

The R logo is a large, stylized graphic in the background of the slide. It is composed of several blue and yellow rectangular blocks arranged in a way that suggests the letter 'R'.

# ANÁLISIS DE DATO CON R



contáctenos: [enei@inei.gob.pe](mailto:enei@inei.gob.pe) / 433-3127



Pasaje Hernán Velarde 285 Lima.

Entre la cuadra 01 y 02 de la Av. Arequipa.

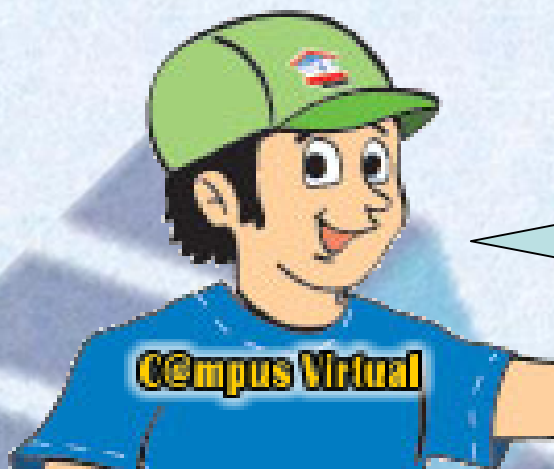
Correo: [enei@inei.gob.pe](mailto:enei@inei.gob.pe) / [campusvirtual@inei.gob.pe](mailto:campusvirtual@inei.gob.pe)

Teléfonos: 433-3127 - 332-4650

## Centro Andino de Formación y Capacitación en Estadística

Cursos Especializados en Estadística e Informática

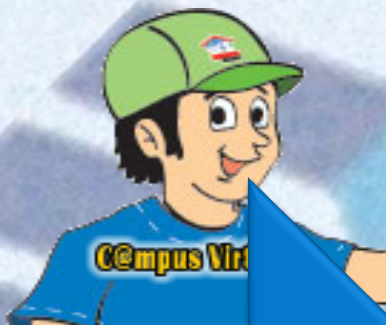




**Estimado alumno,  
buen día. Cualquier  
consulta no dudes en  
comentarme o  
avisarme.**

**Este curso es  
netamente práctico y  
se que lograremos  
objetivos importantes.**

**Recuerda siempre nuestro correo  
para cualquier consulta**



**campusvirtual@inei.gob.pe**





# **DISTRIBUCIONES DE PROBABILIDAD**



- Distribuciones de probabilidad
- Gráficos para distribuciones de probabilidad
- Generación de muestras aleatorias
- Estudio de la distribución de un conjunto de datos
- Estimación de densidades
- Ejercicios propuestos





# Distribuciones de probabilidad (1)

R proporciona funciones que permiten obtener la función de densidad de probabilidad (o de probabilidad para variables discretas) y la función de distribución de probabilidad (acumulada), así como calcular cuantiles y simular la distribución.

Distribución	Nombre en R	Parámetros
beta	beta	shape1, shape2, ncp
binomial	binom	size, prob
Cauchy	cauchy	location, scale
ji-cuadrado	chisq	df, ncp
exponencial	exp	rate
F	f	df1,df2, ncp
gamma	gamma	shape, scale
geométrica	geom	prob
hipergeométrica	hyper	m, n, k
log-normal	lnorm	meanlog, sdlog
logística	logis	location, scale
binomial negativa	nbinom	size, prob
normal	norm	mean, sd
Poisson	pois	lambda
T de Student	t	df, ncp
uniforme	unif	min,max
Weibull	weibull	shape, scale
Wilcoxon	wilcos	m, n

Las funciones se obtienen añadiendo los prefijos *d* para la densidad, *p* para la distribución, *q* para la función cuantil y *r* para la generación aleatoria.

Ej. Para la distribución normal disponemos de cuatro funciones:

```
dnorm(x, mean=0, sd=1)
pnorm(q, mean=0, sd=1)
qnorm(p, mean=0, sd=1)
rnorm(n, mean=0, sd=1)
```

# Distribuciones de probabilidad (2)

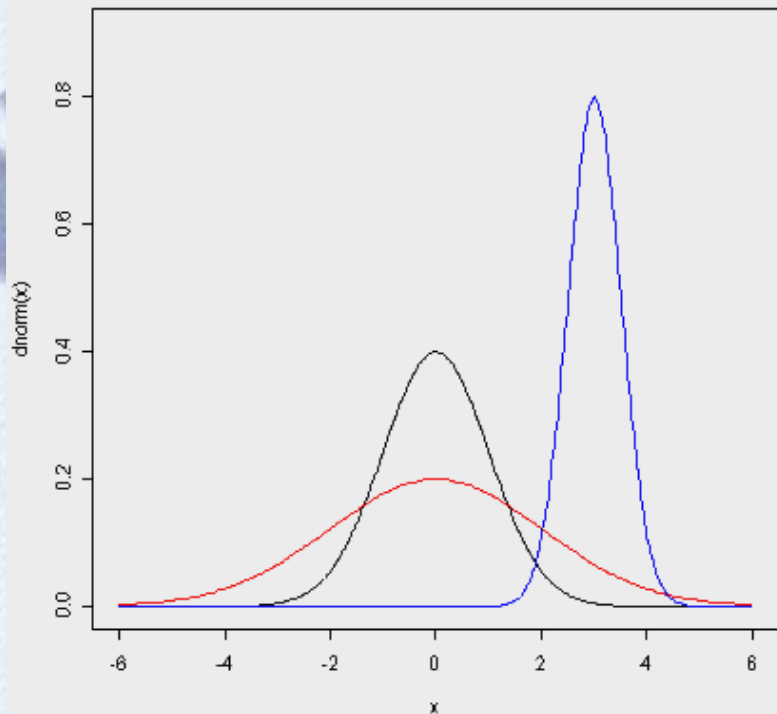
```
> # probabilidad que una binomial(10,0.2) tome el valor 2
> dbinom(2,size=10,prob=0.2) [1]
0.3019899
> #probabilidad que una binomial(10,0.2) tome un valor inferior a
2
> pbinom(2,size=10,prob=0.2) [1]
0.6777995
> #¿que valor de una binomial(10,0.2) presenta una probabilidad acumulada de 0.9 ?
> qbinom(0.9,size=10,prob=0.2) [1] 4
> #Generación de 20 valores aleatorios de una distribucion binomial(10,0.2)
> rbinom(20,size=10,prob=0.2)
[1] 0 7 3 1 1 2 2 2 3 1 2 1 2 1 0 1 1 2 2 2
> #Función de probabilidad y de distribución de probabilidad de una binomial(10,0.2)
> x<-seq(0,10,by=1)
> data.frame(x,p=dbinom(x,size=10,prob=0.2),F=pbinom(x,size=10,prob=0.2))
```

	x	p	F
1	0	0.1073741824	0.1073742
2	1	0.2684354560	0.3758096
3	2	0.3019898880	0.6777995
4	3	0.2013265920	0.8791261
5	4	0.0880803840	0.9672065
6	5	0.0264241152	0.9936306
7	6	0.0055050240	0.9991356
8	7	0.0007864320	0.9999221
9	8	0.0000737280	0.9999958
10	9	0.0000040960	0.9999999
11	10	0.0000001024	1.0000000

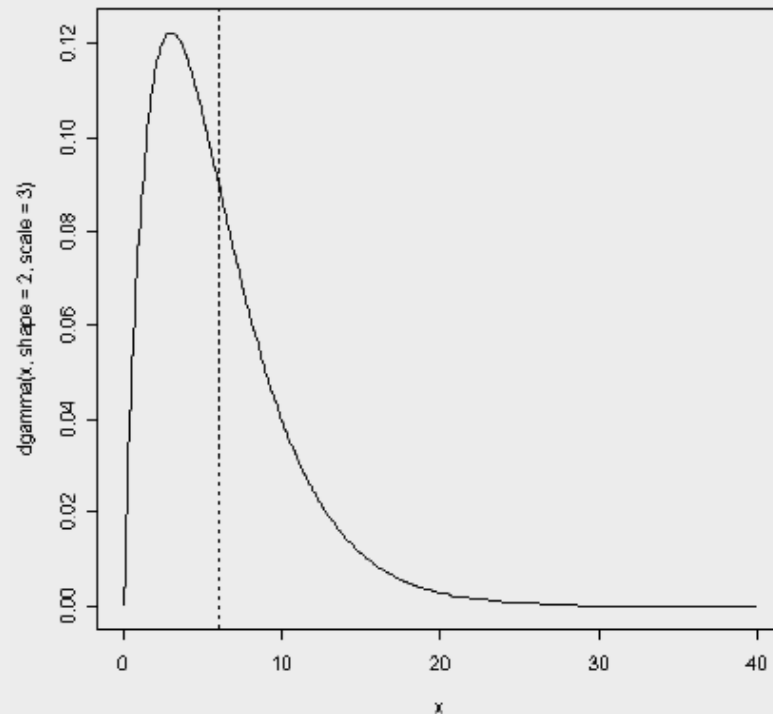


# Gráficos para distribución de probabilidad (1)

```
> #Funciones de densidad de diferentes normales  
> x<-seq(-6,6,by=0.1)  
> plot(x,dnorm(x),type="l",xlim=c(-6,6),ylim=c(0,0.9))  
> lines(x,dnorm(x,mean=0,sd=2),col="red")  
> lines(x,dnorm(x,mean=3,sd=0.5),col="blue")
```

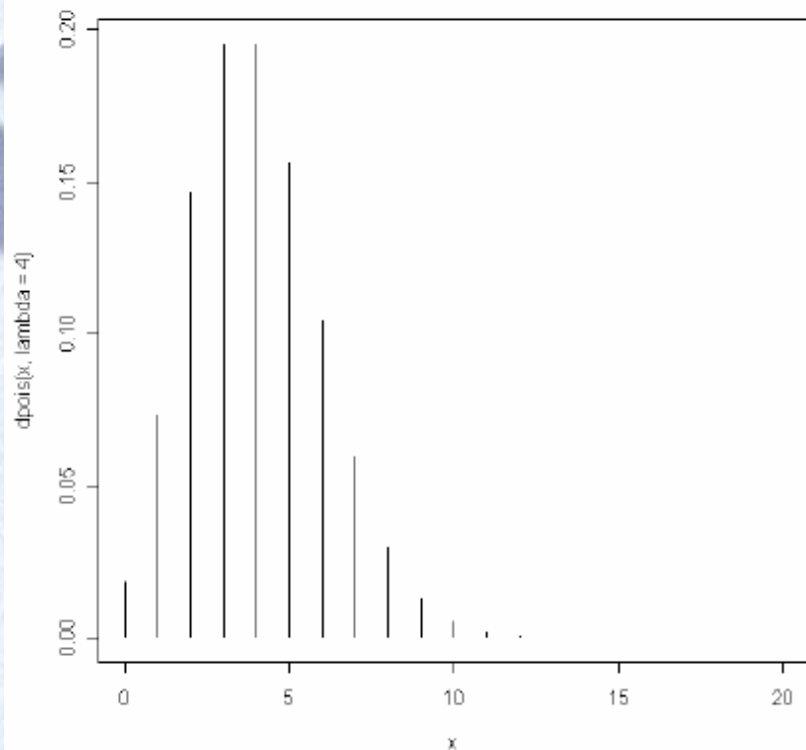


```
> #Funcion de densidad para una gamma  
> x<-seq(0,40,by=0.01)  
> curve(dgamma(x,shape=2,scale=3),from=0,to=40)  
> abline(v=2*3,lty=2) #esperanza de la gamma
```

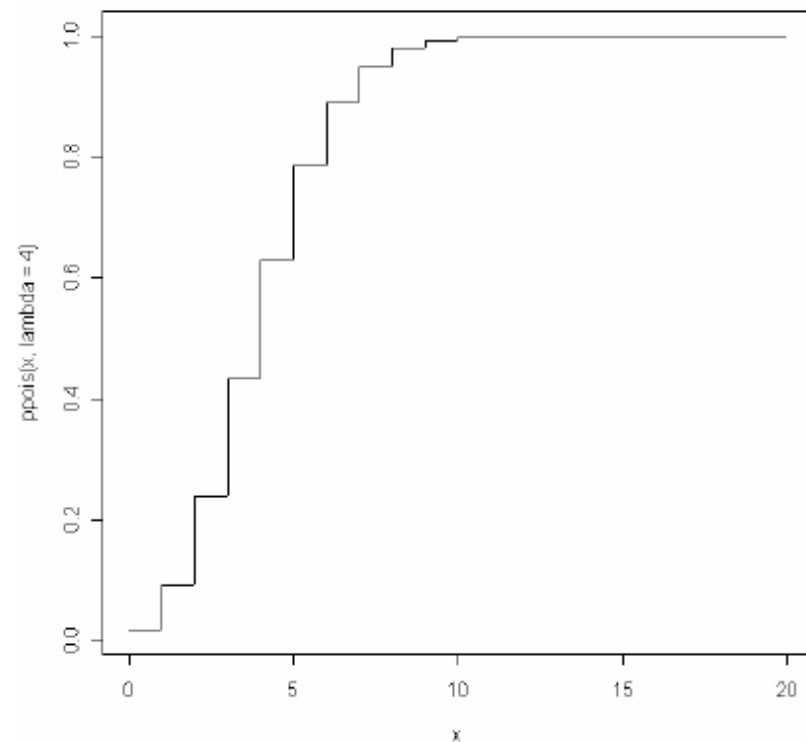


# Distribuciones de probabilidad (2)

```
> #Función de probabilidad para una poisson  
de media 4  
> x<-0:20  
> plot(x,dpois(x,lambda=4),type="h")
```



```
> #Función de distribución de probabilidad  
para una poisson de media 4  
> x<-0:20  
> plot(x,ppois(x,lambda=4),type="s")
```





# Generación de números aleatorios

```
> #Generación de 100 muestras de tamaño 10 de una normal tipificada
> muestras<-matrix(rnorm(1000),nrow=100,byrow=T)
> medias<-apply(muestras,1,mean)
> var(medias)
[1] 0.1169421

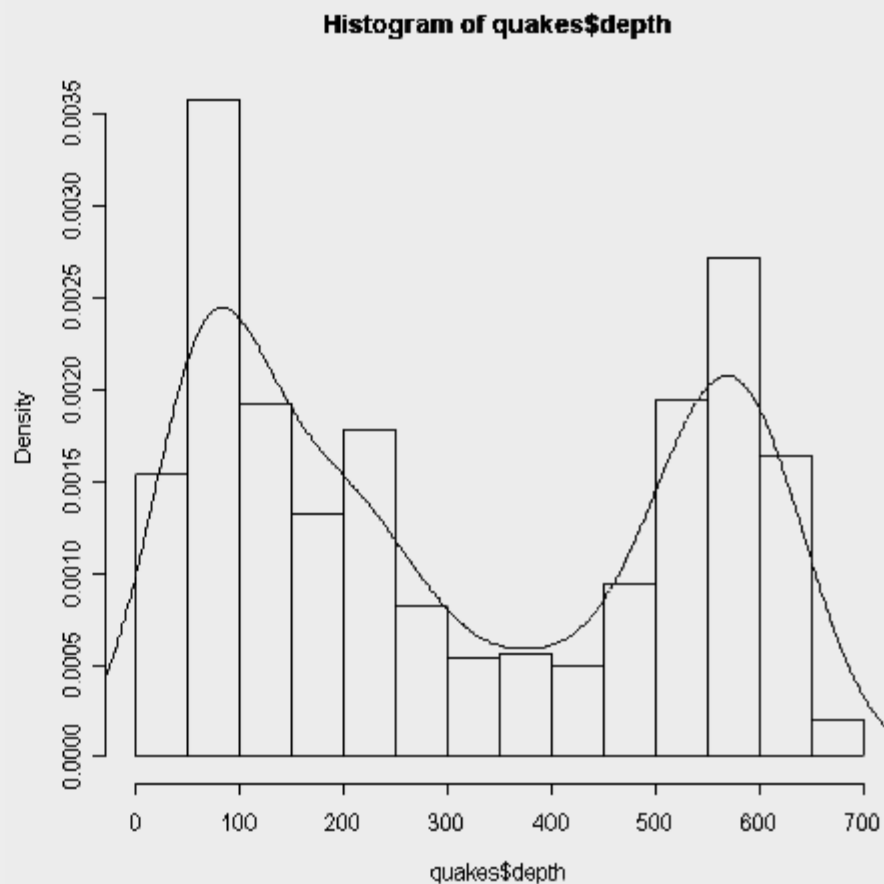
> #Generación de una variable factor segun una binomial
> x<-rbinom(20,size=1,prob=0.2)
> x
[1] 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0
> x<-factor(x)
> levels(x)<-c("fracaso","exito")
> table(x)
x
fracaso  exito
      17      3

> #Generación de una multinomial
> x<-sample(c("A","B","C"),200,replace=T,prob=c(0.5,0.4,0.1))
> table(x)
x
  A   B   C
104  79  17
```

# Estudio de la distribución de un conjunto de datos

Dado un conjunto de datos, su histograma nos informa acerca de la distribución. Podemos emplear una estimación tipo kernel para estimar la densidad de probabilidad.

```
> #Estudio de la distribución de los  
terremotos ocurridos cerca de Fiji  
> data(quakes)  
> hist(quakes$depth, prob=TRUE)  
> lines(density(quakes$depth, bw=50))
```





# Estimación de densidades (1)

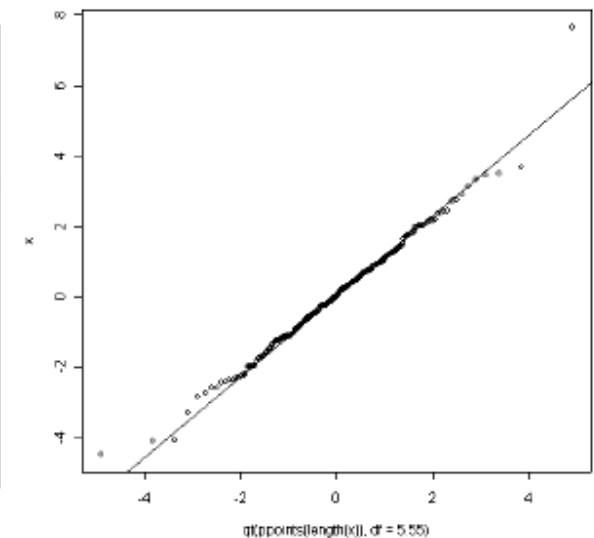
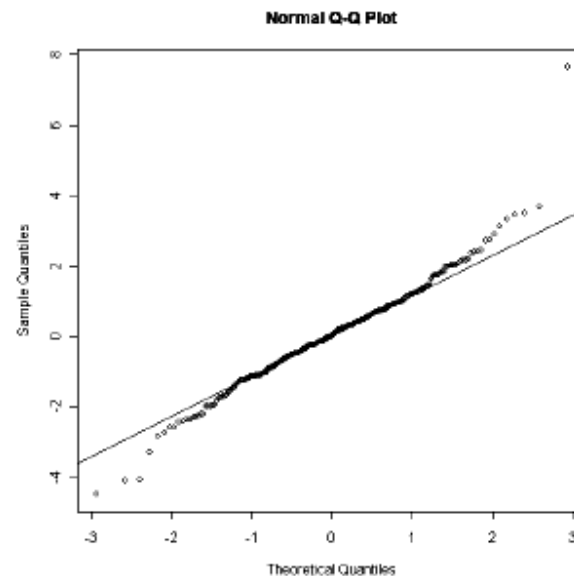
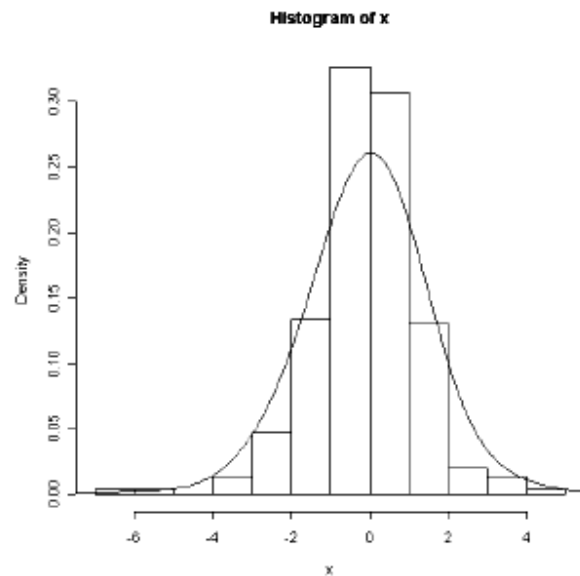
La función `fitdistr` de la librería `MASS` permite obtener una estimación máximo-verosímil de distribuciones univariantes.

#Generación y estimación de una distribución  
t de Student

```
> x<-rt(300,df=5)  
> fitdistr(x,"t")
```

m	s	df
0.03345688	1.09292734	5.55902260
(0.07209231)	(0.07572894)	(1.72774691)

```
> qqnorm(x); qqline(x)  
> qqplot(qt(ppoints(length(x))),df=5.55),x)  
> qqline(x)
```



# Estimación de densidades (2)

La función `ks.test` permite resolver la prueba de Kolmogorov-Smirnov (suponiendo los parámetros conocidos)

```
> #Se distribuyen los datos anteriores según una t de student?
```

```
> ks.test(x, "pt", df=5.55)
```

One-sample Kolmogorov-Smirnov test

```
data: x
```

```
D = 0.0458, p-value = 0.5558
```

```
alternative hypothesis: two.sided
```

```
> #Comparación de dos muestras
```

```
> x <- rnorm(50)
```

```
> y <- runif(30)
```

```
> ks.test(x, y)
```

Two-sample Kolmogorov-Smirnov test

```
data: x and y
```

```
D = 0.52, p-value = 3.885e-05
```



# Ejercicios propuestos

1. Calcular la probabilidad de: (a) que una normal tipificada sea mayor que 3. (b) que una variable normal con media 35 y desviación 6 sea mayor que 42. (c) obtener 10 éxitos en las 10 pruebas de una variable binomial con probabilidad de éxito 0.8. (d) una variable uniforme estándar (entre 0 y 1) tome un valor inferior a 0.9. (e) una variable ji-cuadrado con dos grados de libertad tome un valor superior a 6.5.
2. Calcular los intervalos de confianza al 90, 95 y 99% de una distribución normal de media 5 y desviación 2.
3. Generar 100 valores aleatorios de una distribución t con 10 grados de libertad. Suponga que los valores obtenidos provienen de la realización de una T-Test. Calcular los p-valores y estudia su distribución. Repetir el análisis para una muestra de tamaño 1000.
4. Visualizar las funciones de densidad de diferentes distribuciones de ji-cuadrado con 5, 20 y 50 grados de libertad respectivamente.
5. Analizar la distribución del tiempo de las erupciones del géiser OldFaithful, del parque de Yellostwtone (datos faithful).
6. Analizar la normalidad del tiempo observado entre erupciones del géiser OldFaithful. Considera la existencia de una mixtura de distribuciones normales.

## Comunicación constante con la Escuela del INEI

**Correo de la Escuela del INEI**  
**[enei@inei.gob.pe](mailto:enei@inei.gob.pe)**

**Área de Educación Virtual**  
**[campusvirtual@inei.gob.pe](mailto:campusvirtual@inei.gob.pe)**

