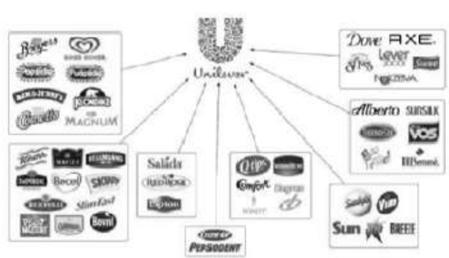
UNIDAD 2

REVISIÓN DE CONCEPTOS ESTADÍSTICOS

TEMA 3: EXPERIMENTOS COMPARATIVOS SIMPLES: MEDIAS, PROPORCIONES Y VARIANZAS CON DOS POBLACIONES

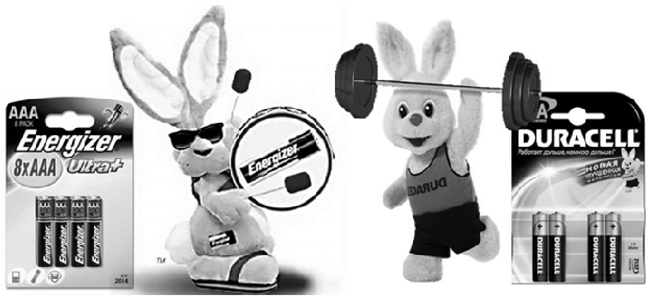
Técnicas para comparar dos poblaciones



Juan Pablo Sucre Reyes

Técnicas para comparar dos poblaciones

- Existen formas muy distintas de estimar por intervalo y realizar pruebas de hipótesis cuando se tienen dos poblaciones: *diferencia* entre dos medias poblacionales o entre dos proporciones poblacionales (*diferencia entre rendimiento de dos marcas A y B*).



Juan Pablo Sucre Reyes

Técnicas para comparar dos poblaciones

- En las pruebas de hipótesis se hace un supuesto tentativo acerca del parámetro poblacional = *hipótesis nula* (H_0). Luego se define la *hipótesis alternativa* (H_a), que contradice lo que establece la H_0.
- En este procedimiento se usan datos de una muestra para probar dos afirmaciones contrarias indicadas por H_0 y H_a.

Juan Pablo Sucre Reyes

1. Inferencia sobre la diferencia entre 2 medias poblacionales: σ_1 y σ_2 conocidas

• Inferencias acerca de: μ_1-μ_2; donde se elige una muestra aleatoria simple n_1 de la población N_1 y otra n_2 de la población N_2 *(muestras aleatorias simples independientes)*.

•*a) Estimación por intervalo para μ_1-μ_2:* su estimación puntual es: $\bar{x}_1 - \bar{x}_2$ con su error estándar $\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ y aproximando una distrib. normal se tiene: $\bar{x}_1 - \bar{x}_2 \pm$ margen de error

• De donde:

ESTIMACIÓN POR INTERVALO DE LA DIFERENCIA ENTRE DOS MEDIAS POBLACIONALES: σ_1 Y σ_2 CONOCIDAS

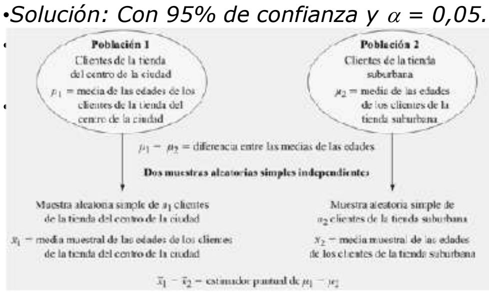
$$\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

donde $1 - \alpha$ es el coeficiente de confianza.

•*Ejemplo: Greystone Stores, tiene 2 tiendas en Buffalo NY: una en el centro y otra en un centro comercial suburbano. Los productos que se venden bien en una tienda no se venden bien en la otra (diferencias demográficas). Se investigará con 95% de confianza la diferencia entre las medias de las edades de los clientes de ambas tiendas dado σ_1=9 y σ_2=10 años. Sea:*

	Tienda del centro de la ciudad	Tienda suburbana
Tamaño de la muestra	$n_1 = 36$	$n_2 = 49$
Media muestral	$\bar{x}_1 = 40$ años	$\bar{x}_2 = 35$ años

•*Solución: Con 95% de confianza y α = 0,05. En la tabla z* $z_{0.025} = 1.96$, *con lo que:*



$$40 - 35 \pm 1.96\sqrt{\frac{9^2}{36} + \frac{10^2}{49}} \qquad 5 \pm 4.06$$

o de 0,94 a 9,06 años.

Juan Pablo Sucre Reyes

1.1 Pruebas de hipótesis acerca de μ_1-μ_2

• Las 3 formas son: $H_0: \mu_1 - \mu_2 \geq D_0$, $H_a: \mu_1 - \mu_2 < D_0$; $H_0: \mu_1 - \mu_2 \leq D_0$, $H_a: \mu_1 - \mu_2 > D_0$; $H_0: \mu_1 - \mu_2 = D_0$, $H_a: \mu_1 - \mu_2 \neq D_0$, pudiendo D_0=0 (iguales μ)

• Los 5 pasos desarrollados son válidos *(α, estadística de prueba z, valor-p, etc.)*

ESTADÍSTICO DE PRUEBA PARA PRUEBAS DE HIPÓTESIS ACERCA DE $\mu_1 - \mu_2$: σ_1 Y σ_2 CONOCIDAS

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

•*Ejemplo: Se evalúa las diferencias en la calidad educativa entre 2 centros, y se aplica un examen estandarizado a los individuos de ambos centros. La diferencia de calidad se evalúa al comparar las medias de las puntuaciones obtenidas. Se parte del supuesto tentativo a un α =0,05 de que no existe diferencia. Se conoce que σ_1=σ_2=10 y se tiene que n_1=30 (centro A) y n_2=40 individuos (centro B), además que $\bar{x}_1$=1,82 y $\bar{x}_2$=2,78.*

•*Solución: : Paso 1: Establecer H_0 y H_a:* $H_0: \mu_1 - \mu_2 = 0$, $H_a: \mu_1 - \mu_2 \neq 0$ donde D_0 =0.

• *Paso 2: Especificar el nivel de significancia: siendo α = 0,05.*

• *Paso 3: Recabar datos muestrales y calcular el valor del estadístico de prueba z:*

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(82 - 78) - 0}{\sqrt{\frac{10^2}{30} + \frac{10^2}{40}}} = 1.66$$

•*Método del valor-p*

•*Paso 4. Emplear el valor del estadístico de prueba para calcular el valor-p:*

• *El valor-p = 2(0,0485) = 0,0970. (Tablas z, área a la derecha de 1,66 = 1-0,9515)*

•*Paso 5. Regla de decisión:*

• *Como 0,0970 >α=0,05$\Rightarrow$ H_0 no es rechazada (no hay evidencia para que $\exists$ diferencia)*

Juan Pablo Sucre Reyes

1.1 Pruebas de hipótesis acerca de μ_1-μ_2

•*Ejemplo: Se evalúa las diferencias en la calidad educativa entre 2 centros, y se aplica un examen estandarizado a los individuos de ambos centros. La diferencia de calidad se evalúa al comparar las medias de las puntuaciones obtenidas. Se parte del supuesto tentativo a un α =0,05 de que no existe diferencia. Se conoce que $\sigma_1=\sigma_2=10$ y se tiene que $n_1=30$ (centro A) y $n_2=40$ individuos (centro B), además que $\bar{x}_1=1,82$ y $\bar{x}_2=2,78$.*

•*Solución: : Paso 1: Establecer H_0 y H_a:* $H_0: \mu_1 - \mu_2 = 0$, $H_a: \mu_1 - \mu_2 \neq 0$ donde D_0 =0.

• *Paso 2: Especificar el nivel de significancia: siendo α = 0,05.*

• *Paso 3: Recabar datos muestrales y calcular el valor del estadístico de prueba z:*

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(82 - 78) - 0}{\sqrt{\frac{10^2}{30} + \frac{10^2}{40}}} = 1.66$$

•*Método del valor crítico*

Rechazar H_0 si $z \le -z_{\alpha/2}$ o si $z \ge z_{\alpha/2}$

•*Paso 4. Utilizar α para determinar los valores críticos y la regla de rechazo:*

•*Si α =0,05; en cada cola, el área más allá del valor crítico es $\alpha/2$=0,025. En tabla z los valores críticos para el estadístico de prueba son $-z_{0,025}$=-1,96 y $z_{0,025}$=1,96.*

•*Paso 5. Emplear el valor del estadístico de prueba z y la regla de rechazo para determinar si H_0 es rechazada.*

• *Como z=1,66 no es < -1,96, ni tampoco es > 1,96 $\Rightarrow H_0$ no es rechazada (no existe evidencia estadística para afirmar la diferencia en la calidad educativa de A y B).*

•*NOTA: Generalmente $n_i \ge 30$ es adecuado. Si alguna es < 30, debe existir la suposición razonable que N_i tiene distribución normal o al menos aproximada.*

Juan Pablo Sucre Reyes

2. Inferencia sobre la diferencia entre 2 medias poblacionales: σ_1 y σ_2 desconocidas

• Inferencias acerca de: μ_1-μ_2; para estimar σ_1 y σ_2 desconocidas se usan S_1 y $S_2 \Rightarrow$ las estimaciones por intervalo y pruebas de hipótesis se basan en la distribución t.

•*a) Estimación por intervalo para μ_1-μ_2:* Dada por:

ESTIMACIÓN POR INTERVALO PARA LA DIFERENCIA ENTRE DOS MEDIAS POBLACIONALES: σ_1 Y σ_2 DESCONOCIDAS

$$\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

donde $1 - \alpha$ es el coeficiente de confianza.

• y los g.l. para $t_{\alpha/2}$:
•con redondeo hacia
•abajo (*t mayor*).

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

•*Ejemplo: Clearwater Bank realiza un estudio para hallar diferencias entre las cuentas de cheques de sus clientes en 2 de sus sucursales: Cherry Grove y Beechmont. Toma una muestra aleatoria simple de 28 cuentas de la 1ª y otra de 22 de la 2ª. Con 95% de confianza, se desea estimar la diferencia entre el saldo medio en dichas cuentas.*

	Cherry Grove	Beechmont
Tamaño de la muestra	$n_1 = 28$	$n_2 = 22$
Media muestral	$\bar{x}_1 = \$1\,025$	$\bar{x}_2 = \$910$
Desviación estándar muestral	$s_1 = \$150$	$s_2 = \$125$

•*Solución: Los* g.l. para $t_{\alpha/2}$:

$$gl = \frac{\left(\frac{150^2}{28} + \frac{125^2}{22}\right)^2}{\frac{1}{28 - 1}\left(\frac{150^2}{28}\right)^2 + \frac{1}{22 - 1}\left(\frac{125^2}{22}\right)^2} = 47.8$$

y se redondea a 47.

•*Con 47 g.l. y α = 0,05; en tabla t* $t_{0.025} = 2.012$ *, la estimación por intervalo de 95% de confianza para μ_1-μ_2 es:* $1\,025 - 910 \pm 2.012\sqrt{\frac{150^2}{28} + \frac{125^2}{22}}$ $\Rightarrow 115 \pm 78$ *o de 37$ a 193$.*

Juan Pablo Sucre Reyes

2.1 Pruebas de hipótesis acerca de μ_1-μ_2

• Las 3 formas son: $\begin{matrix} H_0: \mu_1 - \mu_2 \geq D_0 & H_0: \mu_1 - \mu_2 \leq D_0 & H_0: \mu_1 - \mu_2 = D_0 \\ H_a: \mu_1 - \mu_2 < D_0 & H_a: \mu_1 - \mu_2 > D_0 & H_a: \mu_1 - \mu_2 \neq D_0 \end{matrix}$, pudiendo $D_0=0$ (iguales μ)

• Los 5 pasos desarrollados son válidos (*α, estadística de prueba z, valor-p, etc.)*

• ESTADÍSTICO DE PRUEBA PARA PRUEBAS DE HIPÓTESIS ACERCA DE $\mu_1 - \mu_2$

σ_1 Y σ_2 DESCONOCIDAS $\quad t = \dfrac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$ y para los g.l. se usa tb.:

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

•*Ejemplo: Para evaluar un nuevo software, se toma una muestra aleatoria de 24 analistas de sistemas. A cada analista se le pide elaborar un sistema de información hipotético: 12 de ellos usan tecnología actual, los otros 12 usan el nuevo software y se mide el tiempo necesario para completar el proyecto. Se espera probar al 5% que con el nuevo software se tarda mucho menos que antes.*

$n_1 = 12$	$n_2 = 12$
$\bar{x}_1 = 325$ horas	$\bar{x}_2 = 286$ horas
$s_1 = 40$	$s_2 = 44$

•*Solución: : Paso 1: Establecer H_0 y H_a* : $\begin{matrix} H_0: \mu_1 - \mu_2 \leq 0 \\ H_a: \mu_1 - \mu_2 > 0 \end{matrix}$ donde $D_0 = 0$.

• *Paso 2: Especificar el nivel de significancia: siendo α = 0,05.*

• *Paso 3: Recabar datos muestrales y calcular el valor del estadístico de prueba t:*

•*Método del valor-p*

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \frac{(325 - 286) - 0}{\sqrt{\dfrac{40^2}{12} + \dfrac{44^2}{12}}} = 2.27 \qquad gl = \frac{\left(\dfrac{40^2}{12} + \dfrac{44^2}{12}\right)^2}{\dfrac{1}{12-1}\left(\dfrac{40^2}{12}\right)^2 + \dfrac{1}{12-1}\left(\dfrac{44^2}{12}\right)^2} = 21.8$$

•*Paso 4. Emplear el valor del estadístico de prueba para calcular el valor-p:*

• *0,025 < valor-p < 0,01 (Tablas t: 21 g.l. área a la derecha de 2,27)*

•*Paso 5. Regla de decisión:*

Área en la cola superior	0.20	0.10	0.05	0.025	0.01	0.005
Valor-*t* (21 *gl*)	0.859	1.323	1.721	2.080	2.518	2.831

• *Como valor-p < α=0,05$\Rightarrow H_0$ es rechazada (el nuevo software es más rápido)*

2.1 Pruebas de hipótesis acerca de μ_1-μ_2

•*Ejemplo: Para evaluar un nuevo software, se toma una muestra aleatoria de 24 analistas de sistemas. A cada analista se le pide elaborar un sistema de información hipotético: 12 de ellos usan tecnología actual, los otros 12 usan el nuevo software y se mide el tiempo necesario para completar el proyecto. Se espera probar al 5% que con el nuevo software se tarda mucho menos que antes.*

$n_1 = 12$	$n_2 = 12$
$\bar{x}_1 = 325$ horas	$\bar{x}_2 = 286$ horas
$s_1 = 40$	$s_2 = 44$

•*Solución: : Paso 1: Establecer H_0 y H_a* : $\begin{matrix} H_0: \mu_1 - \mu_2 \leq 0 \\ H_a: \mu_1 - \mu_2 > 0 \end{matrix}$ donde $D_0 = 0$.

• *Paso 2: Especificar el nivel de significancia: siendo α = 0,05.*

• *Paso 3: Recabar datos muestrales y calcular el valor del estadístico de prueba t:*

•*Método del valor crítico*

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \frac{(325 - 286) - 0}{\sqrt{\dfrac{40^2}{12} + \dfrac{44^2}{12}}} = 2.27 \qquad gl = \frac{\left(\dfrac{40^2}{12} + \dfrac{44^2}{12}\right)^2}{\dfrac{1}{12-1}\left(\dfrac{40^2}{12}\right)^2 + \dfrac{1}{12-1}\left(\dfrac{44^2}{12}\right)^2} = 21.8$$

•*Paso 4. Utilizar α para determinar el valor crítico (t_α) y la regla de rechazo:* Rechazar H_0 si $t \geq t_\alpha$

• *t_α (21 g.l.; área de α = 0,05 en la cola superior)= 1,721*

•*Paso 5. Regla de decisión:*

Área en la cola superior	0.20	0.10	0.05	0.025	0.01	0.005
Valor-*t* (21 *gl*)	0.859	1.323	1.721	2.080	2.518	2.831

• *Como 2,27 > 1,721 $\Rightarrow H_0$ es rechazada (el nuevo software es más rápido)*

•*NOTA:* Usar $n_1=n_2$; de forma que $n_1+n_2 \leq 20$ (aún con N no normales, sino > 20)

3. Inferencias acerca de la diferencia entre dos proporciones poblacionales

• Inferencias acerca de: p_1-p_2; donde se elige una muestra aleatoria simple n_1 de la población N_1 y otra n_2 de la población N_2 *(muestras aleatorias simples independientes).*

•*a) Estimación por intervalo para p_1-p_2:* su estimación puntual es: $\bar{p}_1 - \bar{p}_2$ con su error estándar $\sigma_{\bar{p}_1-\bar{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$ y aproximando una distrib. normal: $\bar{p}_1 - \bar{p}_2 \pm \text{margen de error}$

• De donde:

ESTIMACIÓN POR INTERVALO DE LA DIFERENCIA ENTRE DOS PROPORCIONES POBLACIONALES

$$\bar{p}_1 - \bar{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}}$$

donde $1 - \alpha$ es el coeficiente de confianza.

•*Ejemplo: Una empresa que elabora declaraciones de impuestos desea comparar la calidad del trabajo en 2 de sus oficinas regionales. Con muestras aleatorias del trabajo en dichas oficinas (exactitud), se podrá estimar la proporción de declaraciones con errores en que incurrió cada una de estas oficinas. Al 90% de confianza interesa el intervalo de p_1-p_2.*

Oficina 1	Oficina 2
$n_1 = 250$	$n_2 = 300$
Número de declaraciones con errores = 35	Número de declaraciones con errores = 27

•*Solución: Las proporciones muestrales en cada oficina:* $\bar{p}_1 = \frac{35}{250} = 0.14$ $\bar{p}_2 = \frac{27}{300} = 0.09$

• *Con 90% de confianza y α = 0,10. En la tabla z* $z_{0.05} = 1.645$*, con lo que:*

$0.14 - 0.09 \pm 1.645\sqrt{\frac{0.14(1-0.14)}{250} + \frac{0.09(1-0.09)}{300}}$ $\Rightarrow$ 0.05 ± 0.045 *o 0,005 a 0,095.*

3.1 Pruebas de hipótesis acerca de p_1-p_2

• Las 3 formas son: $H_0: p_1 - p_2 \geq 0$, $H_a: p_1 - p_2 < 0$; $H_0: p_1 - p_2 \leq 0$, $H_a: p_1 - p_2 > 0$; $H_0: p_1 - p_2 = 0$, $H_a: p_1 - p_2 \neq 0$, pudiendo D_0=0 (iguales p)

• Los 5 pasos desarrollados son válidos *(α, estadística de prueba z, valor-p, etc.).*

• *Si se supone H_0 verdadera como igualdad, p_1= p_2 = p. De ahí:*

• ERROR ESTÁNDAR DE $p_1 - p_2$ CUANDO $p_1 = p_2 = p$

$$\sigma_{\bar{p}_1-\bar{p}_2} = \sqrt{\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}} = \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

y ESTIMADOR COMBINADO DE p CUANDO $p_1 = p_2 = p$

$$\bar{p} = \frac{n_1\bar{p}_1 + n_2\bar{p}_2}{n_1 + n_2}$$

ESTADÍSTICO DE PRUEBA PARA PRUEBAS DE HIPÓTESIS ACERCA DE $p_1 - p_2$

$$z = \frac{(\bar{p}_1 - \bar{p}_2)}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

•*Ejemplo Una empresa que elabora declaraciones de impuestos desea comparar la calidad del trabajo en 2 de sus oficinas regionales. Con 10% de significancia probar que las proporciones de errores en las 2 oficinas son diferentes.*

•*Solución: : Paso 1: Establecer H_0 y H_a :* $H_0: p_1 - p_2 = 0$, $H_a: p_1 - p_2 \neq 0$ donde D_0 =0.

• *Paso 2: Especificar el nivel de significancia: siendo α = 0,10.*

• *Paso 3: Recabar datos muestrales y calcular el valor del estadístico de prueba z:*

$$\bar{p} = \frac{n_1\bar{p}_1 + n_2\bar{p}_2}{n_1 + n_2} = \frac{250(0.14) + 300(0.09)}{250 + 300} = 0.1127$$

$$z = \frac{(\bar{p}_1 - \bar{p}_2)}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(0.14 - 0.09)}{\sqrt{0.1127(1-0.1127)\left(\frac{1}{250} + \frac{1}{300}\right)}} = 1.85$$

•*Método del valor-p*

•*Paso 4. Emplear el valor del estadístico de prueba para calcular el valor-p:*

• *valor-p = (2* área a la derecha de 1,85: Tabla z)= 2(0,0322)=0,0644.*

•*Paso 5. Regla de decisión:*

• *Como valor-p <α=0,10 $\Rightarrow$ H_0 es rechazada (las p de errores difieren en las 2 oficinas)*

3.1 Pruebas de hipótesis acerca de p_1-p_2

•*Ejemplo Una empresa que elabora declaraciones de impuestos desea comparar la calidad del trabajo en 2 de sus oficinas regionales. Con 10% de significancia probar que las proporciones de errores en las 2 oficinas son diferentes.*

•*Solución: : Paso 1: Establecer H_0 y H_a :* $H_0: p_1 - p_2 = 0$ donde $D_0 = 0$.
$H_a: p_1 - p_2 \neq 0$

• *Paso 2: Especificar el nivel de significancia: siendo α = 0,10.*

• *Paso 3: Recabar datos muestrales y calcular el valor del estadístico de prueba z:*

$$\bar{p} = \frac{n_1\bar{p}_1 + n_2\bar{p}_2}{n_1 + n_2} = \frac{250(0.14) + 300(0.09)}{250 + 300} = 0.1127$$

$$z = \frac{(\bar{p}_1 - \bar{p}_2)}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(0.14 - 0.09)}{\sqrt{0.1127(1 - 0.1127)\left(\frac{1}{250} + \frac{1}{300}\right)}} = 1.85$$

•*Método del valor crítico*

•*Paso 4. Utilizar α para determinar los valores críticos y la regla de rechazo:*

Rechazar H_0 si
$z \leq -z_{\alpha/2}$
o si $z \geq z_{\alpha/2}$

•*Si α =0,10; en cada cola, el área más allá del valor crítico es $\alpha/2$=0,05. En tabla z los valores críticos para el estadístico de prueba son $-z_{0,05}$=-1,645 y $z_{0,05}$=1,645.*

•*Paso 5. Emplear el valor del estadístico de prueba z y la regla de rechazo para determinar si H_0 es rechazada.*

• *Como z=1,85 es > 1,645 $\Rightarrow$ H_0 es rechazada (las p de errores difieren en las 2 oficinas)*

Inferencias acerca de las varianzas poblacionales

• En muchas aplicaciones de fabricación y/o experimentación, controlar la varianza del proceso es de suma importancia para conservar la calidad.

HONDA
The Power of Dreams

4. Inferencias acerca de una varianza poblacional

• Inferencias acerca de: σ^2; a partir de la varianza muestral: $s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1}$

•Donde se vuelve muy útil:

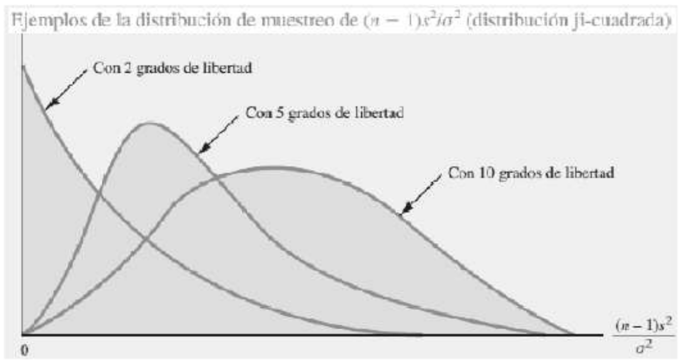
DISTRIBUCIÓN DE MUESTREO DE $(n-1)s^2/\sigma^2$

Siempre que de una población normal se tome una muestra aleatoria simple de tamaño n, la distribución de muestreo de

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

tendrá una distribución ji-cuadrada con $n-1$ grados de libertad.

• Ésta sirve para realizar estimación por intervalo y pruebas de hipótesis acerca de σ^2.



Juan Pablo Sucre Reyes

4.1 Estimación por intervalo de σ^2

• Proceso del uso de la distribución ji-cuadrada para obtener una estimación por intervalo de la varianza poblacional y la desviación estándar poblacional:

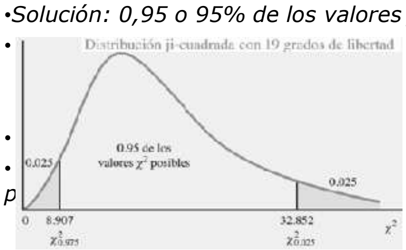
ESTIMACIÓN POR INTERVALO PARA UNA VARIANZA POBLACIONAL

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{(1-\alpha/2)}}$$

donde los valores χ^2 están basados en una distribución ji-cuadrada con $n-1$ grados de libertad, y $1-\alpha$ es el coeficiente de confianza.

•*Ejemplo: Se desea estimar la σ^2 del proceso industrial de llenado en una fábrica. En una n = 20 envases se encuentra que la S^2 = 0,0025 (cantidades de llenado). Obtener una estimación por intervalo para la varianza poblacional del 95%.*

•*Solución: 0,95 o 95% de los valores ji-cuadrada están entre $\chi^2_{0.975}$ y $\chi^2_{0.025}$ (tabla X^2).*



Se tiene así: $\frac{(19)(0.0025)}{32.852} \leq \sigma^2 \leq \frac{(19)(0.0025)}{8.907}$ *ó*

$$0.0014 \leq \sigma^2 \leq 0.0053$$

de ahí, se obtiene el siguiente intervalo de 95% de confianza para la desviación estándar

p

$$0.0380 \leq \sigma \leq 0.0730$$

Juan Pablo Sucre Reyes

4.2 Pruebas de hipótesis de σ^2

• Con para denotar el valor hipotético de la varianza poblacional y calcular z:
• y para el valor del estadístico X^2

	Prueba de cola inferior	Prueba de cola superior	Prueba de dos colas
Hipótesis	$H_0: \sigma^2 \geq \sigma_0^2$ $H_a: \sigma^2 < \sigma_0^2$	$H_0: \sigma^2 \leq \sigma_0^2$ $H_a: \sigma^2 > \sigma_0^2$	$H_0: \sigma^2 = \sigma_0^2$ $H_a: \sigma^2 \neq \sigma_0^2$
Estadístico de prueba	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$
Regla de rechazo: método del valor-p	Rechazar H_0 si el valor-$p \leq \alpha$	Rechazar H_0 si el valor-$p \leq \alpha$	Rechazar H_0 si el valor-$p \leq \alpha$
Regla de rechazo: método del valor crítico	Rechazar H_0 si $\chi^2 \leq \chi^2_{(1-\alpha)}$	Rechazar H_0 si $\chi^2 \geq \chi^2_{\alpha}$	Rechazar H_0 si $\chi^2 \leq \chi^2_{(1-\alpha/2)}$ o si $\chi^2 \geq \chi^2_{\alpha/2}$

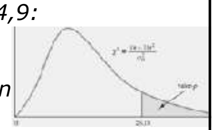
ESTADÍSTICO DE PRUEBA PARA PRUEBAS DE HIPÓTESIS ACERCA DE LA VARIANZA POBLACIONAL

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \quad (11.8)$$

donde χ^2 tiene una distribución ji-cuadrada con $n-1$ grados de libertad.

•*Ejemplo: Una empresa de buses, desea promover una imagen de confiabilidad con conductores sean puntuales en los horarios de llegada a las estaciones. La empresa desea que haya poca varianza en dichos tiempos de arribo a las paradas: 4 minutos o menos. Se tiene así una muestra aleatoria de 24 llegadas, y un S^2 =4,9:*

•*Solución: : Paso 1: Establecer H_0 y H_a :* $H_0: \sigma^2 \leq 4$, $H_a: \sigma^2 > 4$ donde σ_0^2 = 4.



• *Paso 2: Especificar el nivel de significancia: siendo α = 0,05 el están*

• *Paso 3: Recabar datos muestrales y calcular el valor de X^2 :*

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(24-1)(4.9)}{4} = 28.18$$

Área en la cola superior	0.10	0.05	0.025	0.01
Valor χ^2 (23 gl) $\chi^2 = 28.18$	32.007	35.172	38.076	41.638

•*Método del valor-p*

•*Paso 4. Emplear el valor del estadístico de prueba X^2 para calcular el valor-p:*

• *valor-p = (área a la derecha de 28,18: Tabla X^2) > 0,10*

•*Paso 5. Regla de decisión:*

• *Como valor-p> 0,05$\Rightarrow$ H_0 no es rechazada (varianza poblacional no es excesiva).*

Juan Pablo Sucre Reyes

4.2 Pruebas de hipótesis de σ^2

• Con para denotar el valor hipotético de la varianza poblacional y calcular z:
• y para el valor del estadístico X^2

	Prueba de cola inferior	Prueba de cola superior	Prueba de dos colas
Hipótesis	$H_0: \sigma^2 \geq \sigma_0^2$ $H_a: \sigma^2 < \sigma_0^2$	$H_0: \sigma^2 \leq \sigma_0^2$ $H_a: \sigma^2 > \sigma_0^2$	$H_0: \sigma^2 = \sigma_0^2$ $H_a: \sigma^2 \neq \sigma_0^2$
Estadístico de prueba	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$
Regla de rechazo: método del valor-p	Rechazar H_0 si el valor-$p \leq \alpha$	Rechazar H_0 si el valor-$p \leq \alpha$	Rechazar H_0 si el valor-$p \leq \alpha$
Regla de rechazo: método del valor crítico	Rechazar H_0 si $\chi^2 \leq \chi^2_{(1-\alpha)}$	Rechazar H_0 si $\chi^2 \geq \chi^2_{\alpha}$	Rechazar H_0 si $\chi^2 \leq \chi^2_{(1-\alpha/2)}$ o si $\chi^2 \geq \chi^2_{\alpha/2}$

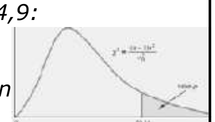
ESTADÍSTICO DE PRUEBA PARA PRUEBAS DE HIPÓTESIS ACERCA DE LA VARIANZA POBLACIONAL

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \quad (11.8)$$

donde χ^2 tiene una distribución ji-cuadrada con $n-1$ grados de libertad.

•*Ejemplo: Una empresa de buses, desea promover una imagen de confiabilidad con conductores sean puntuales en los horarios de llegada a las estaciones. La empresa desea que haya poca varianza en dichos tiempos de arribo a las paradas: 4 minutos o menos. Se tiene así una muestra aleatoria de 24 llegadas, y un S^2 =4,9:*

•*Solución: : Paso 1: Establecer H_0 y H_a :* $H_0: \sigma^2 \leq 4$, $H_a: \sigma^2 > 4$ donde σ_0^2 = 4.



• *Paso 2: Especificar el nivel de significancia: siendo α = 0,05 el están*

• *Paso 3: Recabar datos muestrales y calcular el valor de X^2 :*

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(24-1)(4.9)}{4} = 28.18$$

Área en la cola superior	0.10	0.05	0.025	0.01
Valor χ^2 (23 gl) $\chi^2 = 28.18$	32.007	35.172	38.076	41.638

•*Método del valor crítico*

•*Paso 4. Emplear el nivel de significancia α para calcular el valor X^2 crítico:* $\chi^2_{0.05} = 35.172$

•*Paso 5. Regla de decisión:*

• *Como valor X^2 de prueba < 35,172 $\Rightarrow$ H_0 no es rechazada (varianza poblacional no es excesiva).*

Juan Pablo Sucre Reyes

5. Inferencias acerca de dos varianzas poblacionales

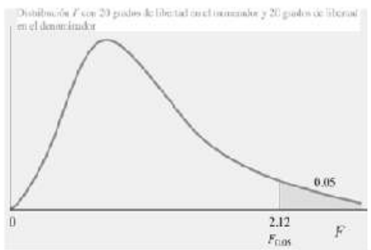
- Para comparar las varianzas de 2 poblaciones se usan datos de 2 muestras aleatorias independientes: n_1 y n_2. Para hacer inferencias acerca de σ_1^2 y σ_2^2 se usan S_1^2 y S_2^2.
- Cuando $\sigma_1^2 = \sigma_2^2$ la distribución de muestreo de S_1^2/S_2^2 es:

DISTRIBUCIÓN DE MUESTREO DE s_1^2/s_2^2 CUANDO $\sigma_1^2 = \sigma_2^2$

Cuando se toman muestras aleatorias simples independientes de tamaños n_1 y n_2 de dos poblaciones normales con varianzas iguales, la distribución de muestreo de

$\dfrac{s_1^2}{s_2^2}$ es una distribución F con $n_1 - 1$ grados de libertad en el numerador, y $n_2 - 1$ grados de libertad en el denominador

- La tabla F da valores de F en áreas de 0,10; 0,05; 0,025 y 0,01 en la cola superior.



Grados de libertad en el denominador	Área en la cola superior	Grados de libertad en el numerador				
		10	15	20	25	30
20	0.10	1.94	1.84	1.79	1.76	1.74
	0.05	2.35	2.20	2.12	2.07	2.04
	0.025	2.77	2.57	2.46	2.40	2.35
	0.01	3.37	3.09	2.94	2.84	2.78

Juan Pablo Sucre Reyes

5.1 Pruebas de hipótesis acerca de σ_1^2 y σ_2^2

- Las 2 formas son:

	Prueba de cola superior	Prueba de dos colas
Hipótesis	$H_0: \sigma_1^2 \le \sigma_2^2$ $H_a: \sigma_1^2 > \sigma_2^2$	$H_0: \sigma_1^2 = \sigma_2^2$ $H_a: \sigma_1^2 \ne \sigma_2^2$ *Nota.* La población 1 tiene la varianza muestral más grande
Estadístico de prueba	$F = \dfrac{s_1^2}{s_2^2}$	$F = \dfrac{s_1^2}{s_2^2}$
Regla de rechazo: método del valor-*p*	Rechazar H_0 si el valor-$p \le \alpha$	Rechazar H_0 si el valor-$p \le \alpha$
Regla de rechazo: método del valor crítico	Rechazar H_0 si $F \ge F_\alpha$	Rechazar H_0 si $F \ge F_{\alpha/2}$

- Los 5 pasos desarrollados son válidos *(α, estadística de prueba F, valor-p, etc.)*
- *Ejemplo: Una escuela debe decidirse entre 2 empresas de autobús: Milbank y Gulf Park. Como medida de la calidad del servicio se usa la varianza de los tiempos en que el autobús llega a recoger/dejar a las personas (baja = servicio más constante y de mayor calidad). Si las varianzas son iguales, se optará por la de mejores finanzas. Sino, se elige a la empresa con menor varianza (mejor servicio). Se parte con α =0,10 de que no existe diferencia. Se conoce que de n_1=48, S_1^2= 48 y de n_2=16, S_2^2=20.*
- *Solución: : Paso 1: Establecer H_0 y H_a :* $H_0: \sigma_1^2 = \sigma_2^2$ $H_a: \sigma_1^2 \ne \sigma_2^2$
- *Paso 2: Especificar el nivel de significancia: siendo α = 0,10.* $F = \dfrac{s_1^2}{s_2^2} = \dfrac{48}{20} = 2.40$
- *Paso 3: Recabar datos muestrales y calcular el valor del estadístico de prueba F:*
- *Método del valor-p*
- *Paso 4. Emplear el valor F para calcular el valor-p:*

Área en la cola superior	0.10	0.05	0.025	0.01
Valor F ($gl_1 = 25$; $gl_2 = 15$)	1.89	2.28	2.69	3.28

$F = 2.40$

- *El valor-p = 2(0,025 a 0,05) = 0,05 a 0,10. (Tablas F)* valor-p = 0.0811 Minitab15
- *Paso 5. Regla de decisión: Como 0,05-0,10 $\le \alpha$=0,10$\Rightarrow H_0$ es rechazada (elegir la 2ª*

Juan Pablo Sucre Reyes

5.1 Pruebas de hipótesis acerca de σ_1^2 y σ_2^2

• Las 2 formas son:

	Prueba de cola superior	Prueba de dos colas
Hipótesis	$H_0: \sigma_1^2 \le \sigma_2^2$ $H_a: \sigma_1^2 > \sigma_2^2$	$H_0: \sigma_1^2 = \sigma_2^2$ $H_a: \sigma_1^2 \ne \sigma_2^2$ *Nota.* La población 1 tiene la varianza muestral más grande
Estadístico de prueba	$F = \frac{s_1^2}{s_2^2}$	$F = \frac{s_1^2}{s_2^2}$
Regla de rechazo: método del valor-*p*	Rechazar H_0 si el valor-$p \le \alpha$	Rechazar H_0 si el valor-$p \le \alpha$
Regla de rechazo: método del valor crítico	Rechazar H_0 si $F \ge F_\alpha$	Rechazar H_0 si $F \ge F_{\alpha/2}$

• Los 5 pasos desarrollados son válidos *(α, estadística de prueba F, valor-p, etc.)*

•*Ejemplo: Una escuela debe decidirse entre 2 empresas de autobús: Milbank y Gulf Park. Como medida de la calidad del servicio se usa la varianza de los tiempos en que el autobús llega a recoger/dejar a las personas (baja = servicio más constante y de mayor calidad). Si las varianzas son iguales, se optará por la de mejores finanzas. Sino, se elige a la empresa con menor varianza (mejor servicio). Se parte con α =0,10 de que no existe diferencia. Se conoce que de n_1=48, S_1^2= 48 y de n_2=16, S_2^2=20.*

•*Solución: : Paso 1: Establecer H_0 y H_a :* $H_0: \sigma_1^2 = \sigma_2^2$
$H_a: \sigma_1^2 \ne \sigma_2^2$

• *Paso 2: Especificar el nivel de significancia: siendo α = 0,10.*

$$F = \frac{s_1^2}{s_2^2} = \frac{48}{20} = 2.40$$

• *Paso 3: Recabar datos muestrales y calcular el valor del estadístico de prueba F:*

•*Método del valor crítico*

•*Paso 4. Emplear el nivel de significancia α para calcular el valor $F_{\alpha/2}$ crítico:*

•*Paso 5. Regla de decisión:*

Área en la cola superior	0.10	0.05	0.025	0.01
Valor F ($gl_1 = 25; gl_2 = 15$)	1.89	2.28	2.69	3.28

• *Como valor F de prueba > 2,28 $\Rightarrow H_0$ es rechazada (elegir la 2ª)*

Juan Pablo Sucre Reyes

5.1 Pruebas de hipótesis acerca de σ_1^2 y σ_2^2

• Las 2 formas son:

	Prueba de cola superior	Prueba de dos colas
Hipótesis	$H_0: \sigma_1^2 \le \sigma_2^2$ $H_a: \sigma_1^2 > \sigma_2^2$	$H_0: \sigma_1^2 = \sigma_2^2$ $H_a: \sigma_1^2 \ne \sigma_2^2$ *Nota.* La población 1 tiene la varianza muestral más grande
Estadístico de prueba	$F = \frac{s_1^2}{s_2^2}$	$F = \frac{s_1^2}{s_2^2}$
Regla de rechazo: método del valor-*p*	Rechazar H_0 si el valor-$p \le \alpha$	Rechazar H_0 si el valor-$p \le \alpha$
Regla de rechazo: método del valor crítico	Rechazar H_0 si $F \ge F_\alpha$	Rechazar H_0 si $F \ge F_{\alpha/2}$

• Los 5 pasos desarrollados son válidos *(α, estadística de prueba F, valor-p, etc.)*

•*Ejemplo: Para estudiar las actitudes frente a los asuntos políticos actuales se elige una muestra de 31 hombres y otra de 41 mujeres. Se piensa que entre las mujeres hay mayor variación en las actitudes respecto de los asuntos políticos. Se parte con α =0,05 y se conoce que S_1^2= 120 (mujeres) y S_2^2= 80 (hombres).*

•*Solución: : Paso 1: Establecer H_0 y H_a :* $H_0: \sigma^2_{\text{mujer}} \le \sigma^2_{\text{hombre}}$
$H_a: \sigma^2_{\text{mujer}} > \sigma^2_{\text{hombre}}$

• *Paso 2: Especificar el nivel de significancia: siendo α = 0,05.*

$$F = \frac{s_1^2}{s_2^2} = \frac{120}{80} = 1.50$$

• *Paso 3: Recabar datos muestrales y calcular el valor del estadístico de prueba F:*

•*Método del valor-p*

•*Paso 4. Emplear el valor F para calcular el valor-p:*

• *El valor-p = > 0,10. (Tablas F)* valor-p = 0.1256. Minitab 15 ►

• *Paso 5. Regla de decisión: Como 0,1256 > α=0,05$\Rightarrow H_0$ no es rechazada (entre las mujeres no hay mayor variación en la actitud frente a los asuntos políticos)*

Juan Pablo Sucre Reyes

5.1 Pruebas de hipótesis acerca de σ_1^2 y σ_2^2

• Las 2 formas son:

	Prueba de cola superior	Prueba de dos colas
Hipótesis	$H_0: \sigma_1^2 \le \sigma_2^2$ $H_a: \sigma_1^2 > \sigma_2^2$	$H_0: \sigma_1^2 = \sigma_2^2$ $H_a: \sigma_1^2 \ne \sigma_2^2$ *Nota*. La población 1 tiene la varianza muestral más grande
Estadístico de prueba	$F = \frac{s_1^2}{s_2^2}$	$F = \frac{s_1^2}{s_2^2}$
Regla de rechazo: método del valor-*p*	Rechazar H_0 si el valor-$p \le \alpha$	Rechazar H_0 si el valor-$p \le \alpha$
Regla de rechazo: método del valor crítico	Rechazar H_0 si $F \ge F_\alpha$	Rechazar H_0 si $F \ge F_{\alpha/2}$

• Los 5 pasos desarrollados son válidos (*α, estadística de prueba F, valor-p, etc.*)

•*Ejemplo: Para estudiar las actitudes frente a los asuntos políticos actuales se elige una muestra de 31 hombres y otra de 41 mujeres. Se piensa que entre las mujeres hay mayor variación en las actitudes respecto de los asuntos políticos. Se parte con α =0,05 y se conoce que S_1^2= 120 (mujeres) y S_2^2= 80 (hombres).*

•*Solución: : Paso 1: Establecer H_0 y H_a :* $H_0: \sigma^2_{\text{mujer}} \le \sigma^2_{\text{hombre}}$ $H_a: \sigma^2_{\text{mujer}} > \sigma^2_{\text{hombre}}$

• *Paso 2: Especificar el nivel de significancia: siendo α = 0,05.*

$$F = \frac{s_1^2}{s_2^2} = \frac{120}{80} = 1.50$$

• *Paso 3: Recabar datos muestrales y calcular el valor del estadístico de prueba F:*

•*Método del valor crítico*

•*Paso 4. Emplear el nivel de significancia α para calcular el valor F_α crítico:*

•*Paso 5. Regla de decisión: Como F = 1,50 < 1,79 (Tabla F, 40:30 g.l.)$\Rightarrow$ H_0 no es rechazada (entre las mujeres no hay mayor variación en la actitud frente a los asuntos políticos)*

Juan Pablo Sucre Reyes

GRACIAS POR SU ATENCIÓN…..

Juan Pablo Sucre Reyes