H_o

Hipótesis nula: $=\ ,\ \geq\ ,\ \leq$ $\qquad = \qquad =$ Hipótesis

Hipótesis alterna o de investigación: $\neq\ ,\ <\ ,\ >$ $\qquad < \qquad >$ Parámetros

H_1

* $H_o : \bar{X} \not> 5$ es incorrecto porque $\bar{X}$ es un estimador. Además, el signo es incorrecto

μ

Hipótesis unilateral con cola a la derecha	Hipótesis bilateral o de dos colas	Hipótesis unilateral con cola a la izquierda
$H_0 : \theta \leq \theta_0$ $H_1 : \theta > \theta_0$	$H_0 : \theta = \theta_0$ $H_1 : \theta \neq \theta_0$	$H_0 : \theta \geq \theta_0$ $H_1 : \theta < \theta_0$

Valor numérico

Parámetro

Aplicación 1

La gerencia de producción ha establecido que, para mantener la competitividad en el mercado, el tiempo promedio de producción por unidad debe ser de 45 minutos o menos. Sin embargo, han recibido reportes de que alguno de los equipos podría estar tardando más de lo esperado, lo que afectaría los tiempos de entrega y la satisfacción del cliente.

¿Los equipos realmente están cumpliendo con el estándar de 45 minutos, o en alguno (o ambos) los tiempos son mayores y se requiere una intervención?

nivel de significancia

$$H_0 : \mu \leq 45 \qquad H_1 : \mu > 45 \qquad \boxed{\alpha = 0.05}$$

muestra	H_0 falsa	H_0 verdadera
Rechazar H_0	correcto	Error Tipo I
No rechazar H_0	Error Tipo II	correcto

$$* \ \alpha = P(\underbrace{\text{Rechazar } H_0 \mid H_0 \text{ verdadera}}_{\text{Error Tipo I}}) \quad 0.01,\ 0.05,\ 0.10$$

$$\beta = P(\underbrace{\text{No rechazar } H_0 \mid H_0 \text{ Falsa}}_{\text{Error Tipo II}}), \quad 1-\beta = \text{Potencia de prueba}$$

Pasos:

1. Planteamiento: Hipótesis y Nivel de significancia
2. Desarrollo: Estadístico de prueba (en base a la evidencia muestral)
3. Decisión: pvalor vs alfa. Rechazar H0 si pvalor < alfa (n.s.)
4. Conclusión

1. $H_0 : \mu \leq 45 \qquad H_1 : \mu > 45 \qquad \alpha = 0.05$

2.

```
datos |>
  filter(Grupo == "Equipo 1") |>
  pull(Tiempo) |>
  t.test(alternative = "greater", mu = 45)
```

$<$: less

$\neq$: Two sided

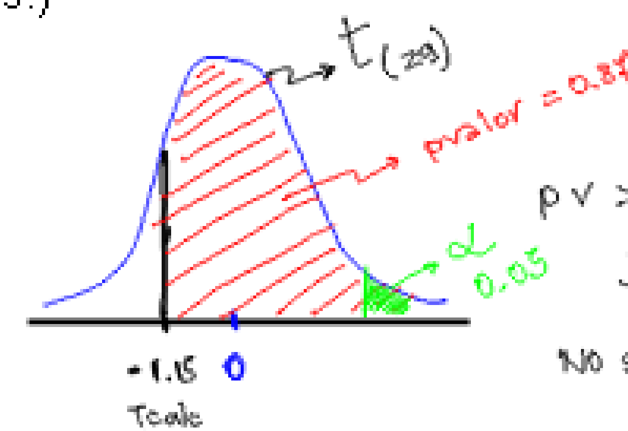
valor hipotético

Tcalc (Estad prueba)

degrees of freedom = grados de libertad

```
One Sample t-test

data:  pull(filter(datos, Grupo == "Equipo 1"), Tiempo)
t = -1.1487, df = 29, p-value = 0.87
alternative hypothesis: true mean is greater than 45
95 percent confidence interval:
 42.66133      Inf
sample estimates:
mean of x
 44.05667
```



$pv > \alpha$

$\Downarrow$

No se rechaza H_0

$\Downarrow$

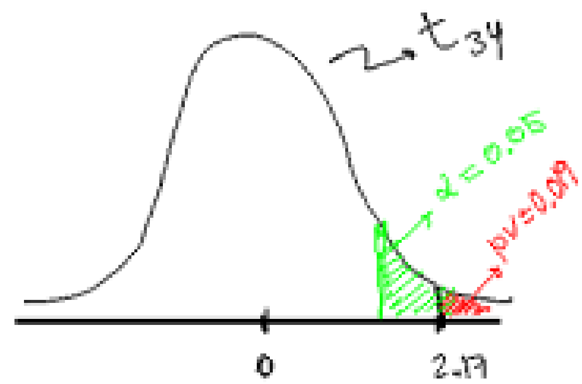
$\mu \leq 45$

```r
datos |>
  filter(Grupo == "Equipo 2") |>
  pull(Tiempo) |>
  t.test(alternative = "greater", mu = 45)
```

Test (estad. muestral)

```
One Sample t-test

data:  pull(filter(datos, Grupo == "Equipo 2"), Tiempo)
t = 2.17, df = 34, p-value = 0.01855
alternative hypothesis: true mean is greater than 45
95 percent confidence interval:
 45.44277      Inf
sample estimates:
mean of x
 47.00571
```



$pv < \alpha \Rightarrow$ Rechazo H_0

Aplicación 2

$\sigma < 3 \Rightarrow \sigma^2 < 9$

La gerencia ha establecido que la desviación estándar debe ser menor a 3 dólares para ser considerada aceptable. Sin embargo, hay indicios de que el equipo 2 está experimentando una variabilidad en costos fuera de lo esperado, lo que puede afectar la planificación financiera y el control de presupuesto. Verificar esta afirmación.

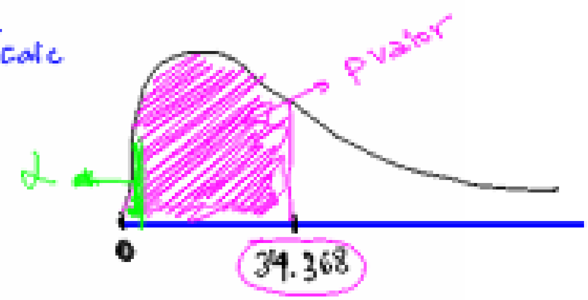
$$H_0 : \sigma^2 \geq 9 \qquad H_1 : \sigma^2 < 9 \qquad \alpha = 0.05$$

```
library(EnvStats)
datos |>
  filter(Grupo == "Equipo 2") |>
  pull(Costo) |>
  varTest(alternative = "less", sigma.squared = 9)
```

$<$ $\qquad$ σ^2 hipotético (H_0)

Results of Hypothesis Test

Null Hypothesis:	variance = 9	$H_0: \sigma^2 \geq 9$
Alternative Hypothesis:	True variance is less than 9	$H_1: \sigma^2 < 9$
Test Name:	Chi-Squared Test on Variance	
Estimated Parameter(s):	variance = 9.097422 $= s^2$ (varianza muestral)	
Data:	pull(filter(datos, Grupo == "Equipo 2"), Costo)	
Test Statistic:	Chi-Squared = 34.36804 $= \chi^2_{calc}$	
Test Statistic Parameter:	df = 34 → grados de libertad de χ^2	
P-value:	0.549879 ← p-valor	
95% Confidence Interval:	LCL = 0.00000 UCL = 14.27753 } IC a la izquierda (porque H1 es hacia la izq.)	



p valor > α ⇒ No se rechaza H_0

Aplicación 3

La empresa de manufactura tiene un estándar de calidad que establece que el porcentaje de productos defectuosos debe ser menor al 3%. Sin embargo, hay sospechas de que se podría estar generando una tasa de defectos diferente a la esperada, lo que podría afectar la satisfacción del cliente y aumentar los costos de reproceso.

$$H_0 : \pi \geq 0.03 \qquad H_1 : \pi < 0.03 \qquad \alpha = 0.05$$

```r
datos |> nrow() -> n
datos |> filter(Defectuoso == 1) |> nrow() -> x
prop.test(x, n, p = 0.03, alternative = "less", correct = FALSE)
```

→ tamaño de muestra

→ # de defectuosos

3 65 valor hipotético (π) <

```
	1-sample proportions test without continuity correction

data:  x out of n, null probability 0.03
X-squared = 0.58287, df = 1, p-value = 0.7774
alternative hypothesis: true p is less than 0.03
95 percent confidence interval:
 0.0000000 0.1099856
sample estimates:
         p
0.04615385
```

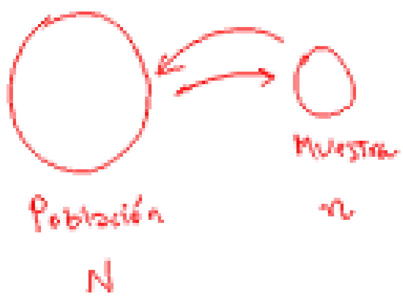
$> 0.05 \Rightarrow$ No se rechaza H_0

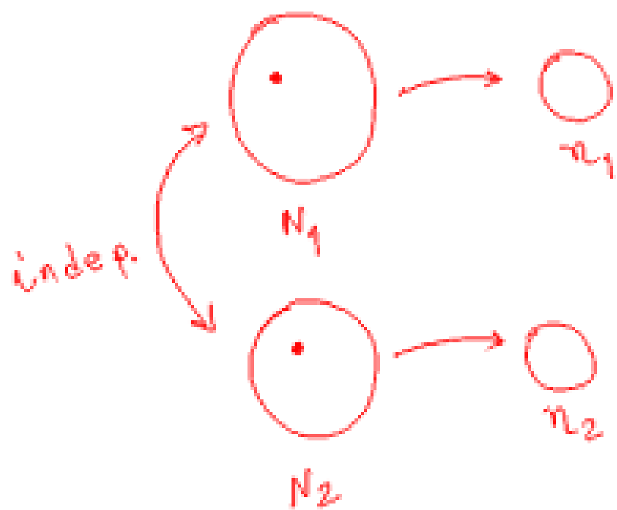
α

$$Z_{calc} = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{3/65 - 0.03}{\sqrt{\frac{0.03 \times 0.97}{65}}} = 0.76346$$

$$Z \sim N(0,1) \rightarrow Z^2 \sim \chi^2_{(1)}$$

$$0.76346^2 = 0.58287$$





Pareados

Estudiante	Edad Madre	Edad Padre
1	41	46
2	32	30
3	38	49

Independientes

Equipo1	Equipo2
20	19
15	19
14	25
18	11

Aplicación 4

En la empresa de manufactura, ambos equipos de producción que fabrican el mismo producto. La gerencia de calidad ha detectado que uno de los equipos podría estar generando más productos defectuosos que el otro, lo que podría afectar la rentabilidad y la satisfacción del cliente.

¿Existe una diferencia significativa en la tasa de defectos entre los dos equipos o las diferencias muestrales se deben al azar?

$$H_0 : \pi_1 - \pi_2 = 0 \qquad H_1 : \pi_1 - \pi_2 \neq 0 \qquad \alpha = 0.05$$

<u>Muestra</u>:

$$\begin{array}{l|c|c|c} p_1 = 0.17 & 0.15 & 0.165 & 0.17 \\ p_2 = 0.16 & 0.16 & 0.17 & 0.18 \end{array} \Bigg\} \ \pi_1 = \pi_2$$

```r
datos |> group_by(Grupo) |> count(Defectuoso) |>
  filter(Defectuoso == 1) |> pull(n) -> x
x
```

Sí defectuosos

```
[1] 1 2
```

→ Vector de defectuosos

```r
datos |> count(Grupo) |>
  pull(n) -> n
n
```

```
[1] 30 35
```

→ Vector de tamaños de muestra

1° equipo: $p_1 = 1/30$

2° equipo: $p_2 = 2/35$

$\Big\}$ ¿$\pi_1 = \pi_2$?

```
prop.test(x, n, alternative = "two.sided")
```

x: 1 2 $\quad$ n: 30 35 $\quad$ $\neq$

```
    2-sample test for equality of proportions with continuity correction

data:  x out of n
X-squared = 8.8709e-31, df = 1, p-value = 1
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.1478158  0.1001968
sample estimates:
    prop 1     prop 2
0.03333333 0.05714286
```

$> \alpha \Rightarrow$ No se rechaza H_0

$$\Downarrow$$

$$\pi_1 - \pi_2 = 0$$

$$\Downarrow$$

$$\pi_1 = \pi_2$$

p_1 $\quad$ p_2

$p_1 = 1/30 = 0.033$

$p_2 = 2/35 = 0.057$

prop.test(x,n,...)
- Si x y n son números, por ejemplo **x = 10** y **n = 80**, entiende que es una muestra.
- Si x y n son vectores de dos elementos, por ejemplo **x = c(14,10)** y **n = c(100,120)**, entiende que se trata de dos muestras.

μ_1 vs μ_2

Cuando las varianzas son homogéneas

grupo 1 = 12,15,14,15,18,14,17,16
grupo 2 = 25,28,25,24,25,22,27,26

Cuando las varianzas son heterogéneas

grupo 1 = 12,15,14,15,18,14,17,16
grupo 2 = 5,28,25,64,25,12,47,26

Aplicación 5

Existen indicios de que uno de los equipos podría estar tardando más en completar sus tareas, lo que impactaría la eficiencia y los tiempos de entrega. ¿El tiempo promedio de producción por unidad es el mismo en ambos equipos, o hay una diferencia significativa?

Sol.

Primero se verificará si las varianzas son iguales

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1 \qquad H_1 : \frac{\sigma_1^2}{\sigma_2^2} \neq 1 \qquad \alpha = 0.05$$

$H_0: \sigma_1^2 = \sigma_2^2$ $\qquad$ $H_1: \sigma_1^2 \neq \sigma_2^2$

VARIANZAS HOMOGÉNEAS $\qquad$ VARIANZAS HETEROGÉNEAS

Y   Variable de grupo   data frame

```r
var.test(Tiempo ~ Grupo, datos, alternative = "two.sided")
```

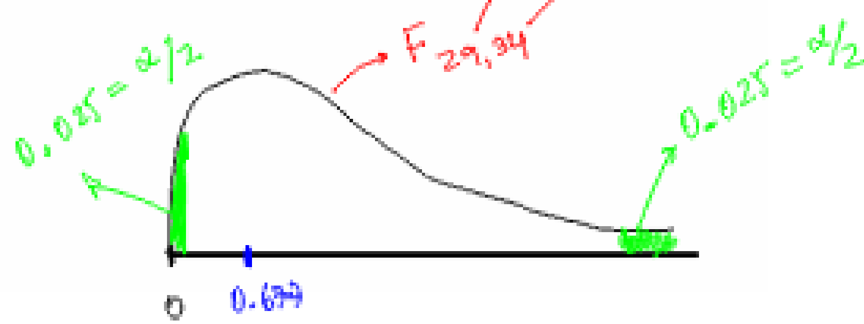
Eq.1   Eq.2   ≠

```
    F test to compare two variances

data:  Tiempo by Grupo
F = 0.67659, num df = 29, denom df = 34, p-value = 0.2865
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.3348314 1.3952971
sample estimates:
ratio of variances
         0.6765865
```

F calc

30-1   35-1



$pv > \alpha$

$0.2865 > 0.05 \Rightarrow$ No Rech H_0

$\Downarrow$

VARIANZAS HOMOG.

Ahora, se prueba la diferencia de medias considerando que las varianzas son homogéneas:

$$H_0 : \mu_1 - \mu_2 = 0 \qquad H_1 : \mu_1 - \mu_2 \neq 0 \qquad \alpha = 0.05$$

$$H_0 : \mu_1 = \mu_2 \qquad H_1 : \mu_1 \neq \mu_2$$

```
    Two Sample t-test

data:  Tiempo by Grupo
t = -2.3495, df = 63, p-value = 0.02195
alternative hypothesis: true difference in means between group Equipo 1 ar
95 percent confidence interval:
 -5.4573575 -0.4407378
sample estimates:
mean in group Equipo 1 mean in group Equipo 2
              44.05667               47.00571
```

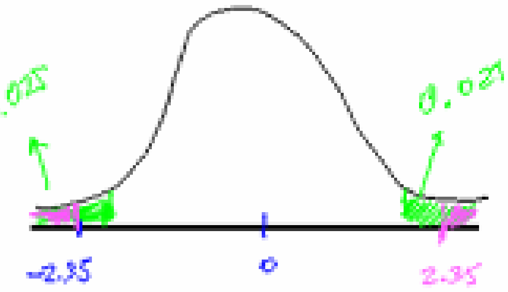
```
> t.test(Tiempo ~ Grupo, datos, alternative = "two.sided",
+        mu = 0, var.equal = TRUE)

        Two Sample t-test

data:  Tiempo by Grupo
t = -2.3495, df = 63, p-value = 0.02195
alternative hypothesis: true difference in means between group
p Equipo 2 is not equal to 0
95 percent confidence interval:
 -5.4573575 -0.4407378
sample estimates:
mean in group Equipo 1 mean in group Equipo 2
              44.05667               47.00571
```

T_{calc}

G.L.



pvalor

α

$pv < \alpha$

$0.02195 < \alpha$

$\Downarrow$

Se rechaza H_0

$\bar{X}_1$ $\bar{X}_2$

$\bar{X}_1 \longleftrightarrow \bar{X}_2$

$\Downarrow$

$\mu_1 \neq \mu_2$

```r
tiempo1 <- datos |> filter(Grupo == "Equipo 1") |> pull(Tiempo)
tiempo2 <- datos |> filter(Grupo == "Equipo 2") |> pull(Tiempo)
t.test(tiempo1, tiempo2, alternative = "two.sided", var.equal = TRUE)
```

```
	Two Sample t-test

data:  tiempo1 and tiempo2
t = -2.3495, df = 63, p-value = 0.02195
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -5.4573575 -0.4407378
sample estimates:
mean of x mean of y 
 44.05667  47.00571 
```

Se rechaza H_0, ya que $pv < \alpha$, lo que indica que hay diferencias significativas entre los tiempos promedio de producción de los equipos.

Tiempo 1 Tiempo2

FORMATO ANCHO

$<$

1. La empresa ha establecido que el tiempo promedio de entrega debe ser menor a 24 horas para cumplir con los estándares de servicio. Sin embargo, hay sospechas de que el centro 1 no está cumpliendo esta especificación, lo que podría generar retrasos en la entrega.

$$H_0: \mu \geq 24 \quad \rightarrow \quad \text{t.test}$$

$$H_1: \mu < 24$$

2. La empresa busca mantener estabilidad en los costos de distribución. Se ha establecido que la desviación estándar no debe ser mayor a 2 dólares. Hay indicios de que el centro 2 tiene costos con alta variabilidad. Verifique esta afirmación.

$$H_0: \sigma^2 \leq 4 \quad \rightarrow \quad \text{varTest}$$

$$H_1: \sigma^2 > 4$$

3. La empresa establece que la tasa máxima aceptable de entregas tardías es del 6%. Se sospecha que uno de los centros está superando esta tasa, lo que afectaría la percepción del servicio por parte de los clientes.

$$\left.\begin{aligned} H_0&: \pi \leq 0.06 \\ H_1&: \pi > 0.06 \end{aligned}\right\} \text{prop.test}(\underbrace{x, n}_{\text{números}}, \ldots)$$

4. ¿Existe una diferencia significativa en la tasa de entregas tardías entre los dos centros, o las diferencias observadas son solo aleatorias?

$$\left.\begin{aligned} H_0&: \pi_1 - \pi_2 = 0 \\ H_1&: \pi_1 - \pi_2 \neq 0 \end{aligned}\right\} \text{prop.test}(\underbrace{x, n}_{\text{vectores}}, \ldots)$$

5. ¿El tiempo promedio de entrega en el centro 1 es más de 3 horas mayor que en el otro, o la diferencia observada es solo variabilidad aleatoria?

$H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1$ $\quad H_1: \frac{\sigma_1^2}{\sigma_2^2} \neq 1 \quad \longrightarrow$ var.Test

$H_0: \mu_1 - \mu_2 \leq 3$ $\quad H_1: \mu_1 - \mu_2 > 3 \quad \longrightarrow$ t.test

$$\begin{array}{cc} 20 & 14 \\ \downarrow & \downarrow \\ c_1 & c_2 \end{array}$$

```
> t.test(Tiempo ~ Grupo, datos, alternative = "two.sided",
+        mu = 0, var.equal = TRUE)

        Two Sample t-test

data:  Tiempo by Grupo
t = -2.3495, df = 63, p-value = 0.02195
alternative hypothesis: true difference in means between group
p Equipo 2 is not equal to 0
95 percent confidence interval:
 -5.4573575 -0.4407378
sample estimates:
mean in group Equipo 1 mean in group Equipo 2
              44.05667               47.00571
```

Tiempo	Grupo
10	Eq 1
15	Eq 1
⋮	⋮
18	Eq2

FORMATO LARGO