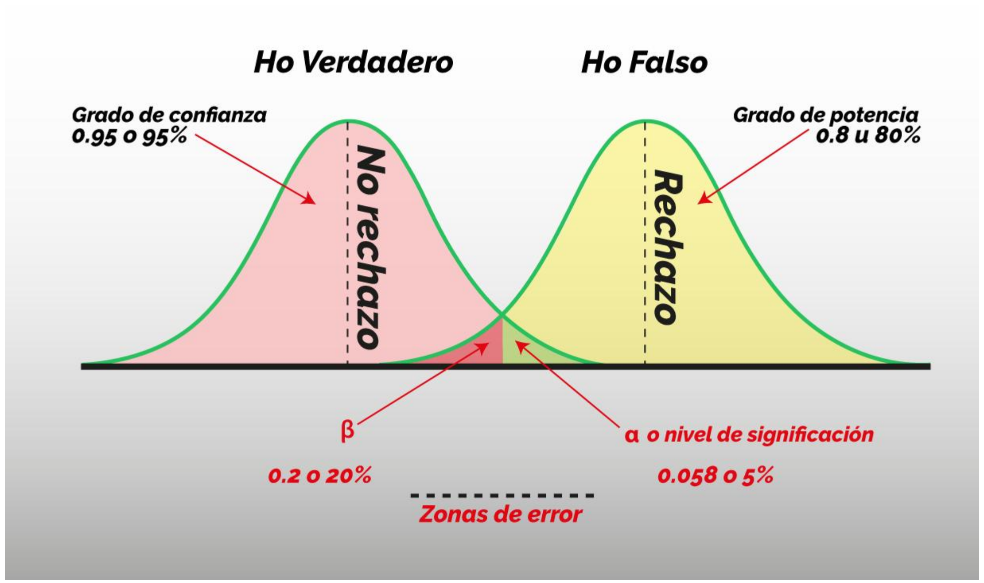
Test de hipótesis



Si beben de mi elixir ustedes adelgazarán de media 2 kg al mes.

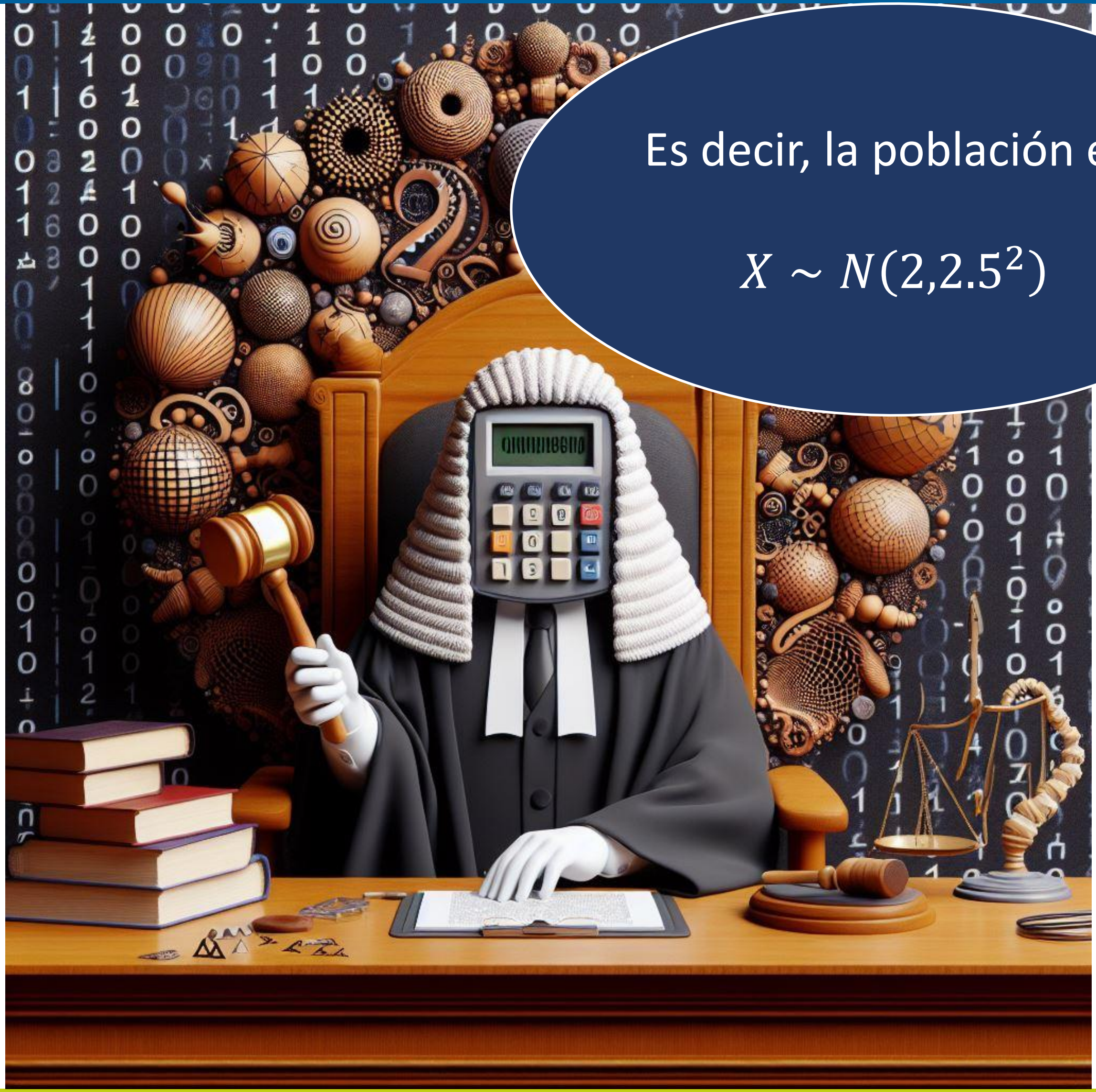
Va usted a ser juzgado por las matemáticas. ¿Jura decir la verdad?

Afirmo que $\mu = 2$.

En ese caso daremos a probar su elixir a 50 personas. Después, apuntaremos la pérdida de peso de todos.


30 Days Later

Los datos obtenidos siguen una distribución normal. Además, según expertos en la materia, es sabido $\sigma^2 = 2.5^2$

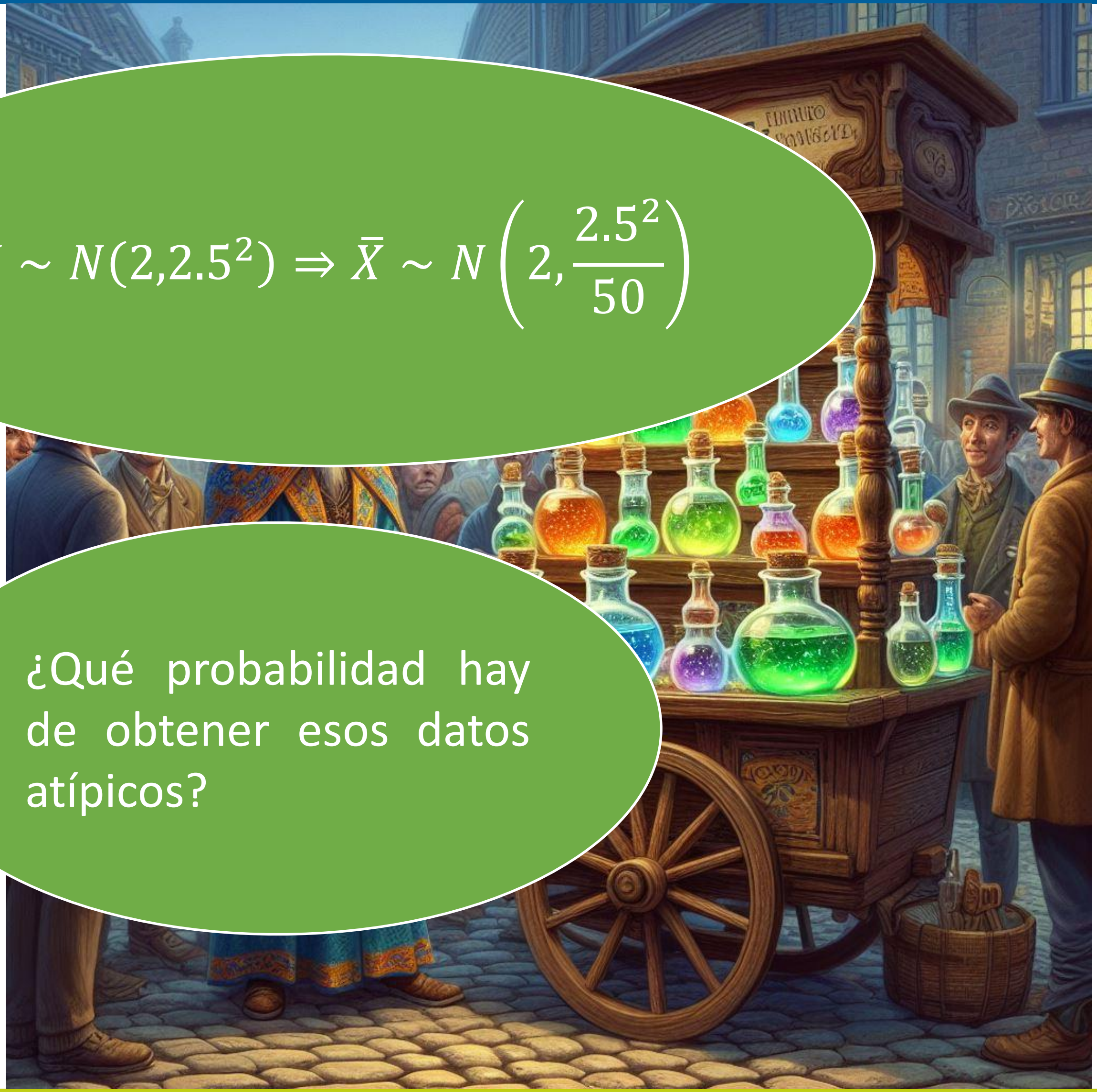


Es decir, la población es:

$$X \sim N(2{,}2.5^2)$$

¿Cómo explica usted que la media muestral sea $\bar{x} = 0.037$?

Un caso desafortunado, en esta ocasión hemos recogido 50 muestras atípicas…



$$X \sim N(2{,}2.5^2) \Rightarrow \bar{X} \sim N\left(2, \frac{2.5^2}{50}\right)$$

¿Qué probabilidad hay de obtener esos datos atípicos?

Le aseguro que es muy común obtener datos atípicos en estudios estadísticos…

$$P(\bar{X} \leq 0.037) \approx 0$$

¡Se rechaza la defensa!

Espero no haber cometido un error…

CEU Universidad San Pablo

Procedimiento

1. Formulamos nuestra hipótesis nula (lo contrario de H_a lo que queremos *probar*).

Ejemplo

Sabiendo que $X \sim N(\mu, 2.5^2)$:

$$H_0 : \mu = 2$$

Igualdad $\Rightarrow$ Dos colas

Procedimiento

1. Formulamos nuestra hipótesis nula (lo contrario de H_a lo que queremos *probar*).

Ejemplo

Sabiendo que $X \sim N(\mu, 2.5^2)$:

$$H_0 : \mu \leq 2$$

Desigualdad ⇒ Una cola

Procedimiento

1. Formulamos nuestra hipótesis nula (lo contrario de H_a lo que queremos *probar*).
2. Consideramos el estadístico de contraste T en relación con H_0.

Ejemplo

Sabiendo que $X \sim N(\mu, 2.5^2)$:

$$H_0 : \mu = 2$$

Consideramos el estadístico Z.

Procedimiento

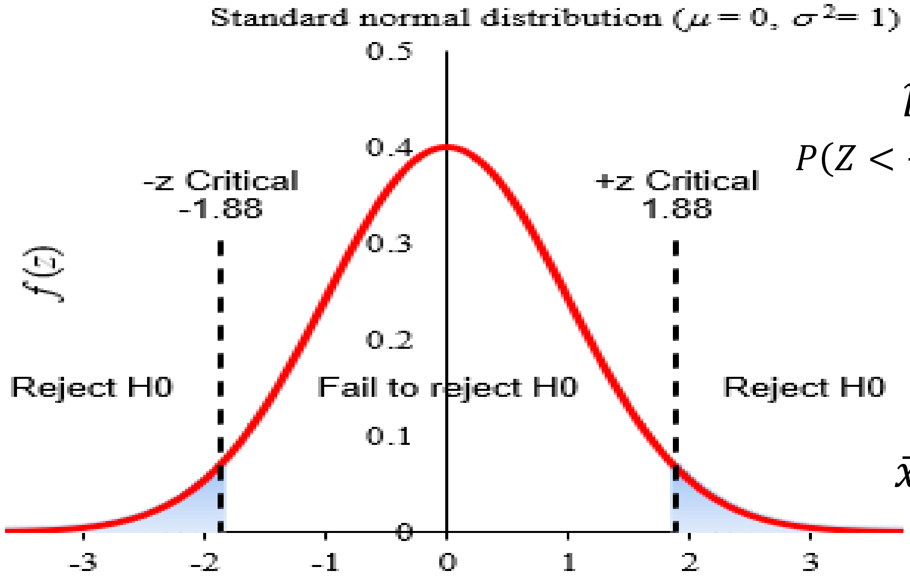
1. Formulamos nuestra hipótesis nula (lo contrario de H_a lo que queremos *probar*).
2. Consideramos el estadístico de contraste T en relación con H_0.
3. Calculamos el p-valor: Según H_0 calculamos la probabilidad de obtener un valor del estadístico como los datos indican.

Ejemplo

Sabiendo que $X \sim N(\mu, 2.5^2)$:

$$H_0 : \mu = 2$$

Consideramos el estadístico Z.



p −valor de dos colas

$P(Z < -5.552 \,\&\, Z > 5.552) = 2.8e-08$

$\bar{x} = 0.037 \Rightarrow z = -5.552$

Procedimiento

1. Formulamos nuestra hipótesis nula (lo contrario de H_a lo que queremos *probar*).
2. Consideramos el estadístico de contraste T en relación con H_0.
3. Calculamos el p-valor: Según H_0 calculamos la probabilidad de obtener un valor del estadístico como los datos indican.
4. Si p-valor $< \alpha$, entonces rechazamos la H_0 con una significancia del $100 \cdot \alpha$ %.

Ejemplo

Sabiendo que $X \sim N(\mu, 2.5^2)$:

$$H_0 : \mu = 2$$

Consideramos el estadístico Z.

$$\boxed{p = 2.8e - 08 < 0.01 \Rightarrow \text{SE RECHAZA } H_0}$$

Significancia del 1%

	H_0 es cierta	H_0 es falsa
H_0 no es rechazada	No error	Falso negativo (β)
H_0 es rechazada	Falso positivo (α)	No error

$$\begin{cases} \text{Confianza del test: } 1-\alpha \\ \text{Potencia del test: } 1-\beta \end{cases}$$

Tamaños

Tamaño del efecto	Uso	Pequeño	Mediano	Grande
Cohen’s h	Proporciones	0.20	0.50	0.80
η^2	ANOVA	0.01	0.06	0.14
ω^2	ANOVA	0.01	0.06	0.14
Cohen’s f	ANCOVA	0.10	0.25	0.40
Cohen’s d	T-test	0.20	0.50	0.80

Identificar hipótesis → **Calcular p-valor** → **Rechazar H_0** → **Informe detallado**

$$\begin{cases} H_0 \\ H_a \end{cases} \rightarrow p \rightarrow p < \alpha \rightarrow$$

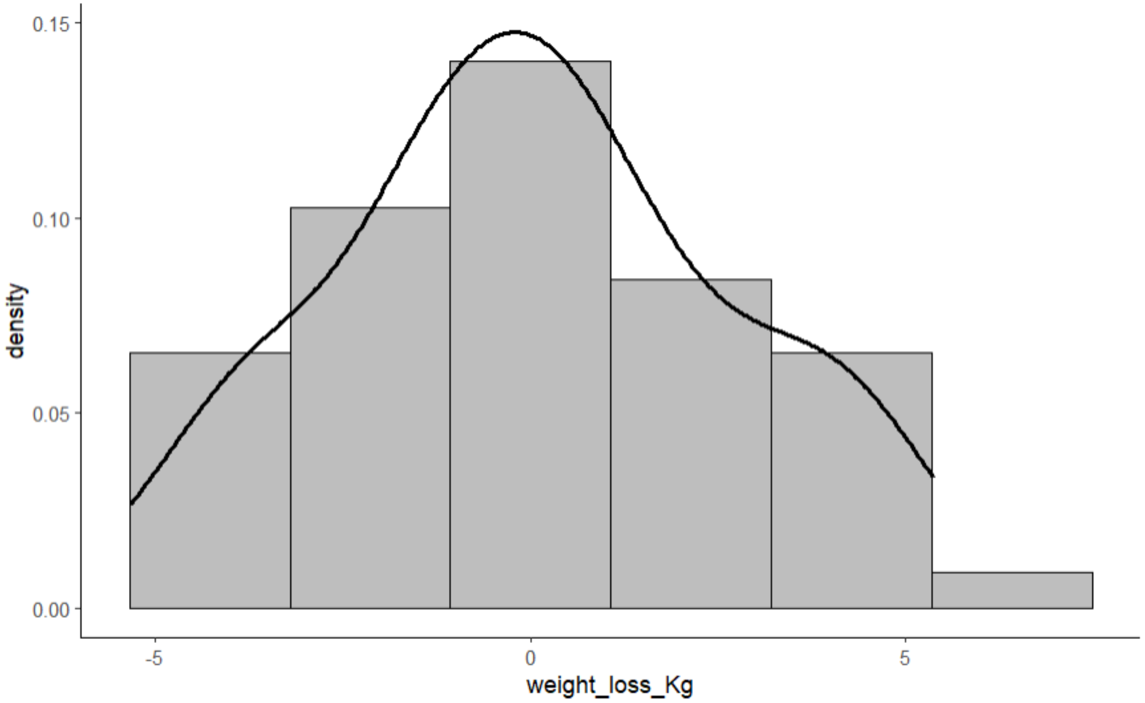
- p-valor
- significancia
- potencia
- tamaño de la muestra
- intervalo de confianza
- tamaño del efecto
- asunciones
- estadístico

En la práctica, normalmente tenemos acceso a los datos. Además, el ejemplo introductorio se basa en el conocimiento de otros parámetros poblacionales que habitualmente no conoceremos.

Ejemplo

Un producto homeopático afirma que "gracias a su uso, perderás 2kg en un mes". Escéptico ante esa afirmación, reclutas a 50 personas de tu ciudad para probar el producto, datos en "homeo_weight_los.csv". Realiza un test de hipótesis (sin conocer la desviación típica poblacional). *Utiliza t.test(), pwr.t.test() y effectsize::effectsize().*

Representa gráficamente los datos.



Observamos una normalidad en los datos.

Realiza un T-test.

```
        One Sample t-test

data:  data$weight_loss_Kg
t = -5.2553, df = 49, p-value = 3.205e-06
alternative hypothesis: true mean is not equal to 2
95 percent confidence interval:
 -0.7136358  0.7876358
sample estimates:
mean of x
    0.037
```

Datos para nuestro informe.

Criterios de calidad	Valores
normalidad	SI
Tamaño de muestra	50
Estadístico	??
P-valor < significancia	??
Intervalo de confianza	??
Tamaño del efecto	??
Potencia	??

Realiza un T-test.

```
        One Sample t-test

data:  data$weight_loss_Kg
t = -5.2553, df = 49, p-value = 3.205e-06
alternative hypothesis: true mean is not equal to 2
95 percent confidence interval:
 -0.7136358  0.7876358
sample estimates:
mean of x 
    0.037 
```

Datos para nuestro informe.

Criterios de calidad	Valores
normalidad	SI
Tamaño de muestra	50
Estadístico	-5.25
P-valor < significancia	??
Intervalo de confianza	??
Tamaño del efecto	??
Potencia	??

CEU Universidad San Pablo

Realiza un T-test.

```
        One Sample t-test

data:  data$weight_loss_Kg
t = -5.2553, df = 49, p-value = 3.205e-06
alternative hypothesis: true mean is not equal to 2
95 percent confidence interval:
 -0.7136358  0.7876358
sample estimates:
mean of x
    0.037
```

Datos para nuestro informe.

Criterios de calidad	Valores
normalidad	SI
Tamaño de muestra	50
Estadístico	-5.25
P-valor < significancia	3.2e-06 < 0.05
Intervalo de confianza	??
Tamaño del efecto	??
Potencia	??

CEU Universidad San Pablo

Realiza un T-test.

```
        One Sample t-test

data:  data$weight_loss_Kg
t = -5.2553, df = 49, p-value = 3.205e-06
alternative hypothesis: true mean is not equal to 2
95 percent confidence interval:
 -0.7136358  0.7876358
sample estimates:
mean of x
    0.037
```

Datos para nuestro informe.

Criterios de calidad	Valores
normalidad	SI
Tamaño de muestra	50
Estadístico	-5.25
P-valor < significancia	3.2e-06 < 0.05
Intervalo de confianza	(-0.71,0.79)
Tamaño del efecto	??
Potencia	??

Tamaño del efecto

```
Cohen's d |         95% CI
--------------------------
-0.74     | [-1.05, -0.43]

- Deviation from a difference of 2.
```

Datos para nuestro informe.

Criterios de calidad	Valores
normalidad	SI
Tamaño de muestra	50
Estadístico	-5.25
P-valor < significancia	3.2e-06 < 0.05
Intervalo de confianza	(-0.71,0.79)
Tamaño del efecto	0.74
Potencia	??

Potencia

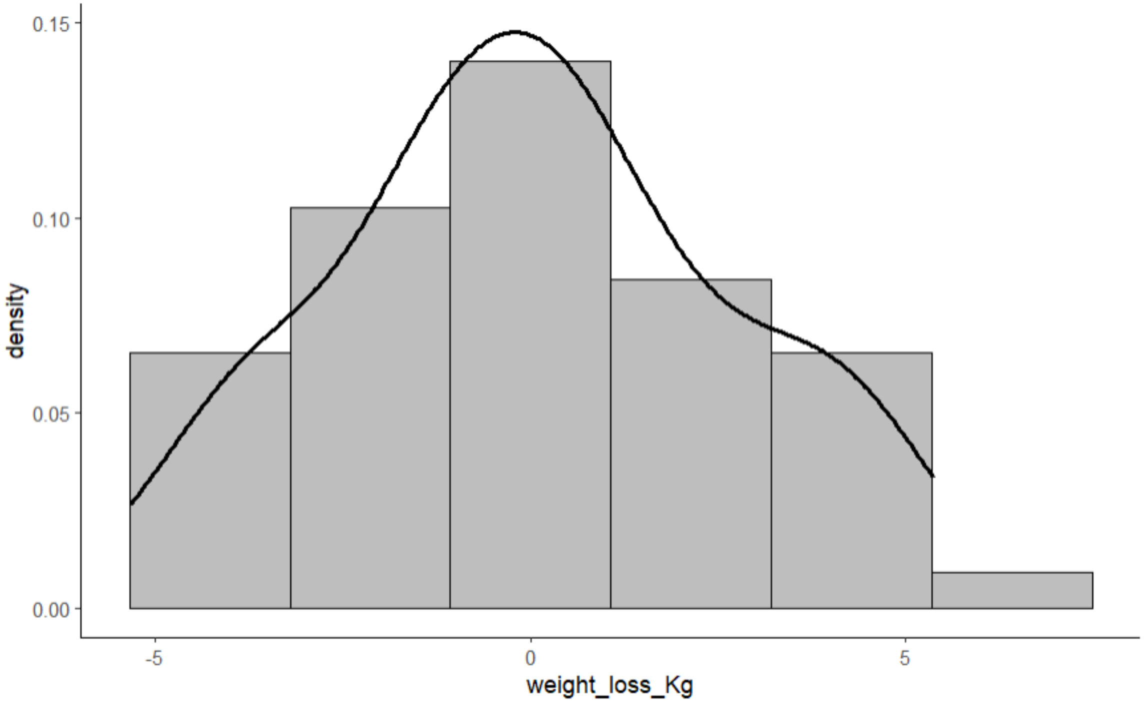
```
One-sample t test power calculation

              n = 50
              d = 0.7432079
      sig.level = 0.05
          power = 0.9992885
    alternative = two.sided
```

Datos para nuestro informe.

Criterios de calidad	Valores
normalidad	SI
Tamaño de muestra	50
Estadístico	-5.25
P-valor < significancia	3.2e-06 < 0.05
Intervalo de confianza	(-0.71,0.79)
Tamaño del efecto	0.74
Potencia	0.999

INFORME

Observamos una normalidad en los datos, por lo que nos disponemos a realizar un test T de Student para contrastar la hipótesis nula de que la media es igual a 2. Los resultados del T-test de dos colas sobre una muestra de 50 sujetos evidencian un rechazo sobre la hipótesis nula con un tamaño del efecto mediano en la media de pérdida de peso $[\, t = -5.25,\ p < 0.00001;\ d = 0.74 \,]$ con una potencia de 0.999. De hecho, hemos obtenido el intervalo $(-0.71, 0.79)$ para la media con una significancia de 0.05.



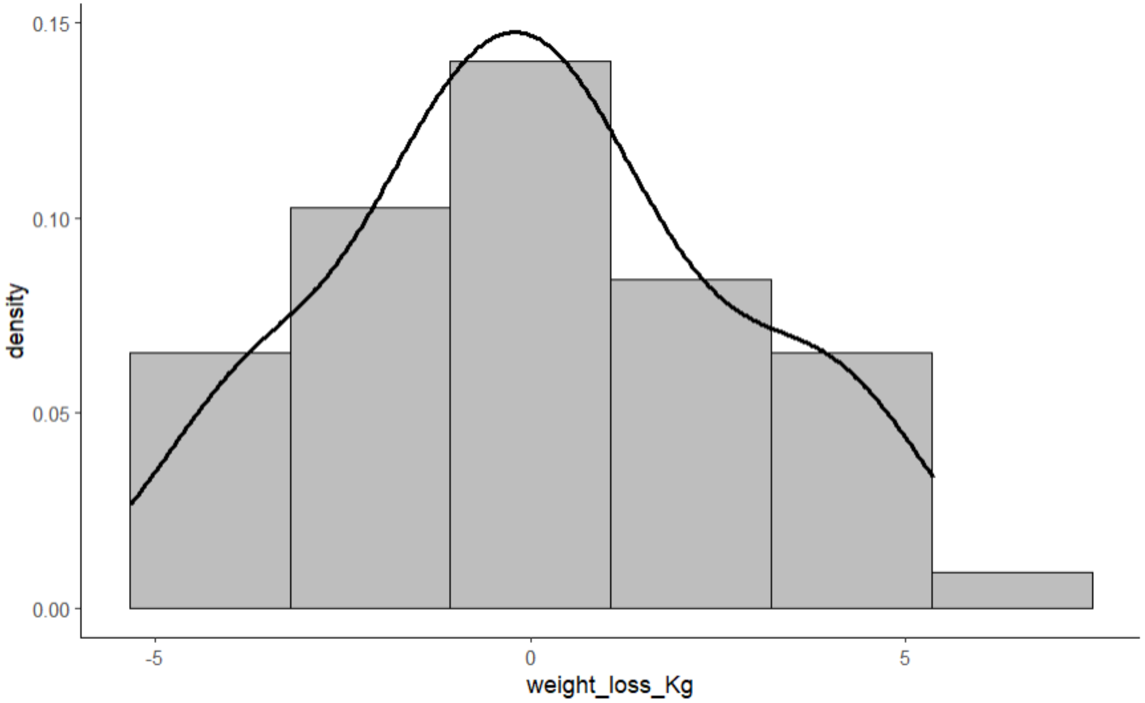
Ejemplo

Un producto homeopático afirma que “gracias a su uso, perderás **al menos** 2kg en un mes”. Escéptico ante esa afirmación, reclutas a 50 personas de tu ciudad para probar el producto, datos en “homeo_weight_los.csv”. Realiza un test de hipótesis (sin conocer la desviación típica poblacional). *Utiliza t.test(), pwr.t.test() y effectsize::effectsize().*

$$\begin{cases} H_0 : \mu \geq 2 \\ H_a : \mu < 2 \end{cases}$$

INFORME

Observamos una normalidad en los datos, por lo que nos disponemos a realizar un test T de Student para contrastar la hipótesis nula de que la media es mayor o igual que 2. Los resultados del T-test de una cola sobre una muestra de 50 observaciones evidencian un rechazo sobre la hipótesis nula con un tamaño del efecto mediano en la media de pérdida de peso $[\,t = -5.25,\ p < 0.00001;\ d = -0.74\,]$ con una potencia de 0.999. De hecho, hemos obtenido el intervalo $(-\infty, 0.66)$ para la media con una significancia de 0.05.



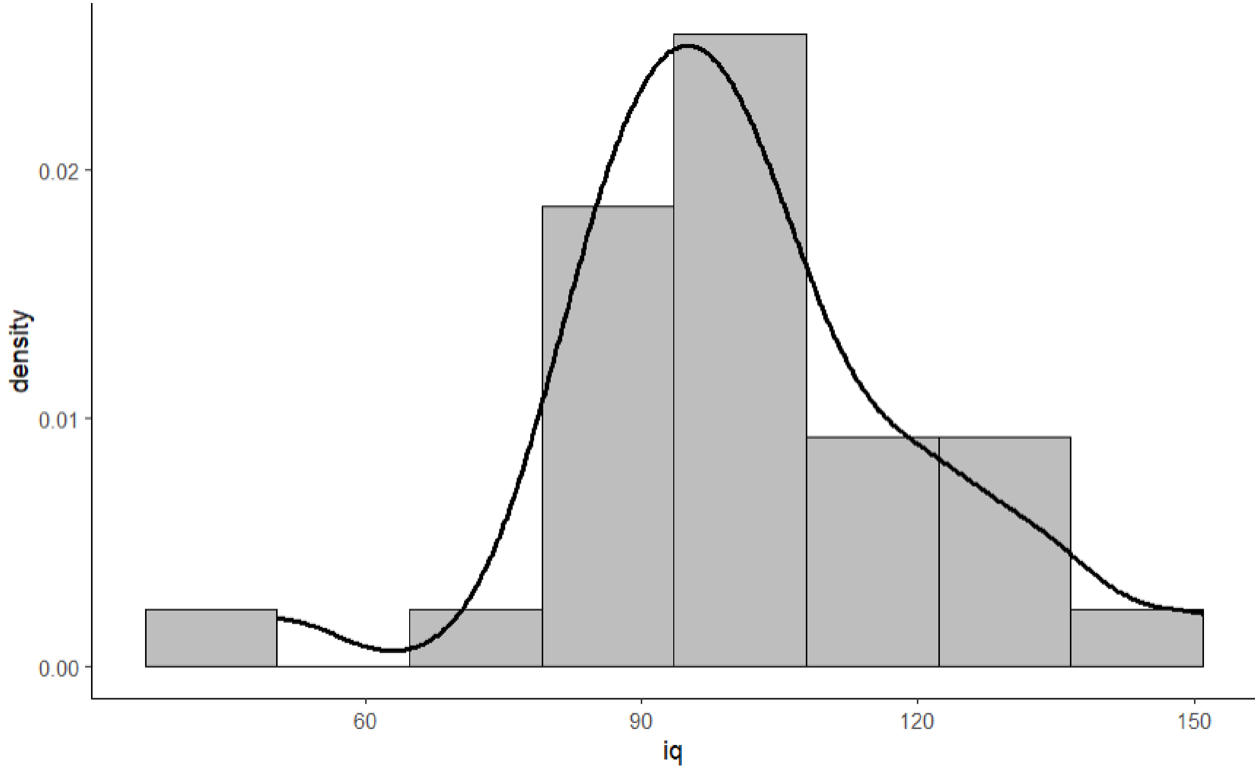
CEU Universidad San Pablo

Ejemplo

Los test de cociente intelectual (CI) están diseñados para que la desviación típica poblacional sea de 15 puntos. Sin embargo, en los procesos de traducción de un test "oficial" de CI pueden surgir desajustes. Por ejemplo, "iq_spanish.csv" tiene los resultados de un test de CI traducido del inglés al español. ¿Hay evidencia de que la desviación típica es distinta de 15 y, por tanto, debe revisarse la traducción? Usa un nivel de significación del 5 %.

$$\begin{cases} H_0 : \sigma = 15 \\ H_a : \sigma \neq 15 \end{cases}$$

Representa gráficamente los datos.



Observamos una normalidad en los datos.

χ^2-test

```
Results of Hypothesis Test
--------------------------

Null Hypothesis:                 variance = 225

Alternative Hypothesis:          True variance is not equal to 225

Test Name:                       Chi-Squared Test on Variance

Estimated Parameter(s):          variance = 388.6902

Data:                            data$iq

Test Statistic:                  Chi-Squared = 50.09785

Test Statistic Parameter:        df = 29

P-value:                         0.01762976

95% Confidence Interval:         LCL = 246.5322
                                 UCL = 702.4345
```

Datos para nuestro informe.

Criterios de calidad	Valores
normalidad	SI
Tamaño de muestra	30
Estadístico	50.098
P-valor < significancia	??
Intervalo de confianza	??

χ^2-test

```
Results of Hypothesis Test
--------------------------

Null Hypothesis:                 variance = 225

Alternative Hypothesis:          True variance is not equal to 225

Test Name:                       Chi-Squared Test on Variance

Estimated Parameter(s):          variance = 388.6902

Data:                            data$iq

Test Statistic:                  Chi-Squared = 50.09785

Test Statistic Parameter:        df = 29

P-value:                         0.01762976

95% Confidence Interval:         LCL = 246.5322
                                 UCL = 702.4345
```

Datos para nuestro informe.

Criterios de calidad	Valores
normalidad	SI
Tamaño de muestra	30
Estadístico	50.098
P-valor < significancia	0.018 < 0.05
Intervalo de confianza	??

χ^2-test

```
Results of Hypothesis Test
--------------------------

Null Hypothesis:                 variance = 225

Alternative Hypothesis:          True variance is not equal to 225

Test Name:                       Chi-Squared Test on Variance

Estimated Parameter(s):          variance = 388.6902

Data:                            data$iq

Test Statistic:                  Chi-Squared = 50.09785

Test Statistic Parameter:        df = 29

P-value:                         0.01762976

95% Confidence Interval:         LCL = 246.5322
                                 UCL = 702.4345
```

Datos para nuestro informe.

Criterios de calidad	Valores
normalidad	SI
Tamaño de muestra	30
Estadístico	50.098
P-valor < significancia	0.018 < 0.05
Intervalo de confianza	(15.70,26.50) (desviación típica)

OJO: Una significancia de 1% no rechazaría H_0

¿Potencia?

CEU Universidad San Pablo

$$\beta = P(\text{No rechazamos } H_0 \mid H_0 \text{ falsa}) \xrightarrow{\text{Simulaciones}} 1 - \beta \approx 0.60$$

H_0 falsa:

$$\boxed{\sigma = 19.72}$$

Hacemos esa suposición en base a los datos muestrales

¿Cómo podríamos aumentar la potencia?

Tomando muestras más grandes.

¿Cuántas observaciones necesitaríamos para una potencia del 90%?

$$n \approx 70$$

Ejemplo

Según estadísticas oficiales, la media de peso de las mujeres de cierto país es de 63.5 Kg (con desviación típica 4.1). Sin embargo, un equipo de investigadores cree que debido a cambios en la alimentación la media se ha incrementado. ¿Cuántas muestras necesitarán para poder detectar un incremento de medio Kg con un nivel de significación del 1% y una potencia del 90 %? Utiliza power.t.test()

```
     One-sample t test power calculation

              n = 1003.81
          delta = 0.5
             sd = 4.1
      sig.level = 0.01
          power = 0.9
    alternative = two.sided
```

Necesitamos 1004 muestras.

Ejemplo

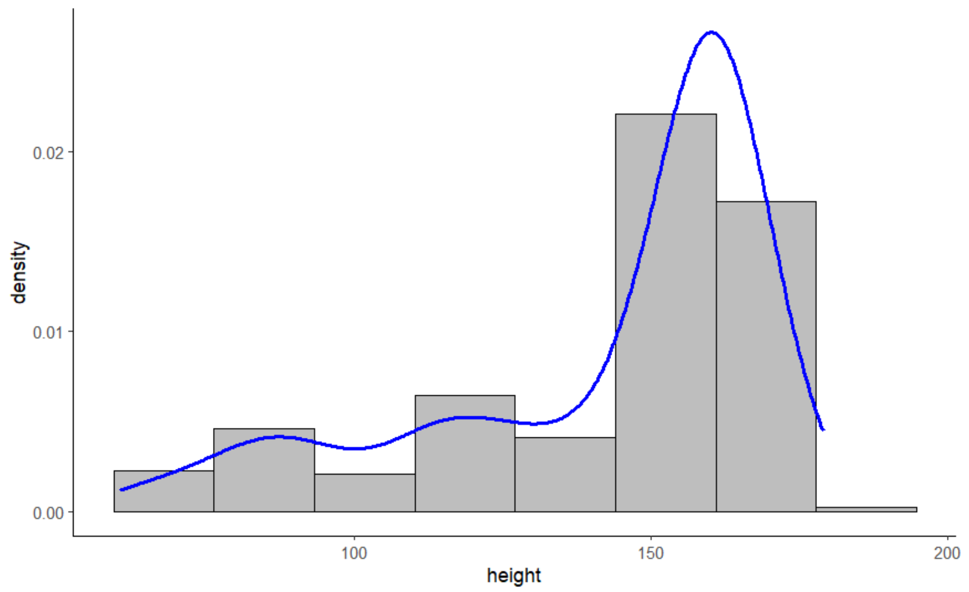
Los datos contenidos en "howell1.csv" son datos censales parciales del área Kung San compilados a partir de entrevistas realizadas a finales de la década de 1960. ¿Depende la altura de los Kung adultos del sexo del individuo? (α = 0.01).

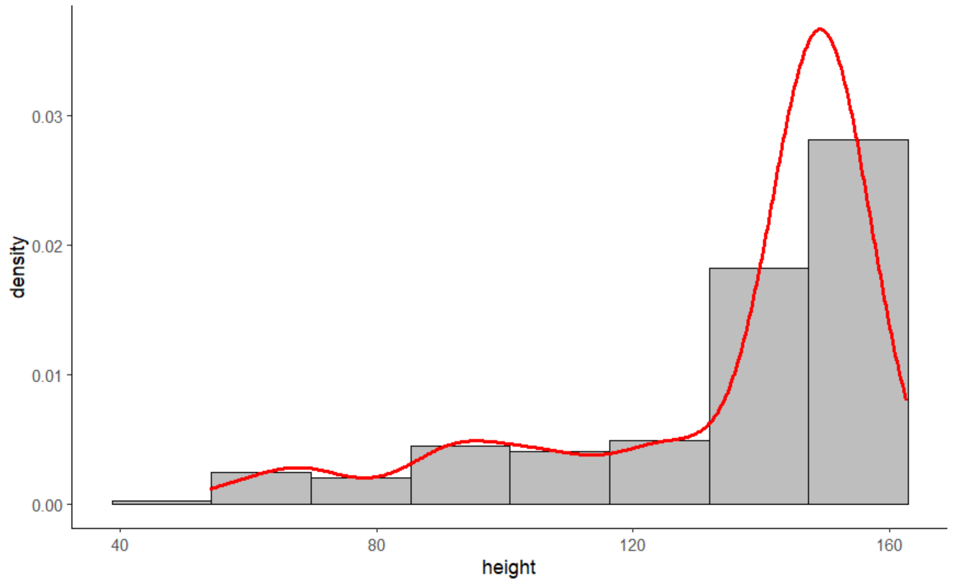
$X_1 =$ "altura hombres"

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_a : \mu_1 \neq \mu_2 \end{cases}$$

$X_2 =$ "altura mujeres"

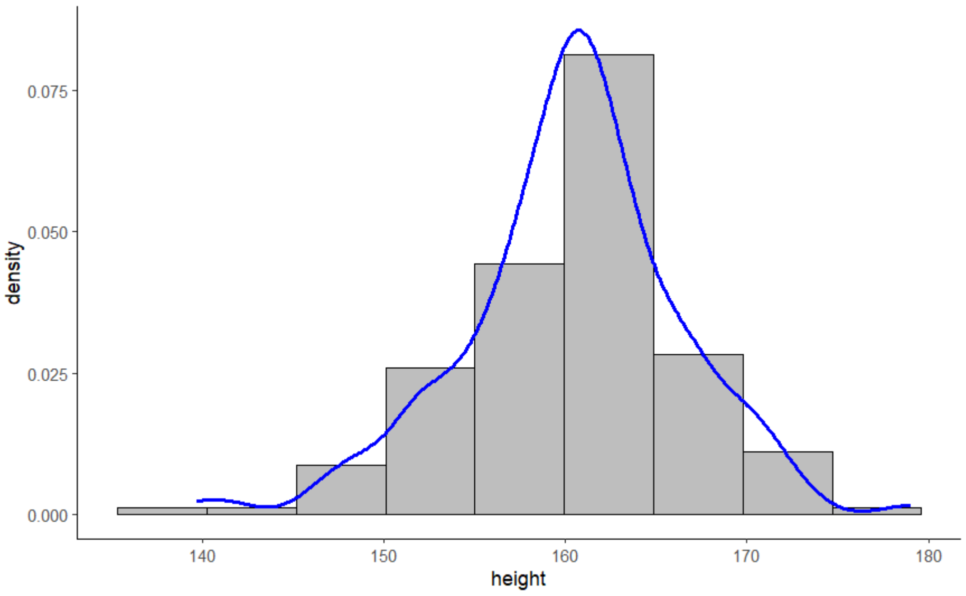
Representa gráficamente los datos.

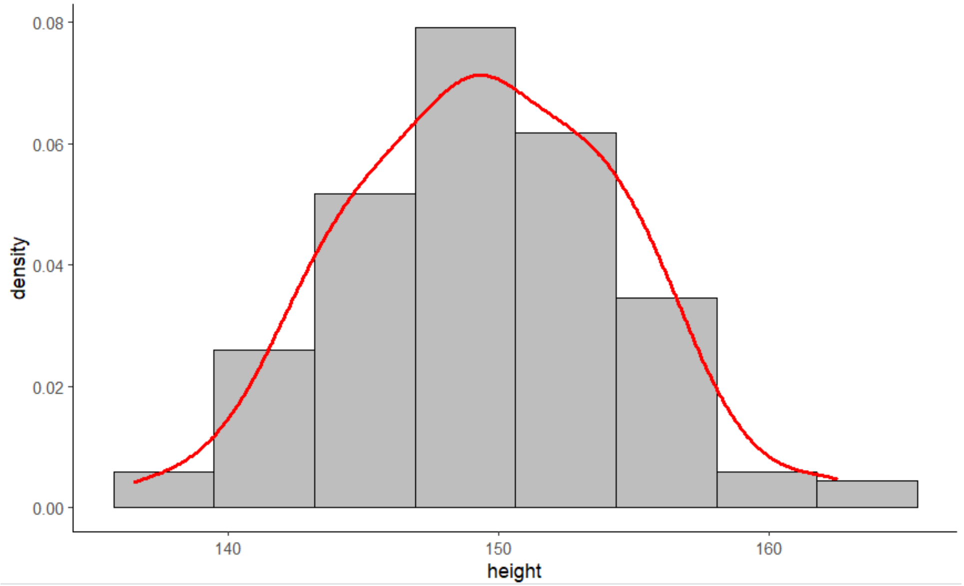




¡ DATOS NO NORMALES !

ELIMINA MENORES DE EDAD





☺ ¡ DATOS NORMALES ! ☺

T-test

```
        Welch Two Sample t-test

data:  male and female
t = 18.148, df = 323, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
99 percent confidence interval:
  9.296549 12.393366
sample estimates:
mean of x mean of y
 160.3585  149.5135
```

Datos para nuestro informe.

Criterios de calidad	Valores
normalidad	SI
Tamaño de muestra	$n_1 = 165, n_2 = 187$
Estadístico	18.148
P-valor < significancia	2.2e-16 < 0.01
Intervalo de confianza	$\mu_1 - \mu_2 \in (9.3,12.4)$
Tamaño del efecto	??
Potencia	??

Datos para nuestro informe.

Tamaño del efecto

```
Cohen's d |         99% CI
-------------------------
1.95      | [1.61, 2.29]

- Estimated using un-pooled SD.
```

Criterios de calidad	Valores
normalidad	SI
Tamaño de muestra	$n_1 = 165, n_2 = 187$
Estadístico	18.148
P-valor < significancia	2.2e-16 < 0.01
Intervalo de confianza	$\mu_1 - \mu_2 \in (9.3,12.4)$
Tamaño del efecto	1.95
Potencia	??

Datos para nuestro informe.

Potencia

```
t test power calculation

         n1 = 165
         n2 = 187
          d = 1.948391
  sig.level = 0.99
      power = 1
alternative = two.sided
```

Criterios de calidad	Valores
normalidad	SI
Tamaño de muestra	$n_1 = 165, n_2 = 187$
Estadístico	18.148
P-valor < significancia	2.2e-16 < 0.01
Intervalo de confianza	$\mu_1 - \mu_2 \in (9.3,12.4)$
Tamaño del efecto	1.95
Potencia	0.999

CEU Universidad San Pablo

Ejemplo

Unos científicos examinaron la función de la vesícula biliar antes y después de una cirugía para detener el reflujo. Los autores midieron la funcionalidad de la vesícula biliar calculando la fracción de eyección de la vesícula biliar (GBEF) antes y después de la operación, cuyo objetivo es aumentar la GBEF. ¿Hay evidencia para concluir que la operación aumenta el GBEF? Datos en “gbef long.txt”

$X_1 =$ “GBEF antes”

$$\begin{cases} H_0 : \mu_1 \geq \mu_2 \\ H_a : \mu_1 < \mu_2 \end{cases}$$

$X_2 =$ “GBEF después”

T-test de medidas repetidas

```
	Paired t-test

data:  antes and despues
t = -1.9159, df = 11, p-value = 0.04086
alternative hypothesis: true mean difference is less than 0
95 percent confidence interval:
      -Inf -1.131919
sample estimates:
mean difference 
        -18.075 
```

Datos para nuestro informe.

Criterios de calidad	Valores
normalidad	…
Tamaño de muestra	$n = 12$
Estadístico	-1.92
P-valor < significancia	0.04 < 0.05
Intervalo de confianza	$\mu_1 - \mu_2 \in (-\infty, -1.13)$
Tamaño del efecto	??
Potencia	??

Tamaño del efecto

```
Cohen's d |         95% CI
--------------------------
-0.55     | [-Inf, -0.03]

- One-sided CIs: lower bound fixed at [-Inf].
```

Datos para nuestro informe.

Criterios de calidad	Valores
normalidad	…
Tamaño de muestra	$n = 12$
Estadístico	-1.92
P-valor < significancia	0.04 < 0.05
Intervalo de confianza	$\mu_1 - \mu_2 \in (-\infty, -1.13)$
Tamaño del efecto	0.55
Potencia	??

CEU Universidad San Pablo

Potencia

```
     Paired t test power calculation

              n = 12
              d = -0.55
      sig.level = 0.05
          power = 0.5563109
    alternative = less

NOTE: n is number of *pairs*
```

Potencia muy baja… ¿Cómo obtener 95%?

Datos para nuestro informe.

Criterios de calidad	Valores
normalidad	…
Tamaño de muestra	$n = 12$
Estadístico	-1.92
P-valor < significancia	0.04 < 0.05
Intervalo de confianza	$\mu_1 - \mu_2 \in (-\infty, -1.13)$
Tamaño del efecto	0.55
Potencia	0.56

Los datos no son del todo concluyentes.

Potencia

```
     Paired t test power calculation

              n = 37.175
              d = -0.55
      sig.level = 0.05
          power = 0.95
    alternative = less

NOTE: n is number of *pairs*
```

Necesitamos 38 muestras para una buena potencia

Criterios de calidad	Valores
normalidad	$\boxed{\ldots}$
Tamaño de muestra	$\boxed{n = 12}$
Estadístico	-1.92
P-valor < significancia	$\boxed{0.04 < 0.05}$
Intervalo de confianza	$\mu_1 - \mu_2 \in (-\infty, -1.13)$
Tamaño del efecto	0.55
Potencia	$\boxed{0.56}$

CEU Universidad San Pablo

Ejemplo

Usa un test de ratio de varianzas para discutir si es razonable asumir igualdad de varianzas en el ejercicio de los Kung (¿Existe evidencia de que las varianzas por sexo son distintas?)

$X_1 =$ "altura hombres"

$$\begin{cases} H_0 : \dfrac{\sigma_1^2}{\sigma_2^2} = 1 \\ H_a : \dfrac{\sigma_1^2}{\sigma_2^2} \neq 1 \end{cases}$$

$X_2 =$ "altura mujeres"

F-test

```
        F test to compare two variances

data:  male and female
F = 1.3968, num df = 164, denom df = 186, p-value = 0.02721
alternative hypothesis: true ratio of variances is not equal to 1
99 percent confidence interval:
 0.9457624 2.0714536
sample estimates:
ratio of variances
          1.396756
```

Evidencias al 5% pero no al 1%

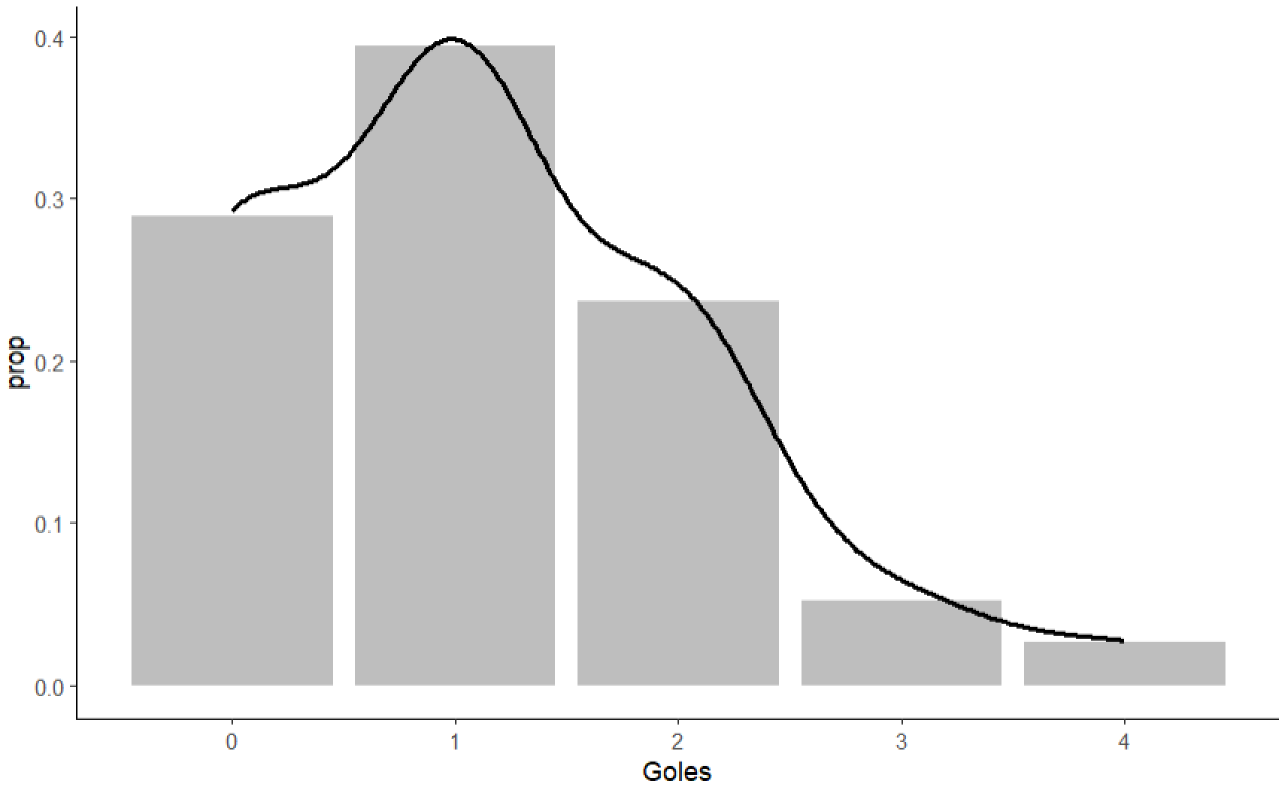
Criterios de calidad	Valores
normalidad	SI
Tamaño de muestra	$n_1 = 165, n_2 = 187$
Estadístico	1.397
P-valor < significancia	0.02 < 0.05
Intervalo de confianza	(0.94,2.07) al 99%

Ejemplo

Un experto deportivo te asegura que un determinado equipo de futbol marca de media 2 o más goles por partido. Realiza un test de hipótesis con los datos del año pasado “goles.csv” y una significancia del 5%.

$$\begin{cases} H_0 : \lambda \geq 2 \\ H_a : \lambda < 2 \end{cases}$$

Representa gráficamente los datos.



Observamos una tendencia a una distribución de Poisson

Poisson test

```
        Exact Poisson test

data:  suma time base: n
number of events = 43, time base = 38, p-value = 2.721e-05
alternative hypothesis: true event rate is less than 2
95 percent confidence interval:
 0.000000 1.459184
sample estimates:
event rate
  1.131579
```

Datos para nuestro informe.

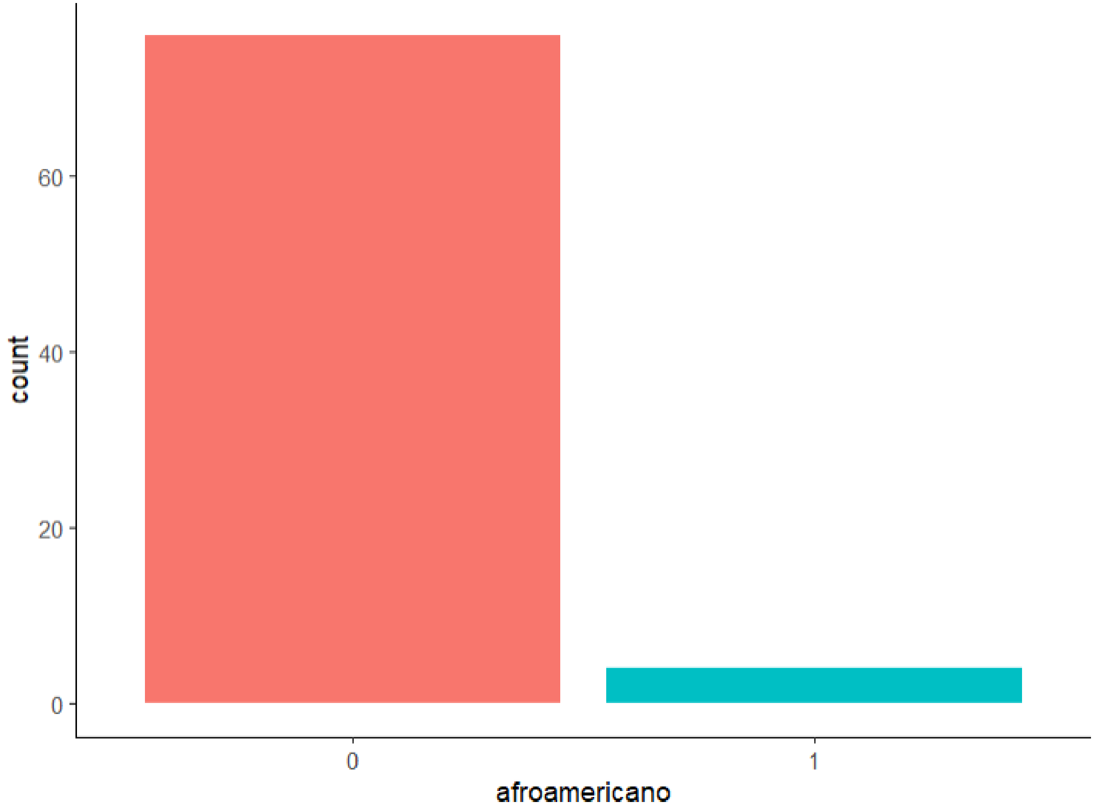
Criterios de calidad	Valores
normalidad	NO, Poisson
Tamaño de muestra	$n = 38$
P-valor < significancia	2.72e-05 < 0.05
Intervalo de confianza	(0,1.46)

Ejemplo

En una selección al azar de jurados las autoridades afirman que la probabilidad de que un elegido sea afroamericano es de $p = 0.5$ (coincidiendo con la proporción de afroamericanos en la población). Dispuesto a rechazar dicha afirmación, decides recoger los datos obtenidos en un juicio relacionado con el racismo. Datos en “juicio.xlsx”. Realiza un test de hipótesis con una significancia del 5%.

$$\begin{cases} H_0 : p = 0.5 \\ H_a : p \neq 0.5 \end{cases}$$

Representa gráficamente los datos.



Distribución Bernoulli.

Binomial test

```
        Exact binomial test

data:  suma and n
number of successes = 4, number of trials = 80, p-value < 2.2e-16
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.01378939 0.12309874
sample estimates:
probability of success
                  0.05
```

Datos para nuestro informe.

Criterios de calidad	Valores
normalidad	NO, Bernoulli.
Tamaño de muestra	$n = 80$
P-valor < significancia	2.2e-16 < 0.05
Intervalo de confianza	(0.01,0.12)

CEU Universidad San Pablo

Ejemplo

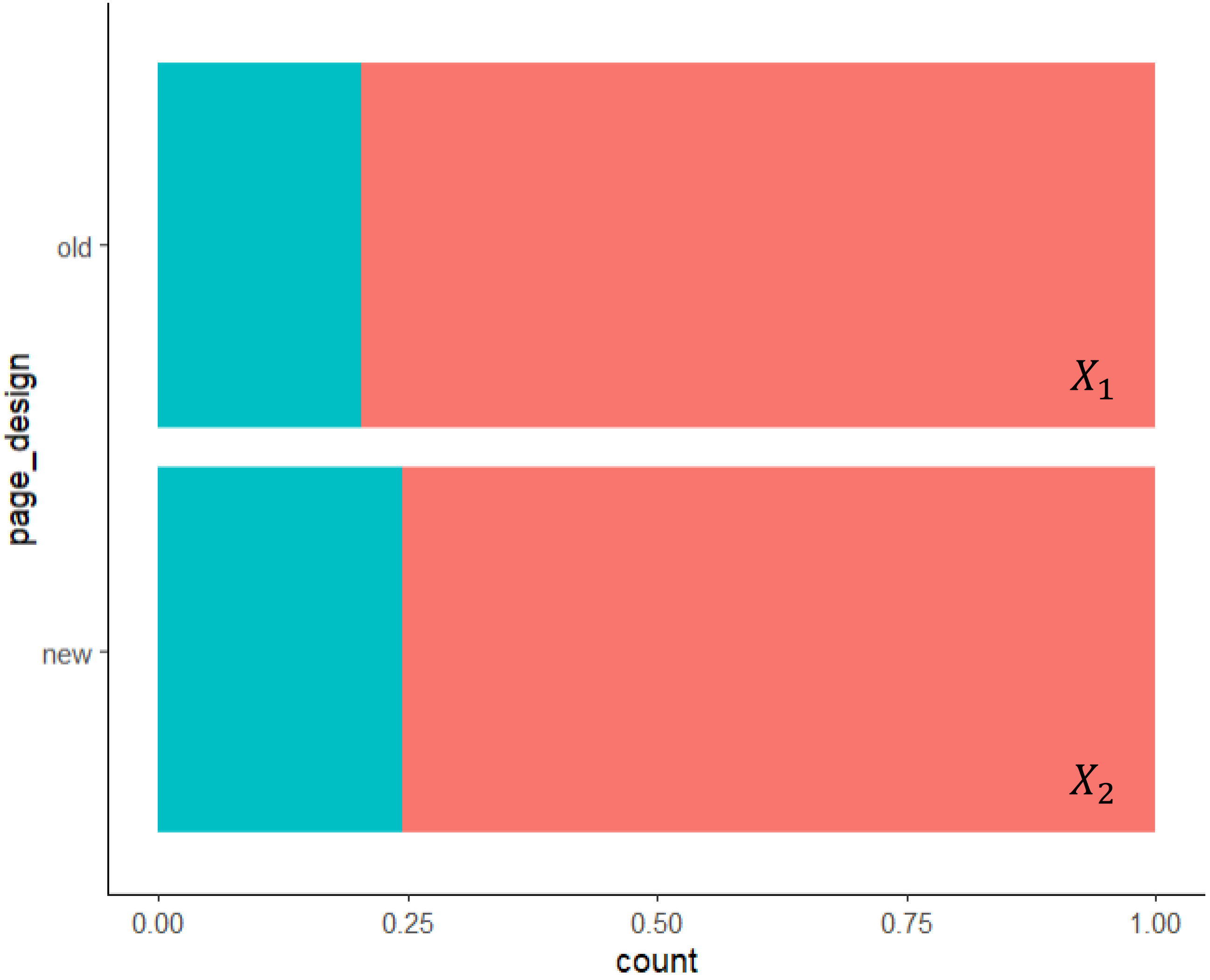
Una página web de venta de productos ha estudiado el número de conversiones de su página web actual (conversión = el usuario hace click en "comprar ahora"). Para aumentar el número de conversiones, rediseña el aspecto de su página web. La nueva página se prueba con un nuevo conjunto de usuarios, midiendo el número de conversiones. Datos en "ab_testing.csv". ¿Se puede concluir que la nueva página incrementa el número de conversiones?

$X_1 = Ber(p_1)$

$$\begin{cases} H_0 : p_1 \geq p_2 \\ H_a : p_1 < p_2 \end{cases}$$

$X_2 = Ber(p_2)$

Representa gráficamente los datos.



A/B test

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

data:  c(suma1, suma2) out of c(n1, n2)
X-squared = 5.0366, df = 1, p-value = 0.01241
alternative hypothesis: less
95 percent confidence interval:
 -1.00000000 -0.01101437
sample estimates:
   prop 1    prop 2 
0.2035573 0.2452660 
```

Datos para nuestro informe.

Criterios de calidad	Valores
normalidad	NO, Bernoullis.
Tamaño de muestra	$n_1 = 1012, n_2 = 1109$
P-valor < significancia	0.012 < 0.05
Intervalo de confianza	(-1,-0.1)