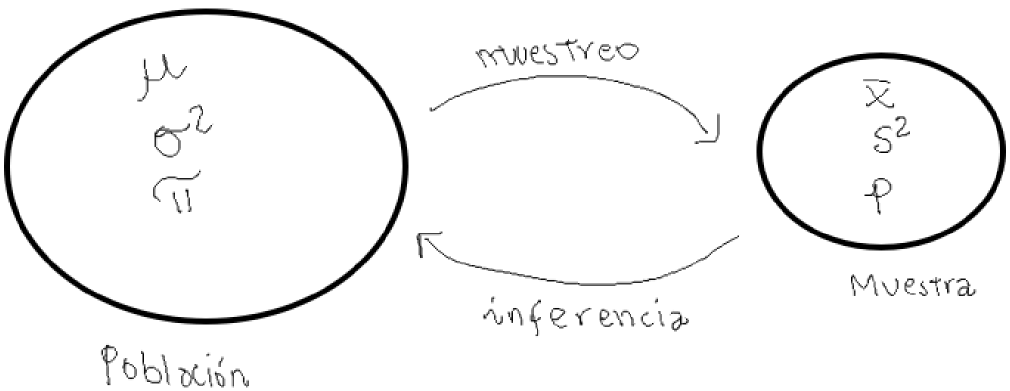


- Paramétrica (sujeta a supuestos) *
  - 1 par.
  - 2 par.
- No paramétrica (no se cumple algún supuesto)

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 → $n_{eff} \downarrow$

$\checkmark\checkmark$ $t \to \mu$ desc, σ^2 desconocida

• $N \to \mu$ desc y σ^2 conocida

media muestral

media hipotética (constante)

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

desv. est. muestral

tamaño muestral

Decisión:

* t_{calc} vs t_{crit}
* α vs pv

} equivalentes

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

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

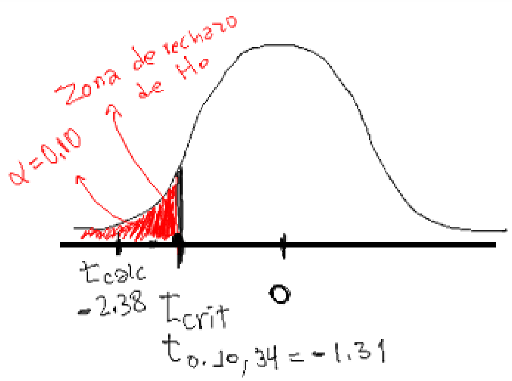
Ejemplo

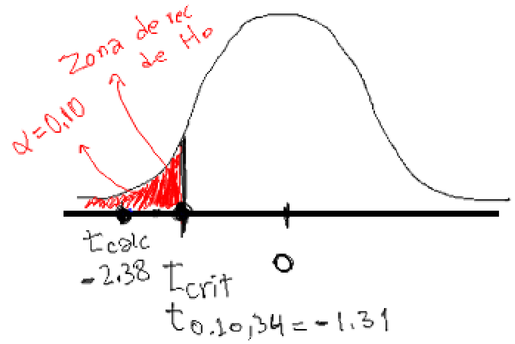
Una universidad afirma que sus estudiantes dedican en promedio 4 horas diarias al estudio. Se sospecha que el promedio es realmente menor, por lo que se toma una muestra aleatoria de 35 estudiantes, quienes reportan las siguientes horas de estudio:

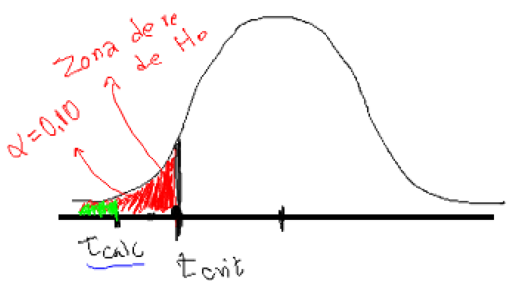
`2.8, 3.9, 4.2, 2.1, 3.7, 4.5, 3.8, 4.1, 4.6, 3.6, 4.3, 4.0, 2.9, 1.4, 3.9, 4.5, 4.2, 4.0, 3.8, 1.0, 4.7, 3.5, 2.1, 2.2, 3.6, 4.3, 4.8, 4.2, 3.9, 3.7, 4.4, 3.0, 3.8, 4.6, 2.5`

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 <

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

"greater" >

"two.sided" ≠

tcalc 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%?

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

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

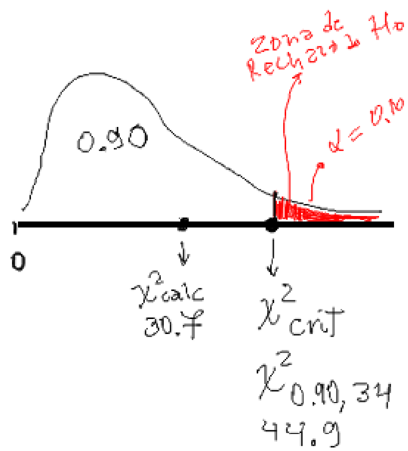
$\frac{(n-1)s^2}{1}$

```
[1] 30.70971
```

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

```
[1] 44.90316
```

```
library(EnvStats)
```



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

σ_0^2 $\quad$ $>$

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

$parameters
df
34 $\rightarrow$ gl

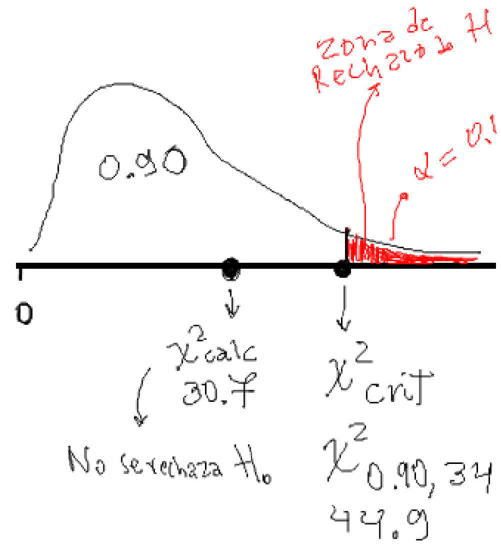
$p.value
[1] 0.6296898 $\rightarrow$ pvalor

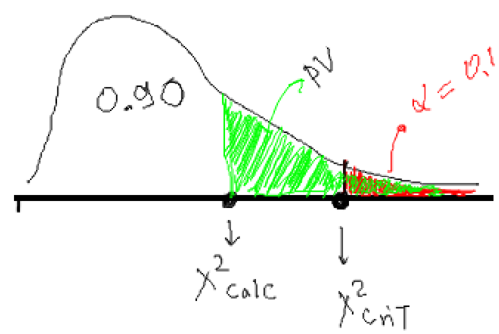
$estimate
variance
0.9032269 $\rightarrow s^2$

$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

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

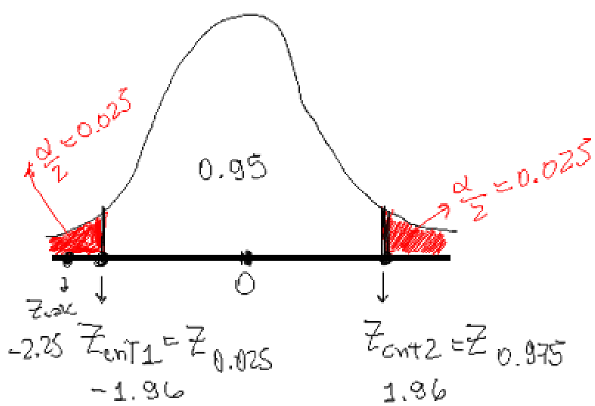
$\pi_0 = 0.80$ $\quad$ $1-\pi_0 = 0.20$

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

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%.

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




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

```
[1] -2.25
```

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

```
[1] -1.959964
```

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

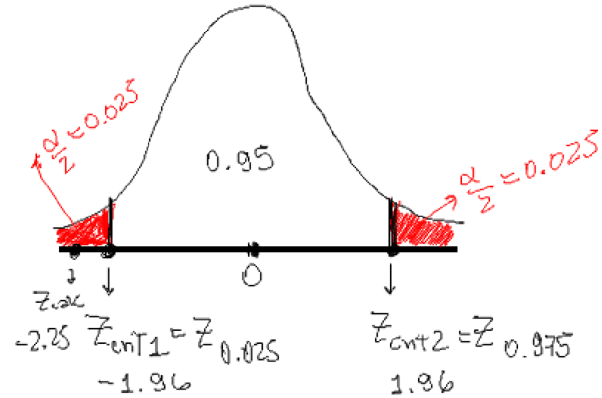
```
[1] 1.959964
```

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

muestra $\quad \pi_0 \quad \neq \quad$ aprox. Normal

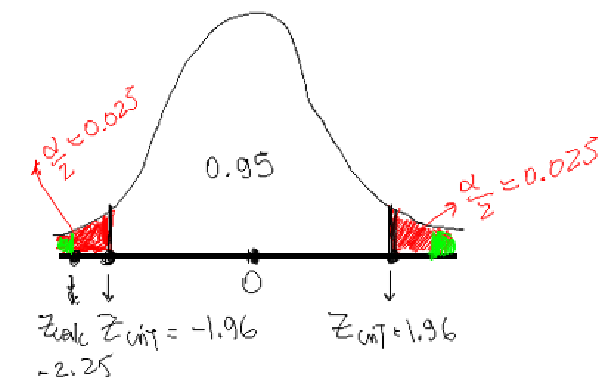
```
	1-sample proportions test without continuity correction

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



Rech. Ho

$\underbrace{pv < \alpha}$

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

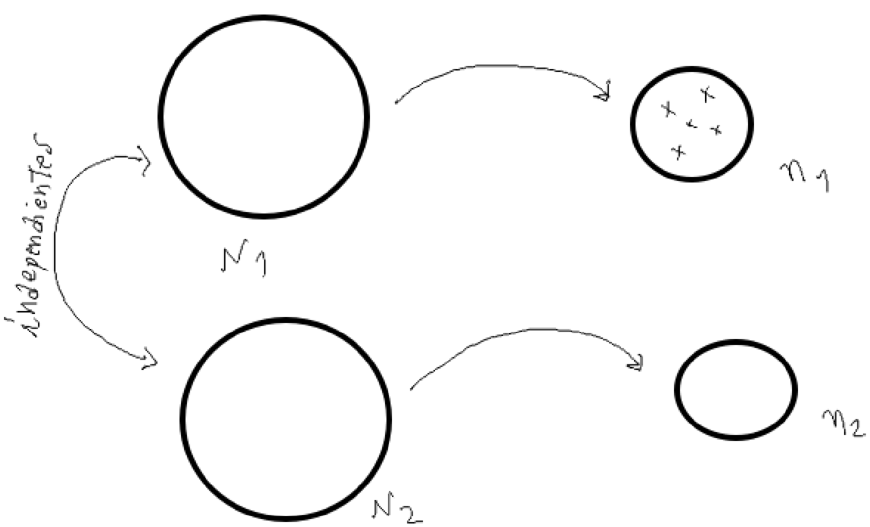
```
> 2*pnorm(-2.25)
[1] 0.02444895
```

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

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

$\Big\}$ 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 pareadas

	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

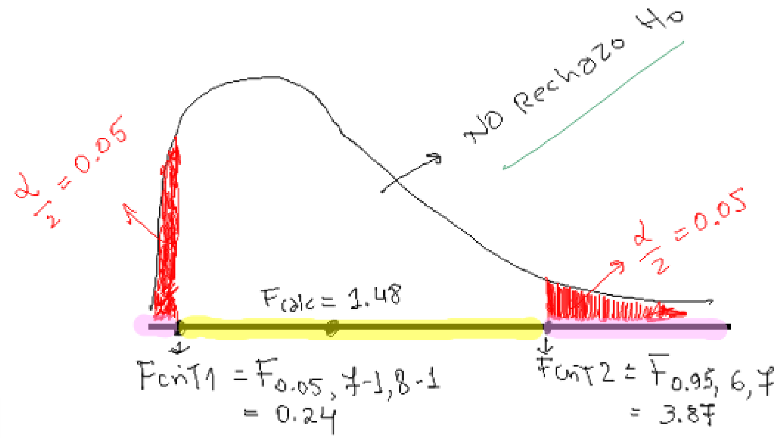
$$F_{calc} = \frac{s_1^2}{s_2^2} \sim F_{(n_1-1,n_2-1)}$$

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`.

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




```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))
```

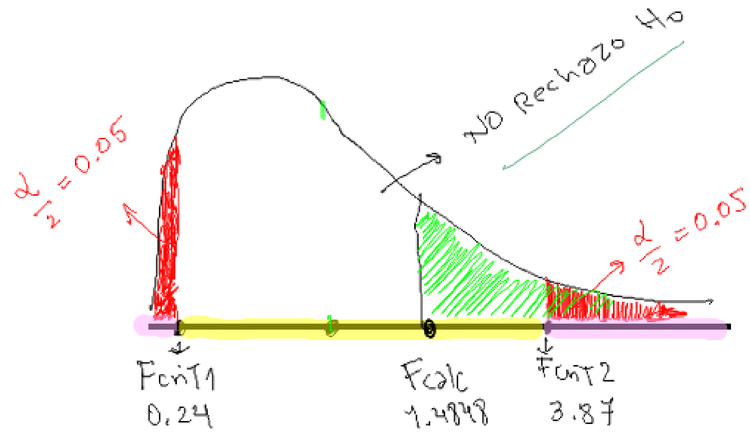
```
[1] 0.2377184
```

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

```
[1] 3.865969
```

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

```
    F test to compare two variances

                  Fcalc            n1-1                 n2-1             pvalor
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

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

```r
> 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

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

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 > 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

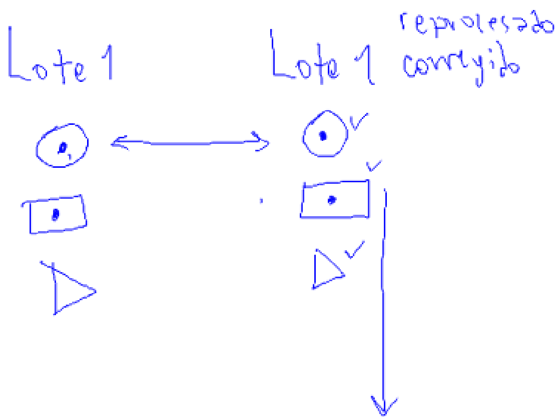
$H_0 : \mu_D \leq 3$ $\quad$ $H_1 : \mu_D > 3$ $\quad$ $\alpha = 0.05$

$pV = 0.0003$

$pV < \alpha$

$\Downarrow$

Se rechaza H_0


Lote 1
Lote 1
reprocesado
corregido