

Estadística I

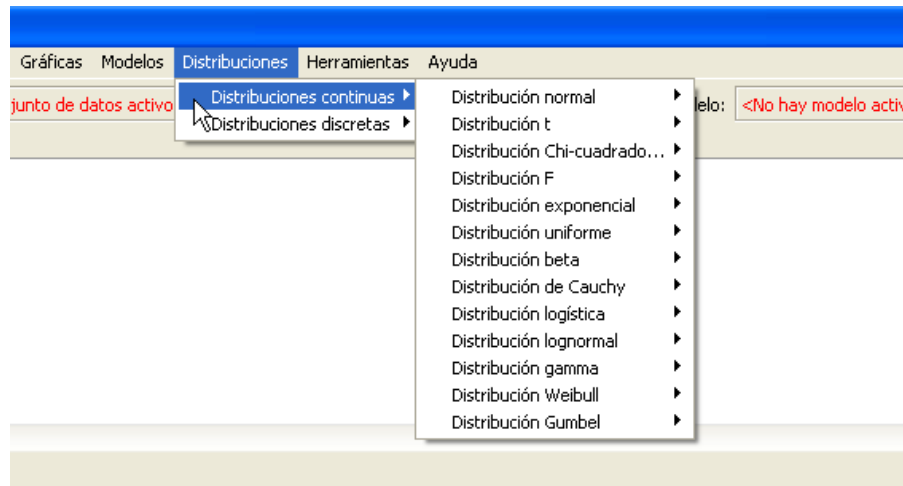
Práctica para los temas 4 y 5:
Variables aleatorias e Inferencia Estadística

Práctica para los temas 4 y 5: Variables aleatorias e Inferencia Estadística

CONTENIDOS

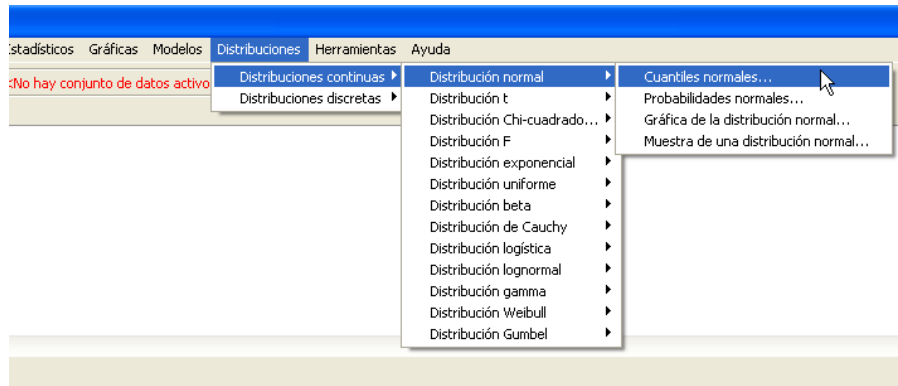
- ▶ **Variables aleatorias (continuas y discretas)**
 - ▶ Cálculo de cuantiles y probabilidades.
 - ▶ Gráficas de funciones de masa y de densidad y de funciones de distribución.
 - ▶ Generación de muestras.
 - ▶ Teorema Central del límite usando el paquete *Teaching Demos*.
- ▶ **Inferencia Estadística**
 - ▶ Intervalos de confianza para la media μ cuando la población es normal con σ conocido - *Teaching Demos*.
 - ▶ Contrastes de hipótesis para μ cuando la población es normal con σ desconocida.
 - ▶ Contrastes de hipótesis e intervalos de confianza para la proporción p de una población en muestras grandes.

Distribuciones de variables aleatorias continuas y discretas



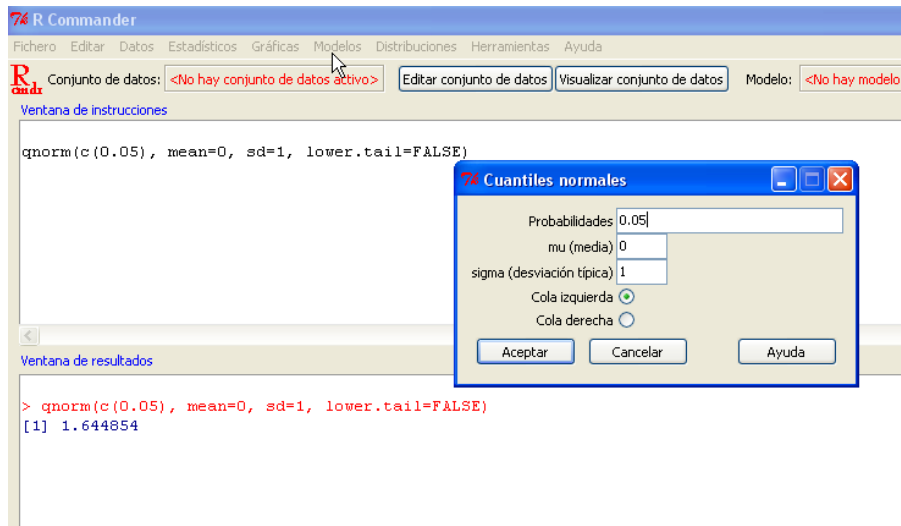
Cálculo de los cuantiles (superiores) de una distribución

- ▶ Encuentras el cuantil (superior) α de una **distribución estándar normal**, es decir, encuentras el número z_α tal que $P(Z > z_\alpha) = \alpha$ para $\alpha = 0.05$



Cálculo de los cuantiles (superiores) de una distribución (cont.)

- ▶ $P(Z > z_{0.05}) = 0.05$ se cumple para $z_{0.05} = 1.64$



The screenshot shows the R Commander interface. The main window displays the command `qnorm(c(0.05), mean=0, sd=1, lower.tail=FALSE)` in the instruction window. A dialog box titled "Cuantiles normales" is open, showing the input values: Probabilidades: 0.05, mu (media): 0, sigma (desviación típica): 1, Cola izquierda: selected, and Cola derecha: unselected. The results window shows the output: `> qnorm(c(0.05), mean=0, sd=1, lower.tail=FALSE)` and `[1] 1.644854`.

R Commander

Fichero Editar Datos Estadísticos Gráficas Modelos Distribuciones Herramientas Ayuda

Conjunto de datos: <No hay conjunto de datos activo> Editar conjunto de datos Visualizar conjunto de datos Modelo: <No hay modelo activo>

Ventana de instrucciones

```
qnorm(c(0.05), mean=0, sd=1, lower.tail=FALSE)
```

Ventana de resultados

```
> qnorm(c(0.05), mean=0, sd=1, lower.tail=FALSE)
[1] 1.644854
```

Cuantiles normales

Probabilidades: 0.05

mu (media): 0

sigma (desviación típica): 1

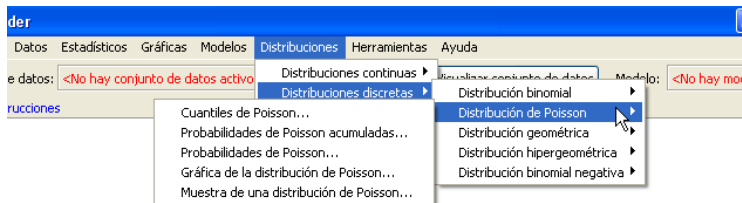
Cola izquierda:

Cola derecha:

Aceptar Cancelar Ayuda

Cálculo de las probabilidades de una distribución

- ▶ Para $X \sim \text{Poisson}(\lambda = 2)$ $E[X] = 1/\lambda = 0.5$ encuentras $P(X > 3)$
- ▶ Las probabilidades superiores aparecen con desigualdad estricta $>$.
- ▶ Las probabilidades inferiores aparecen con \leq



Cálculo de las probabilidades de una distribución (cont.)

- ▶ Para $X \sim \text{Poisson}(\lambda = 2)$, $P(X > 3) = 0.001752$
- ▶ Para calcular $P(X \geq 3) = P(X > 2)$, escribes 2 en Valor(es) de la variable

The screenshot shows the R Commander interface. The main window displays the command `ppois(c(3), lambda=0.5, lower.tail=FALSE)` in the console. A dialog box titled "Probabilidades de Poisson" is open, showing the input values: "Valor(es) de la variable" is 3, "Media" is 0.5, and "Cola derecha" is selected. The console output shows the result: `[1] 0.001751623`.

R Commander

Fichero Editar Datos Estadísticos Gráficas Modelos Distribuciones Herramientas Ayuda

Conjunto de datos: <No hay conjunto de datos activo> Editar conjunto de datos Visualizar conjunto de datos Modelos

Ventana de instrucciones

```
ppois(c(3), lambda=0.5, lower.tail=FALSE)
```

74 Probabilidades de Poisson

Valor(es) de la variable 3

Media 0.5

Cola izquierda

Cola derecha

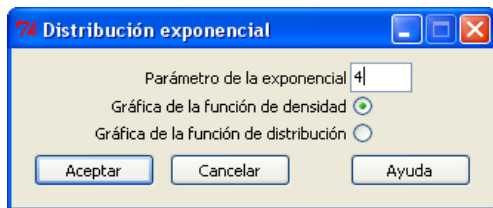
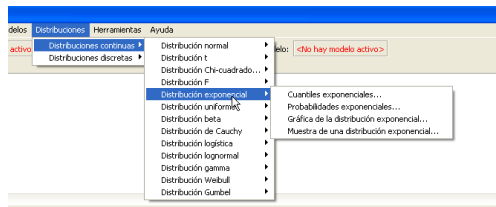
Aceptar Cancelar Ayuda

Ventana de resultados

```
> ppois(c(3), lambda=0.5, lower.tail=FALSE)
[1] 0.001751623
```

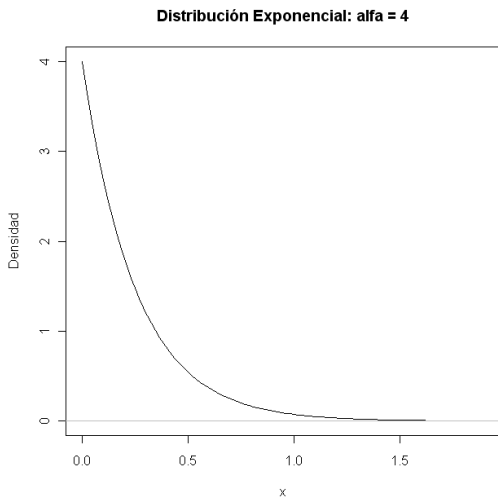
Gráficas de una función de densidad, pdf

- ▶ Para $X \sim \text{Exp}(\alpha = 4)$



Gráficas de una función de densidad, pdf (cont.)

- ▶ Para $X \sim \text{Exp}(\alpha = 4)$



Gráficas (acumuladas) de una función de distribución, cdf

- ▶ Para $X \sim \text{Binomial}(n = 100, p = 0.2)$

74 Distribución binomial

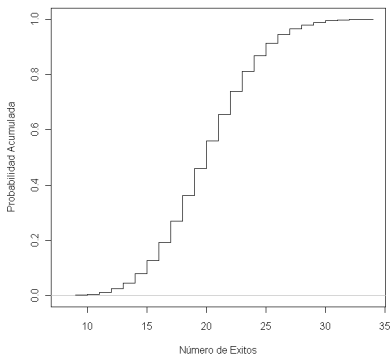
Ensayos binomiales:

Probabilidad de éxito:

Gráfica de la función de probabilidad:

Gráfica de la función de distribución:

Distribución Binomial : Ensayos = 100, Probabilidad de éxito = 0.2



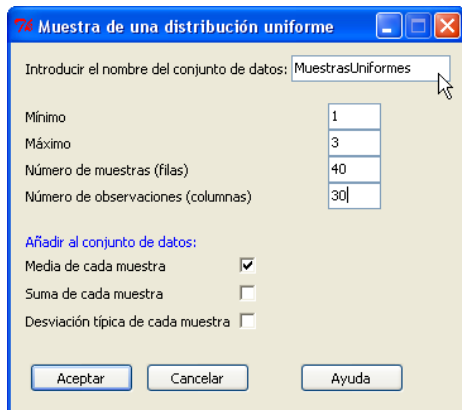
Generación de muestras de una distribución

- ▶ Para $X \sim \text{Uniforme}(a = 1, b = 3)$ generas $m = 40$ muestras, cada una de tamaño $n = 30$
- ▶ **Calculas la media muestral** para cada muestra

The screenshot shows a software interface with a menu bar containing 'Modelos', 'Distribuciones', 'Herramientas', and 'Ayuda'. The 'Distribuciones' menu is open, showing a list of distribution types. The 'Distribución uniforme' option is selected, which has opened a sub-menu. In this sub-menu, the option 'Muestra de una distribución uniforme...' is highlighted by a mouse cursor. Other options in the sub-menu include 'Cuantiles uniformes...', 'Probabilidades uniformes...', and 'Gráfica de la distribución uniforme...'. In the background, a text box labeled 'Modelo:' contains the text '<No hay modelo activo>'. At the bottom of the screen, there are navigation icons for back, forward, and search.

Generación de muestras de una distribución (cont.)

- ▶ Para $X \sim \text{Uniforme}(a = 1, b = 3)$ generas $m = 40$ muestras, cada una de tamaño $n = 30$
- ▶ **Calculas la media muestral** para cada muestra



The image shows a software dialog box titled "Muestra de una distribución uniforme". It contains the following fields and options:

- Introducir el nombre del conjunto de datos: MuestrasUniformes
- Mínimo: 1
- Máximo: 3
- Número de muestras (filas): 40
- Número de observaciones (columnas): 30
- Añadir al conjunto de datos:
 - Media de cada muestra:
 - Suma de cada muestra:
 - Desviación típica de cada muestra:
- Buttons: Aceptar, Cancelar, Ayuda

Generación de muestras de una distribución (cont.)

Interpretación de la tabla de datos:

Muestra 1: $\dots, x_{28}^{(1)} = 2.87, x_{29}^{(1)} = 2.64, x_{30}^{(1)} = 2.77, \bar{x}^{(1)} = 1.99$

...

Muestra 10: $\dots, x_{28}^{(10)} = 1.29, x_{29}^{(10)} = 1.60, \dots, x_{30}^{(10)} = 2.97, \bar{x}^{(10)} = 1.98$

...

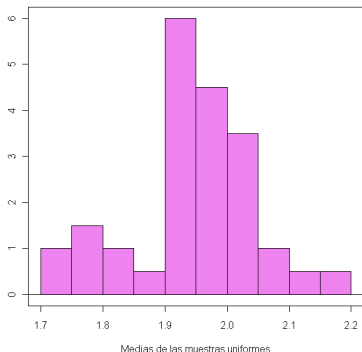
	obs23	obs24	obs25	obs26	obs27	obs28	obs29	obs30	mean
sample1	1.357840	1.408764	1.144718	2.471585	2.822461	2.874870	2.643988	2.774838	1.999130
sample2	2.904315	2.869983	1.068353	2.003333	2.826331	2.440492	1.044722	2.307417	1.964241
sample3	1.157321	1.354268	2.482794	1.596685	1.262029	2.328448	2.487440	2.302827	1.904619
sample4	1.385535	1.844477	1.881504	1.464712	2.241625	1.738727	2.257847	2.391045	1.913762
sample5	1.619741	2.235270	1.084672	2.876547	2.334880	1.132085	1.374358	2.578989	1.930766
sample6	1.317809	1.511673	2.478177	1.069703	2.082996	1.248338	2.032739	1.566174	2.009353
sample7	1.273327	1.079250	2.693483	2.293953	2.423859	2.698841	1.860031	1.750597	2.039123
sample8	1.401991	2.367814	2.273720	1.657166	1.920973	1.318235	2.382029	1.715908	1.774292
sample9	2.881143	1.034715	2.019409	2.678303	1.495054	2.773771	2.985617	2.052342	2.065943
sample10	1.923524	2.927068	1.581044	1.075256	2.840149	1.293293	1.608898	2.978922	1.985182
sample11	1.711150	1.231207	2.880598	1.852499	2.610776	2.383123	1.259893	1.528350	2.172635
sample12	2.801028	2.124749	1.795517	2.508314	2.529156	1.566374	2.432166	2.669524	2.027017
sample13	2.572519	2.303891	2.199450	1.455730	2.018649	1.320304	2.910802	2.294508	1.907983
sample14	1.523888	1.645083	2.882663	2.201425	2.061199	2.765709	1.087543	2.930208	2.145418
sample15	1.399046	2.926242	1.300070	2.133184	2.234044	2.320385	2.374877	2.019161	2.008263
sample16	2.072924	1.565787	2.816490	1.637458	1.561562	1.361107	2.704109	2.721886	1.943728
sample17	1.584174	1.829234	1.166516	1.910012	2.265944	2.068894	1.033363	2.954221	1.921916
sample18	2.043262	1.204295	1.225776	1.919850	2.455724	1.042948	1.184136	1.009368	1.748720
sample19	1.839308	1.696926	2.458488	2.558628	2.721728	1.458565	1.619819	2.645969	2.018694
sample20	1.089997	1.941245	2.034379	2.427345	1.299335	2.522895	1.145551	2.095193	1.835560
sample21	2.115960	2.246721	2.971987	2.730483	1.192233	2.752186	2.706991	2.718870	1.936279
sample22	2.662029	1.094014	2.467784	1.353407	1.946870	2.341636	1.800270	2.346908	1.943334
sample23	1.212303	1.363367	2.307600	1.684011	2.702403	1.121788	1.044576	2.466314	1.910572
sample24	1.707628	2.610576	2.418533	1.706630	1.842839	1.098789	1.050033	2.938759	1.989147
sample25	1.618305	1.714244	2.476367	1.301933	2.156350	2.698494	2.127403	1.553248	2.047222
sample26	1.853760	2.426991	2.089876	1.790196	1.489029	1.843244	1.828816	1.365980	1.861318
sample27	2.922820	2.590785	1.116137	2.612464	2.880108	1.302046	1.378231	2.196453	1.788721
sample28	2.238583	1.971198	1.424201	2.545949	1.481054	2.668033	1.064072	2.685823	1.965094
sample29	1.210739	2.341534	1.122771	2.501800	1.336164	2.988374	2.412130	1.233608	1.721874
sample30	2.814880	1.108977	2.823313	1.950381	2.877600	2.075963	2.092260	2.352235	1.942730

Introducción al Teorema Central del Límite

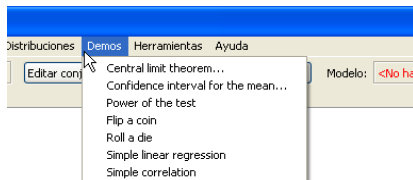
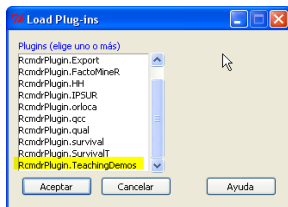
- ▶ Haces un histograma de las $m = 40$ medias muestrales de la transparencia anterior
- ▶ De acuerdo al TCL, ¿cuál debería ser su forma y su centro?

Normal, centrado en la media poblacional

$$\mu = E[X] = \frac{a+b}{2} = 2$$



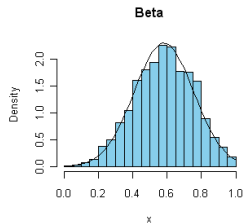
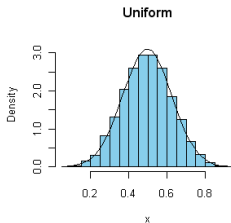
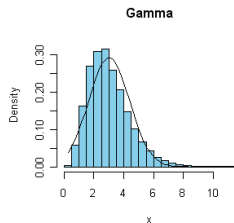
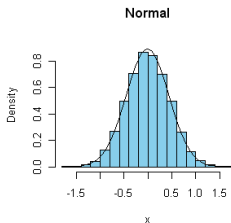
Teorema Central del Límite con Teaching Demos



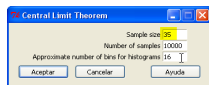
Teorema Central del Límite con Teaching Demos, $n = 5$



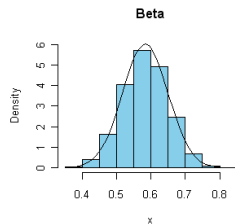
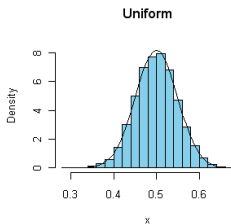
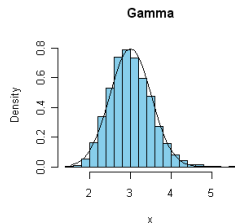
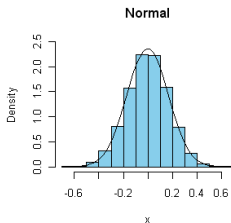
sample size = 5



Teorema Central del Límite con Teaching Demos, $n = 35$

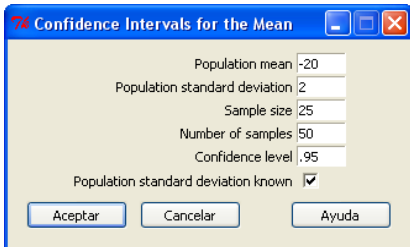
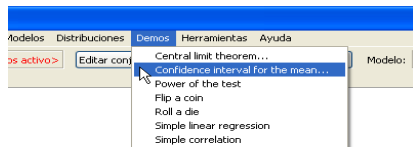


sample size = 35



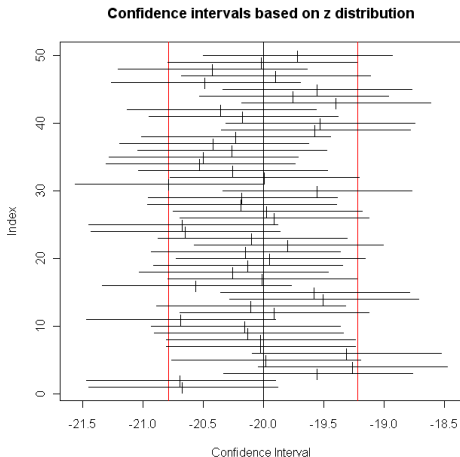
Intervalos de confianza para μ , $n = 35$

- ▶ Sea $X \sim N(\mu = -20, \sigma = 2)$ pero asumimos que μ es desconocido y que sólo σ es conocida
- ▶ Generas $m = 50$ muestras de tamaño $n = 25$ de X y basado en cada una de ellas, construyes un intervalo de confianza al 95% para μ



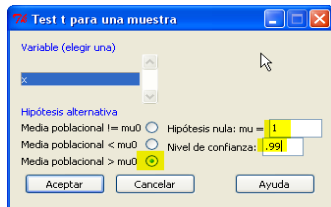
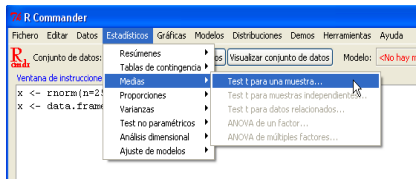
Intervalos de confianza para μ , $n = 35$

- ▶ ¿Cuántos intervalos (aproximadamente) **no incluirán** a μ ?
 $(100\% - 95\%)m \approx 2.5 \approx 3$



Test de hipótesis para μ

- ▶ Generas una muestra de tamaño $n = 20$ de $X \sim N(\mu = 1, \sigma = 2)$ pero asumes que no se conocen los parámetros de la distribución:
Ejecutas en la *Ventana de Instrucciones*:
`x = rnorm(n=25, mean=1, sd=2)`
`x = data.frame(x)`
- ▶ Haces de `x` tu *conjunto de datos activo*
- ▶ Contrastas la siguiente hipótesis: $H_0 : \mu = 1$ frente a $\mu > 1$ con $\alpha = 0.01$



Test de hipótesis para μ (cont.)

- ▶ p-value: 0.2561 ($\geq \alpha$ así no se rechaza H_0)
- ▶ estadístico del test: 0.6654
- ▶ $\bar{x} = 1.2214$
- ▶ Para obtener intervalos de confianza simétricos, se rehace el procedimiento con la alternativa $\neq \mu_0$

Ventana de resultados

```
> t.test(x$x, alternative='greater', mu=1, conf.level=.99)
```

One Sample t-test

```
data: x$x
```

```
t = 0.6654, df = 24, p-value = 0.2561
```

```
alternative hypothesis: true mean is greater than 1
```

```
99 percent confidence interval:
```

```
0.3921308      Inf
```

```
sample estimates:
```

```
mean of x
```

```
1.221414
```

Test de hipótesis e intervalos de confianza para p

- ▶ Generas una muestra de tamaño $n = 80$ de la variable categórica $Y = \text{ingresos}$, con tres niveles: bajo, medio, alto, que tienen probabilidad 0.2, 0.4, 0.4.
Para ello, ejecutas en la *Ventana de instrucciones*:

```
levels = c("bajo", "medio", "alto")  
y = sample(levels, size=80, replace=TRUE, prob=c(0.2,  
0.4, 0.4))
```
- ▶ Nos interesa calcular la proporción p de las personas de la población que tienen ingresos *bajos*
- ▶ Para contrastar $H_0 : p = 0.4$ frente a $H_1 : p \neq 0.4$ para $\alpha = 0.05$ ejecutas en la *Ventana de instrucciones*

```
z.test(x=y=="bajo", stdev=sqrt(0.4*(1-0.4)),  
alternative="two.sided", mu=0.4, conf.level=0.95)
```
- ▶ $y=="bajo"$ es un vector de longitud n de 0's y 1's, 1 indica que el valor era *bajo* y 0 en otro caso

Test de hipótesis e intervalos de confianza para p (cont.)

- ▶ p-value: 0.0006187 ($< \alpha$ así pues, se rechaza H_0)
- ▶ estadístico del test: -3.4233
- ▶ proporción muestral $\hat{p} = 0.2125$
- ▶ Para construir un intervalo de confianza al 95% para p se reemplaza en `stdev` 0.4 por el valor de \hat{p} , esto es, 0.2125 y se vuelve a repetir la instrucción.

```
> z.test(x=y=="bajo",stdev=sqrt(0.4*(1-0.4)),alternative="two.sided",
+ mu=0.4, conf.level=0.95)
```

```
One Sample z-test
```

```
data: y == "bajo"
```

```
z = -3.4233, n = 80.000, Std. Dev. = 0.490, Std. Dev. of the sample
mean = 0.055, p-value = 0.0006187
```

```
alternative hypothesis: true mean is not equal to 0.4
```

```
95 percent confidence interval:
```

```
0.1051484 0.3198516
```

```
sample estimates:
```

```
mean of y == "bajo"
```

```
0.2125
```