

Universidad Nacional de La Plata
Facultad de Ciencias Agrarias y Forestales



**CÁLCULO ESTADÍSTICO Y
BIOMETRÍA**

Pruebas de Hipótesis

INFERENCIA ESTADÍSTICA

```
graph TD; A[INFERENCIA ESTADÍSTICA] --> B[Estimación de Parámetros]; A --> C[Pruebas de Hipótesis]
```

The diagram illustrates the structure of Statistical Inference. At the top, a grey rectangular box contains the text 'INFERENCIA ESTADÍSTICA'. Two red arrows point downwards from this box to two separate grey rectangular boxes below. The left box is labeled 'Estimación de Parámetros' and the right box is labeled 'Pruebas de Hipótesis'.

Estimación de Parámetros

Pruebas de Hipótesis

¿ Qué es una hipótesis estadística ?

Es una proposición o afirmación sobre la distribución de probabilidad de una variable aleatoria.

Parámetros de la distribución

Asociadas a una o más poblaciones

Forma de la distribución

Ejemplos:

La edad media de los alumnos del curso es de 22 años.

Con la aplicación de un determinado nutriente se obtendrán rendimientos medios superiores a 2.000 kg/ha, que es la producción usual de la zona.

El peso promedio al nacimiento de ovejas hembras es menor al de los machos

Hay diferencias en el aumento de peso promedio de novillos alimentados con diferentes pasturas

La altura de álamos sigue una distribución normal

El número de semillas germinadas por placa responde a una distribución binomial

Prueba de Hipótesis

Procedimiento (o experiencia) que conduce a una toma de decisión en cuanto a optar por una u otra hipótesis, en base a la información proporcionada por una muestra aleatoria extraída de la población bajo estudio

la **Hipótesis Nula**
(H_0) y

la **Hipótesis Alternativa** (H_1)

Prueba de Hipótesis

H_0 \rightarrow Hipótesis estadística que se somete a prueba

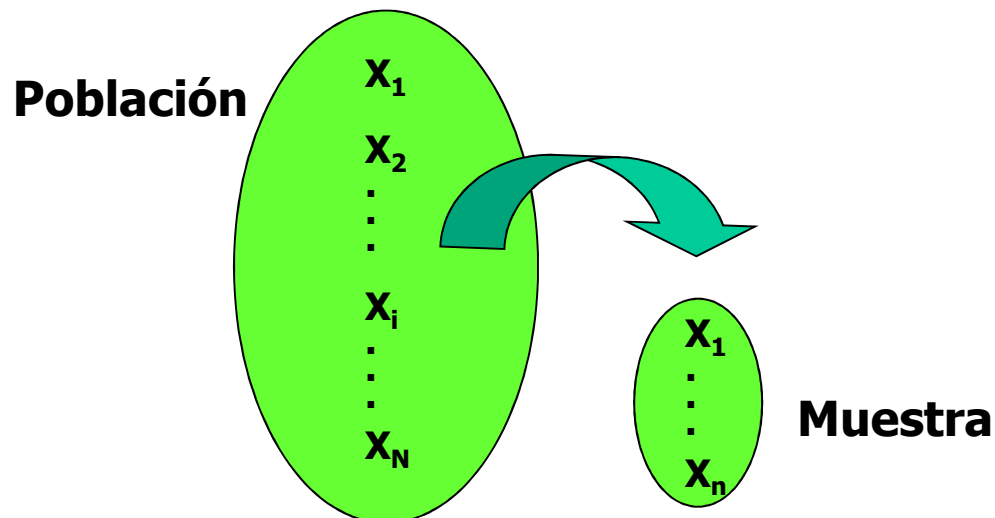
H_1 \rightarrow Hipótesis estadística que postula lo que el investigador esta interesado en probar

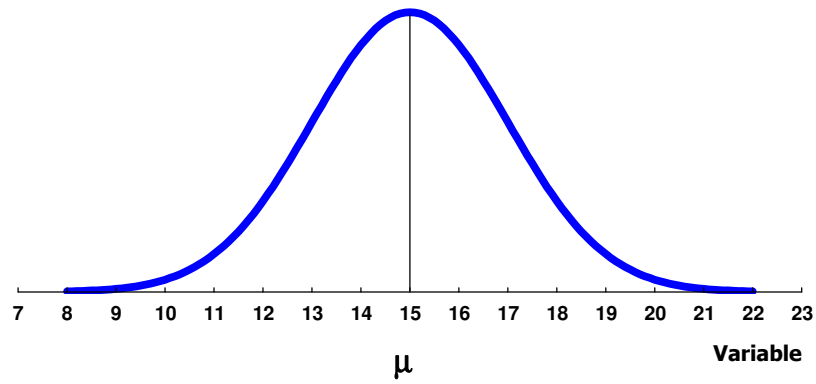
Tanto cuando no se rechaza la hipótesis nula como cuando se rechaza, es posible cometer errores.

Estadístico de prueba y Regla de Decisión

$H_0: \mu = 15$

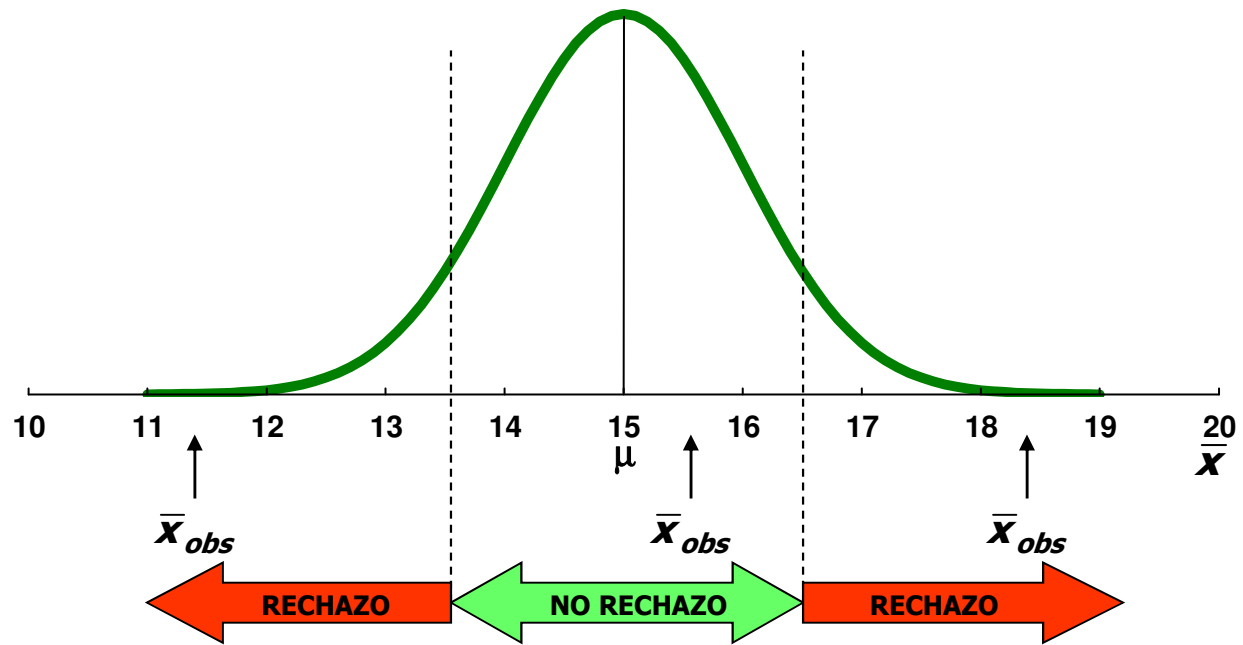
$H_1: \mu \neq 15$





Pruebas

- Bilaterales
- Unilaterales



Conclusión sujeta a error

Errores en una Prueba de hipótesis

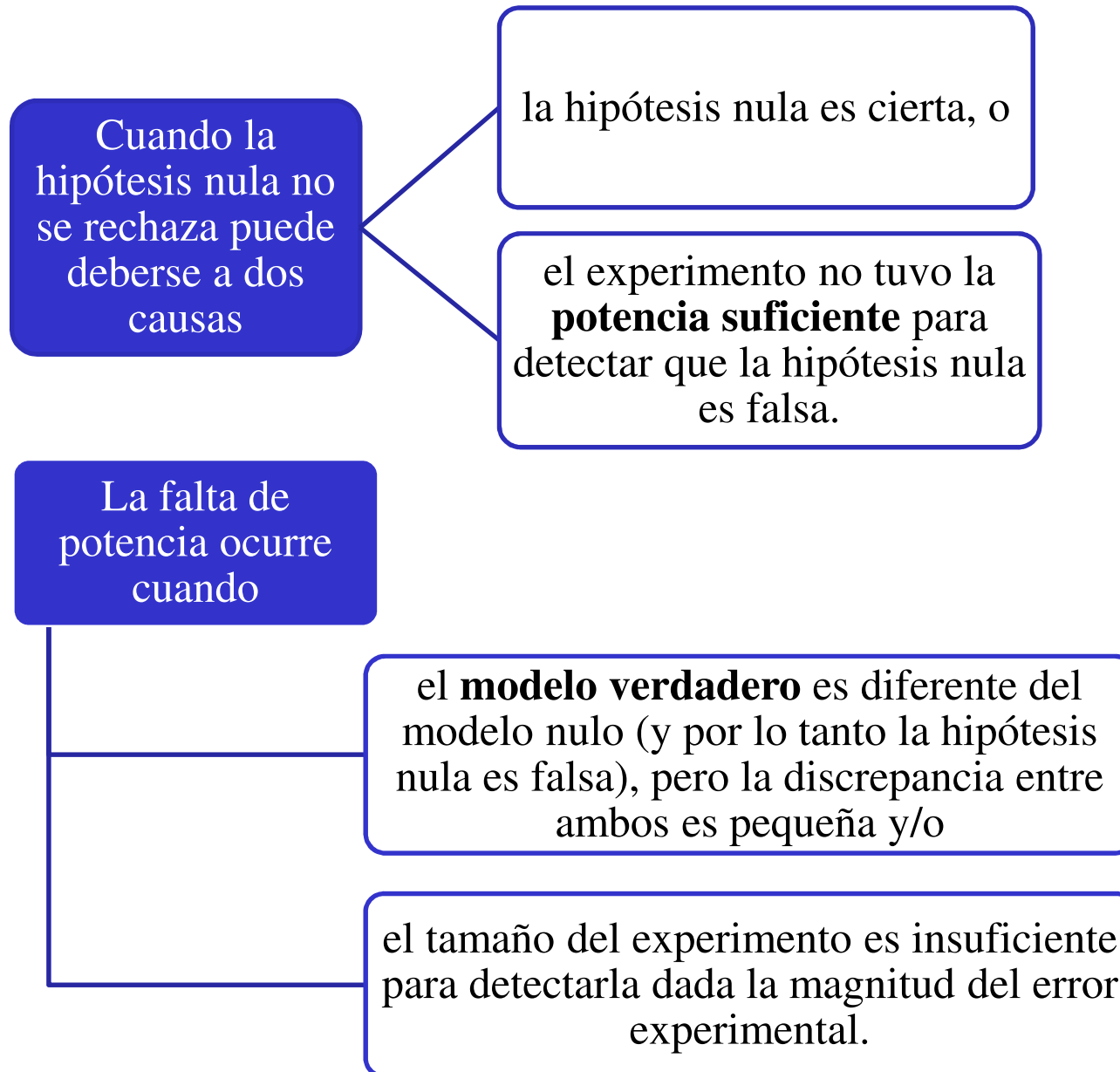
| | Ho es Verdadera | Ho es Falsa |
|---------------------------|-------------------|-------------------|
| DECISIÓN Rechazo H_0 | Error de Tipo I | Decisión correcta |
| No Rechazo H_0 | Decisión correcta | Error de Tipo II |

$$P(\text{Error Tipo I}) = P(\text{Rechazar } H_0 \mid H_0 \text{ es Verdadera}) = \alpha$$

$$P(\text{Error Tipo II}) = P(\text{No Rechazar } H_0 \mid H_0 \text{ es Falsa}) = \beta$$

$$\text{Potencia} = 1 - \beta = P(\text{Rechazar } H_0 \mid H_0 \text{ es Falsa})$$

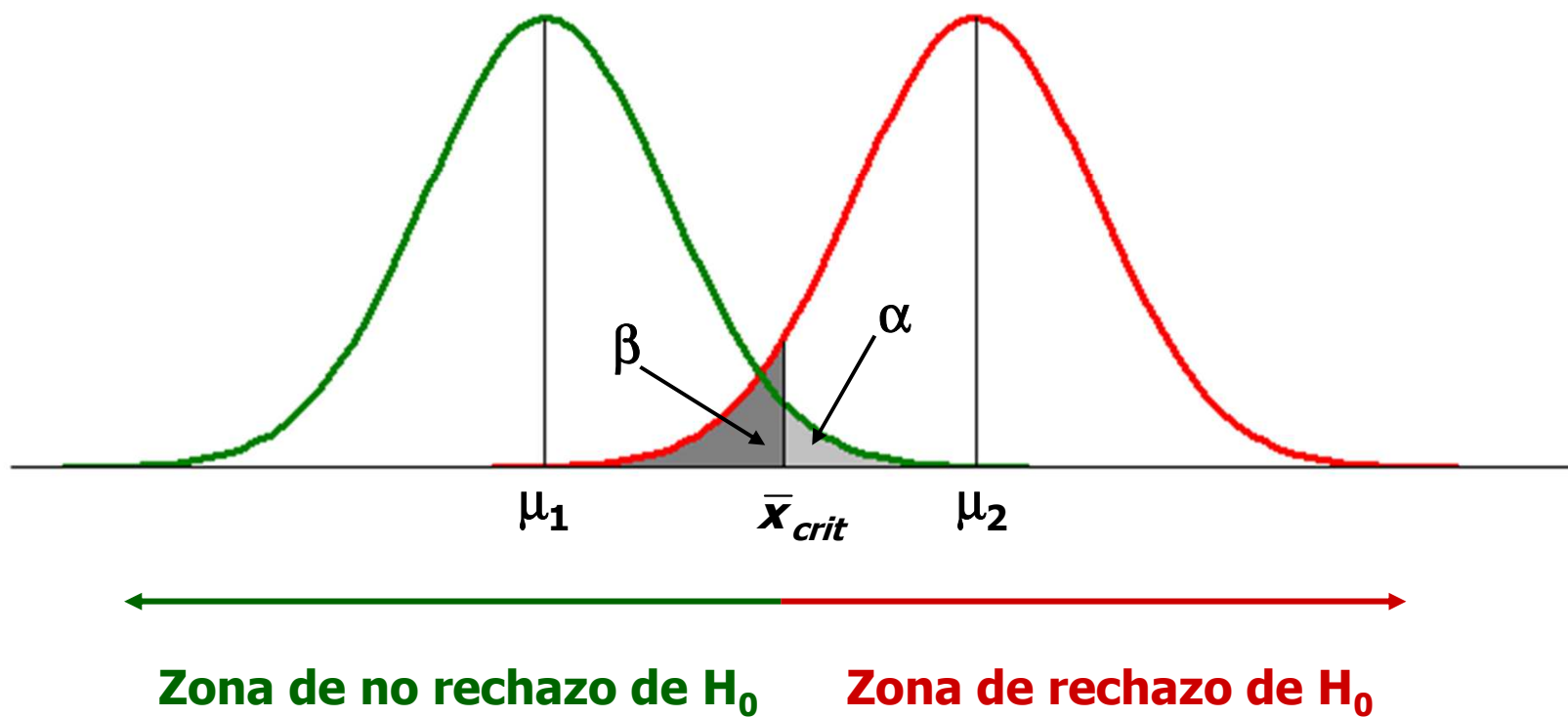
Potencia de una Prueba



Representación esquemática de los errores

H₀: $\mu = \mu_1$

H₁: $\mu = \mu_2$

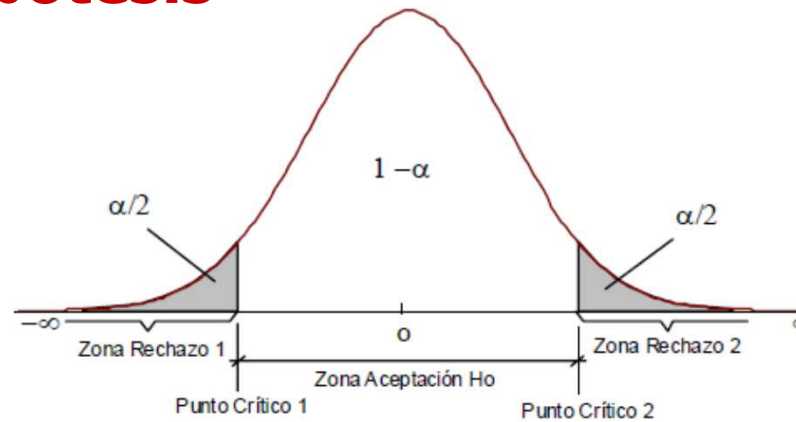


Tipos de Pruebas de Hipótesis

Bilateral

$$H_0: \theta = \theta_0$$

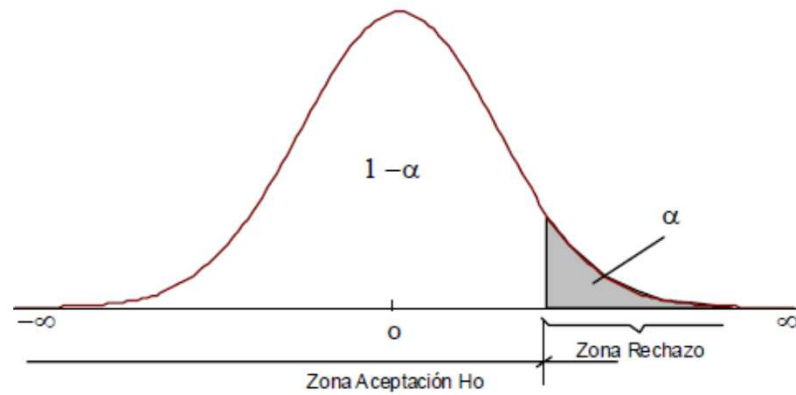
$$H_1: \theta \neq \theta_0$$



Unilateral a la derecha

$$H_0: \theta \leq \theta_0$$

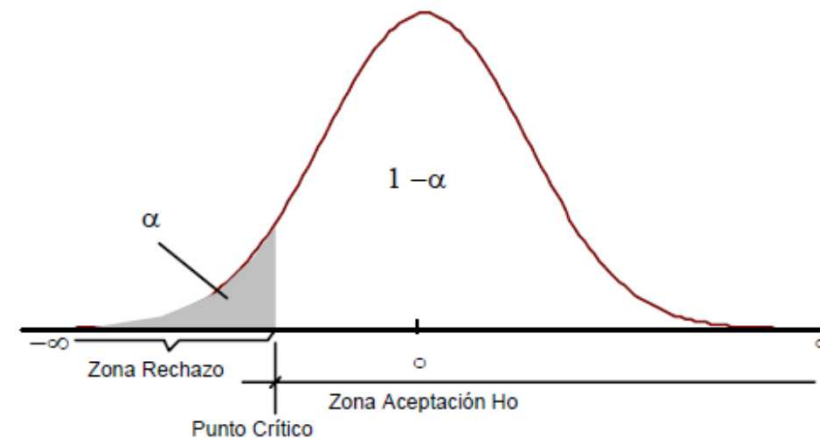
$$H_1: \theta > \theta_0$$



Unilateral a la Izquierda

$$H_0: \theta \geq \theta_0$$

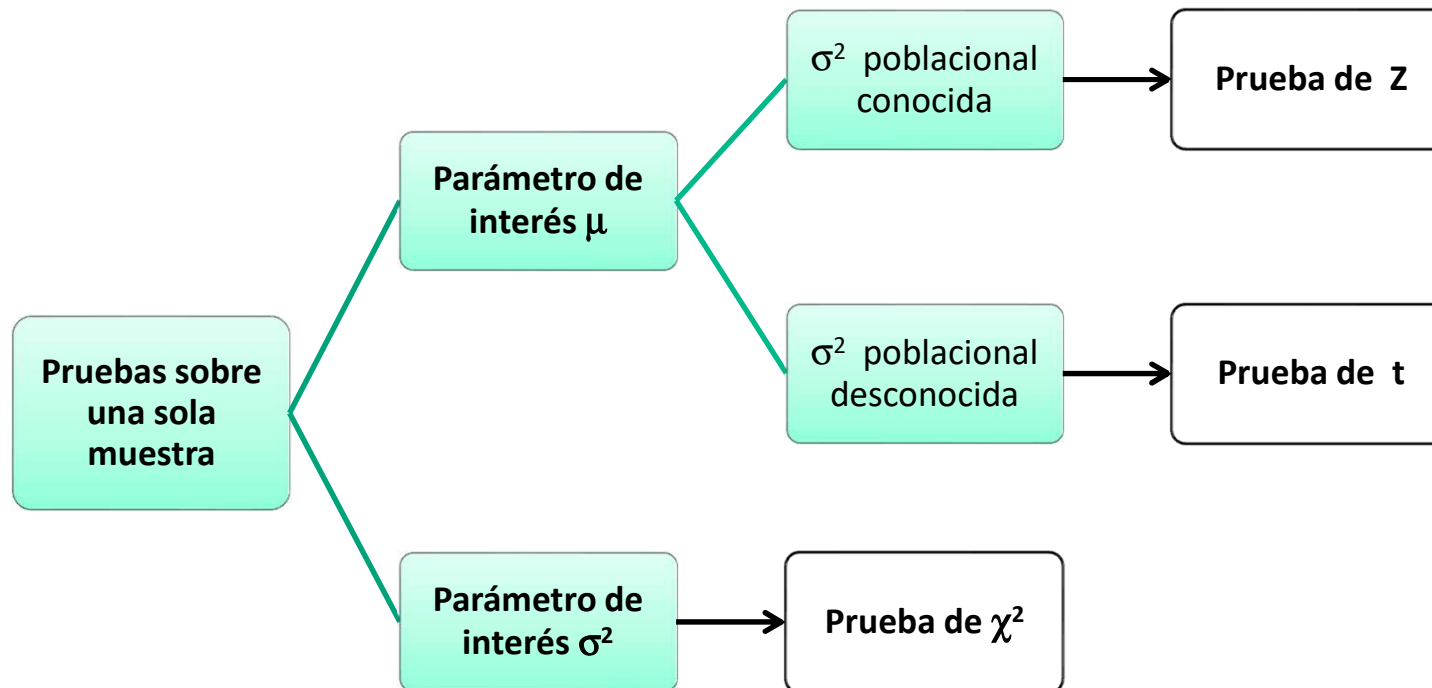
$$H_1: \theta < \theta_0$$



Procedimiento general para una prueba de hipótesis:

- **Del contexto del problema identificar el parámetro de interés**
- **Plantear H_0 y H_1**
- **Planificar una experiencia para la extracción de la muestra**
- **Establecer el nivel de significación de la prueba**
- **Seleccionar un estadístico de prueba e identificar su distribución bajo H_0**
- **Establecer regiones de rechazo y no rechazo para el estadístico de prueba**
- **Calcular de la muestra el valor del estadístico**
- **Decidir si debe o no rechazarse H_0 e interpretar esto en el contexto del problema**

Pruebas de Hipótesis para la μ y la σ^2



A. Pruebas sobre una sola muestra

A1. Pruebas de hipótesis acerca de la media de una distribución normal con variancia conocida

Problema: A un productor se le ofrece un nutriente con el que obtendrá rendimientos medios superiores a 2.000 kg/ha, que es la producción usual de la zona. Se sabe que en esa población los rendimientos tienen un desvío estándar de 210 kg/ha. El productor decide realizar una prueba sobre 9 parcelas, en las que pretende observar los rendimientos, promediarlos y de acuerdo al resultado optar o no por el nuevo producto.

DATOS: 2150 1950 2170 1860 2050 2120 1920 1850 2230

Rendimiento promedio: 2033,33 kg/ha

- a. El parámetro de interés es μ , rendimiento promedio con aplicación de nutriente.
- b. Información: σ^2 , VA distribución normal, rendimiento promedio sin aplicación de nutriente, $n=9$.

c. Planteo H_0 y H_1

$$H_0: \mu \leq 2.000$$

$$H_1: \mu > 2.000$$

~~$$H_0: \mu \geq 2.000$$~~

~~$$H_1: \mu < 2.000$$~~

d. Se acuerda correr un riesgo de rechazar el nutriente cuando en realidad éste cumple con las especificaciones promocionales, fijando para ello una probabilidad de error de 5% ($\alpha = 0,05$)

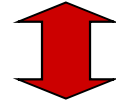
e. Dado que se conoce σ^2 , el estadístico de prueba es:

$$Z_0 = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

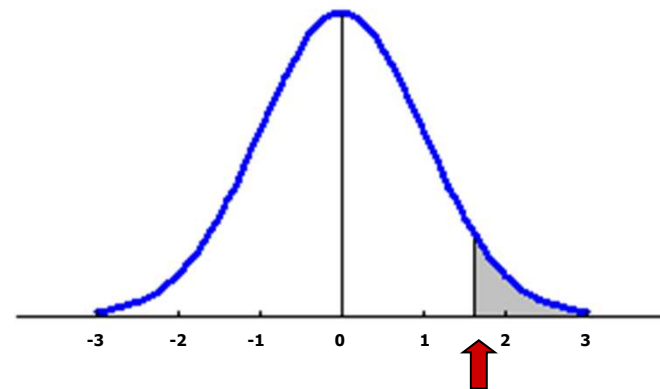
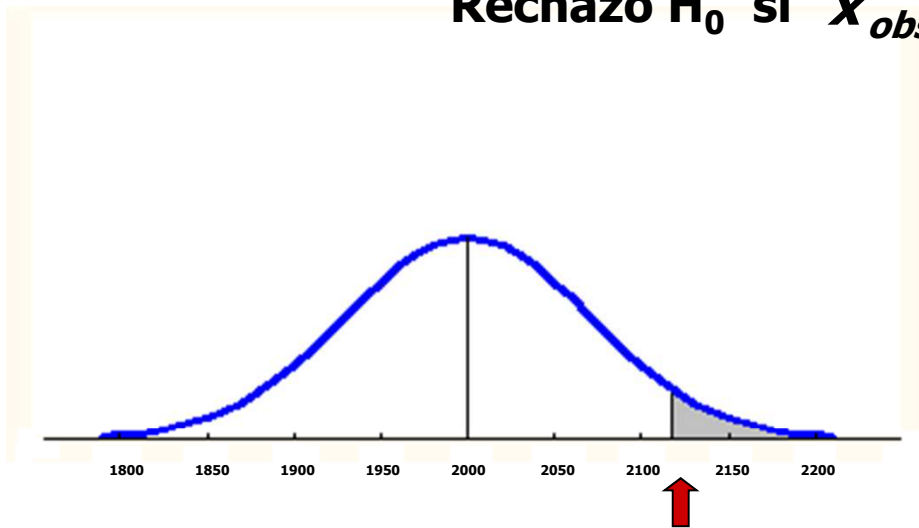
Con $Z_0 \sim N(0;1)$

f. Regiones de rechazo:

Rechazo H_0 si $z_{obs} > 1,645$



Rechazo H_0 si $\bar{x}_{obs} > 2.115,15$ kgs/ha



g. Cálculo del estadístico observado:

$$\bar{x}_{obs} = 2.033,33$$

$$z_{obs} = 0,476$$

h. Conclusión:

No se rechaza H_0 , es decir, no existe suficiente evidencia como para decir que efectivamente el nutriente aumenta los rendimientos.

Comentarios:

* P-value de un estadístico de prueba observado es la probabilidad de que la VA del estadístico de prueba tome un valor al menos tan extremo como el observado dado que la hipótesis nula es verdadera.

En nuestro ejemplo: **P-value = $P(Z > z_{obs} / H_0 \text{ es Verdadera})$**

* Si la VA bajo estudio no se distribuye normalmente deberán considerarse las condiciones para la aplicación del Teorema Central del Límite

A2. Pruebas de hipótesis acerca de la media de una distribución normal con variancia desconocida

$$t_0 = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

Con $t_0 \sim t(n-1)$

A3. Pruebas de hipótesis sobre la variancia poblacional

$$\mathbf{H_0: \sigma^2 = \sigma_0^2}$$

$$\mathbf{H_1: \sigma^2 \neq \sigma_0^2}$$

$$\mathbf{H_0: \sigma^2 = \sigma_0^2}$$

$$\mathbf{H_1: \sigma^2 > \sigma_0^2}$$

$$\mathbf{H_0: \sigma^2 = \sigma_0^2}$$

$$\mathbf{H_1: \sigma^2 < \sigma_0^2}$$

$$\chi_0^2 = \frac{(n-1) \cdot S^2}{\sigma_0^2}$$

Con $\chi_0^2 \sim \chi^2(n-1)$

Ejercicio 1

Un experimentador avícola considera que luego de cuatro semanas de alimentación con una ración especial a pollitos de la raza Cornich, ha de lograr un peso medio superior a 700 gr. por animal. Para verificarlo alimenta con la ración a un lote de 20 pollitos y a los 28 días obtiene un peso promedio de 730 gr. con una desviación estándar de 40.21 gr.

- a) Establecer las hipótesis nula y alternativa y realizar el contraste correspondiente utilizando $\alpha=0.05$.
- b) Construir un intervalo de confianza para μ .
- c) Probar la hipótesis que la varianza poblacional es 1700 gr.²

a)

- **Planteo de hipótesis** (test unilateral)

- **Cálculo del valor de "t"**

$$t_0 = \frac{\bar{X} - \mu_0}{s / \sqrt{n}} = \frac{730 - 700}{40.21 / \sqrt{20}} = 3.34$$

- **Obtención del valor de "t" (Crítico o tabulado)**

$$gl = n - 1 = 20 - 1 = 19$$

nivel de significancia (α) = 0.05

"t" tabulado = 1.73

b)

$$P\left(\mu < \bar{X} + t_0 \frac{S}{\sqrt{n}}\right) = \gamma = 95\%$$

c)

- **Planteo de hipótesis** (test unilateral)

- **Cálculo del valor de**

$$\chi_0^2 = \frac{(n-1) \cdot S^2}{\sigma_0^2} = \frac{19 \cdot 1616.84}{1700} = 18,07$$

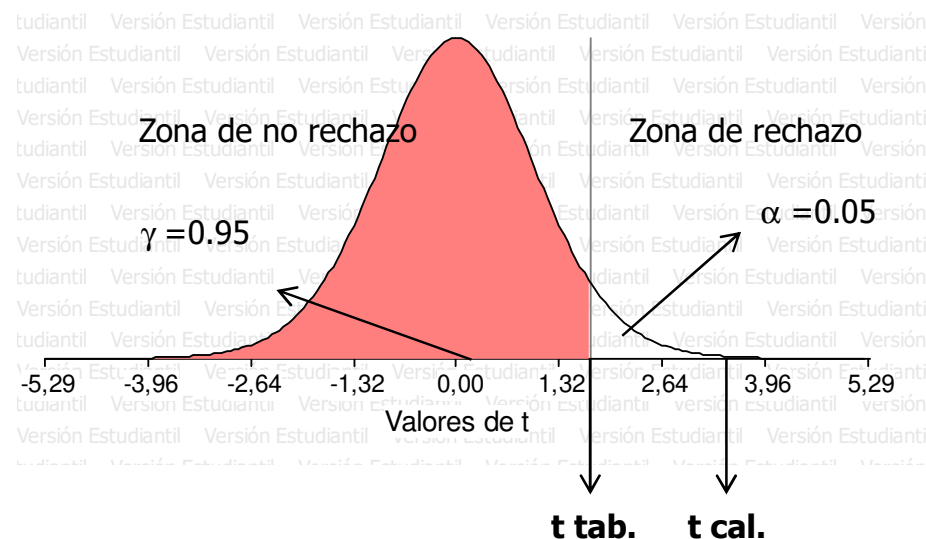
- **Obtención de los valores de χ_{crit}^2 .**

$$gl = n - 1 = 20 - 1 = 19$$

nivel de significancia (α) = 0.05

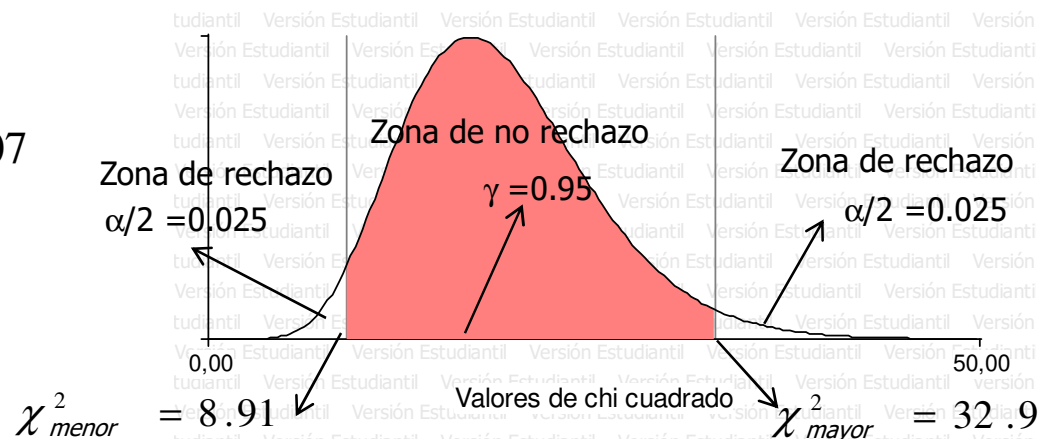
$H_0 : \mu \leq 700$ gr.

$H_1 : \mu > 700$ gr.

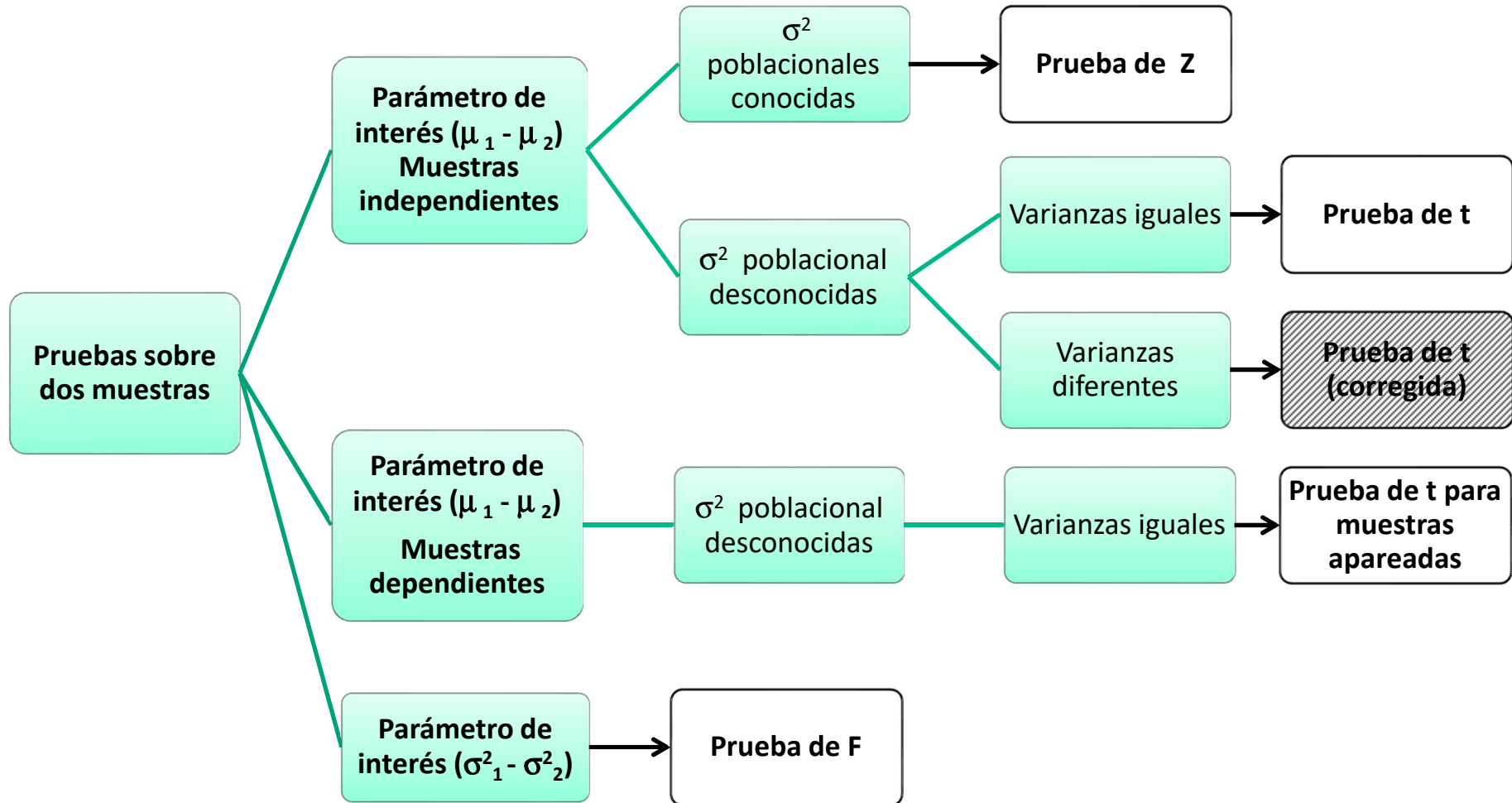


$H_0 : \sigma^2 = 1700$

$H_1 : \sigma^2 \neq 1700$



Pruebas de Hipótesis para la μ y la σ^2



B. Pruebas sobre dos muestras

B1. Pruebas de hipótesis para la igualdad de dos varianzas

Esta prueba es un paso intermedio necesario en las pruebas de comparaciones de medias cuando se desconocen las varianzas poblacionales

$$\begin{array}{l} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{array} \quad \rightarrow \quad \text{Bajo } H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$$

$$F_0 = \frac{S_1^2}{S_2^2}$$

$$\text{Con } F_0 \sim F(n_1-1; n_2-1)$$

Las pruebas unilaterales son muy usadas y constituyen la base de las pruebas asociadas a la técnica ANOVA $H_1 : \sigma_1^2 > \sigma_2^2$

B2. Pruebas de hipótesis sobre las medias de dos distribuciones normales con variancias conocidas (muestras independientes)

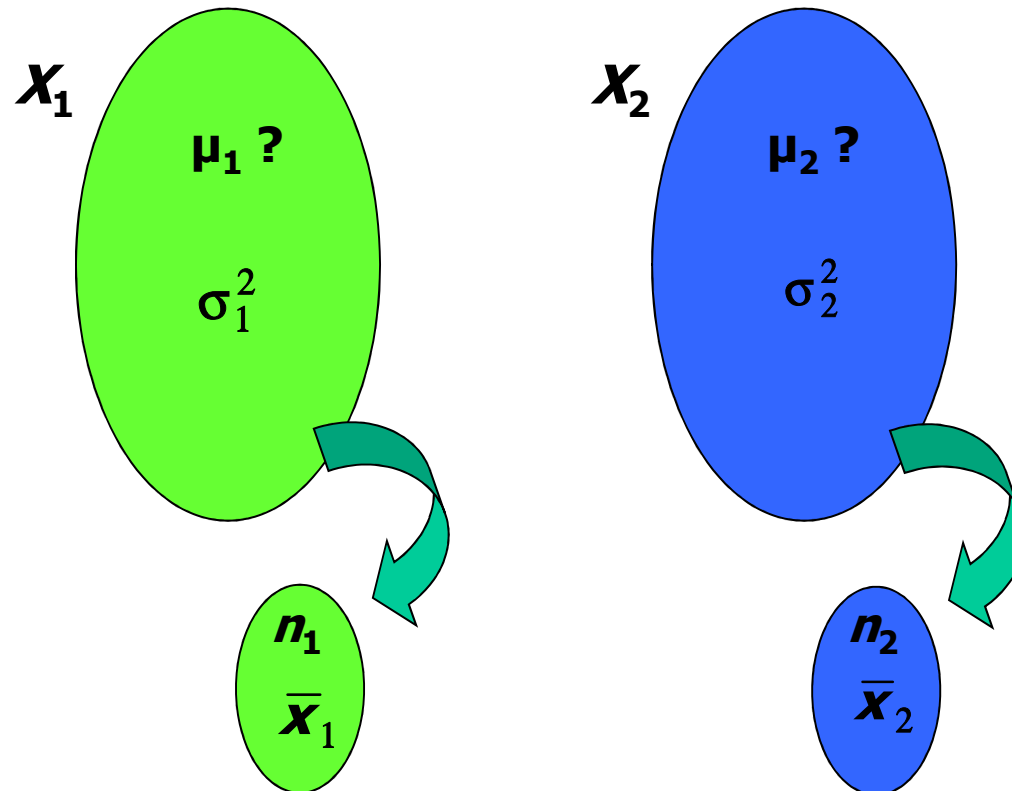
$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$



$$H_0: \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 - \mu_2 \neq 0$$



El estadístico de prueba se basa en la distribución de una diferencia de medias:

$$\bar{X}_1 - \bar{X}_2 \sim N \left(\mu_1 - \mu_2 ; \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right)$$

Bajo H_0 :

$$Z_0 = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Con $Z_0 \sim N(0;1)$

Se rechazará H_0 si:

$$Z_0 > Z_{\alpha/2} \quad \text{o} \quad Z_0 < -Z_{\alpha/2}$$

B3. Pruebas de hipótesis sobre las medias de dos distribuciones normales con variancias desconocidas (muestras independientes)

$$t_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Situación 1: $\sigma_1^2 = \sigma_2^2 = \sigma^2$

Dado que las variancias se estiman de las muestras, antes de calcular el estadístico de prueba para las medias, debemos verificar mediante una Prueba F la igualdad de las variancias poblacionales

Si la conclusión es que las variancias no son distintas, ambas variancias muestrales estiman la variancia común σ^2 , entonces podemos combinarlas para producir una sola estimación, digamos:

$$s_p^2 = \frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{n_1 + n_2 - 2}$$

El estadístico de prueba será:

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2 \left[\frac{1}{n_1} + \frac{1}{n_2} \right]}}$$

Con $t_0 \sim t(n_1+n_2-2)$

Si las hipótesis son de la forma:

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

Se rechazará H_0 si:

$$t_0 > t_{\alpha/2; n_1+n_2-2} \quad \text{o} \quad t_0 < -t_{\alpha/2; n_1+n_2-2}$$

También pueden plantearse pruebas unilaterales y todas las combinaciones entre H_0 y H_1 vistas anteriormente:

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 < \mu_2$$

Situación 2: $\sigma_1^2 \neq \sigma_2^2$

Si la homocedasticidad no es sustentable, no hay un estadístico de prueba exacto disponible para **H₀: $\mu_1 = \mu_2$** , sin embargo el estadístico:

$$t_0^* = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Se distribuye aproximadamente como t con v grados de libertad, donde :

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2 / n_1)^2}{n_1 + 1} + \frac{(s_2^2 / n_2)^2}{n_2 + 1}} - 2$$

B4. Pruebas de hipótesis para muestras apareadas

Se dicen muestras apareadas cuando las observaciones en las dos poblaciones de interés se recaban de a pares, con la premisa que cada par se toma en condiciones homogéneas, aunque estas condiciones pueden cambiar de un par a otro.

Ejemplo: considere que estamos interesados en comparar dos tipos de alimentos para cerdos (A y B). Para ello se planea una experiencia donde se suministrará cada alimento a 10 cerdos para luego comparar el peso ganado al cabo de cierto tiempo. Pero no contamos con un lote homogéneo de 20 cerdos en cuanto a su tamaño y pensamos que el peso inicial de los animales es un factor que podría incidir en los aumentos de peso. Para evitar que este efecto 'peso inicial' se pueda confundir con el efecto 'alimento' se seleccionan 'parejas' de cerdos de igual peso, y en cada una de ellas se asignan ambos alimentos al azar. Se han obtenido 20 observaciones pero estas están ***apareadas***.

Para que la hipótesis contemple el efecto de **apareo**, se calcula una nueva variable ***d*** que consiste en las diferencias observadas en cada par. La hipótesis nula a probar es que la media de las diferencias es cero

$$H_0: \mu_d = 0$$

$$H_1: \mu_d \neq 0$$

El estadístico asociado a la prueba es:

$$t_0 = \frac{\bar{d}}{s_d / \sqrt{n}}$$

$$\text{Con } t_0 \sim t(n-1)$$

La comparación de medias dependientes surge de un muestreo que se realizó con restricciones en la aleatorización, esto constituye el paso elemental en un diseño experimental en el que las unidades experimentales no son homogéneas (DBCA)

Ejercicio 2

Dos lotes de pollos de la misma raza y edad fueron alimentados durante 30 días con dos tipos diferentes de alimento balanceado. Los aumentos de peso, en gramos, fueron:

| | | | | | | | | | | | |
|--------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Balanceado A | 329 | 363 | 298 | 243 | 391 | 333 | 369 | 332 | 340 | 397 | 409 |
| Balanceado B | 353 | 405 | 372 | 345 | 377 | 409 | 428 | 421 | 357 | 372 | 409 |

- Probar la hipótesis que ambos alimentos provienen de poblaciones con varianzas iguales ($\alpha=0.01$).
- Probar si existen diferencias estadísticamente significativas entre los aumentos de peso promedio de los dos lotes. Trabaje con un nivel de significación de 5%.

Sabiendo:

| Variable | n | Media | S | S ² |
|--------------|----|--------|-------|----------------|
| Balanceado A | 11 | 345,82 | 47,97 | 2301,56 |
| Balanceado B | 11 | 386,18 | 29,14 | 849,16 |

1. Planteo las hipótesis:

$$H_0 : \sigma_A^2 = \sigma_B^2$$

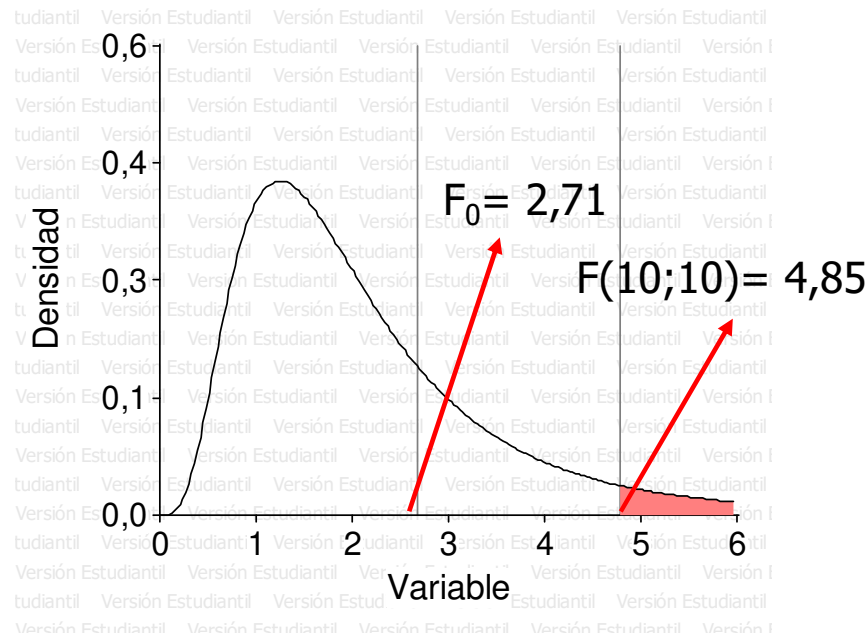
$$H_1 : \sigma_A^2 \neq \sigma_B^2$$

3. Obtención del Estadístico "F"

$$F_0 = \frac{S_A^2}{S_B^2} = \frac{2301,56}{849,16} = 2,71$$

2. Establezco las zonas de rechazo

4. Conclusión



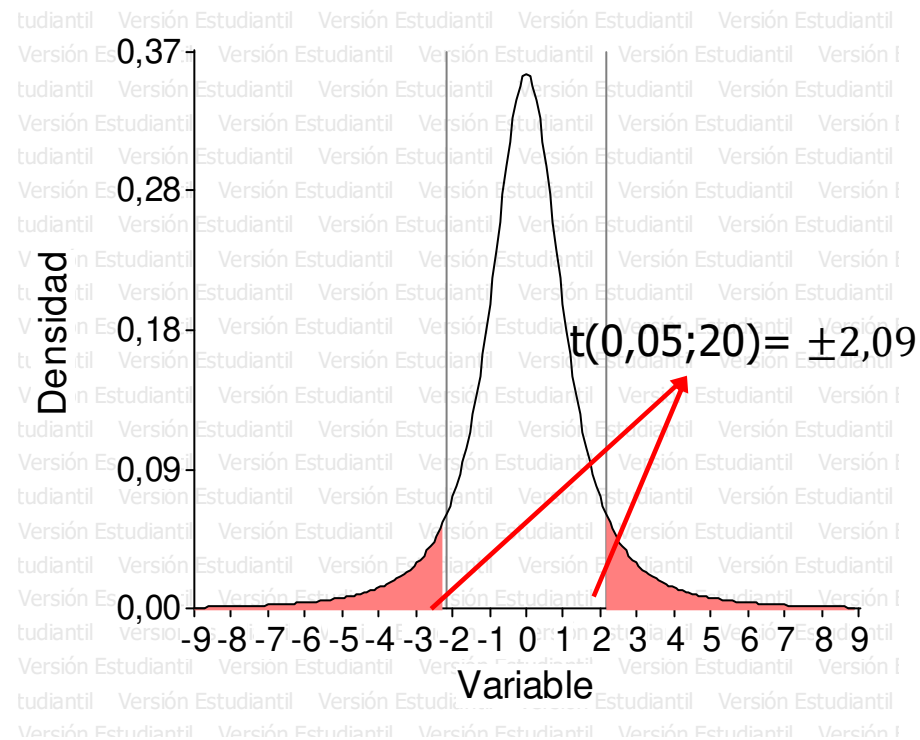
Como el estadístico F calculado cae en la zona de no rechazo la H₀ no es rechazada, por lo tanto puede concluirse para un nivel de significación del 1% que ambas alimentos provienen de poblaciones con varianzas iguales

1. Planteo las hipótesis:

$$H_0: \mu_A = \mu_B$$

$$H_1: \mu_A \neq \mu_B$$

2. Establezco las zonas de rechazo



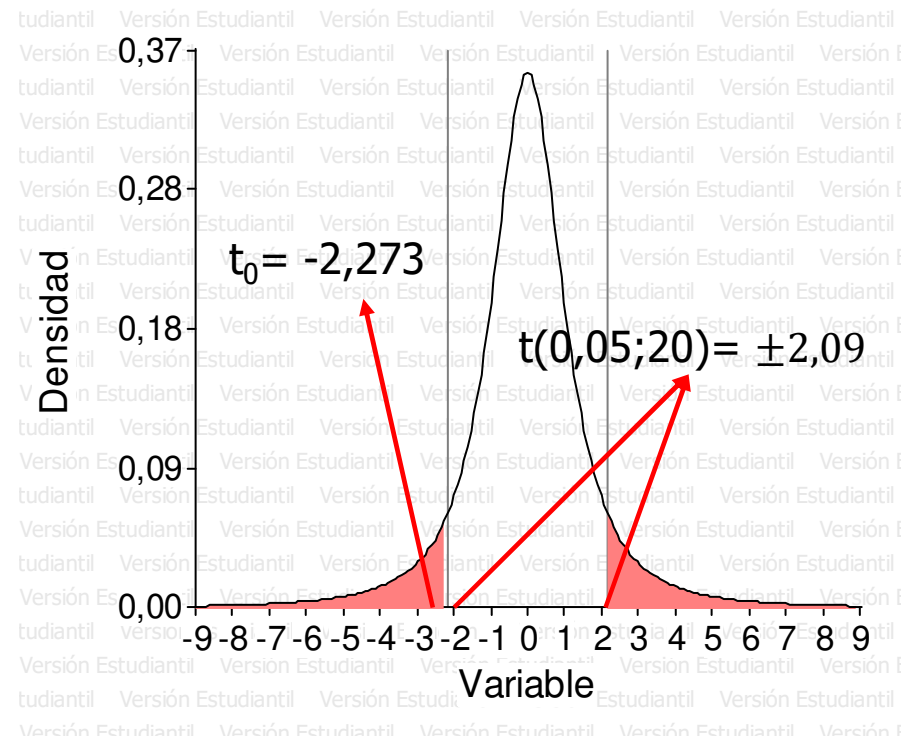
3. Obtención del Estadístico "t"

$$s_p^2 = \frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{n_1 + n_2 - 2} = \frac{10 \times 2301,56 + 10 \times 849,16}{20} = 1575,36$$

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2 \left[\frac{1}{n_1} + \frac{1}{n_2} \right]}} = \frac{345,82 - 386,18}{17,75} = -2,273$$

4. Conclusión

Como el estadístico t calculado cae en la zona de rechazo la H_0 es rechazada, por lo tanto puede concluirse para un nivel de significación del 5% que el aumento de peso difiere de un alimento a otro.



Ejercicio 3

Una empresa semillera quiere comparar el desempeño de dos variedades de maíz en una amplia región. Para realizar el ensayo se dispone que en cada una de las 6 estaciones experimentales que la empresa tiene en la zona se siembren dos parcelas, una para cada variedad. Al final del ciclo del cultivo se obtuvieron los siguientes rendimientos (qq/ha):

| Estación experimental | 1 | 2 | 3 | 4 | 5 | 6 |
|-----------------------|----|----|----|----|----|----|
| Variedad A | 70 | 80 | 75 | 60 | 68 | 72 |
| Variedad B | 72 | 81 | 77 | 62 | 68 | 74 |

- Para hacer el contraste ¿utilizaría una prueba t para muestras independientes o una prueba t apareada?
- Probar si existen diferencias estadísticamente significativas entre el rendimiento de ambas Variedades. Trabaje con un nivel de significación de 5%.