CAPÍTULO VII

PRUEBA DE HIPOTESIS

El propósito de la prueba o docimasia de hipótesis es determinar si el valor supuesto (hipotético) de un parámetro poblacional, debe aceptarse como verosímil con base en evidencias muestrales. Por ejemplo en la medicina si un medicamento disminuye o no el tiempo de restablecimiento de un paciente; en la agricultura cuando se quiere decidir si un nuevo fertilizante eleva el rendimiento o no; en la técnica si una maquina aumenta o no su tiempo de servicio sin roturas; en los servicios si el tiempo de espera en un centro de prestación de servicios a variado o no; etc.

7.1. HIPÓTESIS

Hipótesis es una proposición enunciada para contestar una pregunta sin saber aún si las observaciones, hechos, datos la comprobarán o rechazarán. Es una proposición comprobable que podría ser la solución de un problema. La función de la hipótesis es orientar nuestra búsqueda de orden en los hechos. *Se emplea una hipótesis para establecer una relación entre dos o más variables.* Debe tener las siguientes cualidades*:*

1. Ser una respuesta probable a un problema.
2. Relacionar dos o más variables.
3. Tener una redacción clara.
4. Debe fundamentarse en la realidad para poder ser observada y comprobada.
5. Disponer de referencias teóricas.
6. Estar al alcance del investigador. Disponer de conocimientos y recursos.

Partes de una hipótesis.- Se tiene generalmente una variable independiente y una dependiente (*Causa y efecto, estímulo y respuesta, antecedente y consecuente).*

Ejemplo 9.4: el suministro de una mala alimentación influye en el bajo peso del niño.
Variable independiente (x): "El suministro de una mala alimentación" (causa)
Variable dependiente (y): "Bajo peso del niño" (efecto)

EJEMPLOS TOMADOS DE ALGUNAS INVESTIGACIONES

HIPOTESIS	VARIABLE INDEPENDIENTE (X)	VARIABLE DEPENDIENTE (Y)
Los lactantes de madres adictas a la heroína tienen menor peso neonatal que los hijos de no adictas.	Adicción o no adicción de la madre	Peso neonatal del lactante
Los adolescentes que sufrieron abuso sexual durante la infancia tienen un riesgo mayor de depresión y suicidio que aquellos que no lo sufrieron.	Historia de abuso sexual	Riesgo de depresión, riesgo de suicidio
Existe una relación positiva entre el lazo madre-hija y un lazo madre-feto de la embarazada. (Zachariach,1994)	Grado de apego entre madre e hija	Grado de apego entre madre y feto

7.2. Hipótesis Estadísticas

Una hipótesis estadística es una aseveración sobre los parámetros de una o más poblaciones. Pueden ser probadas o rechazadas en base a datos cuantitativos: números, porcentajes, promedios, dispersiones u otros parámetros de una o más poblaciones dadas. Distinguiremos dos tipos de hipótesis estadísticas: la hipótesis nula (**H_o**) y la hipótesis alternativa (**H_a**).

a) **Hipótesis nula (H_o).-** Se especifica en forma opuesta a la que se supone cierta. Generalmente se plantea con el propósito de rechazarla [1] y aceptar la de investigación. **Ejemplo:**

[1] Si no hay prueba suficiente para rechazar la **hipótesis nula**, ésta se retiene pero no se "**acepta**"

H_o: "La desnutrición de los estudiantes no es un factor que provoca dificultades de aprendizaje".

Nota: Las hipótesis nulas pueden ser expresadas como una función de los parámetros igualados a cero y de aquí la terminología de hipótesis nula (**$\mu_1 - \mu_2 = 0$, $p - 0.60 = 0$,** $\sigma_1^2 - \sigma_2^2 = 0$**)**.

b) Hipótesis alternativa (H_a).- La alternativa es cualquier hipótesis que excluya la hipótesis nula, a menudo, es la hipótesis contraria o la negación de H_o. **Ejemplo**:
H_a: "La desnutrición de los estudiantes es un factor que provoca dificultades de aprendizaje".

- Si H_o resulta verdadera, el investigador rechaza H_a
- Si H_o resulta falsa, el investigador acepta H_a

Ejemplos varios:

1. La hipótesis experimental es que un medicamento determinado bajará la presión sanguínea. Los pacientes son asignados al azar en dos grupos: el grupo 1 recibe el medicamento y el 2 actúa como control sin medicamento.
   H_o: $\mu_1 = \mu_2$; La presión sanguínea media del grupo 1 (**μ_1**), será igual que la del grupo 2 (**μ_2**).
   H_a: $\mu_1 < \mu_2$; La presión sanguínea media del grupo 1 (**μ_1**), será menor del grupo 2 (**μ_2**).

2. En una investigación se afirma que un paquete turístico nuevo lo acepta el 60% de turistas.
   H_o: $p = 0.60$; El paquete turístico nuevo lo aceptan el 60% de turistas.
   H_a: $p \neq 0.60$; El paquete turístico nuevo no lo aceptan el 60% de turistas.

3. Un docente investigador realiza el siguiente estudio: compara dos métodos de enseñanza de Estadística a estudiantes de Turismo con el criterio de efectividad de los puntajes y establece las hipótesis:
   H_o: Los dos métodos de enseñanza de la anatomía no son diferentes; o
   H_o: $\mu_1 = \mu_2$; El puntaje del promedio del primer método es igual al puntaje del promedio del segundo método.
   H_a: $\mu_1 \neq \mu_2$; El primer método de enseñanza de la estadística es diferente al segundo.
   H_a: $\mu_1 > \mu_2$; El primer método es mejor o más efectivo que el segundo.
   H_a: $\mu_1 < \mu_2$; El primer método es menos efectivo que el segundo.

7.3. Tipo de error

Al aceptar o rechazar una hipótesis se puede cometer dos tipos de error:

a) **Error tipo I (α):** Rechazar la hipótesis cuando ha debido aceptarse.

b) **Error tipo II (β):** Aceptar la hipótesis cuando ha debido rechazarse.

Tabla No. 7.1. Tipos de error

		Verdadera	Falsa
DECISIONES	**Aceptar**	**Decisión Correcta** $(1-\alpha)$	**Error tipo II** (β)
	Rechazar	**Error tipo I** (α)	**Decisión Correcta** $(1-\beta)$

- Si se acepta una hipótesis verdadera la decisión es correcta
- Si se acepta una hipótesis falsa, cometemos un error de tipo II
- Si rechazamos una hipótesis verdadera, cometemos error de tipo I
- Si rechazamos una hipótesis falsa, la decisión es correcta.

Nota: La probabilidad de cometer el error tipo I ocurre mientras más alto sea su valor, es entonces más probable que la hipótesis nula sea rechazada equivocadamente. La probabilidad de cometer un error del tipo II aumenta cuando el tamaño de la muestra es pequeño, a medida que el tamaño de la muestra se incrementa esta probabilidad disminuirá.

7.4. Prueba unilateral y bilateral:

Prueba de hipótesis unilateral.- la zona de rechazo o zona crítica está comprendida en uno de los extremos de la distribución y son:
Unilateral derecha, la hipótesis alternativa utiliza términos como: mayor, superior, mejor.
Unilateral izquierda, la hipótesis alternativa utiliza términos como: menor, bajo, inferior.

Prueba de hipótesis bilateral.- Cuando las zonas de rechazo están en los dos extremos de la distribución. La hipótesis alternativa es diferente; por lo tanto se omiten los términos: superior, mayor, mejor, inferior, bajo, menor, etc. La hipótesis alternativa puede definirse por los siguientes tres casos:

H_a: $\mu_1 \neq \mu_2$; Define una hipótesis bilateral a dos colas

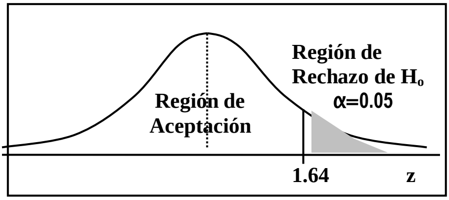
H_a: $\mu_1 > \mu_2$; Define una hipótesis unilateral o a una cola a la derecha.

H_a: $\mu_1 < \mu_2$; Define una hipótesis unilateral o a una cola a la izquierda

7.5. Nivel de significación o puntos críticos

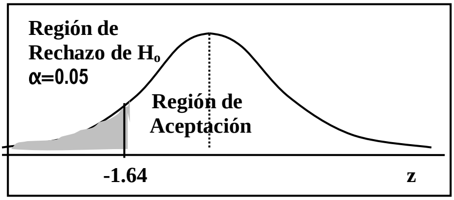
Es el valor de la probabilidad que el investigador escoge por la estimación que hace de la importancia o del posible significado práctico de su investigación. El nivel de significancia se simboliza por alfa (α) siendo generalmente del 1%, 5% ó 10%, pero se puede utilizar cualquier nivel, dependiendo del tipo de investigación. El trabajo es *altamente significativo*, cuando se trabaja con el 1%; es significativo si es el 5% y poco significativo si se trabaja con el 10%.

a) Dócima unilateral hacia la derecha (α=0.05)

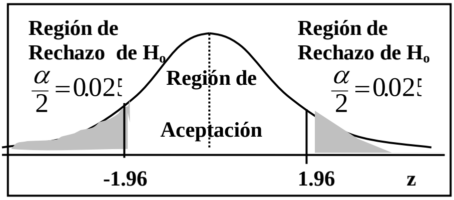


$A = 0.5000 - 0.0500 = 0.4500 \Rightarrow z = 1.64$

b) Dócima unilateral hacia la izquierda (α=0.05)



c) Dócima bilateral (α=0.05)



$A = 0.5000 - 0.0250 = 0.4750 \Rightarrow z = 1.96$

7.6. Etapas en la prueba de hipótesis estadísticas

Se debe tener en cuenta las siguientes alternativas:

1. Formular la hipótesis nula (H_o) y alternativa (H_a)

a) En las medias se escribirá de las siguientes formas:

Ho: $\mu = 40$	Ho: $\mu = 40$	Ho: $\mu = 40$
Ha: $\mu \neq 40$	Ha: $\mu > 40$	Ha: $\mu < 40$
(Dócima bilateral)	*(Unilateral derecha)*	*(Unilateral izquierda)*

b) En una distribución de dos medias muestrales:

Ho: $\mu_1 = \mu_2$	Ho: $\mu_1 = \mu_2$	Ho: $\mu_1 = \mu_2$
Ha: $\mu_1 \neq \mu_2$	Ha: $\mu_1 > \mu_2$	Ha: $\mu_1 < \mu_2$

c) En las proporciones se escribirá para cada caso así:

Ho: $p = 0.60$	Ho: $p = 0.60$	Ho: $p = 0.60$
Ha: $p \neq 0.60$	Ha: > 0.60	Ha: $p < 0.60$
(Dócima bilateral)	*(Unilateral derecha)*	*(Unilateral izquierda)*

d) En la diferencia de proporciones:

Ho: $p_1 = p_2$	Ho: $p_1 = p_2$	Ho: $p_1 = p_2$
Ha: $p_1 \neq p_2$	Ha: $p_1 > p_2$	Ha: $p_1 < p_2$
(Dócima bilateral)	*(Unilateral derecha)*	*(Unilateral izquierda)*

2. Seleccionar el nivel de significación (elegir el riesgo: α= %)

a) α= 1% = 0.01 (Investigación altamente significativa)
b) α= 5% = 0.05 (Investigación significativa)
c) α= 10% = 0.10 (Investigación poco significativa)

3. Conocer o estimar la varianza

a) La muestra es aleatoria
b) La población es normal
c) La varianza poblacional es conocida (en la mayoría de casos se debe estimar)

4. Determine la técnica y la prueba estadística

a) Distribución normal: $z = \dfrac{x_i - \mu}{\sigma}$

b) Distribución de medias muestrales: $z=\dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$ ó $z=\dfrac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$, (n>30)

c) Distribución de proporciones muestrales: $z=\dfrac{p-P}{\sqrt{\frac{p.q}{n}}}$, (n>30)

d) Distribución de diferencias entre dos medias muestrales:

$$z=\frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{\sqrt{\frac{\sigma_x^2}{n_1}+\frac{\sigma_y^2}{n_2}}} \quad \text{ó} \quad z=\frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{\sqrt{\frac{s_x^2}{n_1}+\frac{s_y^2}{n_2}}}$$, (n_1 y n_2 >30)

e) Distribuciónes de diferencias entre dos proporciones muestrales:

$$z=\frac{(p_1-p_2)-(\mu_{p_1}-\mu_{p_2})}{\sqrt{\frac{P_1.Q_1}{n_1}+\frac{P_2.Q_2}{n_2}}}; \quad z=\frac{(p_1-p_2)-(P_1-P_2)}{\sqrt{\frac{p_1.q_1}{n_1}+\frac{p_2.q_2}{n_2}}}$$

5. Determine los valores críticos y sus regiones de rechazo:

a) Para un nivel de significación del 5% de dócima bilateral: z_s = 1.96, z_i = -1.96

b) Para un nivel de significación del 5% de dócima unilateral:
- Dócima unilateral izquierda: z_i = -1.64,
- Dócima unilateral izquierda: z_s = 1.64

6. Calcular los datos muestrales, utilizando las fórmulas correspondientes:

Supongamos que se producen 100 nacimientos, de los cuales 60 son mujeres, se tendrá: $\mu=n.p$=100.(0.50) = 50

$$\sigma=\sqrt{n.p.q}=\sqrt{100(0.50).(0.50)}=5; \quad z=\frac{x_i-\mu}{\sigma}=\frac{60-50}{5}=2$$

7. Tomar la decisión estadística:

Se acepta o se rechaza la hipótesis nula, al nivel de significación dado. En el ejemplo anterior, z = 2, se sitúa en la zona de aceptación, por lo tanto se acepta la hipótesis nula (H_0 : μ = 50), es decir, la diferencia no es significativa.

7.7. PRUEBAS CUANDO SE CONOCE σ O LA MUESTRA ES GRANDE

En el paso 3, se decía que la varianza poblacional (σ^2) es conocida. Si no se conoce debe ser sustituida por la varianza muestral (si n > 30), considerada muestra grande.

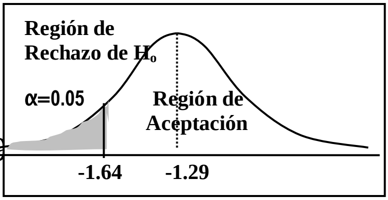
a) Distribución de medias

En general, después de calcular el tamaño de la muestra, y su media, vendrá la identificación de la desviación típica poblacional.

1. En una empresa de lácteos, un inspector de calidad investiga las acusaciones, por el deficiente llenado que debe ser en promedio, de 72.5 cm³. Para ello toma una muestra de 60 botellas, encontrando que el contenido medio es de 71.9 cm³ de líquido. Se sabe que la máquina debe producir un llenado con una desviación típica de 3.6 cm³. ¿Puede el inspector llegar a la conclusión, a un nivel de significancia del 5%, que se están llenando las botellas, por debajo de su especificación de contenido?

   Solución: $\mu = 72.5$ $\sigma = 3.6$ $n = 60$ $\bar{X} = 71.9$

   a) $H_0 : \mu = 72.5$
      $H_a : \mu < 72.5$ (unilateral izquierda)
   b) $\alpha = 0.05$
   c) $\sigma = 3.6$
   d) $$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{71.9 - 72.5}{\frac{3.6}{\sqrt{60}}} = -1.29$$

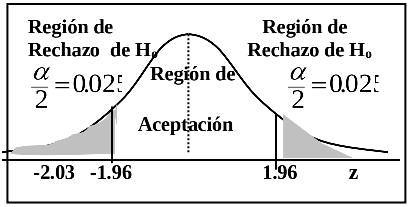


   e) **Decisión**: Observamos que z = -1.29, se sitúa en la región de aceptación, es válida la hipótesis nula. El inspector no puede concluir que se está llenando el producto por debajo de su especificación, al nivel del 5%.

2. Una máquina está programada para empacar la cantidad media de 56 gramos de un producto, se toma la muestra aleatoria de 36 cajas; resulta una media de 54.2 gramos y desviación típica de 5.3 gramos. Al nivel del 5%, se podrá afirmar que no se está cumpliendo con el empaque?

   Solución: $\mu = 56$ $s = 5.3$ $n = 36$ $\bar{X} = 54.2$

a) $H_0 : \mu = 56$
   $H_a : \mu \neq 16$ (Prueba bilateral)
b) $\alpha = 0.05$
c) $s = 5.3$
d) $$z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{54.2 - 56}{\frac{5.3}{\sqrt{36}}} = -2.03$$

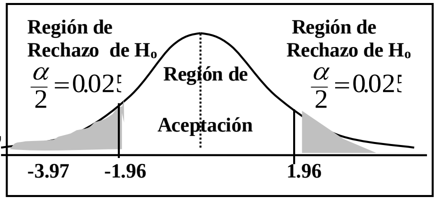


e) **Decisión**: Observamos que z = -2.03, se sitúa en la región de rechazo de la hipótesis nula y se aceptará la hipótesis alternativa. Se concluye que no se está cumpliendo con lo programado, al nivel del 5%.

3. Los investigadores suponen que el gasto diario en alimentación de los turistas tiene una media igual a 9.9 dólares, con desviación estándar de 0.66. Ellos estudiaron 43 turistas y encontraron una media de gasto de 9.5 dólares. Quieren saber si la media de la muestra en estudio es diferente del valor medio de la población de 9.9 dólares.

Solución: $\mu = 9.9$ $\sigma = 0.66$ $n = 43$ $\bar{X} = 9.5$

a) $H_0 : \mu = 9.9$ \$
   $H_a : \mu \neq 9.9$ \$ (Prueba bilateral)
b) $\alpha = 0.05$
c) $\sigma = 0.66$
d) $$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{9.5 - 9.9}{\frac{0.66}{\sqrt{43}}} = -3.97$$



e) **Decisión**: como z = -3.97, se sitúa en la región de rechazo, se rechaza la hipótesis nula y se aceptará la hipótesis alternativa. Se concluye que los turistas tienen gastos diferentes de 9.9 dólares, al nivel del 5%.

EJERCICIOS PROPUESTOS

1. De la "población normal" constituida por 500 turistas que visitan Riobamba, se extrajo una muestra de 16 turistas cuya edad es: 62, 43, 60, 49, 72, 56, 45, 46, 37, 56, 41, 43, 36, 45, 56, 49. Se sabe que la desviación estándar poblacional es de 10 años, pero se desconoce la media poblacional (μ = 50 años verdadera). Cometiendo un riesgo alfa (nivel de significación) del 5%, docimar o probar la hipótesis de que la media poblacional sea igual:
   a) 40 ($H_0 : \mu = 40$; $H_a : \mu \neq 40$)

b) 49 ($H_0 : \mu = 49$; $H_a : \mu \neq 49$)
c) 50 ($H_0 : \mu = 50$; $H_a : \mu \neq 50$)
d) 51 ($H_0 : \mu = 51$; $H_a : \mu \neq 51$)
e) 60 ($H_0 : \mu = 60$; $H_a : \mu \neq 60$)

Para cada uno de los cinco casos anteriores establecer si la decisión es correcta y en caso contrario indicar el tipo de error cometido.

Respuestas:

a) z = 3.9 cae en la región de rechazo ($z_s > 1.96$) por tal motivo se rechaza la hipótesis nula. No hay error, estamos rechazando algo falso.
b) z = 0.30 cae en la zona de aceptación, es decir, aceptamos que $\mu = 49$, como sabemos que la media poblacional verdadera es 50 años. Estamos aceptando algo falso (error de tipo II)
c) z = -0.10 se ubica en la zona de aceptación. No hay error, ya que aceptamos algo verdadero ($\mu = 50$)
d) z = -0.50 se ubica en la región de aceptación, el error es de tipo II, ya que aceptamos algo falso.
e) z = -4.10 cae en la región de rechazo ($z_i < -1.96$) por tal motivo se rechaza la hipótesis nula. No hay error, estamos rechazando algo falso.

2. En una investigación de varios años en un examen de ingreso a la Universidad en matemática arroja una calificación promedio de 64 puntos, con una desviación estándar de 8. Los 64 estudiantes de cierto colegio han obtenido una calificación promedio de 68. ¿Puede tenerse la certeza de que los estudiantes de este colegio sean superiores en matemática?.
Resp. z = 4 se ubica en la zona de rechazo (4 > 1.64) por lo tanto se puede tener la certeza, con un nivel de significancia del 5% que los alumnos de este colegio son superiores en matemática.

3. Cuatrocientos estudiantes, elegidos aleatoriamente, se someten a un "test" de rendimiento, obteniéndose los siguientes resultados: $\overline{X} = 76$ y s = 16, con base a esta información docimar la hipótesis $\mu = 74$ frente a la alternativa $\mu \neq 74$, al nivel de significación del 1%.
Resp. Está en la zona de aceptación; aceptamos que $\mu = 74$, al nivel del 1%.

4. En una investigación revela que los 100 equipos de aire acondicionado, que constituyen una muestra aleatoria, se utilizaron un promedio de 12500 horas, con una desviación estándar de 2400. Con base a esta

información, docimar la hipótesis donde, en promedio, los equipos se utilizaron 12 000 horas, frente a la alternativa que el promedio sea superior. Utilizar el nivel de significancia del 5%.
Resp. Rechazamos la hipótesis de que μ = 12000, luego aceptamos que los equipos se utilizaron en un promedio superior (μ > 12000), al nivel del 5%.

5. Los salarios anuales de un sector del turismo están distribuidos normalmente con una media de \$ 13 200 y una desviación estándar de \$ 2 500. Si una empresa del sector, cuenta con 40 trabajadores, paga, en promedio \$ 12200, ¿puede decirse que esta compañía paga salarios inferiores al nivel de significancia del 1%.
Resp. Se ubica en la región de rechazo, por lo tanto, se puede acusar a la compañía de pagar salarios inferiores (μ < 13 200), al nivel del 1%.

b) Distribución de proporciones

El proceso es similar a lo explicado para las medias, considerando que por lo general la desviación típica y por ende el error estándar de la proporción se calcula con datos obtenidos en la muestra donde n > 30 elementos.

1. Se ha podido establecer en las estadísticas que el 40% de los turistas toman un determinado paquetes. Una muestra aleatoria de 450 turistas reveló que 200 de ellos solían tomar dicho paquete. ¿Cuál podría ser la conclusión al nivel del 1%, acerca de lo que muestran las estadísticas?

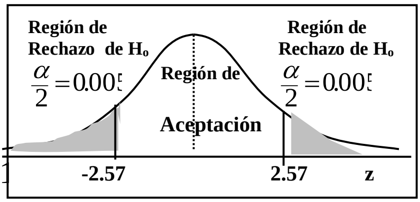
Solución: p = 200/450 = 0.44 = 44%, n = 450, q = 250/450 = 0.56

a) $H_0 : P = 0.40$
   $H_a : P \neq 0.40$

b) $\alpha = 0.01$

c) $s_p = \sqrt{p.q}$

d) $$z = \frac{p-P}{\sqrt{\frac{p.q}{n}}} = \frac{0.44-0.40}{\sqrt{\frac{0.44(0.56)}{450}}} = 1.71$$



e) **Decisión:** z = 1.71 se ubica en la zona de aceptación, se acepta el 40% que arrojan las estadísticas, al nivel del 5%.

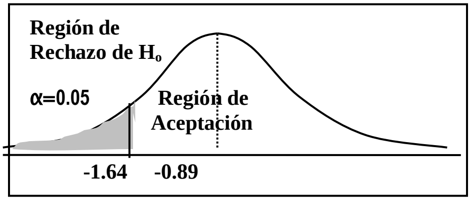
2. El gerente de una empresa afirma que el porcentaje de atrasos en las horas de llegada al trabajo es del 20% de los empleados. Solicita al jefe de personal la revisión de 40 tarjetas marcadas con las horas de llegada, en el mes, y encuentra que 6 han llegado tarde. Al nivel del 5%, ¿hay razón para concluir que el gerente está exagerando?.

   Solución: p = 6/40 = 0.15 = 15%, n = 40, q = 85%

   a) $H_0 : P = 0.15$
      $H_a : P < 0.15$

   b) $\alpha = 0.05$

   c) $s_p = \sqrt{pq}$



   d) $$z = \frac{p - P}{\sqrt{\frac{p.q}{n}}} = \frac{0.15 - 0.20}{\sqrt{\frac{0.15(0.85)}{40}}} = -0.89$$

   e) **Decisión:** z = -0.89 se ubica en la zona de aceptación; por tanto al nivel del 5% el gerente no está exagerando.

EJERCICIOS PROPUESTOS

1. Una empresa turística al seleccionar el personal lo somete a un curso de entrenamiento. Por experiencia, el 76% de los aspirantes aprueban el curso. Se efectúan ciertos cambios en el programa, para el cual se inscriben 40 y 24 aprueban. ¿Podría afirmarse que los cambios introducidos reducen la selección? (1%).

   Resp. Como z = -2.07 cae en la región de aceptación, no reduce la selección los cambios introducidos al nivel del 1%.

2. El Ministro de Trabajo del país asegura que el 80% de los empleados tienen un ingreso mensual superior a $ 370; usted quiere refutar al Ministro con un nivel de confianza del 99% y para hacerlo toma una muestra de 300 empleados, encontrándose 231 con ingresos mayores a $ 370. ¿tiene razón el señor Ministro?.

   Resp. Como z = -1.25, se acepta la hipótesis (P = 0.80) de que los empleados tienen ingresos superiores a $ 370, al nivel del 1%, en una prueba bilateral.

3. El porcentaje de exámenes defectuosos de cierto proceso supervisado es 0.14. Un proveedor ofrece un producto asegurando que reduce la fracción de defectuosos. Con las muestras que el proveedor suministra se hace un ensayo en la producción con un resultado de 48 defectuosos de un total de 360 exámenes. Contrastar si el proveedor tiene o no razón en la calidad del nuevo producto, con el 5% de significación.
   Resp. z = - 0.56, se acepta P = 0.14, el proveedor no tiene razón, el nuevo producto no reduce el porcentaje de defectuosos, al nivel del 5%.

4. Por evidencia experimental se sabe que cierta droga es eficaz en un 80% de los casos, cuando está correctamente administrada. Se aplica dicha droga a 400 personas y se obtiene únicamente 300 resultados positivos. ¿Puede considerarse este resultado como evidencia de que la droga no estuvo bien administrada? (1%)
   Resp. z = -2.27, este resultado puede ser considerado como evidencia que la droga estuvo bien administrada, al nivel del 1%. (Prueba unilateral izquierda).

c) Distribución de diferencias entre dos medias

Se utiliza esta prueba cuando se tiene dos poblaciones independientes y se extrae dos muestras para establecer sus diferencias. Si una media es mayor o menor que la otra, son significativas. Se utiliza cuando se quiere probar:

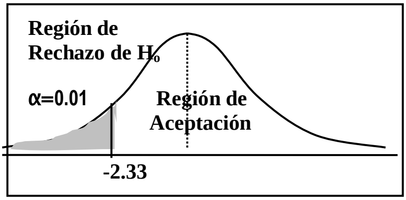
- Si la incidencia de cáncer es mayor en la población masculina que en la femenina.
- Si hay alguna diferencia en los hábitos de fumar de los hombres y las mujeres.
- Si la duración de un producto de la marca A es diferente al de la marca B.

1. El Ministerio de Turismo desarrolló un curso de entrenamiento a guías turísticos, formando dos grupos y aplicando métodos distintos de entrenamiento. Los dos grupos se consideran homogéneos en capacidad. El primer grupo lo componen 36 profesionales que obtuvieron un puntaje de 6 puntos (en una escala de 0 a 10 puntos) y una desviación típica de 4 puntos y el segundo grupo de 40 profesionales cuyo promedio fue de 8.2 y desviación típica de 4.3 puntos. ¿Se puede concluir que el método aplicado al segundo grupo fue superior al primero? (1%).

Solución:

a) $H_0 : \mu_x = \mu_y$
 $H_a : \mu_x < \mu_y$

b) $\alpha = 0.01$



c) $$s_{\bar{x}-\bar{y}} = \sqrt{\frac{s_x^2}{n_1} + \frac{s_y^2}{n_2}}$$

d) $$z = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_x^2}{n_1} + \frac{s_y^2}{n_2}}} = \frac{6-8.2}{\sqrt{\frac{16}{36} + \frac{18.49}{40}}} = -2.31$$

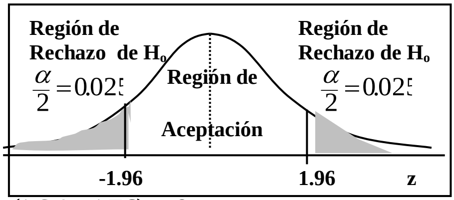
e) **Decisión:** Como z = -2.31, se ubica en la zona de aceptación. No existe una diferencia significativa que permita concluir que el método aplicado en B fue superior al de A, al nivel del 1%.

2. Una empresa tiene dos fábricas y desea establecer el promedio de antigüedad que tienen sus empleados. Se toma de la primera fábrica una muestra de 60 empleados, la cual reflejó un promedio de trabajo de 16.4 años con desviación estándar de 5 años; mientras que en la segunda fábrica una muestra de 40, fue de 15.8 años, con desviación estándar de 4.2 años. ¿Al nivel del 5% se podrá afirmar que hay una diferencia significativa en cuanto a la antigüedad de los empleados?

Solución:

a) $H_0 : \mu_x = \mu_y$
 $H_a : \mu_x \neq \mu_y$

b) $\alpha = 0.05$



c) $$s_{\bar{x}-\bar{y}} = \sqrt{\frac{s_x^2}{n_1} + \frac{s_y^2}{n_2}};$$

d) $$z = \frac{(\bar{x}-\bar{y}) - (\mu_x - \mu_y)}{\sqrt{\frac{s_x^2}{n_1} + \frac{s_y^2}{n_2}}} = \frac{(16.4-15.8)-0}{\sqrt{\frac{5^2}{60} + \frac{4.2^2}{40}}} = 0.65$$

e) **Decisión:** Como z = 0.65, no hay diferencia significativa, al nivel del 5%.

EJERCICIOS PROPUESTOS

1. La distancia promedio alcanzada en los últimos tiempos por un atleta A, lanzador de martillo, es igual a la alcanzada por el lanzador B. Para seleccionar al mejor de los dos se consideran 9 lanzamientos tomados al azar de los realizados en los últimos tiempos por A y 7 de los realizados por B en los que A lanzó el implemento a una distancia promedio de 72.36 m y B una distancia promedio de 74.01 m. ¿Será justo considerar mejor lanzador B con un nivel de significancia del 5%?. La varianza de A es 50 y de B es 74.
   Resp. Como $z = 0.41 < 1.64$ no se rechaza $H_0 : \mu_x = \mu_y$ con un nivel de significación del 5%. No es justo considerar al lanzador B mejor que A en los últimos tiempos.

2. Se comparan dos procesos de fabricación de un producto. Una muestra de 100 cajas del primer proceso tiene una media de 107 y una desviación estándar de 17. En el segundo proceso, una muestra de tamaño 90 tiene una media de 103 y desviación estándar de 16. ¿Existe alguna diferencia significativa entre las medias de ambos procesos?
   Resp. Al nivel del 5%, no existe diferencia significativa entre las medias de los dos procesos.

3. Se requiere comparar el nivel salarial de dos hoteles. El primero reporta en una muestra aleatoria de 46 empleados, un salario promedio de $ 418, con una desviación estándar de 32. Se elige una muestra aleatoria de 60 empleados del segundo hotel, obteniéndose un salario promedio de $ 442 y desviación estándar de 41. ¿Con los anteriores resultados podemos concluir que los salarios del primer hotel son inferiores? (Nivel del 1%).
   Resp. Como $z = -3.38$, podemos concluir que los salarios del segundo hotel son superiores, al nivel del 1%.

4. Un profesor afirma que con un nuevo método de enseñanza los estudiantes logran resolver los ejercicios más rápido que con el método tradicional. Para verificar su afirmación toma dos grupos (muestras aleatorias), uno (experimental) de 15 estudiantes a los que se les aplica un nuevo método y otro (de control) de 17 estudiantes a los que se aplica el método tradicional y se les pide a los estudiantes de ambos grupos que resuelvan una serie de ejercicios y se mide el tiempo que necesita cada estudiante para resolverlos. El grupo de control (grupo 2) demoró un tiempo promedio de 25.7 minutos con una varianza (σ^2 =

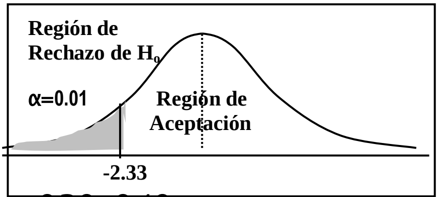
9.1), mientras que el grupo experimental (grupo 1) demoró un tiempo promedio de 20.3 minutos en resolver los ejercicios planteados con una varianza ($\sigma^2 = 4$). ¿Respaldan los resultados obtenidos la afirmación del profesor?. (Nivel del 1%).
Resp. Como z = - 6.02 < -2.33, se rechaza H_0 con el nivel de significación del 1%. Los resultados obtenidos respaldan la afirmación del profesor.

d) Distribución de diferencias entre dos proporciones

Las proporciones son aplicadas como medidas a características cualitativas (atributos). La prueba de hipótesis, que implica el uso de la distribución normal, permite establecer si hay o no diferencia entre dos proporciones obtenidas en dos poblaciones independientes, o si un grupo tuvo una proporción mayor que el otro.

1. Utilizando una muestra aleatoria de 350 profesionales que trabajan y otra muestra independiente de 325 que son profesionales y no trabajan. En el primer caso 105 manifestaron que compran el diario. En el segundo, la respuesta fue de 130 que no lo compraban. ¿Al nivel del 1% se podrá afirmar que los profesionales que trabajan leen menos que los profesionales que no trabajan?.

Solución: $P_x = 105/350 = 0.30$, $P_y = 130/325 = 0.40$



a) $H_0 : P_x = P_y$
$H_a : P_x < P_y$

b) $\alpha = 0.01$

c) $$s_{\bar{P}_x-\bar{P}_y} = \sqrt{\frac{p_x.q_x}{n_1} + \frac{p_y.q_y}{n_2}}$$

d) $$z = \frac{p_x - p_y}{\sqrt{\frac{p_x.q_x}{n_1} + \frac{p_y.q_y}{n_2}}} = \frac{0.30 - 0.40}{\sqrt{\frac{0.3(0.7)}{350} + \frac{0.4(0.6)}{325}}} = -2.73$$

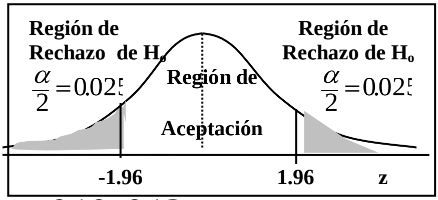
e) **Decisión:** Como z = -2.73, se ubica en la zona de rechazo. Se puede afirmar que los profesionales que trabajan leen menos que las que no trabajan, al nivel del 1%.

2. Un investigador en el campo del turismo realiza dos muestras de tamaño de 120 trabajadores, una en cada hotel, con el fin de determinar el porcentaje de faltas en el trimestre. En el primer hotel se observó que se presentaron 12 casos en el trimestre, mientras que en el segundo 16. ¿Al nivel del 5% se podrá afirmar que los contagios son iguales en los dos hoteles?.

Solución: $P_x = 12/120 = 0.10$, $P_y = 16/120 = 0.13$

$H_0 : P_x = P_y$

$H_a : P_x \neq P_y$



a) $\alpha = 0.05$

b) $$s_{\bar{P}_x - \bar{P}_y} = \sqrt{\frac{p_x . q_x}{n_1} + \frac{p_y . q_y}{n_2}}$$

c) $$z = \frac{p_x - p_y}{\sqrt{\frac{p_x . q_x}{n_1} + \frac{p_y . q_y}{n_2}}} = \frac{0.10 - 0.13}{\sqrt{\frac{0.1(0.9)}{120} + \frac{0.13(0.87)}{120}}} = -0.73$$

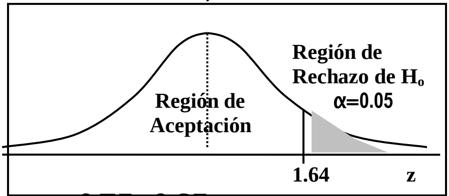
d) **Decisión:** Como z = -0.73, se ubica en la zona de aceptación. Se puede afirmar que el nivel de faltas es igual en los dos hoteles, al nivel del 5%.

3. Dos grupos A y B de 100 personas cada uno tiene determinada enfermedad. Un suero es dado al grupo A, pero no al B. Por otra parte, los grupos son tratados por igual. Si encontramos que en el grupo A, 75 personas se recobran de la enfermedad y en el B, 65, pruebe la hipótesis de que el suero cura la enfermedad? (5%)

Solución: $n_x = 100$, $n_y = 100$, $P_x = 75/100 = 0.75$, $P_y = 65/100 = 0.65$

a) $H_0 : P_x = P_y$

$H_a : P_x > P_y$



b) $\alpha = 0.05$

c) $$s_{\bar{P}_x - \bar{P}_y} = \sqrt{\frac{p_x . q_x}{n_1} + \frac{p_y . q_y}{n_2}}$$

d) $$z = \frac{p_x - p_y}{\sqrt{\frac{p_x . q_x}{n_1} + \frac{p_y . q_y}{n_2}}} = \frac{0.75 - 0.65}{\sqrt{\frac{0.75(0.25)}{100} + \frac{0.65(0.35)}{100}}} = 1.56$$

e) **Decisión:** Como z = 1.56, se ubica en la zona de aceptación ($H_0 : P_x = P_y$). No podemos aceptar que el suero cure la enfermedad, al nivel del 5%.

EJERCICIOS PROPUESTOS

1. Un estudio del consumo de café en el trabajo, por sexo, mostró en una muestra aleatoria de 200 mujeres que 128 lo toman durante el trabajo, mientras que una muestra de 125 hombres reveló que 106 lo toman. ¿Hay alguna diferencia entre la proporción de los dos grupos, en cuánto al hábito de tomar café en el trabajo?.
   Resp. (P_x = 0.64; P_y = 0.71); (z = -1.39 > - 1.96). La diferencia no es significativa. Bilateral.

2. El departamento de matemática desea determinar si la proporción de alumnos que pierden la materia se debe a su estado civil. Se tomó una muestra de 60 alumnos casados de los cuales 12 la habían perdido, mientras que en otra muestra de alumnos solteros de igual número, tan solo 10 la perdieron. ¿Al nivel del 5%, el estado civil influye en el rendimiento?.
   Resp. (P_x = 0.20; P_y = 0.17); (z = 0.42 < 1.96). El estado civil no influye en el rendimiento. Bilateral.

3. El jefe de personal de un hotel quiere comprobar que la ausencia al trabajo es mayor en el turno nocturno que en el diurno. Selecciona de cada jornada 40 y 50 trabajadores y encuentra que en el mes, para el turno diurno 7 trabajadores se excusaron de ir a trabajar, mientras que en la noche faltaron 12. ¿los anteriores resultados dan la razón al jefe de personal?. (nivel del 10%)
   Resp. (z = -0.7 > -1.28). No le dan la razón al jefe de personal. Unilateral izquierda.

4. Dos grupos A y B de 500 personas, cada una, presenta la misma sintomatología. Al grupo A se le da una droga que contiene un nuevo compuesto y la droga del grupo B carece de él. Por otra parte, los grupos son tratados idénticamente. Si encontramos que en el grupo A, 375 personas se recobran de la enfermedad y en el B, 325, pruebe la hipótesis de que la droga aplicada al grupo A es mejor que la del B. (nivel del 5%)
   Resp. (z = 3.47 > 1.64). Al nivel del 5%, la droga aplicada al grupo A es mejor. Unilateral.

7.8 Muestras pequeñas

En pruebas de hipótesis no importa que la muestra sea grande o pequeña cuando se conoce la desviación típica poblacional. Se dice muestra grande (n > 30) y muestra pequeña (n < 30)

a) Distribución "t" de Student

Esta distribución es simétrica en forma de campana, pero la curva es achatada y se extiende a los extremos a las áreas críticas o de rechazo. Existe una familia de distribuciones "t", a medida que se incrementa el tamaño de la muestra se acerca a la normal. La función es:

$$Y = C\left(1 + \frac{t^2}{\upsilon}\right)^{-\frac{\upsilon+1}{2}}$$

Donde: υ = n – 1 (grados de libertad)
υ = n_1 + n_2 – 2 (grados de libertad para la diferencia de medias muestrales)

b) Distribución de medias muestrales

Cuando la muestra es pequeña ($n \leq 30$) la desviación típica se tendrá que corregir, para luego ser aplicada en la variante estadística "t".

$$t = \frac{x - \mu}{\frac{\hat{s}}{\sqrt{n-1}}}; \quad ó \quad t = \frac{x - \mu}{\frac{s}{\sqrt{n}}}$$

Donde: s = desviación típica corregida; $\hat{s}$ = desviación típica sin corregir

$$s = \hat{s}\sqrt{\frac{n}{n-1}} \quad ó \quad s = \sqrt{\frac{\sum (x_i - x)^2}{n-1}} = \sqrt{\frac{\sum x_i^2 - n x^2}{n-1}}$$

Ejercicio: En una empresa de lácteos, un inspector de calidad investiga las acusaciones de no llenar bien los envases, afirmando que contienen 35 onzas de líquido, se muestrean 28 cartones de leche, encontrando un contenido medio de 33,2 onzas, con una desviación estándar de 2,2 onzas.

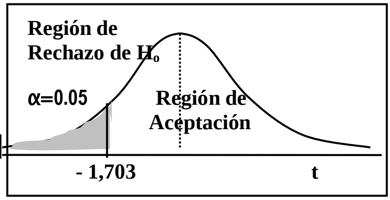
¿Debe llegar el inspector a la conclusión, al nivel del 5%, que se está exagerando su contenido?

Solución: n = 28; $\bar{X}$ =33,2; $\hat{S}$ = 2,2; μ= 35; υ= 28 – 1 = 27

a) H₀: μ= 35

b) Hₐ: μ< 35

c) α = 0.05



d) $$s=\hat{s}\sqrt{\frac{n}{n-1}}=2{,}2\sqrt{\frac{28}{27}}=2{,}24$$

e) $$t=\frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}=\frac{33{,}2-35}{\frac{2{,}24}{\sqrt{28}}}=-4{,}25$$

f) Al nivel del 5%, el inspector si puede llegar a la conclusión de que se está exagerando el llenado.

c) Distribución de una proporción muestral

En una investigación en la Universidad se determinó que el 42% de alumnos tiene dificultad en aprender matemática. El vicerrectorado desea disminuir este porcentaje; para ello debe comprobar ese porcentaje, y decide realizar una investigación por muestreo a 25 estudiantes encontrando que 13 de ellos tienen dificultad de aprendizaje. ¿A nivel del 1% el vicerrectorado puede aceptar el 42% como indicador?

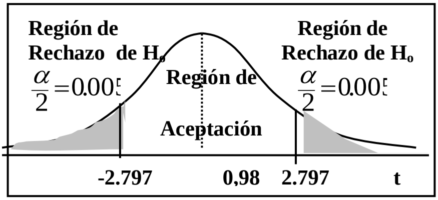
Solución:

μ_p = P = 42%, $p=\frac{13}{25}=0{,}52$, q = 0,48, n = 25, υ= n – 1 = 24

a) H₀: μ_p = 0,42

b) Hₐ: $\mu \neq$ 0,42

c) α = 0.01



d) $$s_p=\sqrt{p.q}$$

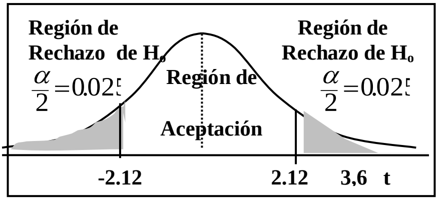
e) $$t=\frac{p-P}{\sqrt{\frac{p.q}{n-1}}}=\frac{0{,}52-0{,}42}{\sqrt{\frac{0{,}52(0{,}48)}{25-1}}}=0{,}98$$

f) Si se debe aceptar el 42% como indicador del problema, al nivel del 1%

d) Distribución de diferencia entre dos medias

Para investigar sobre el rendimiento de los alumnos en la sección diurna y nocturna se obtuvo una muestra de sus calificaciones promedio:
Diurno: 14.0, 16.8, 15.2, 14.4, 16.0, 14.4, 16.8, 19.2, 16.0, 14.4
Nocturno: 12.8, 14.4, 13.6, 15.2, 16.0, 12.8, 14.4, 13,6
¿Los anteriores resultados permiten concluir que hay diferencia en el rendimiento diurno y nocturno?

Solución:



a) H_o: $\mu_x = \mu_y$

H_a: $\mu_x \neq \mu_y$

b) $\alpha = 0.05$

c) $\upsilon = n_1 + n_2 - 2 = 16$

d) $\bar{x} = 15{,}72$; $\bar{y} = 14{,}1$

e) $s^2 = \dfrac{\sum(x_i - \bar{x})^2 + \sum(y_i - \bar{y})^2}{n_1 + n_2 - 2}$; $s^2 = \dfrac{23{,}06 + 8{,}88}{10 + 8 - 2} = 0{,}886$

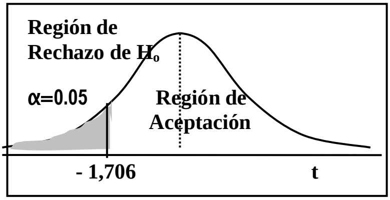
f) $t = \dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{s^2}{n_1} + \dfrac{s^2}{n_2}}} = \dfrac{15{,}72 - 14{,}1}{\sqrt{\dfrac{0{,}886}{10} + \dfrac{0{,}886}{8}}} = \dfrac{1{,}62}{0{,}4465} = 3{,}6$

g) Al nivel del 5% si se puede concluir que existe diferencia en el rendimiento

e) Distribución de diferencia entre dos proporciones

Una empresa desea realizar una investigación sobre el consumo de aceite marca A. En una muestra probabilística de 14 amas de casa, el 20% indicó la preferencia por la marca A. Posterior a una campaña publicitaria intensiva de radio y televisión, se seleccionó una nueva muestra entre amas de casa del mismo tamaño y clase social. En esta muestra el 25% indicó preferencia por la marca A. De acuerdo con estos resultados y a un nivel del 5%, ¿podría rechazarse la hipótesis de que la campaña publicitaria no fue efectiva?

Solución: $n_1 = 14$, $n_2 = 14$, $p_1 = 0.20$, $p_2 = 0.25$, $\alpha = 0.05$,



a) H_o: $\mu_{p_1} = \mu_{p_2}$

H_a: $\mu_{p_1} < \mu_{p_2}$

b) α = 0.05

c) $\upsilon = n_1 + n_2 - 2 = 26$

d) $S_{p_1-p_2} = \sqrt{\dfrac{p_1.q_1}{n_1} + \dfrac{p_2.q_2}{n_2}}$

e) $t = \dfrac{p_1 - p_2}{\sqrt{\dfrac{p_1.q_1}{n_1} + \dfrac{p_2.q_2}{n_2}}} = \dfrac{0{,}20 - 0{,}25}{\sqrt{\dfrac{(0.2)(0.8)}{14} + \dfrac{(0.25)(0.75)}{14}}} = \dfrac{-0.05}{0{,}1575} = -0{,}32$

f) Al nivel del 5% si se puede concluir que existe diferencia en el rendimiento

7.9. OTRAS PRUEBAS DE HIPOTESIS[2]

a) PRUEBA EN OBSERVACIONES APAREADAS

Se debe tomar una muestra aleatoria de pares, de manera que cada observación, esté asociada con alguna observación en particular. La muestra de pares de observaciones da base para considerar dos muestras dependientes, donde las observaciones de un par estarán relacionadas entre si. Sus fórmulas son:

$$\bar{d} = \frac{\sum d_i}{n} \; ; \quad s_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n-1}} \; ; \quad t = \frac{\bar{d} - a_d}{\frac{S_d}{\sqrt{n}}} = \frac{\bar{d} - 0}{\frac{S_d}{\sqrt{n}}} = \frac{\bar{d}}{\frac{S_d}{\sqrt{n}}}$$

Donde: $\bar{d}$ = Media aritmética de las diferencias

s_d = Desviación típica de las observaciones apareadas

a_d = 0 (Medias de las diferencias de la población es cero)

t = Variante estadística

1. En una escuela pública se escogieron 10 pares de niños de primer año para comparar similitud de inteligencia y preparación. Un niño de cada par fue

[2] Estadística y Muestreo. Martínez C. Ecoediciones. Bogotá. 2002

enseñado a leer con un método y el otro niño con otro método. Después del período de aprendizaje, los niños fueron sometidos a una prueba de lectura con los siguientes resultados, (el puntaje fue de 0 a 100). ¿A un nivel de significación del 5%, existe alguna diferencia significativa en la mayor efectividad de alguno de los métodos aplicados?

Solución:

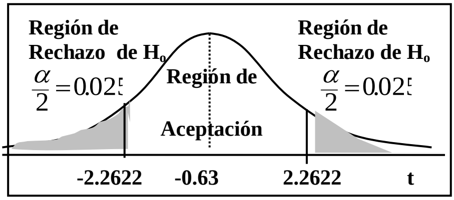
Pareja	**Método I** x_i	**Método II** y_i	$d_i = x_i - y_i$	$d_i - \overline{d}$	$(d_i - \overline{d})^2$
1	65	63	2	2.7	7.29
2	68	68	0	0.7	0.49
3	70	68	2	2.7	7.29
4	63	60	3	3.7	13.69
5	64	68	-4	-3.3	10.89
6	62	66	-4	-3.3	10.89
7	74	70	4	4.7	22.09
8	72	78	-6	-5.3	28.09
9	70	70	0	0.7	0.49
10	66	70	-4	-3.3	10.89
			-7		**112.10**

$$\overline{d} = \frac{\sum d_i}{n} = \frac{-7}{10} = -0.7 \qquad s_d = \sqrt{\frac{\sum (d_i - \overline{d})^2}{n-1}} = \sqrt{\frac{112.1}{10-1}} = 3.53$$

a) $H_o: a_d = 0$
   $H_a: a_d \neq 0$

b) $\alpha = 0.05$

c) $$t = \frac{\overline{d}}{\frac{S_d}{\sqrt{n}}} = \frac{-4.375}{\frac{3.53}{\sqrt{10}}} = -0.63$$



d) gl = v = n – 1 = 10 – 1 = 9 $\Rightarrow$ t = 2.2622

e) **Decisión**: t = -0.63 se ubica en la zona de aceptación y se acepta la hipótesis nula, por lo tanto se concluye que ningún método es superior al otro, al nivel del 5%.

2. Supongamos que se quiere estudiar la efectividad de una dieta y se nos proporciona la siguiente información referente a los pesos, antes y después en una muestra al azar de 8 mujeres adultas con edades de 35 a 40 años o más (dado en libras). Docimar al nivel del 5% que la dieta fue efectiva

Solución:

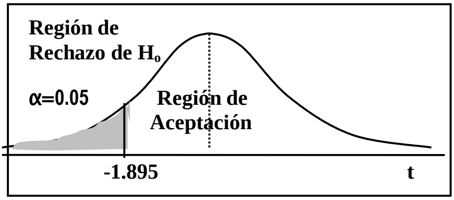
Pareja	Antes x_i	Después y_i	$d_i = x_i - y_i$	$d_i - \bar{d}$	$(d_i - \bar{d})^2$
1	137	132	-5	-0.625	0.39
2	130	121	-9	-4.625	21.39
3	124	126	2	6.375	40.64
4	138	130	-8	-3.625	13.14
5	149	147	-2	2.375	5.64
6	140	141	1	5.375	28.89
7	168	159	-9	-4.625	21.39
8	152	147	-5	-0.625	0.39
			-35	**0**	**131.87**

Nota.- Se hubiera podido trabajar con diferencias positivas y la gráfica sería al contrario, lado derecho. El resultado es el mismo.

$$\bar{d} = \frac{\sum d_i}{n} = \frac{-35}{8} = -4.375 \qquad s_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n-1}} = \sqrt{\frac{131.87}{8-1}} = 4.34$$

$$S_{\bar{d}} = \frac{s_d}{\sqrt{n}} = \frac{4.34}{\sqrt{8}} = 1.53$$

a) $H_o: a_d = 0$
   $H_a: a_d < 0$
b) $\alpha = 0.05$
c) $S_{\bar{d}} = 1.53$



d) $t = \frac{\bar{d}}{S_{\bar{d}}} = \frac{-4.375}{1.53} = -2.86$

e) gl = v = n – 1 = 8 – 1 = 7 ⇒ t = -1.895
f) Decisión: t = -2.86 se ubica en la zona crítica y se rechaza la hipótesis nula, por lo tanto aceptamos que la dieta ha sido efectiva, al nivel del 5%.

EJERCICIOS PROPUESTOS

1. A diez atletas se les sometió a un programa de entrenamiento físico intensivo por parte del preparador físico. Se anotaron sus pesos (en libras) antes y después de dicho entrenamiento con los resultados siguientes:

Atleta	No.	1	2	3	4	5	6	7	8	9	10
Antes	x:	127	195	162	170	143	205	168	175	197	136
Después	y:	135	200	160	182	147	200	172	186	194	141

¿Afecta el programa el peso medio de los atletas?

2. La dirección de una empresa de productos dietéticos para niños desea confrontar las propiedades nutritivas de una de sus especialidades con las de un producto de la competencia. Para estos efectos, encarga a un médico nutrir 6 niños con su producto (x) y después alimentar a estos mismos niños con el producto de la competencia (y). Se ha decidido que el efecto de las propiedades nutritivas de los productos venga medido por el aumento de peso diario de los niños. Al nivel del 1%, ¿existe una diferencia entre el valor nutritivo de los producto?

Niño	No.	1	2	3	4	5	6
Producto A	x:	21	20	20	22	16	21
Producto B	y:	17	18	18	16	14	13

3. Un laboratorio farmacéutico hace una prueba previa con cierto tipo de medicamento, que puede preparar en forma de jarabe (x) o en otra forma de cápsulas (y). para medir la preferencia por uno y otro tipo de medicamento, utiliza una muestra de 10 médicos, quienes manifiestan sus preferencias, por dichos medicamentos, con los siguientes resultados (se utilizó una escala de 1 a 10) en puntos. Al nivel de significación del 5%, ¿se podría concluir que el medicamento (y)

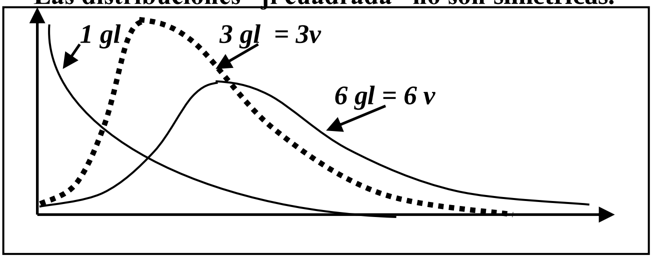
Médico	No.	1	2	3	4	5	6	7	8	9	10
Medicamento	x:	3	1	5	2	0	4	3	3	2	5
Medicamento	y:	2	4	4	7	3	4	6	5	5	8

Respuesta: Se rechaza la hipótesis nula y se acepta la hipótesis alternativa $a_d < 0$. Al nivel del 5%, el tipo de medicamento (y) alcanzó una mayor preferencia en la muestra.

b) PRUEBA DE CHI-CUADRADO χ^2 (Ji-cuadrada)

Esta distribución fue introducida por F.R Helmert en 1876 y redescubierta en 1900 por Kart Pearson. Tiene muchos usos importantes, incluyendo ensayos de hipótesis acerca de proporciones y cálculo de intervalos de confianza para varianzas. Hay una distribución ji cuadrada diferente según el valor de n-1, lo cual representa los grados de libertad (gl). Así:

Las distribuciones "ji cuadrada" no son simétricas.



Nota: Cuando gl es grande (v > 30), la distribución ji cuadrada se aproxima a la distribución normal. La variable $\sqrt{2\chi^2}$ es asintóticamente normal con media $\sqrt{2v-1}$ y varianza 1.

La curva está dada por: $Y = C\left(\chi^2\right)^{\frac{v-2}{2}} . e^{-\frac{\chi^2}{2}}$

Donde. v = n-1 = grados de libertad

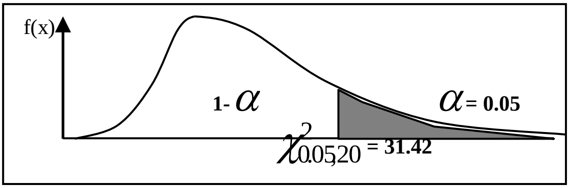
C = constante que depende de v, para que el área bajo la curva sea 1

¿Cómo leer en la tabla?.

Se busca en la primera fila χ^2_v y en la primera columna gl, en la intersección de la fila y la columna correspondiente se encuentra el valor de χ^2 correspondiente.

Ejercicio Si se tiene una variable aleatoria que sigue una distribución χ^2 con 20 grados de libertad, obtener $\chi^2_{\alpha,v}$ para:

a) α = 5% = 0.05 $\Rightarrow$ $\chi^2_{\alpha,v}$ = $\chi^2_{0.05,20}$ = 31.42



Proceso para la prueba Ji-cuadrada

1. Formular la hipótesis
2. Establecer las diferencias entre las frecuencias observadas y las esperadas, se eleva la cuadrado y se divide cada una de ellas para la frecuencia teórica esperada.
3. Se suma y se obtiene Ji-cuadrada

Ecuación sin corregir:

$$\chi^2 = \sum\left[\frac{\left(n_i - n_i^*\right)^2}{n_i^*}\right]$$

Ecuación con corrección de Yates:

$$\chi^2 = \sum\left[\frac{\left(\left|n_i - n_i^*\right| - 0.5\right)^2}{n_i^*}\right]$$

Donde: n_i = Frecuencia observada o real

n_i^* = Frecuencia teórica o esperada

v = (k-1)(j-1) = grados de libertad (k = filas, j = columnas)

Nota: La corrección de yates se utiliza cuando la tabla es de 2x2, es decir, v = 1 y la variables es discreta. En muestras grandes se obtienen prácticamente los mismos resultados. La corrección de Yates hoy es muy poco utilizada por cuanto se ha demostrado que, en la mayoría de casos la hipótesis nula no se rechaza.

1. Durante una epidemia se obtuvieron los siguientes datos sobre la efectividad de una vacuna como medida preventiva para los médicos. Estos datos, ¿indican la efectividad de la vacunación con base en el nivel de significación del 1%?

TRATAMIENTO	ENFERMOS	NO ENFERMOS	TOTAL
Vacunados	192	4	196
No vacunados	113	34	147
TOTAL	305	38	343

Solución:

a) a) Calculamos: $n_i^* = n.p$

$n_1^* = 305\left(\frac{196}{343}\right) = 174.28;$ $n_2^* = 305\left(\frac{147}{343}\right) = 130.71;$

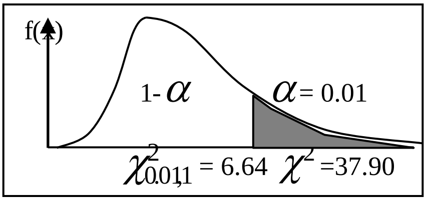
$n_3^* = 38\left(\frac{196}{343}\right) = 21.71;$ $n_4^* = 38\left(\frac{147}{343}\right) = 16.29$

b) Calculamos χ^2

n_i	n_i^*	$n_i - n_i^*$	$(n_i - n_i^*)^2$	$\frac{(n_i - n_i^*)^2}{n_i^*}$
192 113 4 34	174.28 130.71 21.71 16.29	17.72 -17.71 -17.71 17.71	313.99 313.64 313.64 313.64	1.802 2.399 14.45 19.25
343				**37.90**
n				$\chi^2 = \sum\left[\frac{(n_i - n_i^*)^2}{n_i^*}\right]$

c) Hipótesis. H₀ = $n_i = n_i^*$

H$_a$ = $n_i \neq n_i^*$



d) $\chi^2 = \sum\left[\frac{(n_i - n_i^*)^2}{n_i^*}\right]$ = 37.90

e) v = (k-1)(j-1) = (2-1)(2-1) = 1 $\chi^2_{0.01,1} = 6.64$

f) **Decisión**: Como $\chi^2 = 37.90$, cae en el área de rechazo de Ho, Se rechaza la hipótesis nula y se acepta $H_a = n_i \neq n_i^*$. Es decir, la diferencia es significativa.

Aplicando la corrección de Yates:

n_i	n_i^*	$\lvert n_i - n_i^* \rvert$	$\lvert n_i - n_i^* \rvert - 0.5$	$(\lvert n_i - n_i^* \rvert - 0.5)^2$	$\frac{(\lvert n_i - n_i^* \rvert - 0.5)^2}{n_i^*}$
192	174.28	17.72	17.21	296.18	1.699
113	130.71	17.71	17.21	296.18	2.266
4	21.71	17.71	17.21	296.18	13.64
34	16.29	17.71	17.21	296.18	18.18
					35.785
n					$\chi^2 = \sum \left[\frac{(\lvert n_i - n_i^* \rvert - 0.5)^2}{n_i^*} \right]$

$$\chi^2 = \sum \left[\frac{(\lvert n_i - n_i^* \rvert - 0.5)^2}{n_i^*} \right] = 35.785$$ (se llega a la misma conclusión)

c) Dócimas de homogeneidad e independencia

Prueba de independencia, se extrae una sola muestra de la población y sólo nos interesa probar si existe alguna relación entre los criterios de clasificación establecidos.

Prueba de homogeneidad, al contrario de la de independencia, se extraen dos o más muestras provenientes de dos o más poblaciones; algunas veces de una muestra se pueden obtener dos o más categorías y el interés es probar si las poblaciones tienen cierta identidad con respecto a las características analizadas.

1. Un almacén vende lavadoras y aspiradoras eléctricas y tiene tres vendedores. Las ventas realizadas en un mes por dichos vendedores se resume en la tabla. ¿Demuestra este resultado que la habilidad de cada

vendedor depende del tipo de artículo que vende?. Pruebe la hipótesis de homogeneidad al nivel del 5%.

APARATOS	VENDEDORES			TOTAL
	1	2	3	
Lavadoras	20	8	15	43
Aspiradoras	17	16	5	38
TOTAL	37	24	20	81

Solución:

a) Calculamos n_i^* : $\quad n_i^* = n.p$

$$n_1^* = 37\left(\frac{43}{81}\right) = 19.6; \qquad n_2^* = 37\left(\frac{38}{81}\right) = 17.4;$$

$$n_3^* = 24\left(\frac{43}{81}\right) = 12.7; \qquad n_4^* = 24\left(\frac{38}{81}\right) = 11.3;$$

$$n_5^* = 20\left(\frac{43}{81}\right) = 10.6; \qquad n_6^* = 20\left(\frac{38}{81}\right) = 9.4$$

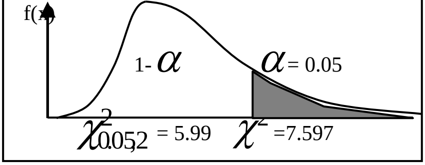
b) Calculamos χ^2

n_i	n_i^*	$n_i - n_i^*$	$(n_i - n_i^*)^2$	$\frac{(n_i - n_i^*)^2}{n_i^*}$
20	19.6	0.4	0.16	0.0082
17	17.4	-0.4	0.16	0.0092
8	12.7	-4.7	22.09	1.7393
16	11.3	4.7	22.09	1.9549
15	10.6	4.4	19.36	1.8264
5	9.4	-4.4	19.36	2.0596
81	**81**			**7.5976**
n	$\sum n_i^*$			$\sum\left[\frac{(n_i - n_i^*)^2}{n_i^*}\right]$

c) Hipótesis: H_0 = Las muestras se extraen de la misma población

H_a = Las muestras se extraen de distintas poblaciones

d) $\chi^2 = \sum\left[\frac{(n_i - n_i^*)^2}{n_i^*}\right]$ = 7.5976



e) v = (k-1)(j-1) = (2-1)(3-1) = 2

f) $\chi^2_{0.05,2} = 5.99$

g) **Decisión**: Como $\chi^2 = 7.5976$, cae en el área de rechazo de Ho, Se rechaza la hipótesis nula y se acepta H_a, las muestras se extraen de distintas poblaciones. Es decir, que la habilidad de cada vendedor no depende del artículo.

EJERCICIOS PROPUESTOS

1. Un grupo de 30 hombres y 80 mujeres expresan su preferencia por dos productos. ¿Existe alguna diferencia significativa en cuanto a las preferencias por el lugar?, al nivel de significación del 5%.

SEXO	PRODUCTO A	PRODUCTO B	TOTAL
Hombres	9	21	30
Mujeres	8	72	80
TOTAL	17	93	110

2. Se eligen 1100 turistas al azar y se les interroga acerca de sus hábitos de fumar y beber, se resume en la tabla. ¿Qué conclusión debe deducirse respecto a los hábitos de fumar y beber entre los turistas?, al nivel de significación del 5%. Probar:
   H_o: No hay relación entre los hábitos de fumar y beber.
   H_a: Si hay relación entre los hábitos de fumar y beber.

Beber	Fumar				Total	%
	Nunca	Ocasionalmente	Moderadamente	Excesivamente		
Nunca	85	23	56	36	200	18.18
Ocasionalmente	153	44	128	75	400	36.36
Moderadamente	128	26	101	45	300	27.28
Excesivamente	134	7	15	44	200	18.18
Total	500	100	300	200	1100	100

3. En una investigación de opinión, se preguntó a 1 000 personas: ¿cómo calificaría el desempeño del Presidente de la República: bueno, regular, malo?. Las respuestas, clasificadas de acuerdo al nivel educativo de los encuestados, se resume en la tabla. Utilizando un nivel del 5%, ¿podemos concluir que la calificación de su desempeño es independiente del nivel educacional de los encuestados?

Desempeño	Nivel educacional			Total
	Elemental	Secundario	Universitario	
Bueno	82	427	191	700
Regular	10	110	60	180
malo	8	63	49	120
Total	100	600	300	1000