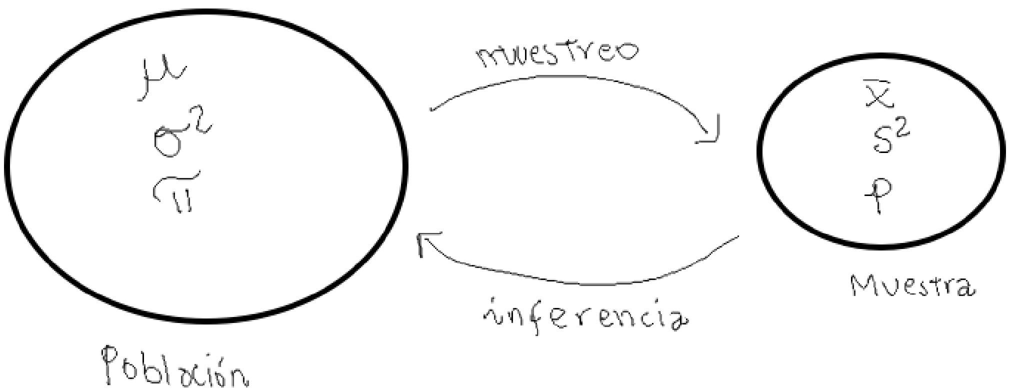


- Paramétrica (sujeta a supuestos) *
  - 1 par.
  - 2 par.
- No paramétrica (no se cumple algún supuesto)

Prueba de hipótesis para una media

* Muestra
  - Si $n < 30 \to X \sim N(\mu, \sigma^2)$
  - Si $n \geq 30 \to X \sim D(\mu, \sigma^2)$ ← cualq. dist.

* Independencia de observ: → observ. asociados → $n_{eff} \downarrow$

$\sqrt{}\sqrt{}$ $t \to \mu$ desc, σ^2 desconocida

• $N \to \mu$ desc y σ^2 conocida

$$t_{calc} = \frac{\overline{x} - \mu_0}{\frac{s}{\sqrt{n}}} \sim t_{(n-1)}$$

$\overline{x}$ → media muestral

μ_0 → media hipotética (constante)

s → desv. est. muestral

n → tamaño muestral

Decisión:

* t_{calc} vs t_{crit}
* α vs pv

} equivalentes

→ Rechazar Ho si t_{calc} cae en la zona de rechazo

→ $\boxed{\text{Rechazar Ho si pv} < \alpha}$

Ejemplo

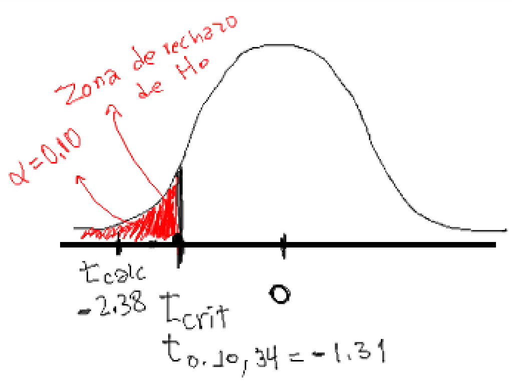
Una universidad afirma que sus estudiantes dedican en promedio 4 horas diarias al estudio. Se sospecha que el promedio es realmente menor, por lo que se toma una muestra aleatoria de 35 estudiantes, quienes reportan las siguientes horas de estudio:

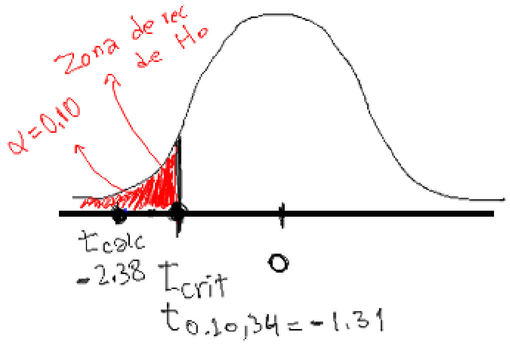
2.8, 3.9, 4.2, 2.1, 3.7, 4.5, 3.8, 4.1, 4.6, 3.6, 4.3, 4.0, 2.9, 1.4, 3.9, 4.5, 4.2, 4.0, 3.8, 1.0, 4.7, 3.5, 2.1, 2.2, 3.6, 4.3, 4.8, 4.2, 3.9, 3.7, 4.4, 3.0, 3.8, 4.6, 2.5

Verificar la afirmación con un nivel de significancia del 10%.

$$H_0 : \mu \geq 4 \qquad H_1 : \mu < 4 \qquad \alpha = 0.10$$

$>$





μ_0 <

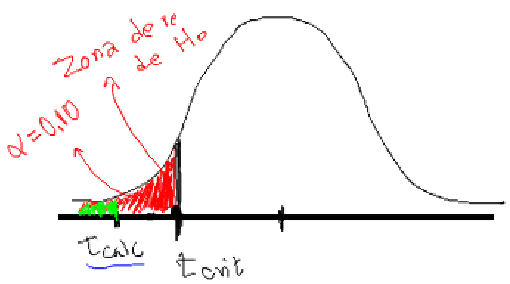
```r
t.test(x = tiempo, mu = 4, alternative = "less")
```

"greater" >

"two.sided" ≠

t calc   gl

```
	One Sample t-test

data:  tiempo
t = -2.3833, df = 34, p-value = 0.01144
alternative hypothesis: true mean is less than 4
95 percent confidence interval:
    -Inf 3.88878
sample estimates:
mean of x
 3.617143
```




```
> pt(q = -2.3833, df = 34)
[1] 0.01144406
```

$t_{crit} \to \alpha \to$ = 0.10

$t_{calc} \to pV \to$ = 0.01144

vemos que $pV < \alpha \Rightarrow$ Rech. Ho

$pv = P(t_{34} < t_{calc}) = P(t_{34} < -2.3833) = 0.01144$

Prueba de hipótesis para una varianza

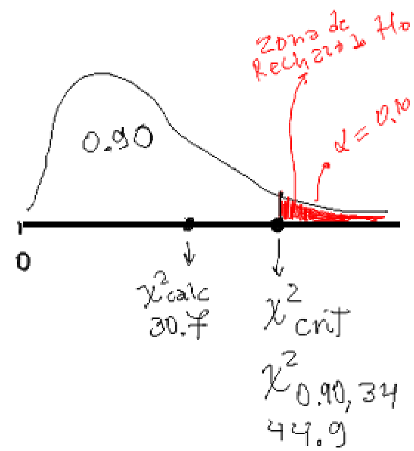
Tam. muestra

Varianza muestral

$$\chi^2_{calc} = \frac{(n-1)s^2}{\underbrace{\sigma_0^2}_{\text{Varianza hipotética}}} \sim \chi^2_{(n-1)}$$

Ejemplo

Para los mismos datos de tiempo de estudio, se sospecha que la varianza es mayor a 1 hora. ¿Se puede verificar dicha afirmación con un nivel de significancia del 10%?



$$H_0 : \boxed{\sigma^2 \leq 1} \qquad H_1 : \sigma^2 > 1 \qquad \alpha = 0.10$$

```r
(chicalc = (length(tiempo)-1)*var(tiempo)/1)
```

$\frac{(n-1)s^2}{1}$

```
[1] 30.70971
```

```r
(chicrit = qchisq(p = 0.90, df = length(tiempo)-1))
```

```
[1] 44.90316
```

```r
library(EnvStats)
```

```
varTest(x = tiempo, sigma.squared = 1, alternative = "greater")
```

σ_0^2 $>$

$statistic
Chi-Squared
30.70971 $\rightarrow \chi^2_{calc}$

$parameters
df
34 $\rightarrow$ gl

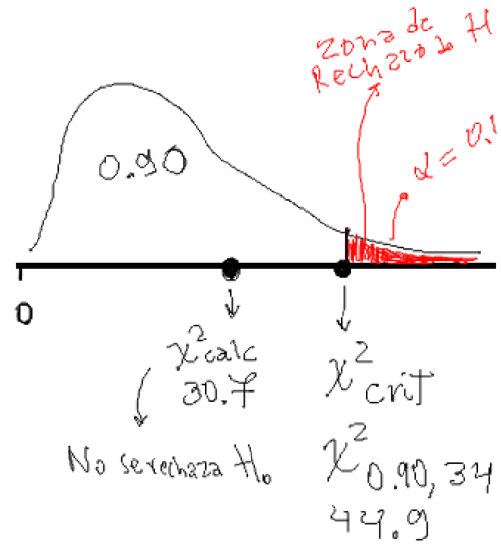
$p.value
[1] 0.6296898 $\rightarrow$ pvalor

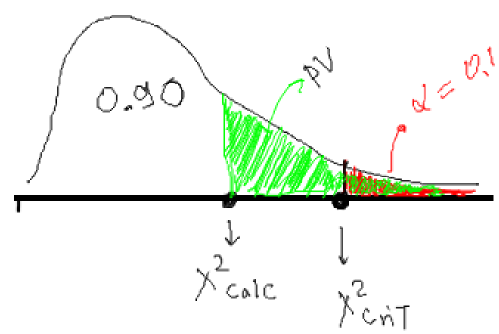
$estimate
variance
0.9032269 $\rightarrow S^2$

$null.value
variance
1 $\rightarrow \sigma_0^2$

$alternative
[1] "greater" $\rightarrow$ $>$

$method
[1] "Chi-Squared Test on Variance"





$pv > \alpha \Rightarrow$ No se rechaza H_0

Prueba de hipótesis para una proporción

proporción muestral

proporción hipotética

$$Z_{calc} = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \sim N(0,1)$$

tamaño muestral

$n = 100$ $\quad$ $n = 100$

$\pi_0 = 0.80$ $\quad$ $1-\pi_0 = 0.20$

$n\pi_0 \geq 5$ $\quad$ $n(1-\pi_0) \geq 5$

Ejemplo

Una universidad sostiene que el 80% de los estudiantes están satisfechos con el servicio de la biblioteca. Se encuesta a 100 estudiantes al azar y 71 dicen estar satisfechos. Verificar si la proporción real difiere de 0.80, con un nivel de significancia del 5%.

$$H_0 : \pi = 0.80 \qquad H_1 : \pi \neq 0.80 \qquad \alpha = 0.05$$

```r
p <- 71/100
(Z_calc <- (p - 0.8) / sqrt(0.8 * (1 - 0.8) / 100))
```

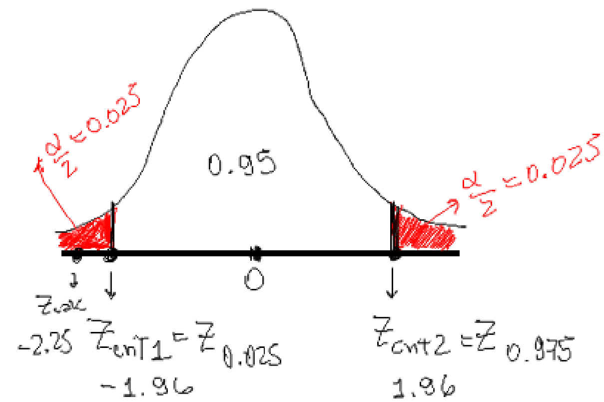
```
[1] -2.25
```

```r
(Z_crit1 <- qnorm(0.025))
```

```
[1] -1.959964
```

```r
(Z_crit2 <- qnorm(0.975))
```

```
[1] 1.959964
```

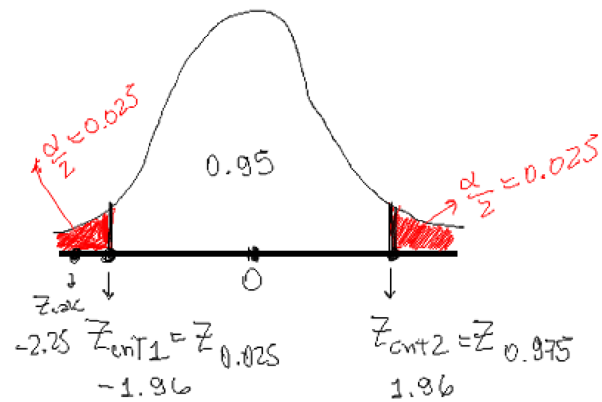


```
prop.test(x=71, n=100, p=0.80, alternative = "two.sided", correct = F)
```

muestra π_0 $\neq$ aprox. Normal

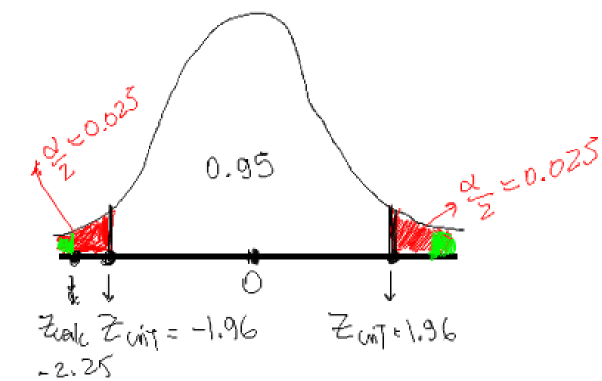
```
	1-sample proportions test without continuity correction

data:  71 out of 100, null probability 0.8
X-squared = 5.0625, df = 1, p-value = 0.02445
alternative hypothesis: true p is not equal to 0.8
95 percent confidence interval:
 0.6146111 0.7898516
sample estimates:
   p
0.71
```

pvalor



$Z_{calc} = -2.25 \rightarrow Z^2_{calc} = 5.0625$

PROPIEDAD: $Z^2 \sim \chi^2_{(1)}$



$\underbrace{pv < \alpha}_{}$ Rech. Ho

$0.0244 < 0.05$ $\Leftarrow$ $pv = 2P(Z < -2.25) = 0.0244$

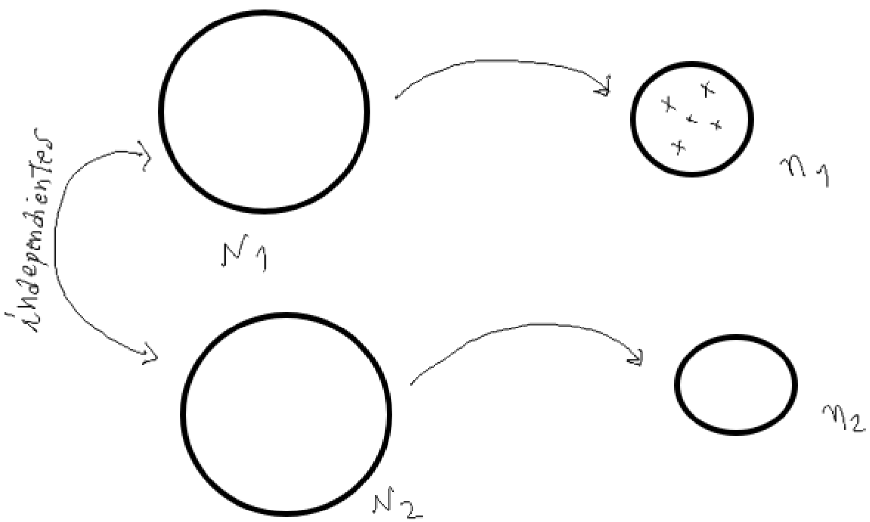
```
> 2*pnorm(-2.25)
[1] 0.02444895
```

G1: 10, 14, 16, 11, 15, 22, 9

G2: 10, 13, 17, 6, 15, 27, 9

} medias iguales, varianzas diferentes

Antes de comparar dos medias, se debe verificar si las varianzas son iguales o no, es decir si existe homogeneidad de varianzas o no.



emparejados o pareados

	ingresos padre	ingresos madre	dif
1	4000	4000	0
2	4500	0	4500
3	0	5000	-5000
4	2500	3500	-1000

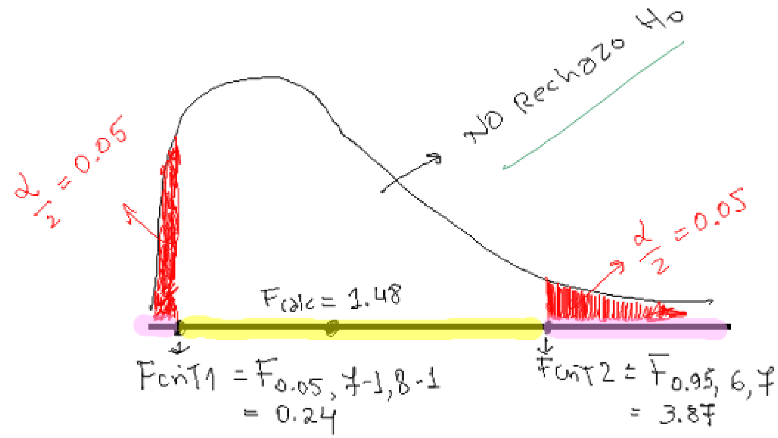
Prueba de hipótesis de homogeneidad de dos varianzas

= vs ≠

varianza muestral 1

$$F_{calc} = \frac{s_1^2}{s_2^2} \sim F_{(n_1-1, n_2-1)}$$

varianza muestral 2



Ejemplo

Un laboratorio desea determinar si la variabilidad de concentración (en mg/L) de un fármaco en la sangre es igual para dos fabricantes distintos, considerando un nivel de significancia del 10%. Los datos de concentración con el fabricante A son: `8.1, 7.9, 8.3, 7.8, 8.0, 8.2, 7.7`, mientras que con el fabricante B: `7.5, 7.2, 7.1, 7.4, 7.3, 7.6, 7.2, 7.5`.

$$\underbrace{H_0: \frac{\sigma_A^2}{\sigma_B^2} = 1}_{\text{var. homog}} \qquad \underbrace{H_1: \frac{\sigma_A^2}{\sigma_B^2} \neq 1}_{\text{var. heterog.}} \qquad \alpha = 0.10$$

```r
A <- c(8.1, 7.9, 8.3, 7.8, 8.0, 8.2, 7.7)
B <- c(7.5, 7.2, 7.1, 7.4, 7.3, 7.6, 7.2, 7.5)
(Fcalc <- var(A)/var(B))
```

```
[1] 1.484848
```

7-1 8-1

```r
(Fcrit1 <- qf(0.05, 6, 7))
```

```
[1] 0.2377184
```

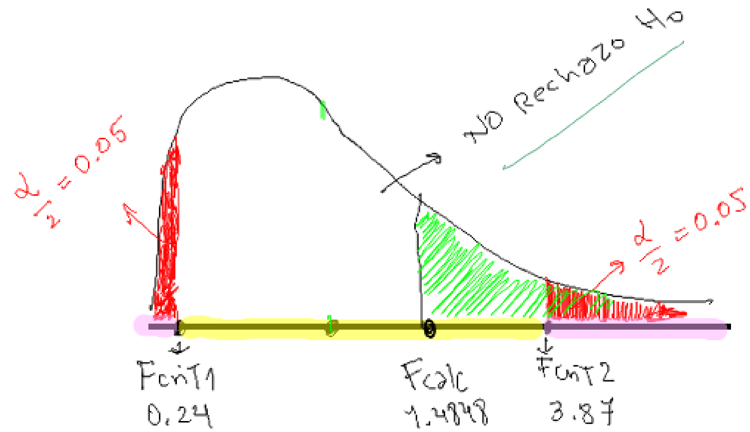
```r
(Fcrit2 <- qf(0.95, 6, 7))
```

```
[1] 3.865969
```

```r
var.test(A, B, alternative = "two.sided", ratio = 1)
```

F_{calc} $\quad n_1-1 \quad n_2-1 \quad$ pvalor

```
	F test to compare two variances

data:  A and B
F = 1.4848, num df = 6, denom df = 7, p-value = 0.6136
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.290089 8.456911
sample estimates:
ratio of variances
          1.484848
```



$pv > \alpha$

$0.61 > 0.10$

No se rechaza Ho

$pv = 2\,P(F > 1.4848)$

```
> 1-pf(1.4848,6,7)
[1] 0.3068118
> 2*(1-pf(1.4848,6,7))
[1] 0.6136236
```

Prueba de hipótesis para dos medias independientes

$$H_0 : \frac{\sigma_A^2}{\sigma_B^2} = 1 \qquad H_1 : \frac{\sigma_A^2}{\sigma_B^2} \neq 1 \Rightarrow \quad t = \frac{\bar{X}_1 - \bar{X}_2 - \mu_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_{\nu-Welch}$$

(t de Welch)

$$\Downarrow$$

$$t = \frac{\bar{X}_1 - \bar{X}_2 - \mu_0}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{\overbrace{(n_1+n_2-2)}^{\nu}} \quad \text{donde} \quad \underbrace{s_p^2}_{\text{varianza ponderada}} = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

(t de Student)

Prueba de hipótesis de homogeneidad de dos varianzas

paso 1:
Comparar σ_1^2 vs σ_2^2

paso 2: Comparar μ_1 vs μ_2

$\sigma_1^2 = \sigma_2^2 \longrightarrow$ para comparar las medias se usa t de Student

$\sigma_1^2 \neq \sigma_2^2 \longrightarrow$ para comparar las medias se usa t de <u>Welch</u>

Prueba de hipótesis para dos medias pareadas

dif. hipotética

$$H_0 : \mu_D \leq 3 \qquad H_1 : \mu_D > 3 \qquad \alpha = 0.05$$

despues - antes

```
antes   = c(120, 122, 121, 119, 118, 123, 121, 120, 122, 119, 115, 123)
despues = c(125, 127, 126, 124, 123, 129, 126, 125, 128, 123, 116, 129)
t.test(despues, antes, mu = 3, alternative = "greater", paired = T)
```

dif. hipotética $\qquad >$ $\qquad$ pareados o emparejados

```
    Paired t-test

data:  despues and antes
t = 4.7497, df = 11, p-value = 3e-04
alternative hypothesis: true mean difference is greater than 3
95 percent confidence interval:
 4.140136      Inf
sample estimates:
mean difference
       4.833333
```

Controles de reunión de pa...

```
> antes = c(120,122,121,119,118,123,121,120,122,119,115,123)
> despues   = c(125,127,126,124,123,129,126,125,128,123,116,129)
> t.test(despues, antes, mu = 3, alternative = "greater", paired = T)

        Paired t-test

data:  despues and antes
t = 4.7497, df = 11, p-value = 3e-04
alternative hypothesis: true mean difference is greater than 3
95 percent confidence interval:
 4.140136      Inf
sample estimates:
mean difference
       4.833333

> diferencia = despues - antes
> t.test(diferencia, mu = 3, alternative = "greater")

        One Sample t-test

data:  diferencia
t = 4.7497, df = 11, p-value = 3e-04
alternative hypothesis: true mean is greater than 3
95 percent confidence interval:
 4.140136      Inf
sample estimates:
mean of x
 4.833333
```

despues - antes

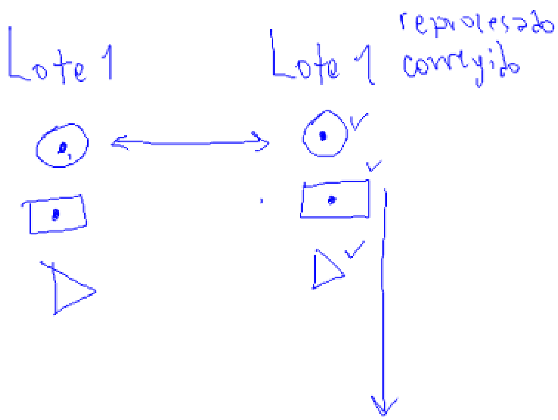
$H_0 : \mu_D \leq 3$ $\quad$ $H_1 : \mu_D > 3$ $\quad$ $\alpha = 0.05$

$pV = 0.0003$

$pV < \alpha$

$\Downarrow$

Se rechaza H_0


Lote 1
Lote 1
reprocesado
corregido

Ejemplo

Se desea comparar la proporción de hogares que hierven el agua antes de consumirla. En zona urbana: 90 de 120 hogares lo hacen; en zona rural: 80 de 110. ¿Las proporciones son las mismas?

$$H_0 : \pi_{urbana} - \pi_{rural} = 0 \qquad H_1 : \pi_{urbana} - \pi_{rural} \neq 0 \qquad \alpha = 0.05$$

cuando $\pi_0 = 0$

```
prop.test(x = c(x1, x2), n = c(n1, n2), alternative = "two.sided", correct = F)

	2-sample test for equality of proportions without continuity correction

data:  c(x1, x2) out of c(n1, n2)
X-squared = 0.15374, df = 1, p-value = 0.695
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.09097861  0.13643315
sample estimates:
   prop 1    prop 2
0.7500000 0.7272727
```

```
x1 <- 90; n1 <- 120; (p1 <- x1/n1)
```

```
[1] 0.75
```

```
x2 <- 80; n2 <- 110; (p2 <- x2/n2)
```

```
[1] 0.7272727
```

```
(p <- (x1+x2)/(n1+n2))
```

```
[1] 0.7391304
```

```
(zcalc <- (p1-p2)/sqrt(p*(1-p)*(1/n1+1/n2)))
```

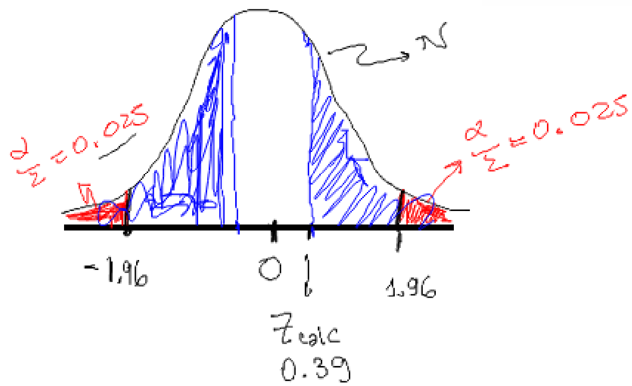
```
[1] 0.3921012
```

```
(zcrit1 <- qnorm(0.025))
```

```
[1] -1.959964
```

```
(zcrit2 <- qnorm(0.975))
```

```
[1] 1.959964
```



No se rechaza H0, por lo tanto **las proporciones** de hogares que hierven el agua antes de consumirla **son estadísticamente iguales**

pvalor $= P(Z > 0.39) = 0.695$ ✓

```
> pnorm(0.3921, lower.tail = FALSE)*2
[1] 0.6949843
```

```
prop.test(x = c(x1, x2), n = c(n1, n2), alternative = "two.sided", correct = F)
```

```
	2-sample test for equality of proportions without continuity correction

data:  c(x1, x2) out of c(n1, n2)
X-squared = 0.15374, df = 1, p-value = 0.695
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.09097861  0.13643315
sample estimates:
   prop 1    prop 2
0.7500000 0.7272727
```

$Z_{calc} = 0.3921 \sim N(0,1)$

PROPIEDAD: Si $A \sim N(0,1) \Rightarrow A^2 \sim \chi^2_{(1)}$

$\Rightarrow Z = 0.3921 \sim N(0,1) \Rightarrow A^2 = 0.15374 \sim \chi^2_{(1)}$

Si $A_i \sim N(0,1)$, $i=1,\dots,N$

$\Rightarrow \sum_{i=1}^{N} A_i^2 \sim \chi^2_{(n)}$

Ejemplo

Un equipo de especialistas en gestión ambiental desea evaluar si una campaña de sensibilización ambiental logra aumentar sustancialmente la proporción de hogares que clasifican adecuadamente sus residuos sólidos.

2 - En el barrio sin campaña, 45 de 100 hogares clasifican correctamente.
1 - En el barrio con campaña, 70 de 100 hogares lo hacen.

El equipo busca determinar si la proporción de hogares que clasifican adecuadamente sus residuos sólidos en el barrio con campaña supera en más de un 20% a la del barrio sin campaña, lo cual justificaría su implementación a mayor escala. Para ello, se emplea un nivel de significancia del 5%

$$H_0 : \pi_{con} - \pi_{sin} \leq 0.20 \qquad H_1 : \pi_{con} - \pi_{sin} > 0.20 \qquad \alpha = 0.05$$

$$H_0 : \pi_{con} - \pi_{sin} \leq 0.20 \qquad H_1 : \pi_{con} - \pi_{sin} > 0.20 \qquad \alpha = 0.05$$

```r
x1 <- 70; n1 <- 100; (p1 <- x1/n1)
```

```
[1] 0.7
```

```r
x2 <- 45; n2 <- 100; (p2 <- x2/n2)
```

```
[1] 0.45
```

```r
(zcalc <- (p1-p2)/sqrt(p1*(1-p1)/n1+p2*(1-p2)/n2))
```

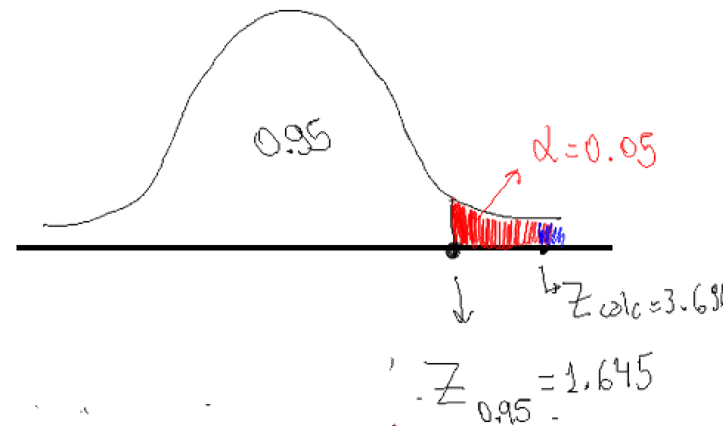
```
[1] 3.696106
```

```r
(zcrit2 <- qnorm(0.95))
```

```
[1] 1.644854
```

```r
(pv <- 1-pnorm(zcalc))
```

```
[1] 0.0001094656
```



$\alpha = P(Z > Z_{0.95})$

$pV = P(Z > Z_{calc})$

$PV < \alpha$

Rechazar H0

Sí existe evidencia estadística para afirmar que en el barrio con campaña de sensibilización ambiental, la proporción supera en más del 20% al barrio sin campaña