.UBA económicas
FACULTAD DE CIENCIAS ECONÓMICAS

Estadística I

Test de Hipótesis

Natalia SALABERRY

Metodología

↓

Conjunto de pasos que me permitirán obtener información a partir de los datos

Diseño → Descripción → Inferencia → Conclusión

Diseño: del plan y desarrollo del proceso de investigación

Descripción: de los datos, a través de la exploración

Inferencia: de un suceso, a partir de los datos (muestra)

Conclusión: toma de decisiones entorno al suceso

↓

Test de hipótesis

Test de Hipótesis

Una prueba o test de hipótesis es un procedimiento que permite decidir, controlando cierto riesgo, la veracidad o falsedad de hipótesis referentes a un parámetro de una población, a partir de examinar una muestra aleatoria de ella (construyendo un estadístico muestral).

Formulación de un Test de Hipótesis

En toda prueba de hipótesis hay cinco conceptos claves:

1. **Formulación de las Hipótesis Nula y Alternativa**

- La **hipótesis alternativa (H1)** es un enunciado acerca de un parámetro de la población.
- La **hipótesis nula (H0)** es la negación de la hipótesis alternativa.

Las posibilidades que se presentan, dado un parámetro poblacional $\boldsymbol{\theta}$ y un número real $\boldsymbol{\theta}_0$, son:

1) H0: $\boldsymbol{\theta} \leq \boldsymbol{\theta}_0$ H1: $\boldsymbol{\theta} > \boldsymbol{\theta}_0$
2) H0: $\boldsymbol{\theta} \geq \boldsymbol{\theta}_0$ H1: $\boldsymbol{\theta} < \boldsymbol{\theta}_0$

Prueba unilateral (o a una cola)

3) H0: $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ H1: $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$

Prueba bilateral (o a dos colas)

2. Estadístico de prueba

Los datos se sintetizan en un **estadístico** de prueba (o de tabla). Dicho estadístico se calcula para ver si es razonablemente compatible con la hipótesis nula. Su cálculo dependerá de la distribución que siga la variable que se está estudiando y los parámetros que se conocen de esa distribución.

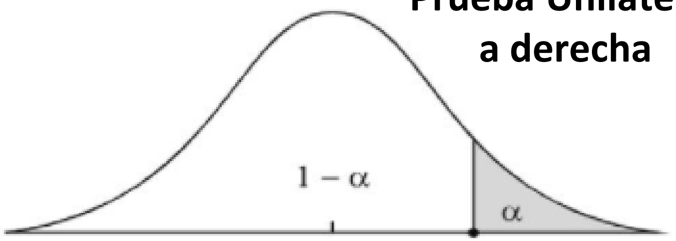
3. Zona de Rechazo

A partir del valor del estadístico de prueba, el cual es sumamente improbable bajo la H0, que define una **zona de rechazo** para el Test y permite tomar una decisión (siempre hablamos de rechazar H0 o No rechazar H0). Para ver en qué valores del estadístico de prueba se comienza a decir que los datos apoyan a la hipótesis alternativa, se requiere conocer la distribución muestral del estadístico de prueba.

4. Criterio de decisión

A partir del valor del estadístico de prueba que delimita la zona de rechazo y el valor del estadístico muestral construido, se determina un **criterio de decisión** para rechazar la H0. Este vendrá asociado al signo establecido en la H1.

Prueba Unilateral a derecha



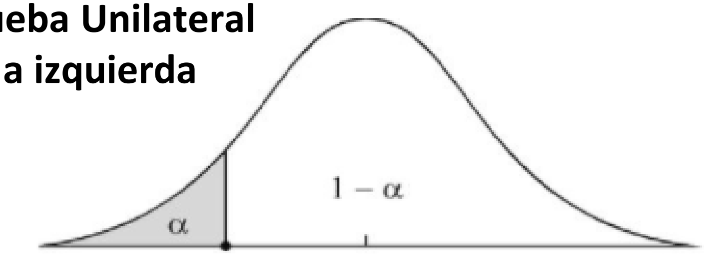
H0: $\boldsymbol{\theta} \leq \boldsymbol{\theta}_0$

H1: $\boldsymbol{\theta} > \boldsymbol{\theta}_0$

Estadístico de prueba

Rechazo H0 si el estadístico muestral cae en esta zona

Prueba Unilateral a izquierda



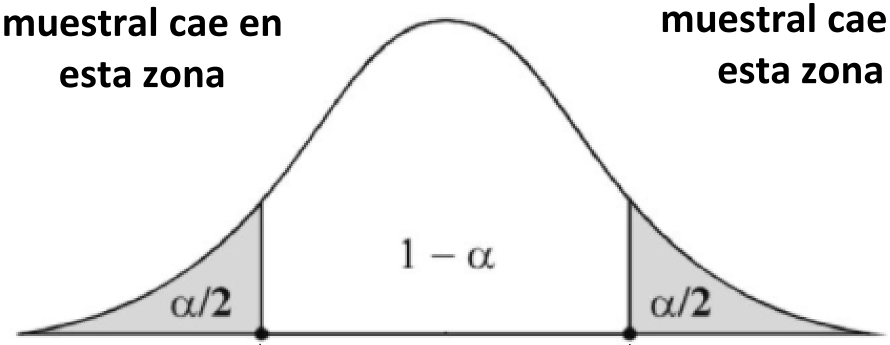
H0: $\boldsymbol{\theta} \geq \boldsymbol{\theta}_0$

H1: $\boldsymbol{\theta} < \boldsymbol{\theta}_0$

Estadístico de prueba

Rechazo H0 si el estadístico muestral cae en esta zona

Prueba Bilateral



H0: $\boldsymbol{\theta} = \boldsymbol{\theta}_0$

H1: $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$

Estadístico de prueba

Rechazo H0

Estadístico de prueba

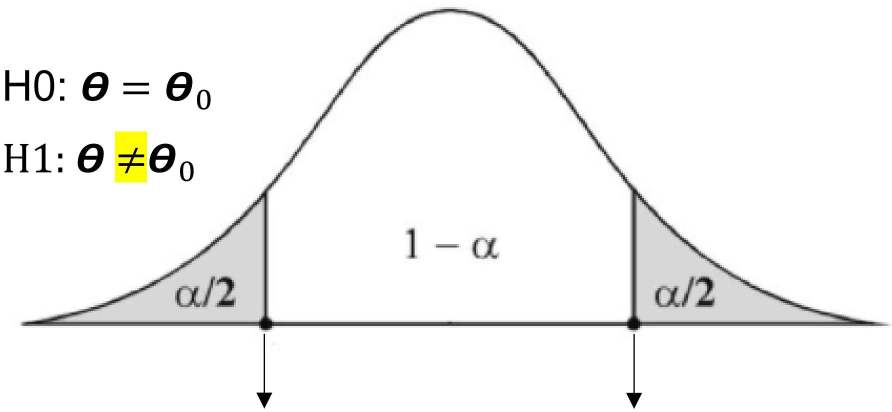
Rechazo H0

Rechazo H0 si el estadístico muestral cae en alguna de estas zonas de rechazo

Relación entre la prueba bilateral y los intervalos de confianza

H0: $\boldsymbol{\theta} = \boldsymbol{\theta}_0$

H1: $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$



Estadístico de prueba — Estadístico de prueba

Rechazo H0 — **Rechazo H0**

Rechazo H0 si el estadístico muestral cae en alguna de estas zonas de rechazo.

Es decir, la probabilidad de que el parámetro no pertenezca a la zona $1-\boldsymbol{\alpha}$ viene dada por:

P($\mu \notin [L_i; L_s]$)=1-**P**($\mu \in [L_i; L_s]$)=1-$[1-\boldsymbol{\alpha}]$= $\boldsymbol{\alpha}$

De aquí que es posible construir una prueba de hipótesis de nivel de significación $\boldsymbol{\alpha}$ para rechazar H0 (se pretende que el estadístico muestral caiga en la zona $1-\boldsymbol{\alpha}$)

5. Error

Cuando se especifica la zona de rechazo se debe admitir la posibilidad de estar cometiendo un **error**. Como en las hipótesis aparecen parámetros poblacionales, que tienen valores fijos desconocidos, se toma una muestra aleatoria con la que se estiman esos parámetros y se decide si la hipótesis nula es verdadera o falsa (siempre que trabajamos con una muestra existe la posibilidad de que la misma no sea muy representativa de la población)

Las posibilidades de error que se pueden presentar se expresan en el siguiente cuadro:

	Si H0 Verdadera	**Si H0 Falsa**
Si No rechazo H0	Decisión correcta	Cometo **Error Tipo II**
Si Rechazo H0	Cometo **Error Tipo I**	Decisión correcta

Error tipo I: es el que se comete cuando el Test rechaza H0 y en realidad H0 es verdadera

Nivel de significancia: es la probabilidad de cometer Error Tipo I

$\boldsymbol{\alpha}$ **= P(cometer Error tipo I) = P(rechazar H0 siendo H0 verdadera)**

Vamos a querer que sea lo más chica posible.

Asociado al nivel significación surge el concepto de **P-valor.** Es la probabilidad de observar un estadístico tan extremo o más que el observado. Cuanto menor es esta probabilidad, mayor es la evidencia contra la hipótesis nula

P-valor = P(Estadístico< estadístico Observado)

El estadístico observado es un cálculo realizado a partir de los datos muestrales, obteniéndose un valor numérico. Este valor es el que se debe buscar en la tabla correspondiente a la distribución muestral asociada para así obtener la probabilidad planteada.
Luego, si el valor de probabilidad obtenido es menor a $\boldsymbol{\alpha}$, entonces se rechaza la H0.

Error tipo II: es el que se comete cuando el Test no rechaza H0 y en realidad H0 es falsa

La probabilidad de cometer Error Tipo II se denomina con la letra $\boldsymbol{\beta}$

$$\boldsymbol{\beta} = \text{P(cometer Error tipo II)} = \text{P(no rechazar H0 siendo H0 falsa)}$$

Vamos a querer que sea lo más chica posible.

Asociado a este surge el concepto de **Potencia del test.** Es la probabilidad de no cometer Error Tipo II

$$\boldsymbol{\pi} = \text{P(no cometer Error tipo II)} = \text{P(rechazar H0 siendo H0 falsa)} = 1 - \boldsymbol{\beta}$$

Es decir que la potencia del test indica la probabilidad de tomar una decisión correcta cuando se está rechazando la hipótesis nula. Vamos a querer que sea lo más grande posible.

Reglas prácticas o a modo de resumen:

- Cuando planteamos hipótesis, siempre lo hacemos sobre parámetros poblacionales
- Por definición, en un test de hipótesis siempre se busca rechazar la H0, es decir que en H0 se plantea lo que "no quiero que suceda" y en consecuencia en H1 se plantea "lo que quiero que suceda" (esto está asociado a lo que se observa en la muestra). La H0 actúa como en un juicio donde se presume inocencia hasta que se demuestre lo contrario. Por tanto, en H1 se postula lo contrario lo que implica que en base a la evidencia (muestra) se demostrará la culpabilidad.
- La evaluación de un test de hipótesis se realiza con datos muestrales, es decir, construimos estadísticos muestrales a partir de datos muestrales.
- En consecuencia, el criterio de decisión para rechazar H0 se realiza comparando el estadístico muestral (u observado) respecto del estadístico crítico (de tabla)
- El "estadístico de prueba (o de tabla)" es quien delimita una zona de rechazo dentro de cada distribución. Hablar de zona de rechazo es hablar del valor ALFA (nivel de significación), siendo una probabilidad.

Reglas prácticas:

- Luego, si mi estadístico observado cae dentro de la zona de rechazo, entonces rechazo la H0, si no, no la rechazo.
- Una regla práctica para recordar cual es el criterio de decisión a utilizar es mirar el signo de la hipótesis alternativa (H1).
  - Si es por mayor, entonces prueba unilateral a derecha => criterio de rechazo de H0 EstadObs > EstadTabla.
  - Si es por menor entonces prueba unilateral a izquierda => criterio de rechazo de H0 EstadObs < EstadTabla.
  - Si es por distinto entonces prueba bilateral. Pero la convertiremos en una prueba unilateral a derecha => criterio de rechazo de H0 |EstadObs| > EstadTabla. Al aplicar módulo, aseguramos que siempre EstadObs sea positivo y por tanto estaremos siempre del lado derecho respecto de la media de la distribución.

Metodología para resolver un Test de hipótesis:

Los pasos que se deben seguir para resolver cualquier problema mediante una prueba de hipótesis son:

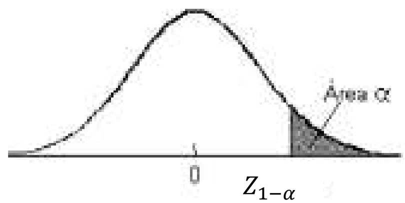
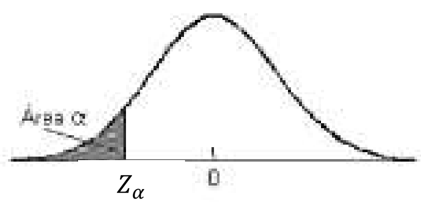
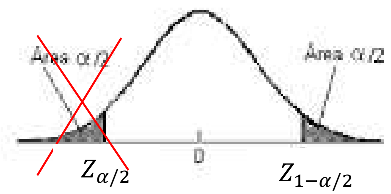
1) Definir que representa/n la/s variable/s aleatoria/s.
2) Establecer las suposiciones necesarias para que se cumpla el modelo.
3) Plantear las hipótesis H0 y H1.
4) Definir el estadístico observado.
5) Indicar la distribución del estadístico observado suponiendo que H0 sea verdadera.
6) Determinar la zona de rechazo (o criterio de decisión) teniendo en cuenta el nivel de significación.
7) Utilizando los datos de la muestra, calcular el valor del estadístico observado y tomar la decisión comparándolo con el de tabla.
8) Interpretar la decisión en términos del problema.

1. Test de Hipótesis para la media de una variable aleatoria normal con varianza conocida

Modelo: **Supuesto** X ~ N($\boldsymbol{\mu}$;$\boldsymbol{\sigma}$) con **σ conocida**

Sea $\{X_1, X_2 \ldots\ldots X_n\}$ una muestra aleatoria de X, entonces:

Estadístico Observado: $Z_{obs} = \dfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ ~ N(0;1) si H0 es verdadera

Hipótesis Nula	**Hipótesis Alternativa**	**Criterio de Decisión. Se rechaza H0 si:**	**Zona de Rechazo**
$\boldsymbol{\mu} \leq \boldsymbol{\mu}_0$	$\boldsymbol{\mu} > \boldsymbol{\mu}_0$	$Z_{obs} \geq Z_{1-\alpha}$	
$\boldsymbol{\mu} \geq \boldsymbol{\mu}_0$	$\boldsymbol{\mu} < \boldsymbol{\mu}_0$	$Z_{obs} \leq Z_{\alpha}$	
$\boldsymbol{\mu} = \boldsymbol{\mu}_0$	$\boldsymbol{\mu} \neq \boldsymbol{\mu}_0$	$\lvert Z_{obs}\rvert \geq Z_{1-\alpha/2}$	

Ejemplo 1:

La compañía A afirma que el contenido del envase del producto que ellos producen tiene un peso medio mínimo de 500gr. Algunas experiencias pasadas han demostrado que la desviación estándar del peso del envase es de 15gr. Un empleado toma una muestra aleatoria de 25 envases producidos por la compañía A y pesa su contenido, obteniendo un peso medio $\bar{X}$ = 490 gr. Entonces el empleado sospecha que la compañía no llena debidamente los envases y quiere decidir si hacer juicio o no, acusando a la compañía de desleal. Con un nivel de significación del 5%, ¿qué decisión tomará el empleado?

1) X_i=“peso del contenido del envase del producto producida por la compañía A“

2) Supuesto: $X_i \sim$ N(μ; σ) con σ=15 (conocida)

3) Hipótesis: H0: $\boldsymbol{\mu} \geq 500$ H1: $\boldsymbol{\mu} < 500$

4) El estadístico observado es: $Z_{obs} = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{X}-500}{\frac{15}{\sqrt{25}}} = \frac{490-500}{3} = \frac{-10}{3}$= - 3,33

5) Z ~N(0,1)

6) Criterio de decisión: Si $Z_{obs} < Z_{\alpha}$ => rechazo H0

7) evaluando en función del estadístico de prueba y la zona de rechazo

$\Rightarrow \boldsymbol{Z_{obs}}$**= -3,33** como es respecto de $Z_{0,05}$

$\Rightarrow Z_{0,05}$ => busco en la tabla de la normal $Z_{0,95}$=1,645 => $\boldsymbol{Z_{0,05}}$**= -1,645**

$\Rightarrow \boldsymbol{Z_{obs} < Z_{0,05}}$ **=> -3,33< -1,645 => se cumple el criterio =>rechazo H0**

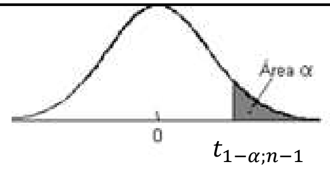
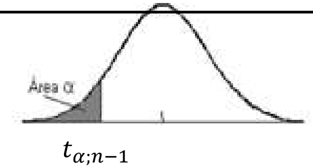
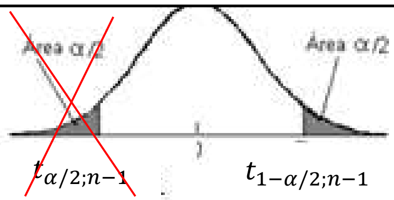
8) Por lo tanto hay evidencia suficiente para suponer que la sospecha del empleado es cierta y el empleado tomará la decisión de hacer juicio a la compañía A.

2. Test de Hipótesis para la media de una variable aleatoria normal con varianza desconocida

Modelo: **Supuesto** X ~ N($\boldsymbol{\mu}$;$\boldsymbol{\sigma}$) con **σ desconocida**

Sea $\{X_1, X_2 \ldots\ldots X_n\}$ una muestra aleatoria de X, entonces:

Estadístico observado: $t_{obs} = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$ $\boldsymbol{\sim t_{n-1}}$ si H0 es verdadera

Hipótesis Nula	**Hipótesis Alternativa**	**Criterio de Decisión Se rechaza H0 si:**	**Zona de Rechazo**
$\boldsymbol{\mu} \leq \boldsymbol{\mu}_0$	$\boldsymbol{\mu} > \boldsymbol{\mu}_0$	$t_{obs} \geq t_{1-\alpha;n-1}$	
$\boldsymbol{\mu} \geq \boldsymbol{\mu}_0$	$\boldsymbol{\mu} < \boldsymbol{\mu}_0$	$t_{obs} \leq t_{\alpha;n-1}$	
$\boldsymbol{\mu} = \boldsymbol{\mu}_0$	$\boldsymbol{\mu} \neq \boldsymbol{\mu}_0$	$\lvert t_{obs}\rvert \geq t_{1-\alpha/2;n-1}$	

Ejemplo 2:

Al ver lo que pasó con la compañía A, la empresa competidora B decide estar muy atento en su línea de producción de los envases de 500gr. Se sabe que tanto el sobrellenado o la falta de llenado son problemas graves y se debe parar la línea de producción en cualquiera de esos casos. El jefe de control de calidad decide tomar muestras aleatorias de 25 unidades cada 2hs con el fin de detectar si debe parar la producción para el ajuste de la maquinaria o bien dejarla funcionando. Si en la última muestra se obtuvo un peso promedio de 491.609 gr con un desvío de 15gr. ¿Qué decisión tomará el jefe de control de calidad si trabaja con un nivel de significación del 5%? Suponer que el peso de las cajas sigue una distribución normal

1) X_i="peso de los envases producidos por la empresa B"

2) Suponemos: $X_i \sim$ N(μ; σ) con σ desconocida

3) H0: $\boldsymbol{\mu}$=500   H1: $\boldsymbol{\mu}$≠500

4) El estadístico observado es: $t_{obs} = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} = \frac{\bar{X}-500}{\frac{15}{\sqrt{5}}} = \frac{\bar{X}-500}{3} = \frac{491.609-500}{3}$

= -2,797

5) $t_{obs} \sim t_{n-1}$

6) Criterio de decisión: Si $|t_{obs}| \geq t_{1-\alpha/2;n-1}$ => rechazo H0

7) Evaluando en función del estadístico de prueba y la zona de rechazo

$\Rightarrow t_{obs}$= - 2,797 como es respecto de $t_{1-\frac{0,05}{2};25-1}$=> busco en la tabla

$\Rightarrow t_{0,975;24}$ = 2,064

$\Rightarrow |t_{obs}| \geq t_{0,975;25-1}$=>| - 2,797| ≥ 2,064 => como se cumple el criterio => rechazo H0

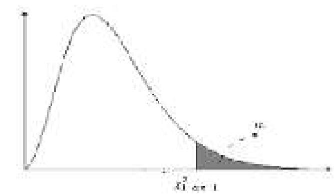
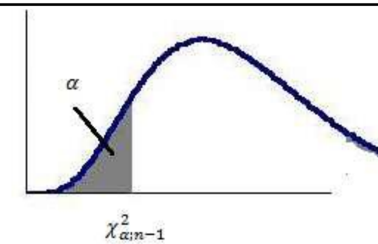
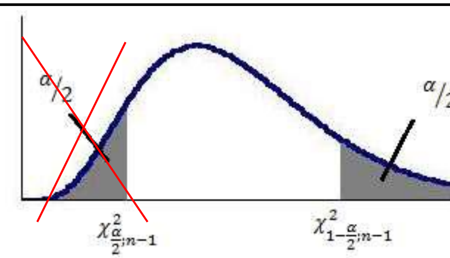
8) Por lo tanto hay evidencia suficiente para suponer que hay que reparar la maquinaria y el jefe tomará la decisión de detener el proceso.

3. Test de Hipótesis para la varianza de una variable aleatoria normal

Modelo: **Supuesto** X ~ N($\boldsymbol{\mu}$;$\boldsymbol{\sigma}$)

Sea $\{X_1, X_2 \ldots\ldots X_n\}$ una muestra aleatoria de X, entonces:

Estadístico observado: $x_{obs} = \frac{(n-1)S^2}{\sigma^2}$ $\boldsymbol{\sim \chi^2_{n-1}}$ si H0 es verdadera

Hipótesis Nula	**Hipótesis Alternativa**	**Criterio de Decisión Se rechaza H0 si:**	**Zona de Rechazo**
$\sigma^2 \leq \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$x_{obs} > x^2_{1-\boldsymbol{\alpha};n-1}$	
$\sigma^2 \geq \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$x_{obs} < x^2_{\boldsymbol{\alpha};n-1}$	
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\|x_{obs}\| > x^2_{1-\boldsymbol{\alpha}/2;n-1}$	

Ejemplo 3:

La empresa A también produce sobres de azúcar individuales. Los sobres se envasan automáticamente. Si la varianza del contenido de los sobres es mayor que 0,2 gr^2 , se debe parar el proceso de envasado para proceder a su ajuste.

Se tomó una muestra de 15 sobres, se pesó su contenido y se obtuvo una media de 10,2g y una varianza de 0,34 gr^2 .

Con un nivel de significación del 5% decidir si se debe ajustar el proceso.

1) X_i="peso en gramos del contenido de los sobres de azúcar"

2) Suponemos: $X_i \sim$ N(μ; σ)

3) H0: σ^2 ≤ 0,2   H1: σ^2 > 0,2

4) El estadístico es: $x = \frac{(n-1)S^2}{\sigma^2}$ = $\frac{(15-1)*0,34}{0,2}$ = $\frac{(15-1)*0,34}{0,2}$ = 23,8

5) $x \sim x^2_{n-1}$

6) Criterio de decisión: Si $x_{obs} > x^2_{1-\boldsymbol{\alpha};n-1}$ => rechazo H0

7) evaluando en función del estadístico de prueba y la zona de rechazo

$\Rightarrow x_{obs}$= 23,8 como es respecto de $x^2_{1-0,05;15-1}$ => buscamos en la tabla $x^2_{0,95;14}$=23,68

$\Rightarrow x_{obs} > x^2_{0,95;14}$=> se cumple el criterio => rechazo H0

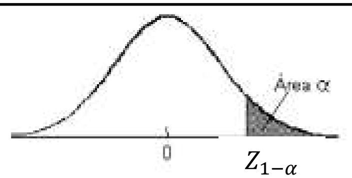
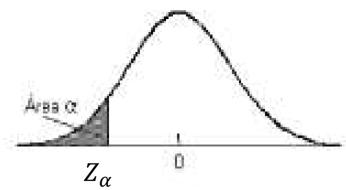
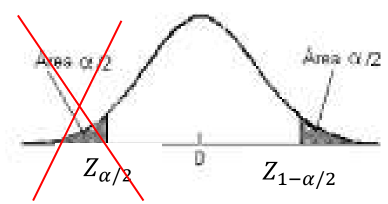
8) Por lo tanto hay evidencia suficiente para detener el proceso para realizar ajustes ya que la varianza del contenido de los sobres es mayor que 0,2 gr^2

4. Test de Hipótesis para una proporción binomial

Modelo: **Supuesto** X ~ Bi(n;p)

Sea $\{X_1, X_2 \ldots\ldots X_n\}$ una muestra aleatoria de X con distribución X ~ Bi(n;p), entonces:

Estadístico observado: $Z_{obs} = \dfrac{\bar{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0;1)$ si H0 es verdadera

Hipótesis Nula	**Hipótesis Alternativa**	**Criterio de Decisión Se rechaza H0 si:**	**Zona de Rechazo**
$p \leq p_0$	$p > p_0$	$z_{obs} \geq Z_{1-\alpha}$	
$p \geq p_0$	$p < p_0$	$z_{obs} \leq Z_{\alpha}$	
$p = p_0$	$p \neq p_0$	$\lvert z_{obs} \rvert \geq Z_{1-\alpha/2}$	

Ejemplo 4:
Estudios de mercado indican que uno de cada 3 consumidores de un producto prefiere los de tipo A. En la empresa A deciden hacer su propia encuesta ya que ellos aseguran que sus productos tipo A está muy bien posicionado en el mercado.
Se consulta a 200 consumidores del producto y se obtiene que 80 compran el de tipo A que produce la empresa A. ¿Qué puede concluirse sobre la proporción de clientes que consumen el producto de tipo A de la empresa A? Utilice un nivel de significación del 5%

1) X_i=“clientes que consumen el producto de tipo A de la empresa A“

2) Suponemos: $X_i \sim$ Bi(n;p)

3) H0: p ≤ 0,33   H1: p > 0,33   Nota: 1 de cada 3 consumidores => 1/3=0,33

4) El estadístico es $Z_{obs}= \frac{\bar{p}-p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0{,}4-0{,}33}{\sqrt{\frac{0{,}33(1-0{,}33)}{200}}} = 2{,}10$ Nota: $\hat{p}$ =80/200=0,4

5) $x \sim N(0;1)$

6) Criterio de decisión: Si $z_{obs} \geq Z_{1-\alpha}$=> rechazo H0

7) evaluando en función del estadístico de prueba y la zona de rechazo

$\Rightarrow Z_{obs}$= 2,10 como es respecto de $Z_{1-\,0,05}$=> busco en la tabla $Z_{0,95}$ =1,645

$\Rightarrow Z_{obs} \geq Z_{0,95}$=> se cumple el criterio => rechazo H0

8) Por lo tanto hay evidencia suficiente para concluir que la proporción de clientes de la empresa A que consumen su producto tipo A es superior a 0,33.

5. Test de Hipótesis para diferencia de medias en dos poblaciones con varianzas conocidas

Modelo: Supuestos X_1 ~ N($\boldsymbol{\mu}_1$; $\boldsymbol{\sigma}_1$) y X_2 ~ N($\boldsymbol{\mu}_2$; $\boldsymbol{\sigma}_2$)

Sea $\{X_1, X_2\}$ independientes y con $\boldsymbol{\sigma}_1$ y $\boldsymbol{\sigma}_2$ conocidas y n_1 y n_2 tamaños de cada muestra , entonces:

Estadístico observado: $z_{obs} = \frac{\overline{X_1} - \overline{X_2} - a}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0;1)$ si H0 es verdadera

Hipótesis Nula	**Hipótesis Alternativa**	**Criterio de Decisión Se rechaza H0 si:**
$\boldsymbol{\mu}_1$-$\boldsymbol{\mu}_2 \leq a$	$\boldsymbol{\mu}_1$-$\boldsymbol{\mu}_2 > a$	$z_{obs} \geq Z_{1-\alpha}$
$\boldsymbol{\mu}_1$-$\boldsymbol{\mu}_2 \geq a$	$\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 < a$	$z_{obs} \leq Z_{\alpha}$
$\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = a$	$\boldsymbol{\mu}_1$-$\boldsymbol{\mu}_2 \neq a$	$\lvert z_{obs} \rvert \geq Z_{1-\frac{\alpha}{2}}$

Ejemplo 5:
Una compañía de seguros médicos reúne datos sobre el tiempo de hospitalización (en días) de pacientes internados después de una cirugía de alto riesgo en dos hospitales A y B. Se sabe, por estudios anteriores, que las desviaciones estándar son 3,6 y 2,9 respectivamente y que los tiempos de internación siguen una distribución normal. En el hospital A se tomó una muestra de 56 pacientes y se obtuvo una media de 8,2 días. En el hospital B se tomó una muestra de 38 pacientes y se obtuvo una media de 9,4 días. Decida con un nivel de significación del 5% si los tiempos medios de internación **difieren** en ambos hospitales.

1) X_1="tiempo de internación de pacientes en el hospital A"
   X_2="tiempo de internación de pacientes en el hospital B"

2) Suponemos: X_1 ~ N($\boldsymbol{\mu}_1$; $\boldsymbol{\sigma}_1$) y X_2 ~ N($\boldsymbol{\mu}_2$; $\boldsymbol{\sigma}_2$) y $\boldsymbol{\sigma}_1$ y $\boldsymbol{\sigma}_2$ conocidas e iguales

3) H0: $\boldsymbol{\mu}_1$-$\boldsymbol{\mu}_2$= 0 H1:$\boldsymbol{\mu}_1$-$\boldsymbol{\mu}_2 \neq$ 0 -> ($\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$)

4) El estadístico es $x = \dfrac{\overline{X_1} - \overline{X_2} - a}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \dfrac{8,2 - 9,4 - 0}{\sqrt{\frac{3,6^2}{56} + \frac{2,9^2}{38}}} = -1,78342447 \cong -1,78$

5) x ~ $N(0; 1)$

6) Criterio de decisión: Si $|z_{obs}| \geq Z_{1-\frac{\alpha}{2}}$=> rechazo H0

7) evaluando en función del estadístico de prueba y la zona de rechazo

$\Rightarrow |z_{obs}|$ como es respecto de $Z_{1-\frac{\alpha}{2}}$ -> $Z_{0,975}$= 1,96

$\Rightarrow |-1{,}78| < Z_{0,975}$=> No rechazo H0

8) Por lo tanto hay evidencia suficiente para concluir que no hay diferencia significativa entre los tiempos medios de internación de ambos hospitales

6. Test de Hipótesis para diferencia de medias en dos poblaciones con varianzas desconocidas

Modelo: Supuestos $X_1 \sim N(\boldsymbol{\mu}_1; \boldsymbol{\sigma}_1)$ y $X_2 \sim N(\boldsymbol{\mu}_2; \boldsymbol{\sigma}_2)$

Sea $\{X_1, X_2\}$ independientes y con $\boldsymbol{\sigma}_1$ y $\boldsymbol{\sigma}_2$ **desconocidas pero iguales** y n_1 y n_2 tamaños de cada muestra , entonces:

Estadístico Observado: $t_{obs} = \dfrac{\overline{X_1} - \overline{X_2} - a}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$ si H0 es verdadera

donde $S_p{}^2 = \dfrac{(n_1-1)\,S_1^2+(n_2-1)\,S_2^2}{n_1+n_2-2}$

Hipótesis Nula	Hipótesis Alternativa	Criterio de Decisión Se rechaza H0 si:
$\boldsymbol{\mu}_1$-$\boldsymbol{\mu}_2 \leq a$	$\boldsymbol{\mu}_1$-$\boldsymbol{\mu}_2 > a$	$t_{obs} \geq t_{n_1+n_2-2;1-\boldsymbol{\alpha}}$
$\boldsymbol{\mu}_1$-$\boldsymbol{\mu}_2 \geq a$	$\boldsymbol{\mu}_1-\boldsymbol{\mu}_2 < a$	$t_{obs} \leq t_{n_1+n_2-2;\boldsymbol{\alpha}}$
$\boldsymbol{\mu}_1-\boldsymbol{\mu}_2 = a$	$\boldsymbol{\mu}_1$-$\boldsymbol{\mu}_2 \neq a$	$\lvert t_{obs}\rvert \geq t_{n_1+n_2-2;1-\frac{\boldsymbol{\alpha}}{2}}$

Ejemplo 6:

Se realizó un estudio para comparar el consumo de combustible en dos tipos de vehículos de características similares de marcas distintas. Se tomaron 12 automóviles Volkswagen y 10 Toyota en pruebas de 90 km/h.

Para los autos Volkswagen se obtuvo un promedio de 16 km por litro con una desviación estándar de 1 km por litro. Para los autos Toyota se obtuvo un promedio de 11 km por litro con una desviación estándar de 1,8 km por litro.

Pruebe la hipótesis que los vehículos Volkswagen consumen en promedio **más de 3.953** km por litro con respecto a los Toyota. Utilice un nivel de significación del 10% y suponga que el consumo de combustible sigue una distribución normal.

1) X_1="consumo de combustible en vehículos Volkswagen"
X_2="consumo de combustible en vehículos Toyota"

2) Suponemos: X_1 ~ N($\boldsymbol{\mu}_1$; $\boldsymbol{\sigma}_1$) y X_2 ~ N($\boldsymbol{\mu}_2$; $\boldsymbol{\sigma}_2$) y $\boldsymbol{\sigma}_1$ y $\boldsymbol{\sigma}_2$ desconocidas y $suponemos$ $\boldsymbol{\sigma}_1 = \boldsymbol{\sigma}_2$

3) H0: $\boldsymbol{\mu}_1$-$\boldsymbol{\mu}_2 \leq$3.953 H1:$\boldsymbol{\mu}_1$-$\boldsymbol{\mu}_2$> 3.953 -> ($\boldsymbol{\mu}_1$> $\boldsymbol{\mu}_2$+ 3.953)

4) El estadístico es $t = \frac{\overline{X_1} - \overline{X_2} - 3.953}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{16 - 11 - 3.953}{1{,}417\sqrt{\frac{1}{12}+\frac{1}{10}}} = 1{,}725$ $con S_p{}^2 = \frac{(12-1)1^2+(10-1)1{,}8^2}{12+10-2} = 2{,}008$

=> $S_p = \sqrt{2{,}008} = 1{,}417$

5) X ~ $t_{n_1+n_2-2}$

6) Criterio de decisión: Si $t_{obs} \geq t_{1-\alpha}$=> rechazo H0

7) evaluando en función del estadístico de prueba y la zona de rechazo

$\Rightarrow t_{obs}$= $1{,}725$ como es respecto de $t_{1-\alpha} \Rightarrow t_{0{,}90;20}$= 1,325

$\Rightarrow t_{obs} \geq t_{0{,}90;20}$=> rechazo H0

8) Por lo tanto hay evidencia suficiente para concluir que los vehículos Volkswagen consumen en promedio más de 3.953 km por litro con respecto a los vehículos Toyota

F de Fisher-Snedecor:

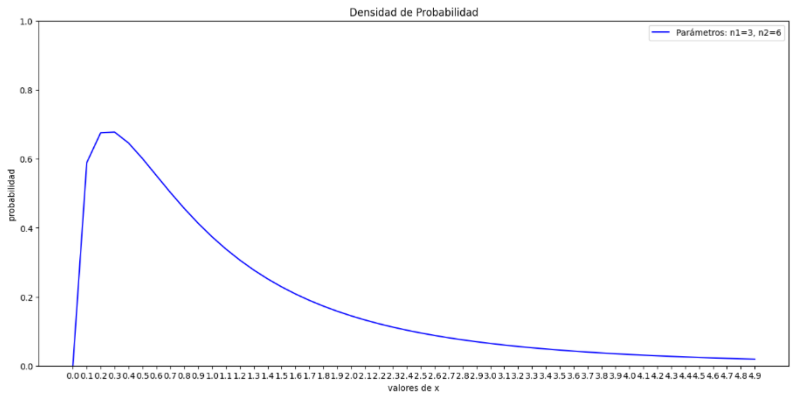
$$X \sim F_{n_1,n_2} \qquad \text{E(X)} = \frac{n_2}{n_2-2}\ para\ n_2 > 2 \qquad \text{V(X)}= \frac{2*n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}$$

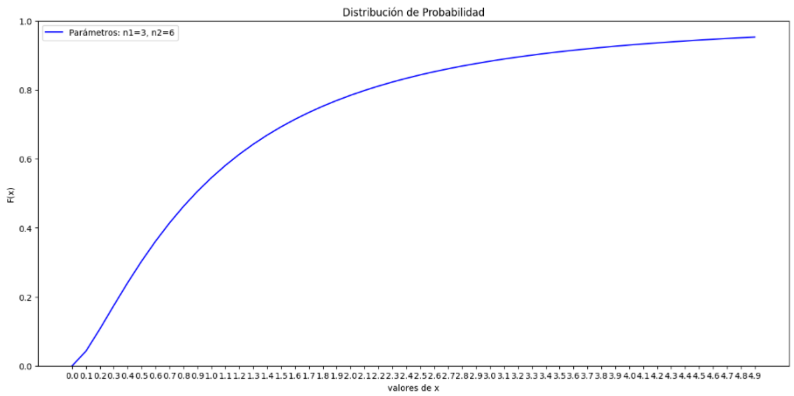
Donde n_1 son los grados de libertad del numerador y n_2 son los grados de libertad del denominador

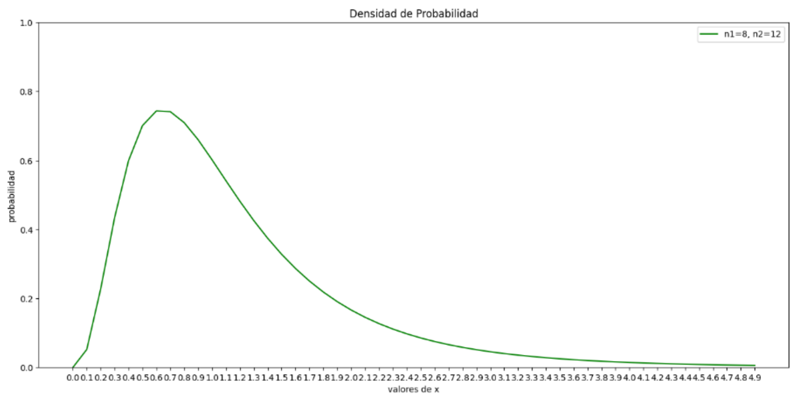
Característica: es una distribución asimétrica a derecha. Requiere uso de tabla

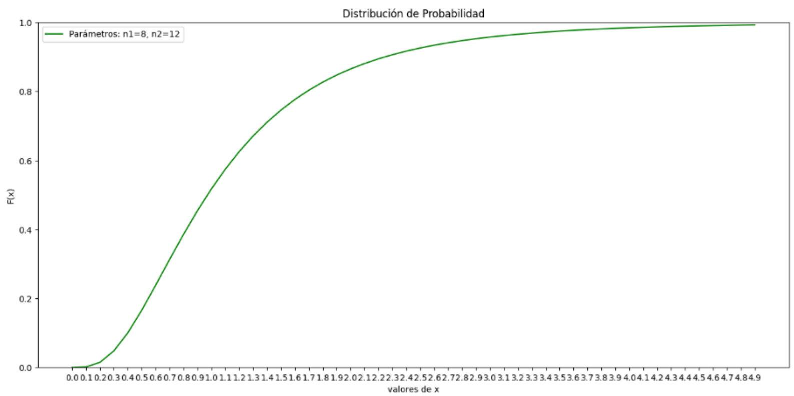
Si $X_1 \sim x_{n1}^2\ \ y\ X_2 \sim x_{n2}^2$ son dos variables aleatorias independientes, entonces la variable f:

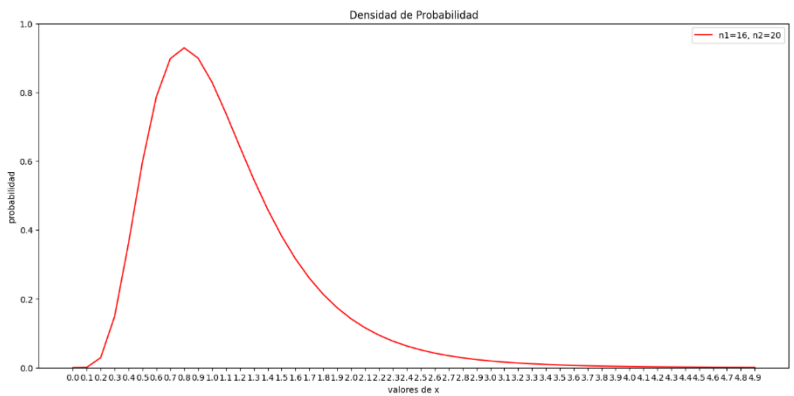
$$f=\frac{\frac{X_1}{n_1}}{\frac{X_2}{n_2}} \sim F_{n_1,n_2}$$

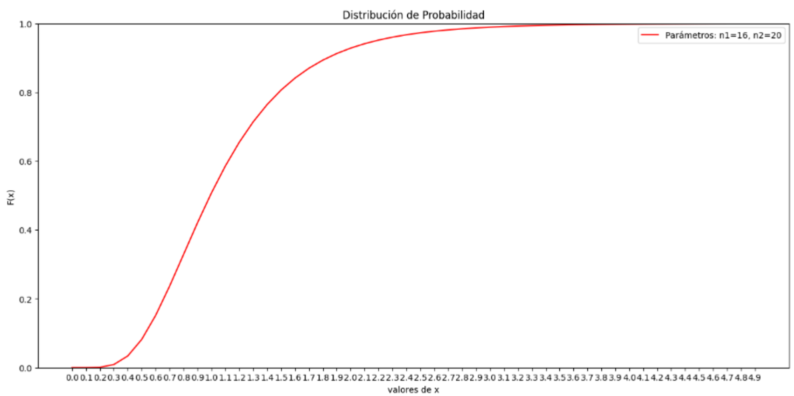

Densidad de Probabilidad
Parámetros: n1=3, n2=6
probabilidad
valores de x


Distribución de Probabilidad
Parámetros: n1=3, n2=6
F(x)
valores de x


Densidad de Probabilidad
n1=8, n2=12
probabilidad
valores de x


Distribución de Probabilidad
Parámetros: n1=8, n2=12
F(x)
valores de x


Densidad de Probabilidad
n1=16, n2=20
probabilidad
valores de x


Distribución de Probabilidad
Parámetros: n1=16, n2=20
F(x)
valores de x

Ejemplos Distribuciones Muestrales uso de tablas

Ejemplo : Supongamos $X \sim F_{n_1,n_2}$ con 7 grados de libertad en el numerador y 10 grados de libertad en el denominador. Hallar la probabilidad de que X sea menor 3,135.

La P(X < 3,135) = 0,95

Tablas. Cátedra: BIANCO, María José.

Distribución F de Snedecor

Área acumulada a izquierda: 0,95

n1 - n2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	19	24	
1	161,45	18,51	10,13	7,709	6,608	5,987	5,591	5,318	5,117	4,965	4,844	4,747	4,667	4,6	4,543	4,381	4,26	4
2	199,5	19	9,552	6,944	5,786	5,143	4,737	4,459	4,256	4,103	3,982	3,885	3,806	3,739	3,682	3,522	3,403	3
3	215,71	19,16	9,277	6,591	5,409	4,757	4,347	4,066	3,863	3,708	3,587	3,49	3,411	3,344	3,287	3,127	3,009	2
4	224,58	19,25	9,117	6,388	5,192	4,534	4,12	3,838	3,633	3,478	3,357	3,259	3,179	3,112	3,056	2,895	2,776	2
5	230,16	19,3	9,013	6,256	5,05	4,387	3,972	3,687	3,482	3,326	3,204	3,106	3,025	2,958	2,901	2,74	2,621	2
6	233,99	19,33	8,941	6,163	4,95	4,284	3,866	3,581	3,374	3,217	3,095	2,996	2,915	2,848	2,79	2,628	2,508	2
7	236,77	19,35	8,887	6,094	4,876	4,207	3,787	3,5	3,293	3,135	3,012	2,913	2,832	2,764	2,707	2,544	2,423	2
8	238,88	19,37	8,845	6,041	4,818	4,147	3,726	3,438	3,23	3,072	2,948	2,849	2,767	2,699	2,641	2,477	2,355	2
9	240.54	19.38	8.812	5.999	4.772	4.099	3.677	3.388	3.179	3.02	2.896	2.796	2.714	2.646	2.588	2.423	2.3	2

7. Test de Hipótesis para igualdad de varianzas

Modelo: $Supuestos\ X_1$ ~ N($\boldsymbol{\mu}_1$; $\boldsymbol{\sigma}_1$) y X_2 ~ N($\boldsymbol{\mu}_2$; $\boldsymbol{\sigma}_2$)

Sea $\{X_1, X_2\}$ independientes y n_1 y n_2 tamaños de cada muestra , entonces:

Estadístico: $F_{obs} = \frac{S_1^2}{S_2^2} \sim F_{\mathrm{n}_1-1;\mathrm{n}_2-1}$ si H0 es verdadera

Hipótesis Nula	**Hipótesis Alternativa**	**Criterio de Decisión Se rechaza H0 si:**
$\sigma_1^2 \leq \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$	$F_{obs} \geq F_{1-\alpha;\mathrm{n}_1-1;\mathrm{n}_2-1}$
$\sigma_1^2 \geq \sigma_2^2$	$\sigma_1^2 < \sigma_2^2$	$F_{obs} \leq F_{\alpha;\mathrm{n}_1-1;\mathrm{n}_2-1}$
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$	$\lvert F_{obs} \rvert \geq F_{\mathrm{n}_1-1;\mathrm{n}_2-1;1-\frac{\alpha}{2}}$

Ejemplo 7:

Estudie la hipótesis acerca de la igualdad de varianzas de las variables del ejemplo 6 para un nivel de significación del 2%

1) X_1="consumo de combustible en vehículos Volkswagen"
X_2="consumo de combustible en vehículos Toyota"

2) Suponemos: X_1 ~ N($\boldsymbol{\mu}_1$; $\boldsymbol{\sigma}_1$) y X_2 ~ N($\boldsymbol{\mu}_2$; $\boldsymbol{\sigma}_2$) y $\boldsymbol{\sigma}_1$ y $\boldsymbol{\sigma}_2$ desconocidas

3) H0: $\sigma_1^2 = \sigma_2^2$ H1: $\sigma_1^2 \neq \sigma_2^2$

4) los valores obtenidos a partir de las muestras eran n_1 =12 y n_2 =10 y S_1 =1 y S_2 =1.8

El estadístico es $F_{obs} = \frac{1^2}{1.8^2}$=0,3086

5) X ~$F_{n_1-1;n_2-1}$

6) Criterio de decisión: Si $|F_{obs}| \geq F_{\mathrm{n}_1-1:\mathrm{n}_2-1;1-\frac{\alpha}{2}}$=> rechazo H0

7) evaluando en función del estadístico de prueba y la zona de rechazo

$\Rightarrow$ $|F_{obs}|$=0,3086 como es respecto de $F_{1-\frac{0,02}{2},\ 11;9}$= $F_{0,99,\ 11;9}$ =5,178

$\Rightarrow$ $|F_{obs}| < F_{0,99,11;9}$=> no rechazo H0

8) Por lo tanto hay evidencia suficiente para concluir que no hay diferencia significativa entre las varianzas de ambas variables (resultó adecuado haber usado el test de hipótesis de diferencia de medias en el ejemplo 6)

8. Test de Hipótesis para diferencia de proporciones

Modelo: **Supuesto** X_1 ~ Bi(n;p) y X_2 ~ Bi(n;p)

Estadístico observado: $Z_{obs} = \dfrac{(\overline{p_1} - \overline{p_2}) - a}{\sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2}}}$ ~ $N(0;1)$ si H0 es verdadera

Hipótesis Nula	**Hipótesis Alternativa**	**Criterio de Decisión Se rechaza H0 si:**
p_1-$p_2 \leq a$	p_1-$p_2 > a$	$z_{obs} \geq Z_{1-\alpha}$
p_1-$p_2 \geq a$	$p_1 - p_2 < a$	$z_{obs} \leq Z_{\alpha}$
$p_1 - p_2 = a$	p_1-$p_2 \neq a$	$\lvert z_{obs} \rvert \geq Z_{1-\alpha/2}$

Ejemplo 8:

Se va a registrar el voto de los residentes de una ciudad y de la zona suburbana para determinar si se debe construir una planta química dentro de los límites de la ciudad. Si 120 de los 200 votantes del pueblo y 240 de los 500 de la zona suburbana están de acuerdo con la propuesta, ¿esto favorece la hipótesis de que la proporción de votantes a favor de la propuesta de la ciudad es **mayor** que la proporción del área suburbana? Utilice un nivel de significancia de 0,025.

1) X_1=“residentes de la ciudad ”
X_2=“residentes de la zona suburbana”

2) Suponemos: X_1 ~ Bi(n;p) y X_2 ~ Bi(n;p)

3) H0:p_1-p_2≤ 0 H1:p_1-p_2> 0 -> (p_1> p_2)

4) El estadístico es $Z_{obs}= \dfrac{(\overline{p_1}-\overline{p_2})-a}{\sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1}+\frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2}}} = \dfrac{(0{,}6-0{,}48)-0}{\sqrt{\frac{0{,}6(1-0{,}6)}{200}+\frac{0{,}48(1-0{,}48)}{500}}}$ =2,911112549

Nota: $\overline{p_1}$=120/200=0,6 $\overline{p_2}$=240/500=0,48

5) Z_{obs}~ $N(0;1)$

6) Criterio de decisión: Si $z_{obs} \geq Z_{1-\alpha}$=> rechazo H0

7) evaluando en función del estadístico de prueba y la zona de rechazo

$\Rightarrow Z_{obs}$= $2{,}911112549$ como es respecto de $Z_{1-\ 0{,}025}$=> busco en la tabla $Z_{0{,}975}$ =1,96

$\Rightarrow Z_{obs} \geq Z_{0{,}95}$=> se cumple el criterio => rechazo H0

8) Por lo tanto hay evidencia suficiente para concluir que la proporción de votantes en al ciudad es mayor a los del área suburbana con un 97,5%.

EJEMPLOS

Ejemplo 1:

Un productor cuenta con dos máquinas cosechadoras y sospecha que la los tiempos medios de cosecha de la máquina A son superiores a los de la B. Para poner a prueba dicha afirmación se toma una muestra del tiempo de 9 cosechas realizadas, obteniendo una media de 15 con un desvío de 2, para la primera máquina, y una media de 9 con un desvío de 1.08 para la segunda. Considerando que ambos tiempos de cosecha se distribuyen normalmente.

Valide los supuestos que considere necesarios y realice el test correspondiente, en ambos casos con una significación del 2%.

Como las varianzas son desconocidas, el supuesto que debemos testar es de igualdad de varianzas

1) X_1="tiempo de cosecha máquina A" 2) Suponemos: X_1 ~ N($\boldsymbol{\mu}_1$; $\boldsymbol{\sigma}_1$) y X_2 ~ N($\boldsymbol{\mu}_2$; $\boldsymbol{\sigma}_2$)
X_2="tiempo de cosecha máquina B"

3) H0: $\sigma_1^2 = \sigma_2^2$ H1: $\sigma_1^2 \neq \sigma_2^2$ 4 y 5) $F_{obs} = \frac{S_1^2}{S_2^2} \sim F_{\mathrm{n}_1-1;\mathrm{n}_2-1}$

6) Criterio: $|F_{obs}| \geq F_{\mathrm{n}_1-1;\mathrm{n}_2-1;1-\frac{\alpha}{2}}$ para rechazar H0

7) Tenemos S_1 =2 y S_2 =1,08

$\Rightarrow F_{obs}$ como es respecto de $F_{1-0,02/2,8;8}$= $F_{0.99;8;8}$ =6,029

$\Rightarrow$ El estadístico es $F_{obs} = \frac{2^2}{1,08^2}$=3,4294

$\Rightarrow F_{obs} < F_{0.99;8;8}$=> no rechazo H0

8) No hay evidencia suficiente para considerar que las varianzas son diferentes. Por tanto suponer que son iguales, es correcto.

Ahora entonces resolvemos el test correspondiente

1) X_1=“tiempo de cosecha máquina A”
X_2=“tiempo cosecha máquina B”

2) Suponemos: X_1 ~ N($\boldsymbol{\mu}_1$; $\boldsymbol{\sigma}_1$) y X_2 ~ N($\boldsymbol{\mu}_2$; $\boldsymbol{\sigma}_2$) y $\boldsymbol{\sigma}_1$ y $\boldsymbol{\sigma}_2$ desconocidas pero iguales ($\boldsymbol{\sigma}_1 = \boldsymbol{\sigma}_2$)

3) H0: $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = 0$
H1: $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 > 0$ => ($\boldsymbol{\mu}_1 > \boldsymbol{\mu}_2$)

4 y 5) $t_{obs} = \dfrac{\overline{X_1} - \overline{X_2} - a}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$

6) Criterio de decisión: Si $t_{obs} \geq t_{n_1+n_2-2;1-\boldsymbol{\alpha}}$=> rechazo H0

7) Evaluamos en función del estadístico de prueba

$$t_{obs} = \frac{\overline{X}_1 - \overline{X}_2 - a}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{15 - 9 - 0}{S\sqrt{\frac{1}{9}+\frac{1}{9}}} = \frac{15 - 9 - 0}{1{,}61\sqrt{\frac{1}{9}+\frac{1}{9}}} = 7{,}9055$$

$$con\ S^2 = \frac{(9-1)2^2+(9-1)1.08^2}{9+9-2} = \frac{41.3312}{16} = 2{,}5832 \Rightarrow S=\sqrt{2{,}5832}=1{,}607233648=1{,}61$$

$$t_{9+9-2;1-0{,}02} = t_{16;\ 0{,}98} = 2{,}235$$

$$\Rightarrow t_{obs} > t_{16;\ 0{,}98} \Rightarrow \text{rechazo H0}$$

8) Por lo tanto hay evidencia suficiente para concluir que el tiempo medio de cosecha de la máquina A es mayor que en de la B

Ejemplo 2:

Considere la cotización diaria de los activos bursátiles de VISA (Xa) y de Mastercard (Xb) en el período comprendido entre el 9/6/2023 y el 7/7/2023. A continuación, las estadísticas muestrales:

	n	**MEDIA**	**DESVÍO**
VISA	25	229,1	5,1
MASTERCRAD	25	379,2	8,2

A) ¿Existe suficiente evidencia para suponer que las varianzas de ambas cotizaciones son iguales? Utilice un $\alpha=0,05$. Escriba la respuesta..

B) ¿Existe suficiente evidencia para asegurar que la cotización media diaria de VISA es más chica en 150,1 dólares respecto de la cotización media diaria de MASTERCARD? Utilice un $\alpha=0,05$. Escriba la respuesta.

A) Como las varianzas son desconocidas, el supuesto que debemos testar es de igualdad de varianzas

1) X_a="cotización de VISA"
X_b="cotización de MASTERCARD"

2) Suponemos: X_a ~ N($\boldsymbol{\mu}_a$; $\boldsymbol{\sigma}_a$) y X_b ~ N($\boldsymbol{\mu}_b$; $\boldsymbol{\sigma}_b$)

3) H0: $\sigma_a^2 = \sigma_b^2$ H1: $\sigma_a^2 \neq \sigma_b^2$ 4 y 5) $F_{obs} = \frac{S_a^2}{S_b^2} \sim F_{\mathrm{n}_a-1;\mathrm{n}_b-1}$

6) Criterio: $|F_{obs}| \geq F_{\mathrm{n}_a-1;\mathrm{n}_b-1;1-\frac{\alpha}{2}}$ para rechazar H0

7)

$\Rightarrow F_{1-0,05/2,24;24} = F_{0.975;24;24}$ =2,269

$\Rightarrow F_{obs} = \frac{5,1^2}{8,2^2}$=0,386823319

$\Rightarrow F_{obs} < F_{0.975;24;24}$=> no rechazo H0

8) No hay evidencia suficiente para considerar que las varianzas de ambas cotizaciones son diferentes. Por lo tanto, suponer que son iguales, es correcto.

B)

1) X_a="cotización de VISA"
X_b="cotización de MASTERCARD"

2) Suponemos: X_a ~ N($\boldsymbol{\mu}_a$; $\boldsymbol{\sigma}_a$) y X_b ~ N($\boldsymbol{\mu}_b$; $\boldsymbol{\sigma}_b$) y $\boldsymbol{\sigma}_a$ y $\boldsymbol{\sigma}_b$ desconocidas pero iguales **($\boldsymbol{\sigma}_a$= $\boldsymbol{\sigma}_b$)**

3) H0: $\boldsymbol{\mu}_a$-$\boldsymbol{\mu}_b$≥150,1
H1: $\boldsymbol{\mu}_a$-$\boldsymbol{\mu}_b$< 150,1 => ($\boldsymbol{\mu}_a$<$\boldsymbol{\mu}_b + 150{,}1$)

4 y 5) $t_{obs} = \dfrac{\overline{X_a} - \overline{X_b} - a}{S\sqrt{\frac{1}{n_a}+\frac{1}{n_b}}} \sim t_{n_a+n_b-2}$

6) Criterio de decisión: Si $t_{obs} < t_{n_a+n_b-2;\boldsymbol{\alpha}}$=> rechazo

7) $t_{obs} = \frac{\overline{X_a} - \overline{X_b} - a}{S\sqrt{\frac{1}{n_a}+\frac{1}{n_b}}} = \frac{20 - 12 - 0}{S\sqrt{\frac{1}{15}+\frac{1}{15}}} = \frac{229,1 - 379,2 - 150,1}{6,828250142\sqrt{\frac{1}{25}+\frac{1}{25}}} = \frac{-300,2}{1,931320792} =$ - 155,437668

$$con\ S^2 = \frac{(25-1)5,1^2 + (25-1)8,2^2}{25+25-2} = \frac{624,24 + 1613,76}{48} = \frac{2238}{48} = 46,625$$

=> S=$\sqrt{46,625}$=6,828250142

$t_{n_1+n_2-2;1-\boldsymbol{\alpha}} = t_{25+25-2;0,05} = t_{48;\ 0,05} =$ -1,677 (en tabla $t_{48;\ 0,95} = 1,677)$

$\Rightarrow t_{obs} > t_{48;\ 0,05}$=> rechazo H0

8) Por lo tanto, hay evidencia suficiente para concluir que la media de la cotización de VISA es más chica en 150,1 dólares que la media de la cotización de MASTERCARD con un 95% de confianza

Ejemplo 3:

Consideremos la inflación mensual durante 2024 (IPC-INDEC) del rubro Alimentos y Servicios básicos. En el siguiente cuadro tenemos las medias y los desvíos estándar correspondientes a los datos desde enero de 2024 hasta diciembre de 2024.

	MEDIA	DESVÍO
IPC Servicios	11,3	9,3
IPC Alimentos	5,8	5,7

a) ¿Existe suficiente evidencia para suponer que las varianzas de ambos grupos de datos son iguales? Utilice un $\alpha=0,05$

b) ¿Existe suficiente evidencia para asegurar que la media mensual del IPC Servicios es significativamente mayor que la media mensual del IPC Alimentos? Utilice un $\alpha=0,05$

a)

1) X_1="IPC Servicios"
X_2="IPC Alimentos"

2) Suponemos: X_1 ~ N($\boldsymbol{\mu}_1$; $\boldsymbol{\sigma}_1$) y X_2 ~ N($\boldsymbol{\mu}_2$; $\boldsymbol{\sigma}_2$)

3) H0: $\sigma_1^2 = \sigma_2^2$
H1: $\sigma_1^2 \neq \sigma_2^2$

4 y 5) $F_{obs} = \frac{S_1^2}{S_2^2} \sim F_{\mathrm{n}_1-1;\mathrm{n}_2-1}$

6) Criterio de decisión: $|F_{obs}| \geq F_{\mathrm{n}_1-1;\mathrm{n}_2-1;1-\boldsymbol{\alpha}/2}$ => rechazo H0

7) $F_{obs} = \frac{9{,}3^{\ 2}}{5{,}7^2}$=2,662049861

$\Rightarrow F_{1-\frac{0{,}05}{2};12-1;12-1}$= $F_{0.975;11;11}$ =3,474

$\Rightarrow F_{obs} < F_{0.975;11;11}$=> no rechazo H0

8) No hay evidencia suficiente para considerar que las varianzas son diferentes. Por tanto suponer que son iguales, es correcto.

b)

1) X_1="IPC Servicios"
X_2="IPC Alimentos"

2) Suponemos: X_1 ~ N($\boldsymbol{\mu}_1$; $\boldsymbol{\sigma}_1$) y X_2 ~ N($\boldsymbol{\mu}_2$; $\boldsymbol{\sigma}_2$) y $\boldsymbol{\sigma}_1$ y $\boldsymbol{\sigma}_2$ desconocidas pero iguales ($\boldsymbol{\sigma}_1 = \boldsymbol{\sigma}_2$)

3) H0: $\boldsymbol{\mu}_1$-$\boldsymbol{\mu}_2$= 0
H1: $\boldsymbol{\mu}_1$-$\boldsymbol{\mu}_2$> 0 => ($\boldsymbol{\mu}_1$> $\boldsymbol{\mu}_2$)

4 y 5) $t_{obs} = \dfrac{\overline{X_1} - \overline{X_2} - a}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$

6) Criterio de decisión: Si $t_{obs} \geq t_{n_1+n_2-2;1-\boldsymbol{\alpha}}$=> rechazo H0

7)

$$t_{obs} = \frac{\overline{X_1} - \overline{X_2} - a}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{11{,}3-5{,}8-0}{S\sqrt{\frac{1}{12}+\frac{1}{12}}} = \frac{5{,}5}{7{,}712976079\sqrt{\frac{1}{12}+\frac{1}{12}}} = \frac{5{,}5}{1{,}285496013} = 4{,}278504129$$

$$con\ S^2 = \frac{(12-1)9{,}3^2+(12-1)5{,}7^2}{12+12-2} = \frac{1308{,}78}{22} = 59{,}49$$

=> $S = \sqrt{59{,}49} = 7{,}712976079$

t_{obs}= 4,107 $\quad t_{n_1+n_2-2;1-\alpha} = t_{12+12-2;1-0{,}05} = t_{22;\,0{,}95}$ = 1,717

$\Rightarrow t_{obs} > t_{22;\,0{,}95}$=> rechazo H0

8) Por lo tanto hay evidencia suficiente para concluir que el IPC medio de Servicios es superior al IPC medio de Alimentos

Ejemplo 4:

Considere la cotización (tomada de https://finance.yahoo.com/cryptocurrencies/) en dólares de las criptomonedas Bitcoin (BTC) y Wrapped Bitcoin (WBTC) para un período de 50 días en cada caso comprendido entre el 8/4/2022 y el 28/5/2022. En este período, se obtuvo una media 35735,45 y un desvío 5031,74 para BTC; y una media de 35732,03 y un desvío de 5033,59 para WBTC.

a) ¿Existe suficiente evidencia para suponer que las varianzas de ambas cotizaciones son iguales? Utilice un $\alpha=0,05$

b) ¿Existe suficiente evidencia para asegurar que la cotización media diaria de BTC **difiere** significativamente de la cotización media diaria de WTBC? Utilice un $\alpha=0,05$

a)

1) X_1="cotización de BTC"

X_2="cotización de WBTC"

2) Suponemos: X_1 ~ N($\boldsymbol{\mu}_1$; $\boldsymbol{\sigma}_1$) y X_2 ~ N($\boldsymbol{\mu}_2$; $\boldsymbol{\sigma}_2$)

3) H0: $\sigma_1^2 = \sigma_2^2$

H1: $\sigma_1^2 \neq \sigma_2^2$

4 y 5) $F_{obs} = \frac{S_1^2}{S_2^2} \sim F_{\mathrm{n}_1-1;\mathrm{n}_2-1}$

6) Criterio de decisión: $|F_{obs}| \geq F_{\mathrm{n}_1-1;\mathrm{n}_2-1;1-\boldsymbol{\alpha}/2}$ => rechazo H0

7) $F_{obs} = \frac{5031{,}74^2}{5033{,}59^2}$=$0{,}999265073 \cong 0{,}999$

$\Rightarrow F_{1-0{,}05/2{,}49;49}$= $F_{0.975;49;49}$ =1,762

$\Rightarrow F_{obs} < F_{0.975;49;49}$=> no rechazo H0

8) No hay evidencia suficiente para considerar que las varianzas son diferentes. Por tanto suponer que son iguales, es correcto.

b)

1) X_1="cotización de BTC"
X_2="cotización WBTC"

2) Suponemos: X_1 ~ N($\boldsymbol{\mu}_1$; $\boldsymbol{\sigma}_1$) y X_2 ~ N($\boldsymbol{\mu}_2$; $\boldsymbol{\sigma}_2$) y $\boldsymbol{\sigma}_1$ y $\boldsymbol{\sigma}_2$ desconocidas pero iguales ($\boldsymbol{\sigma}_1 = \boldsymbol{\sigma}_2$)

3) H0: $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = 0$
H1: $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \neq 0$ => ($\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$)

4 y 5) $t_{obs} = \dfrac{\overline{X_1} - \overline{X_2} - a}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$

6) Criterio de decisión: Si $|t_{obs}| \geq t_{n_1+n_2-2;1-\boldsymbol{\alpha}/2}$ => rechazo H0

7)

$$t_{obs} = \frac{\overline{X_1} - \overline{X_2} - a}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{35735{,}45 - 35732{,}03 - 0}{S\sqrt{\frac{1}{50}+\frac{1}{50}}} = \frac{3{,}42}{5032{,}67\sqrt{\frac{1}{50}+\frac{1}{50}}} = 0{,}0034$$

$$con\ S^2 = \frac{(50-1)5031{,}74^2 + (50-1)5033{,}59^2}{50+50-2} = \frac{2482116350}{98} = 25327717{,}86$$

$$=> S = \sqrt{25327717{,}86} = 5032{,}665085 = 5032{,}67$$

$$t_{n_1+n_2-2;1-\boldsymbol{\alpha}} = t_{50+50-2;1-0{,}05/2} = t_{98;\,0{,}975}$$

(como no están los 98 grados de libertad, tomamos el valor para 99)= 1,984

$\Rightarrow |t_{obs}| < t_{98;\,0{,}975}$=> **No rechazo H0**

8) Por lo tanto hay evidencia suficiente para concluir que las medias de la cotización de ambas criptomonedas son iguales.

Pueden realizar la práctica 7