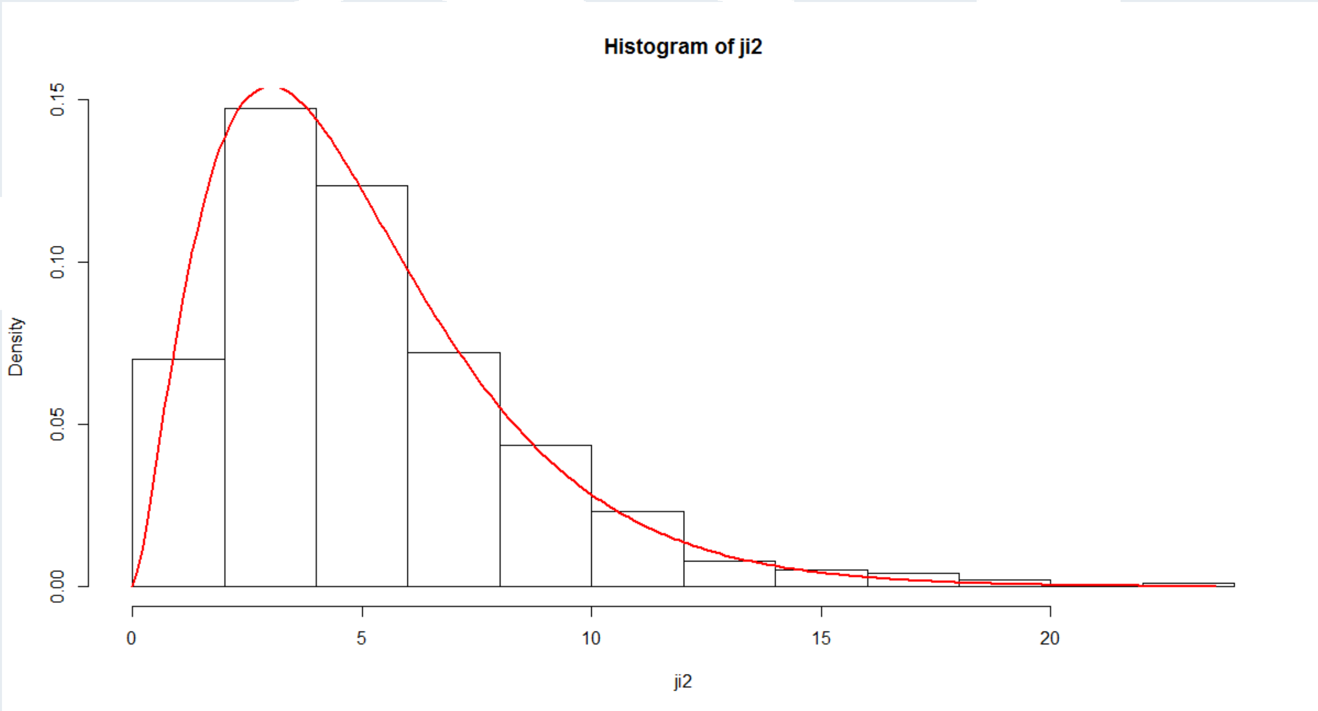
Distribución χ^2

- Recordemos que la suma de los cuadrados de *k* variables aleatorias normales tipificadas es una nueva variable que sigue una distribución χ^2 con *k* grados de libertad.

```
> z1 <- rnorm(1000)
> z2 <- rnorm(1000)
> z3 <- rnorm(1000)
> z4 <- rnorm(1000)
> z5 <- rnorm(1000)
> ji2 <- z1^2 + z2^2 + z3^2 + z4^2 + z5^2
> hist(ji2,freq=FALSE) # Dibujamos el histograma de área total 1
> x <- seq(0,max(ji2),0.1)
> y <- dchisq(x,5)
> lines(x,y,col="red",lwd=2) # Superponemos una χ² teórica con 5
  grados de libertad
```

69

Distribución χ^2



Distribución χ^2

- La función distribución (área debajo de la curva) se calcula con pchisq. Utilizando de nuevo la χ^2 con 5 grados de libertad:

```
> pchisq(3,5)  # P(X<3) si X ~ χ5² = Área bajo la cola izquierda
[1] 0.3000142
> pchisq(8,5,lower.tail=FALSE)  # P(X>8) si X ~ χ5² = Área bajo la cola
  derecha
[1] 0.1562356
```

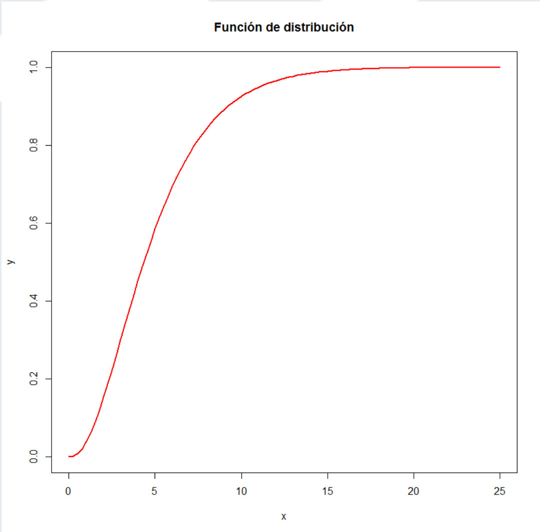
- La función cuantil es qchisq:

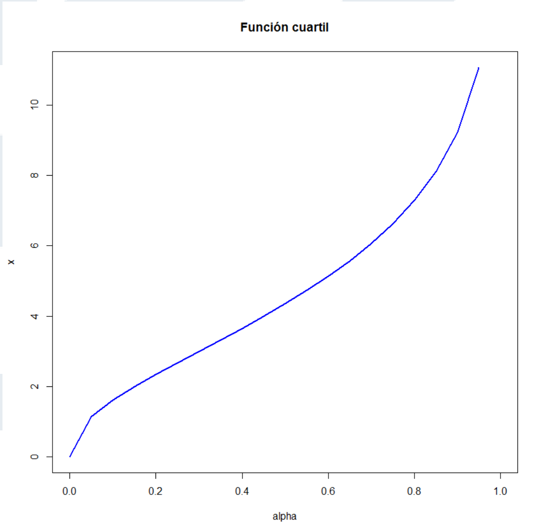
```
> qchisq(0.5,5) # Buscamos el valor abarca el 50% del área bajo la
  cola izquierda de la curva con 5 grados de libertad
[1] 4.35146
```

71

Distribución χ^2

```
> x <- seq(0,25,0.25)
> y <- pchisq(x,5)
> plot(x,y,col="red",lwd=2,type="l",main="Función de distribución")
> alpha <- seq(0,1,.05)
> x2 <- qchisq(alpha,5)
> plot(alpha,x2,col="blue",lwd=2,type="l",main="Función cuantil")
```

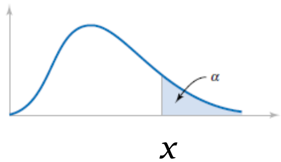




Distribución χ^2

- Recordemos que la función cuantil viene tabulada en los libros de Estadística:

```
> qchisq(0.05,8,lower.tail=FALSE)
[1] 15.50731
```



Función cuartil medida en la cola derecha para la distribución χ^2 según el área α y los grados de libertad k

k \ α	.995	.990	.975	.950	.900	.500	.100	.050	.025	.010	.005
1	.00+	.00+	.00+	.00+	.02	.45	2.71	3.84	5.02	6.63	7.88
2	.01	.02	.05	.10	.21	1.39	4.61	5.99	7.38	9.21	10.60
3	.07	.11	.22	.35	.58	2.37	6.25	7.81	9.35	11.34	12.84
4	.21	.30	.48	.71	1.06	3.36	7.78	9.49	11.14	13.28	14.86
5	.41	.55	.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	.68	.87	1.24	1.64	2.20	5.35	10.65	12.59	14.45	16.81	18.55
7	.99	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	13.36	$\boxed{15.51}$	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19

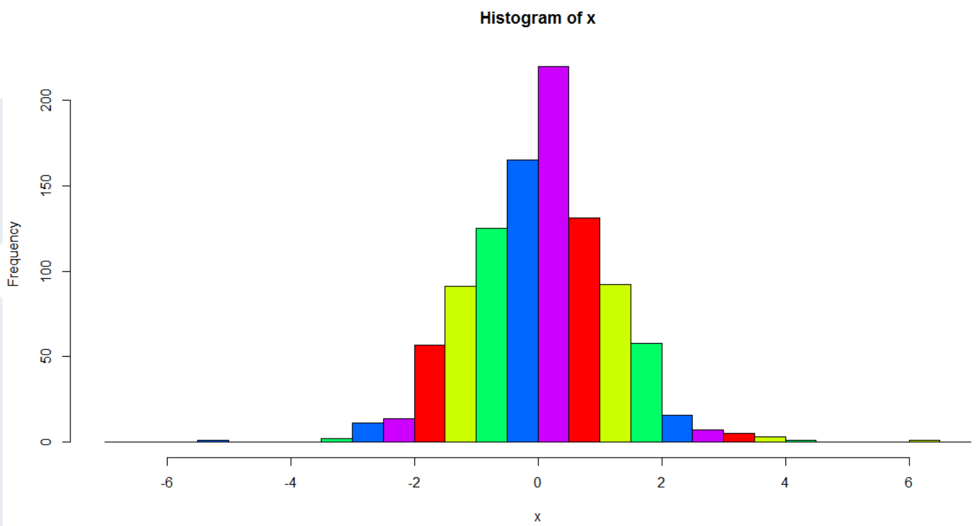
73

Distribución t de Student

Distribución t de Student

- Generación de 1000 números aleatorios según una distribución t de Student con k = 8 grados de libertad:

```
> x <- rt(1000,8)
> hist(x,breaks=seq(-7,7,0.5),col=rainbow(5))
> mean(x)
[1] 0.0398955
> var(x)
[1] 1.270159
```



Valores téoricos:

$\mathrm{E}(X) = 0$

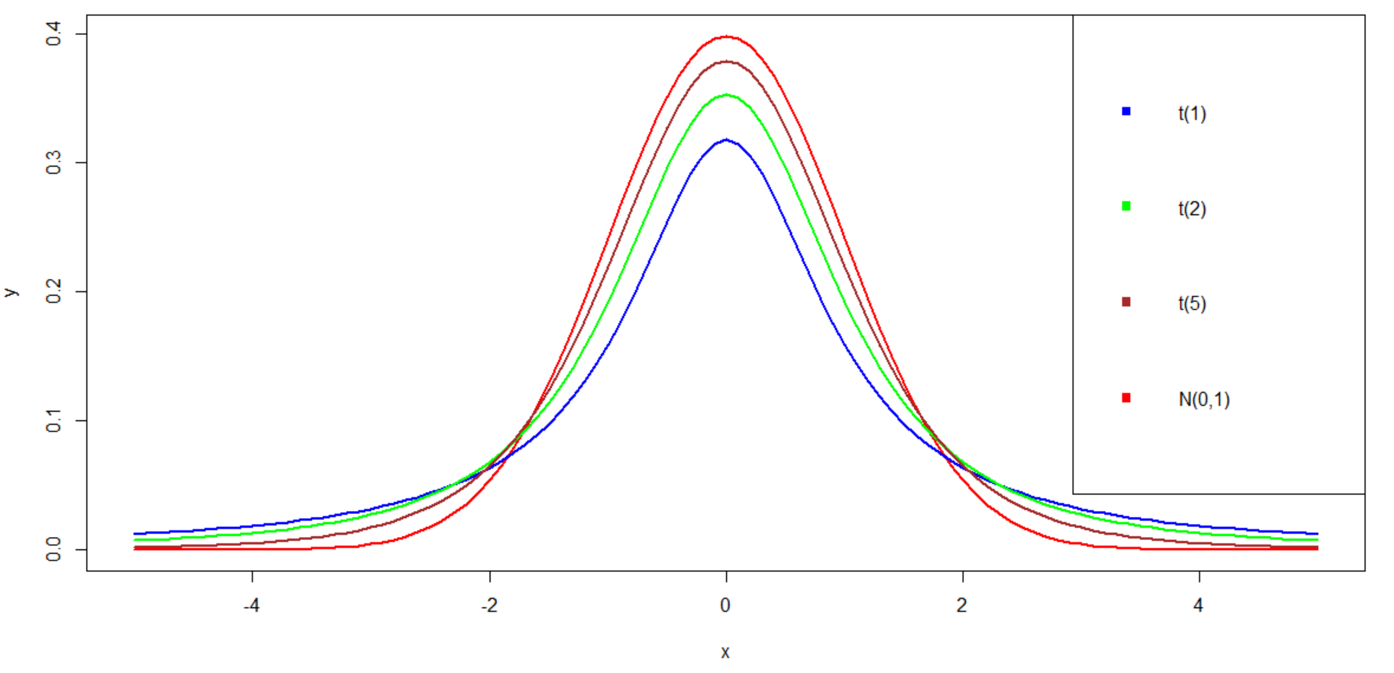
$\mathrm{Var}(X) = k/(k-2)$

Distribución t de Student

- La distribución t de Student es más baja que la Normal pero tiende a cero más lentamente.
- Además tiende a la Normal a medida que los grados de libertad aumentan.

```
> x <- seq(-5,5,0.1)
> t1 <- dt(x,1)
> t2 <- dt(x,2)
> t5 <- dt(x,5)
> z <- dnorm(x)
> plot(x,z,type="l",ylab="y",col="red",lwd=2)
> lines(x,t1,type="l",col="blue",lwd=2)
> lines(x,t2,type="l",col="green",lwd=2)
> lines(x,t5,type="l",col="brown",lwd=2)
> legend("topright",legend=c("t(1)","t(2)","t(5)",
  "N(0,1)"),col=c("blue","green","brown","red"), pch=15)
```

Distribución t de Student



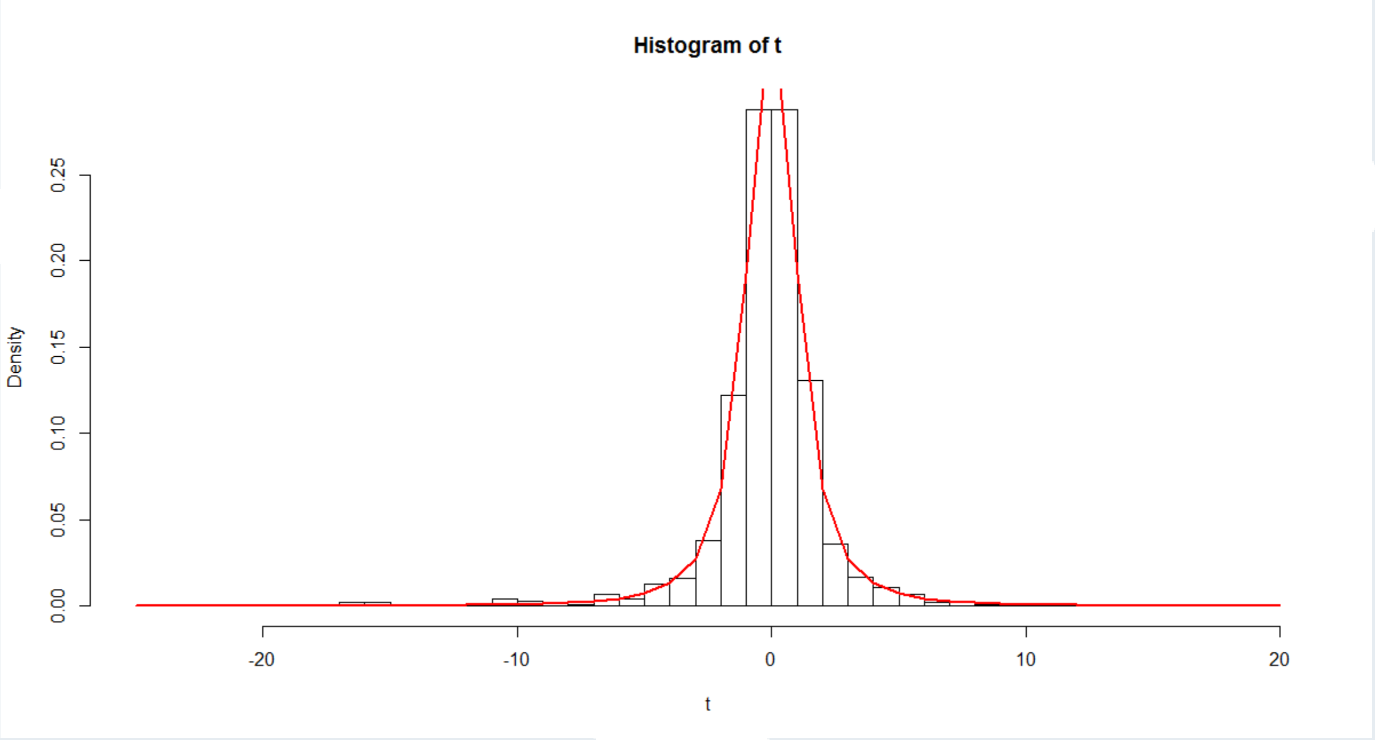
77

Distribución t de Student

- Relación de una t de Student con una distribución Normal estándar y una χ^2:

```
> z <- rnorm(1000)  # Normal estándar
> ji2 <- rchisq(1000,2)  # χ² con 2 grados de libertad
> t <- z/sqrt(ji2/2) # t con 2 grados de libertad
> hist(t,freq=FALSE,breaks=seq(-25,20,1))
> x <- seq(-25,20,1)
> y <- dt(x,2) # Distribución t teórica
> lines(x,y,col="red",lwd=2)
```

Distribución t de Student



Distribución t de Student

- La función distribución (área debajo de la curva) se calcula con pt. Utilizando de nuevo la t con 2 grados de libertad:

```
> pt(-1,2) # P(X<-1) si X ~ t(2) = Área bajo la cola
  izquierda de la curva
[1] 0.2113249
> pt(1,2,lower.tail=FALSE)  # P(X>1) si X ~ t(2) = Área
  bajo la cola derecha
[1] 0.2113249
```

Distribución t de Student

- Función cuantil (tabulada en los libros):

```
> qt(0.01,5,lower.tail=FALSE)
[1] 3.36493
```

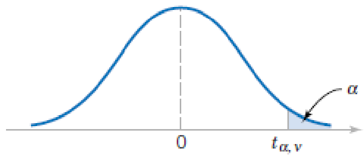


Table IV Percentage Points $t_{\alpha,\nu}$ of the t-Distribution

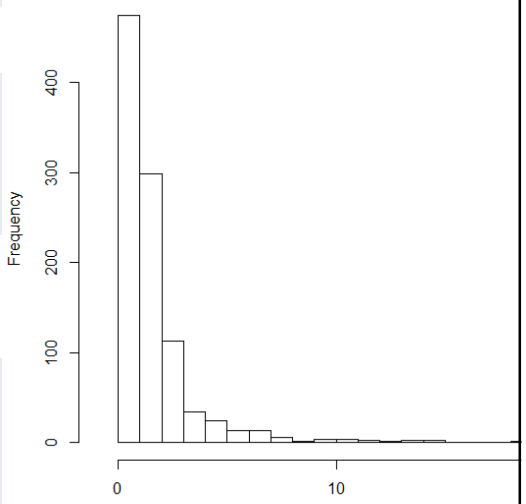
ν \ α	.40	.25	.10	.05	.025	.01	.005	.0025	.001	.0005
1	.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	.289	.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	.277	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	.271	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	.267	.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	.265	.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	.263	.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	.262	.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	.261	.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	.260	.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587

Distribución F

Distribución F

- Generación de 1000 números aleatorios según una distribución F con $k_1 = 8$ grados de libertad en el numerador y $k_2 = 5$ en el denominador:

```
> x <- rf(1000,8,5)
> hist(x,breaks=seq(0,45,1))
> mean(x)
[1] 1.775777
> var(x)
[1] 8.167527
```



- Valores teóricos:

$$E(X) = \frac{k_2}{k_2 - 2}, \quad \text{para } k_2 > 2 \qquad V(X) = \frac{2k_2^2(k_1 + k_2 - 2)}{k_1(k_2 - 4)(k_2 - 2)^2}, \quad \text{para } k_2 > 4$$

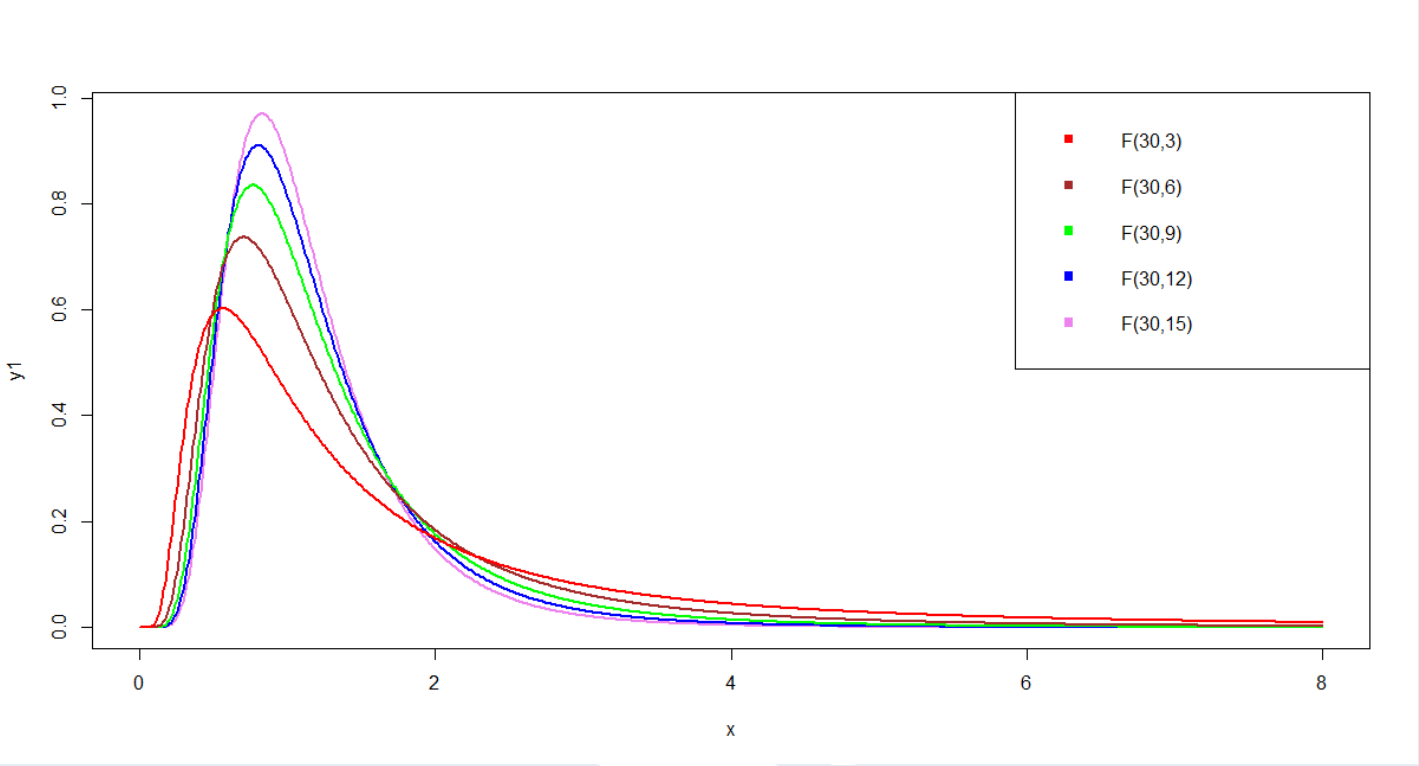
Distribución F

- Al aumentar el valor de k_2, manteniendo k_1 constante, la distribución F resultante tiende a concentrarse entorno a 1 (su valor esperado tiende a 1).

$$E(X) = \frac{k_2}{k_2 - 2}, \quad \text{para } k_2 > 2$$

```
> x <- seq(0,8,0.01)
> y1 <- df(x,30,15)  # F(30,15)
> y2 <- df(x,30,12)  # F(30,12)
> y3 <- df(x,30,9)   # F(30,9)
> y4 <- df(x,30,6)   # F(30,6)
> y5 <- df(x,30,3)   # F(30,3)
> plot(x,y1,col="violet",lwd=2,type="l")
> lines(x,y2,col="blue",lwd=2,type="l")
> lines(x,y3,col="green",lwd=2,type="l")
> lines(x,y4,col="brown",lwd=2,type="l")
> lines(x,y5,col="red",lwd=2,type="l")
> legend("topright",legend=c("F(30,3)","F(30,6)","F(30,9)",
  "F(30,12)","F(30,15)"),col=c("red","brown","green","blue",
  "violet"), pch=15)
```

Distribución F

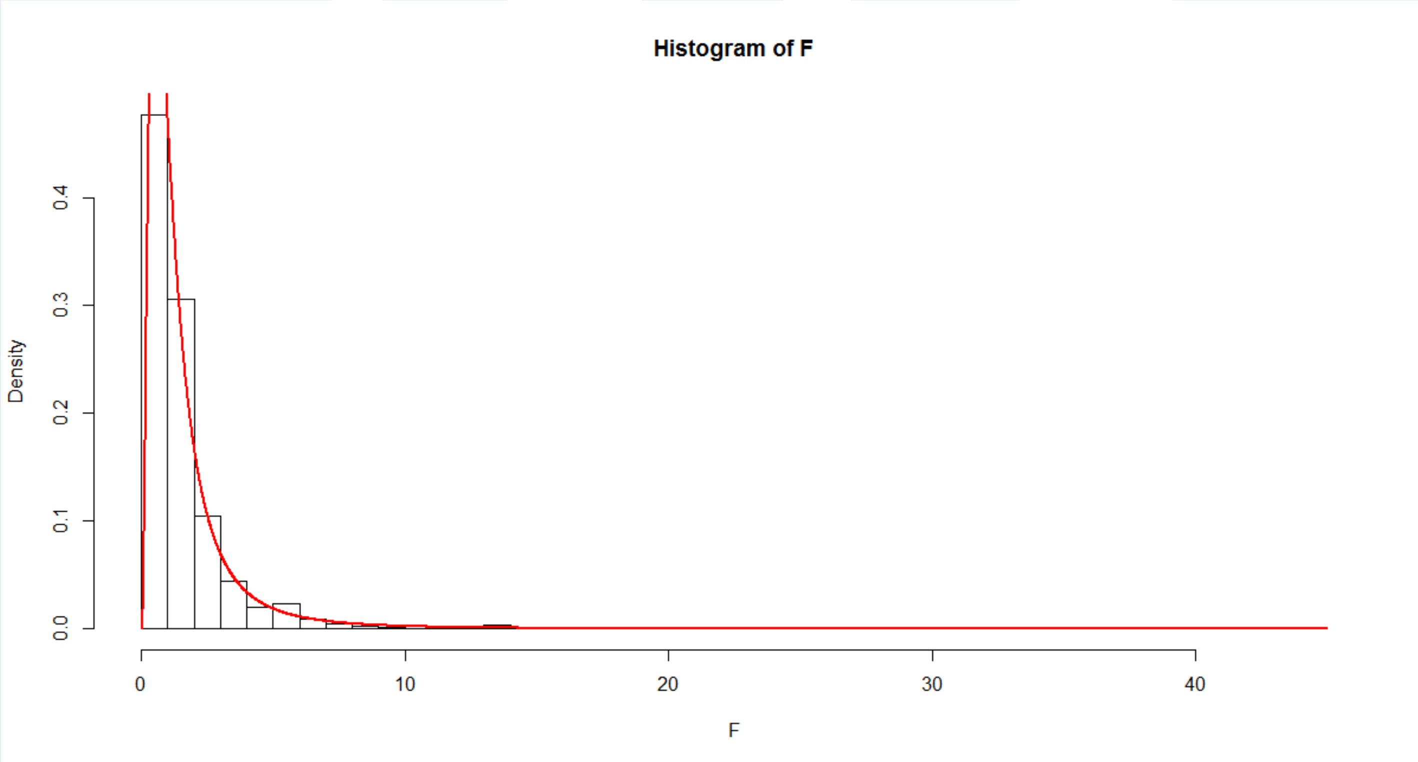


Distribución F

- Relación de la distribución F con dos distribuciones χ^2 con k_1 (numerador) y k_2 (denominador) grados de libertad, respectivamente.

```
> ji2n <- rchisq(1000,8) # χ² con k1 = 8
> ji2d <- rchisq(1000,5) # χ² con k2 = 5
> F <- (ji2n/8)/(ji2d/5) # F con 8 y 5 grados de lib.
> hist(F,freq=FALSE,breaks=seq(0,45,1))
> x <- seq(0,45,0.01)
> y <- df(x,8,5)  # F teórica
> lines(x,y,type="l",col="red",lwd=2)
```

Distribución F



Distribución F

- El área bajo la curva se calcula con la función pf:

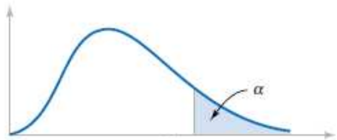
```
> pf(2,30,2) # P(X<2) si X ~ F(30,2) = Área bajo la
  cola izquierda
[1] 0.6114957

> pf(2,30,2,lower.tail=FALSE)  # P(X>2) si X ~ F(30,2)
  = Área bajo la cola derecha
[1] 0.3885043
```

Distribución F

- Función cuantil (tabulada en los libros):

```
> qf(0.25,10,7,lower.tail=FALSE)
[1] 1.689803
```



f_{α,k_1,k_2}

Función cuartil medida en la cola derecha para la distribución F según el área α (en este caso 0,25) y los grados de libertad k_1 (numerador) y k_2 (denominador)

k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
	Degrees of freedom for the numerator																		
1	5.83	7.50	8.20	8.58	8.82	8.98	9.10	9.19	9.26	9.32	9.41	9.49	9.58	9.63	9.67	9.71	9.76	9.80	9.85
2	2.57	3.00	3.15	3.23	3.28	3.31	3.34	3.35	3.37	3.38	3.39	3.41	3.43	3.43	3.44	3.45	3.46	3.47	3.48
3	2.02	2.28	2.36	2.39	2.41	2.42	2.43	2.44	2.44	2.44	2.45	2.46	2.46	2.46	2.47	2.47	2.47	2.47	2.47
4	1.81	2.00	2.05	2.06	2.07	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08
5	1.69	1.85	1.88	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.88	1.88	1.88	1.88	1.87	1.87	1.87
6	1.62	1.76	1.78	1.79	1.79	1.78	1.78	1.78	1.77	1.77	1.77	1.76	1.76	1.75	1.75	1.75	1.74	1.74	1.74
7	1.57	1.70	1.72	1.72	1.71	1.71	1.70	1.70	1.70	1.69	1.68	1.68	1.67	1.67	1.66	1.66	1.65	1.65	1.65
8	1.54	1.66	1.67	1.66	1.66	1.65	1.64	1.64	1.63	1.63	1.62	1.62	1.61	1.60	1.60	1.59	1.59	1.58	1.58
9	1.51	1.62	1.63	1.63	1.62	1.61	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56	1.55	1.54	1.54	1.53	1.53
10	1.49	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56	1.55	1.54	1.53	1.52	1.52	1.51	1.51	1.50	1.49	1.48
11	1.47	1.58	1.58	1.57	1.56	1.55	1.54	1.53	1.53	1.52	1.51	1.50	1.49	1.49	1.48	1.47	1.47	1.46	1.45
12	1.46	1.56	1.56	1.55	1.54	1.53	1.52	1.51	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.45	1.44	1.43	1.42

89