Universidad Manuela Beltrán

Estadística Inferencial

Elaborado por

Milady Astrid Guerrero Velasco

Magíster en Docencia de las Matemáticas (Ciencias Básicas)

Bogotá, Colombia
Noviembre 2024.

Contenido

Capítulo 1

Intervalos de confianza

Los intervalos de confianza (IC) son una herramienta estadística fundamental que permite estimar un rango de valores dentro del cual se espera que se encuentre un parámetro poblacional, con un nivel de confianza determinado. En otras palabras, un intervalo de confianza proporciona una estimación más precisa que un solo valor puntual (como la media o la proporción) y dice cuánta incertidumbre hay en esa estimación.

El intervalo de confianza describe la variabilidad entre la medida obtenida en un estudio y la medida real de la población (el valor real). Corresponde a un rango de valores, cuya distribución es normal y en el cual se encuentra, con alta probabilidad, el valor real de una determinada variable. Esta alta probabilidad se ha establecido por consenso en 95%. Así, un intervalo de confianza de 95% indica que dentro del rango dado se encuentra el valor real de un parámetro con 95% de certeza.

1.1 Para la media de una población con varianza conocida

Sea X_n una variable aleatoria con σ conocida con una muestra suficientemente grande ($n > 30$) y sigue una disbribución normal se tiene que el intervalo de confianza IC se puede hallar:

$$IC = \bar{x} \pm Z(\tfrac{\sigma}{\sqrt{n}})$$

Ejemplo:

Se tiene una muestra de 100 pacientes y la media de presión arterial es 120 mmHg, con una desviación estándar poblacional de 15 mmHg. Se quiere calcular un intervalo de confianza bilateral del 95%.

A continuación se muestra cómo hallar el intervalo de confianza.

```
> # Parámetros
  media_muestral <- 120     # Media de la muestra
  desviacion_estandar <- 15 # Desviación estándar de la población
  n <- 100                  # Tamaño de la muestra
  nivel_confianza <- 0.95   # Nivel de confianza (95%)

  # Valor crítico Z para un nivel de confianza del 95%
  z_critico <- qnorm(1 - (1 - nivel_confianza) / 2)

  # Error estándar
  error_estandar <- desviacion_estandar / sqrt(n)

  # Intervalo de confianza
  limite_inferior <- media_muestral - z_critico * error_estandar
  limite_superior <- media_muestral + z_critico * error_estandar

  # Resultado
  cat("Intervalo de confianza al 95%: [", round(limite_inferior, 2), ",", round(limite_superior, 2), "]\n")
```

Figura 1.1: Código para hallar un intervalo de confianza bilateral varianza conocida.

Y se obtiene que e el intervalo de confianza para la media al 95% es $[117.06, 122.94]$ mmHg.

1.2 Para la media de una población con varianza desconocida

El intervalo de confianza para la media cuando la varianza es desconocida se calcula usando una distribución conocida como T de student, ya que, como la varianza de la población no es conocida se hace una estimación de la misma cuando la muestra es significativamente grande.

Para hallar el intervalo de confianza para la media, se estima desviación estándar de la muestra s y se calcula de la siguiente manera:

$$IC = \bar{x} \pm t_{\alpha/2,df}(\frac{s}{\sqrt{n}})$$

donde $t_{\alpha/2,df}$ es el valor crítico de la distribución T de Student para un nivel de confianza $1-\alpha$, con df grados de libertad.

Ejemplo: Un investigador de salud está estudiando la presión sistólica de un grupo de personas mayores de 60 años. Para ello, toma una muestra de 30 personas mayores que 60 años y halla la presión sistólica de cada paciente. Los datos se presentan a continuación:

120.8	122.3	124	125.2	108
120	127.4	132	115.7	73
139	102.4	83	183	78
142	102.6	83	173	79
120.1	102.4	122.3	125	78
76	106.4	116.2	73.2	122

Se quiere calcular el intervalo de confianza para la media con un nivel de confianza de 99%.

Para determinar el intervalo de confianza en R usando T de Student se puede realizar de dos maneras y esto, depende de la información que brinde el problema.Si es como en este caso, que bridan el conjunto de datos (la información de la muestra) se usa el parámetro t.test, así:

```
Session restored from your saved work on 2024-Nov-19 14:59:06 UTC (1 day ago)
> presion_sistolica <- c(120.8,122.3,124,125.2,108,120,127.4,132,115.7,73,139,102.4,83,183,78,142,102.6,83,173,7
9,120.1,102.4,122.3,125,78,76,106.4,116.2,73.2,122)
> t.test(presion_sistolica,conf.level=0.99)

        One Sample t-test

data:  presion_sistolica
t = 22.482, df = 29, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
99 percent confidence interval:
  98.70732 126.29268
sample estimates:
mean of x
    112.5
```

Figura 1.2: Código para hallar un intervalo de confianza bilateral varianza desconocida.

Pero, si se da el valor de la media y la desviación estándar poblacional, es decir, no se conocen los valores específicos de la muestra, es decir, se tiene que $\bar{x}$=112,17 y σ =27,35 (valores aproximados). Entonces:

```
> # Definir los parámetros
> media_muestral <- 112.17   # Media muestral
> desviacion_muestral <- 27.35  # Desviación estándar muestral
> tamano_muestra <- 30  # Tamaño de la muestra
> nivel_confianza <- 0.99  # Nivel de confianza
>
> # Calcular el valor t crítico para el intervalo de confianza
> alfa <- 1 - nivel_confianza
> grados_libertad <- tamano_muestra - 1
> t_critico <- qt(1 - alfa / 2, grados_libertad)
>
> # Calcular el margen de error
> margen_error <- t_critico * (desviacion_muestral / sqrt(tamano_muestra))
>
> # Calcular el intervalo de confianza
> limite_inferior <- media_muestral - margen_error
> limite_superior <- media_muestral + margen_error
>
> # Imprimir los resultados
> cat("El intervalo de confianza es: [", limite_inferior, ",", limite_superior, "]\n")
El intervalo de confianza es: [ 98.40625 , 125.9337 ]
>
```

Figura 1.3: Código para hallar un intervalo de confianza bilateral varianza desconocida.

Como se puede evidenciar en las dos maneras en cómo se hallan los intervalos de confianza no se obtienen valores exactamente iguales y esto se genera debido a que la media y la desviación estándar estan aproximadas por lo que siempre conviene trabajar con los datos en caso de que sea posible.

1.3 Para la proporción poblacional

El intervalo de confianza para la proporción muestral es una estimación del rango dentro del cual se encuentra el valor verdadero de la proporción en la población. Este intervalo se construye a partir de una proporción calculada a partir de una muestra y tiene un nivel de confianza determinado, como el 95% o el 99%.

El intervalo de confianza para una proporción se determina de la siguiente manera:

$$IC = \hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Donde $\hat{p}$ es la proporción muestral que se define como el número de éxitos dividido el tamaño de la muestra.

Ejemplo:

La mosca adulta del gusano tornillo es de color azul metálico y su tamaño triplica al de la mosca común. El gusano tornillo pone los huevos en las heridas de animales de sangre caliente y produce una grave infección. Se realizó un experimento con el objetivo de controlar esta población. Se expuso a las crisálidas del gusano tornillo a una dosis de radiación de 2500 rad con la esperanza de esterilizar a la mayor parte de los machos. Dado que las hembras se aparean sólo una vez, si lo hacen con un macho estéril producirán huevos estériles. Se encontró que, tras la radiación, 415 de los 500 pareamientos observados dieron como resultado huevos estériles.

Se va a construir un intervalo de confianza del 95% y para ello, lo primero que se encontrar es $\hat{p}$ así:

$$\hat{p} = \frac{415}{500} = 0.83 = 83\%$$

A continuación se presenta una manera para determinar el intervalo de confianza en R.

```
> # Definir los parámetros
> x <- 415  # Número de éxitos
> n <- 500  # Tamaño de la muestra
> nivel_confianza <- 0.95  # Nivel de confianza
>
> # Calcular la proporción muestral
> p_hat <- x / n
>
> # Calcular el valor Z para el nivel de confianza dado (por defecto para 95% es 1.96)
> z <- qnorm(1 - (1 - nivel_confianza) / 2)
>
> # Calcular el margen de error
> error_margin <- z * sqrt(p_hat * (1 - p_hat) / n)
>
> # Calcular el intervalo de confianza
> IC_inferior <- p_hat - error_margin
> IC_superior <- p_hat + error_margin
>
> # Imprimir el resultado
> cat("El intervalo de confianza es: [", IC_inferior, ", ", IC_superior, "]", "\n")
El intervalo de confianza es: [ 0.7970749 ,  0.8629251 ]
```

Figura 1.4: Código para hallar un intervalo de confianza bilateral para la proporción.

1.4 Para la diferencia de dos proporciones

En los estudios médicos y biológicos surge frecuentemente el problema de comparar dos proporciones. La situación general puede describirse así: hay dos poblaciones de interés; en cada población, se estudia el mismo rasgo; cada miembro de cada población puede clasificar se en función de que lo tenga o no, y en cada población es conocida la proporción de los que lo tienen. Por lo que se realizan inferencias para $\hat{p}_1$, $\hat{p}_2$ y $\hat{p}_1 - \hat{p}_2$.

Para determinar el estimador puntual $\hat{p}_1 - \hat{p}_2$ a un intervalo se considera la distribución de probabilidad de esta variable aleatoria, donde las muestras son de un gran tamaño, y el estimador $\hat{p}_1 - \hat{p}_2$ es aproximadamente normal se tiene que el intervalo está dado por:

$$IC = (\hat{p}_1 - \hat{p}_2) \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Ejemplo

En un estudio sobre el uso de la prednisona en el tratamiento de pacientes renales, se utilizaron 72 sujetos en 19 hospitales. De los 34 pacientes tratados con prednisona, sólo uno sufrió insuficiencia renal. Sin embargo, de los 38 que recibieron un placebo, se produjo insuficiencia renal en 10. Basándose en este estudio se tienen que las proporciones muestrales son:

$$\hat{p}_1 = \frac{1}{34} = 0.03 \text{ y } \hat{p}_2 = \frac{10}{38} = 0.26$$

Se va a construir un intervalo de confianza del 95% de la diferencia en la tasa de insuficiencia renal entre los que reciben prednisona y los que no reciben el fármaco.

A continuación se ejemplifica de cómo hacerlo en R.

```
> # Definir los parámetros para ambas muestras
> x1 <- 1  # Número de éxitos en la primera muestra
> n1 <- 34  # Tamaño de la primera muestra
> x2 <- 10  # Número de éxitos en la segunda muestra
> n2 <- 38  # Tamaño de la segunda muestra
> nivel_confianza <- 0.95  # Nivel de confianza
>
> # Calcular las proporciones muestrales
> p_hat1 <- x1 / n1
> p_hat2 <- x2 / n2
>
> # Calcular el valor Z para el nivel de confianza dado
> z <- qnorm(1 - (1 - nivel_confianza) / 2)
>
> # Calcular el margen de error
> error_margin <- z * sqrt((p_hat1 * (1 - p_hat1) / n1) + (p_hat2 * (1 - p_hat2) / n2))
>
> # Calcular la diferencia de las proporciones
> diferencia <- p_hat1 - p_hat2
>
> # Calcular el intervalo de confianza
> IC_inferior <- diferencia - error_margin
> IC_superior <- diferencia + error_margin
>
> # Imprimir el resultado
> cat("El intervalo de confianza para la diferencia entre proporciones es: [", IC_inferior, ", ", IC_superior,
"]", "\n")
El intervalo de confianza para la diferencia entre proporciones es: [ -0.3848337 ,  -0.08265854 ]
```

Figura 1.5: Código para hallar un intervalo de confianza bilateral para la diferencia de dos proporciones.

Capítulo 2

Prueba de hipótesis

La estadística inferencial o estadística inductiva tiene como objetivo estudiar un población tomando como base una o más muestras. Trata de conocer o explicar el comportamiento de la población mediante los datos obtenidos de una muestra. Y en este análisis una de las maneras para conocer el comportamiento de la población es a través de la prueba de hipótesis.

Las pruebas de hipótesis son procedimientos que permiten tomar decisiones o inferencias de una población a partir de una muestra. Donde el proceso implica generar dos hipótesis:

- Hipótesis nula H_0: es la afirmación que se pone a prueba, generalmente afirma que no hay efecto o diferencia.

- Hipótesis alternativa H_1: es la afirmación que se considera si se rechaza la hipótesis nula, generalmente afirma que si hay efecto o diferencia.

A partir de la creación de las hipótesis, es importante tener en cuenta la distribución que se debe usar para desarrollar la prueba de manera general. A continuación se describen tres posibles maneras de hacer una prueba de hipótesis.

2.1 Prueba de hipótesis con varianza conocida

La prueba de hipótesis con varianza conocida se utiliza cuando se quiere evaluar una hipótesis sobre la media de una población, pero se conoce o se tiene una estimación confiable de la varianza de esa población. Este tipo de prueba es común cuando se tiene información previa sobre la varianza, como cuando se han realizado estudios previos o se conoce de antemano la dispersión de la población. Y para determinar el estadístico de prueba dada la μ se tiene:

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

La variable Z se conoce como una variable de distribución normal.

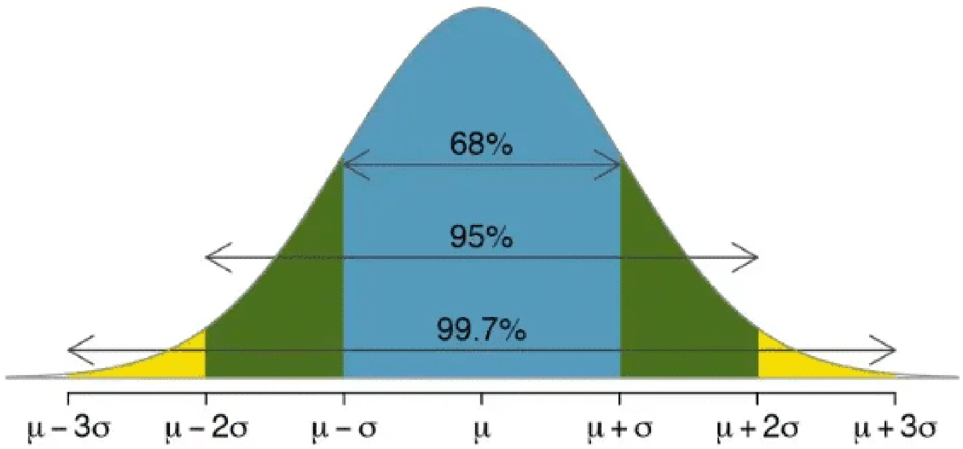

Figura 2.1: Distribución normal con varianza conocida.

Caso 1. Prueba de hipótesis unilateral izquierda. Con las hipótesis:

$$H_0 : \mu \geq \mu_0$$

$$H_1 : \mu < \mu_0$$

Y con un nivel de significancia de α, y obteniendo el etadístico de prueba Z se tiene que la región de rechazo está dada por:

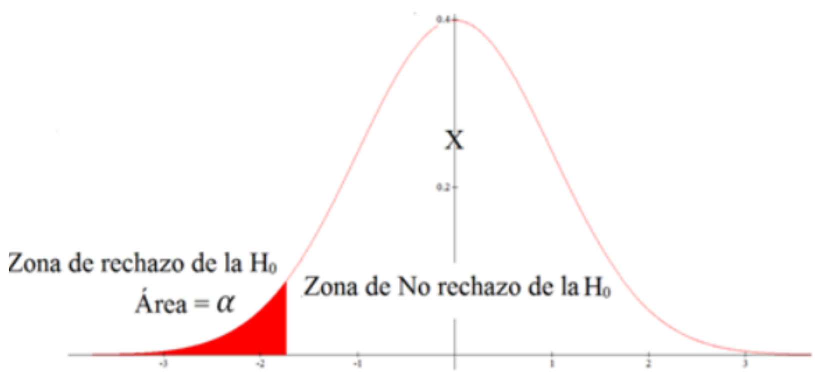

Figura 2.2: Prueba de hipótesis unilateral izquierda.

Tener en cuenta que la hipótesis nula se rechaza si el valor crítico cae en la zona de rechazo, es decir, cuando el valor crítico es menor que el estadístico de prueba.

Ejemplo:

En un informe de investigación, Richard H. Weindruch, de la Escuela de Medicina de la UCLA, afirma que los ratones con una vida promedio de 32 meses vivirán hasta alrededor de 40 meses con desviación estándar de 5.8 si 40% de las calorías en su dieta se reemplazan con vitaminas y proteínas. El investigador cree que, en realidad, los ratones viven menos de lo que se espera, por lo que toma una muestra de 64 ratones que son sometidos a esa dieta y observa una vida promedio de 38 meses, así, con un 99% de confianza realiza una prueba de hipótesis.

- Planteamiento de la hipótesis:

$$H_0 : \mu \geq 40$$

$$H_1 : \mu < 40$$

- Nivel de significancia:

$$\alpha = 1\% = 0,01$$

- Estadístico de prueba:

```
xbarra <- 38     # Media muestral
desvia <- 5.8    # Desviación estándar de la población
n <- 64          # Tamaño de la muestra
mu <- 40         # Media de referencia bajo la hipótesis nula

# Cálculo de la estadística de prueba Z
est <- (xbarra - mu) / (desvia / sqrt(n))
```

Figura 2.3: Código para hallar el estadístico de prueba en R.

El valor estadístico es $Z = -2.7586$

- Valor crítico:

```
qnorm(0.01)
```

Figura 2.4: Código para determinar el valor crítico en prueba de hipótesis unilateral izquierda.

El valor crítico (zona de rechazo) es $Z_\alpha = -2.326$

- Decisión:

  Dado que el valor crítico es mayor que el valor estadístico entonces H_0 se rechaza, es decir, existe suficiente evidencia estadística para determinar que la media de vida de los ratones es menor que 40 meses.

Caso 2. Prueba de hipótesis unilateral derecha. Con las hipótesis:

$$H_0 : \mu < \mu_0$$

$$H_1 : \mu \geq \mu_0$$

Y con un nivel de significancia de α, y obteniendo el etadístico de prueba T se tiene que la región de rechazo está dada por:

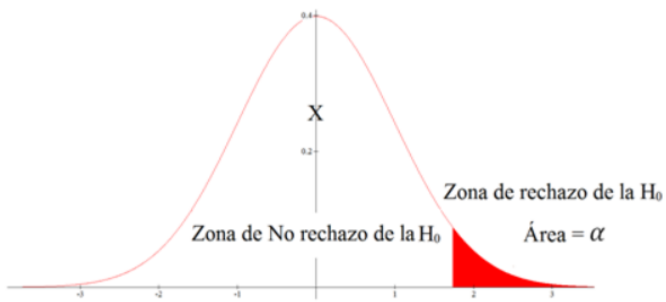

Figura 2.5: Prueba de hipótesis unilateral derecha

Tener en cuenta que la hipótesis nula se rechaza si el valor crítico cae en la zona de rechazo, es decir, cuando el valor crítico es mayor que el estadístico de prueba.

Ejemplo: De acuerdo con un estudio sobre un régimen alimenticio, la ingesta elevada de sodio se relaciona con úlceras, cáncer estomacal y migrañas. El requerimiento humano de sal es de tan sólo 220 miligramos diarios, el cual se rebasa en la mayoría de las porciones individuales de cereales listos para comerse. Si una muestra aleatoria de 20 porciones similares de cierto cereal tiene un contenido medio de 244 miligramos de sodio, ¿esto sugiere, aun nivel de significancia del 5%, que el contenido promedio de sodio para porciones individuales de ese cereal es mayor que 220 miligramos? Suponga que la distribución de contenidos de sodio es normal con una desviación estándar de 24.5 miligramos.

- Planteamiento de la hipótesis:

$$H_0 : \mu < 220$$

$$H_1 : \mu \geq 220$$

- Nivel de significancia:

$$\alpha = 5\% = 0,05$$

- Estadístico de prueba:

```
xbarra <- 244    # Media muestral
desvia <- 24.5   # Desviación estándar de la población
n <- 20          # Tamaño de la muestra
mu <- 220        # Media de referencia bajo la hipótesis nula

# Cálculo de la estadística de prueba Z
est <- (xbarra - mu) / (desvia / sqrt(n))
```

Figura 2.6: Código para hallar el estadístico de prueba en R.

El valor estadístico es $Z = 4.381$

- Valor crítico:

```
qnorm(1 - 0.05)
```

Figura 2.7: Código para determinar el valor crítico en prueba de hipótesis unilateral derecha.

El valor crítico (zona de rechazo) es $Z_{1-\alpha} = 1.645$

- Decisión:

  Dado que el valor crítico es mayor que el valor estadístico entonces H_0 se rechaza, es decir, existe suficiente evidencia estadística para determinar que el contenido medio de sodio para proporciones individuales de ese cereal es mayor que 220 miligramos.

Caso 3. Prueba de hipótesis bilateral. Con las hipótesis:

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

Y con un nivel de significancia de α, y obteniendo el etadístico de prueba T se tiene que la región de rechazo está dada por:

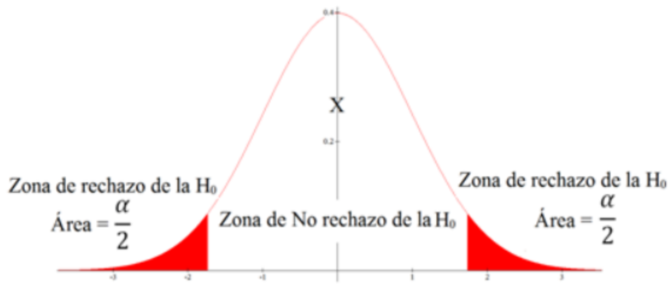

Figura 2.8: Prueba de hipótesis bilaterial

2.2 Prueba de hipótesis con varianza desconocida

Se refiere a un proceso estadístico que se utiliza cuando se quiere hacer una inferencia sobre una población, pero no se conoce la varianza de esa población. En este caso, se debe hacer una estimación de la varianza usando una muestra de datos, lo que implica usar la distribución t de Student en lugar de la distribución normal. Por tal motivo el estadístico que surge para realizar inferencias sobre μ, bajo estas condiciones es el siguiente:

$$T = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$$

La variable aleatoria T se conoce una variable de distribución de T de Student con $n-1$ grados de libertad. La distribución t y la distribución Z son similares en cuanto a que son simétricas alrededor de la media y el valor de la media es cero. Ambas tienen forma de campana, pero la distribución t decae más lentamente. Observe lo anterior en la siguiente imagen. La estructura de la prueba de hipótesis para μ cuando

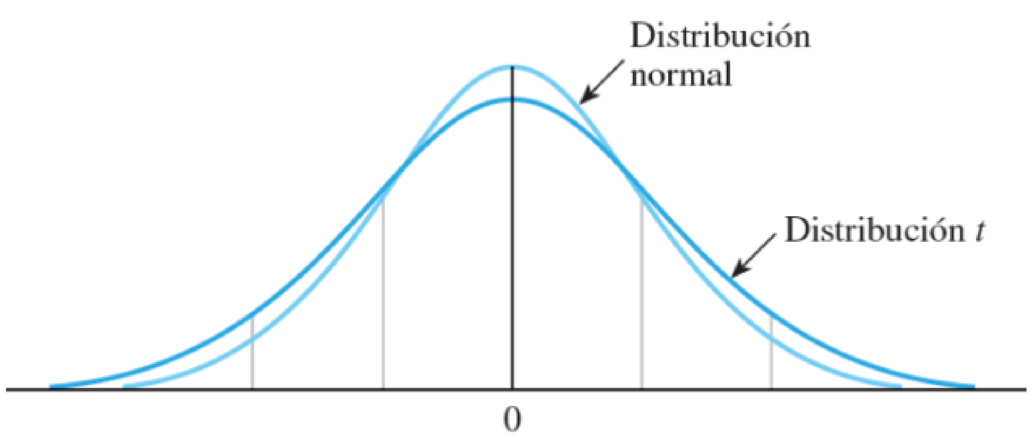

Figura 2.9: Comparación de la distribución normal con la distribución T de Student

se desconoce la varianza es igual a la prueba de hipótesis para μ conociendo σ (en cuanto a los pasos del procedimiento). A continuación, se muestran los posibles casos que se pueden presentar:

Caso 1. Prueba de hipótesis unilateral izquierda. Con las hipótesis:

$$H_0 : \mu \geq \mu_0$$

$$H_1 : \mu < \mu_0$$

Y con un nivel de significancia de α, y obteniendo el etadístico de prueba T se tiene que la región de rechazo está dada por: Tener en cuenta que la hipótesis nula se

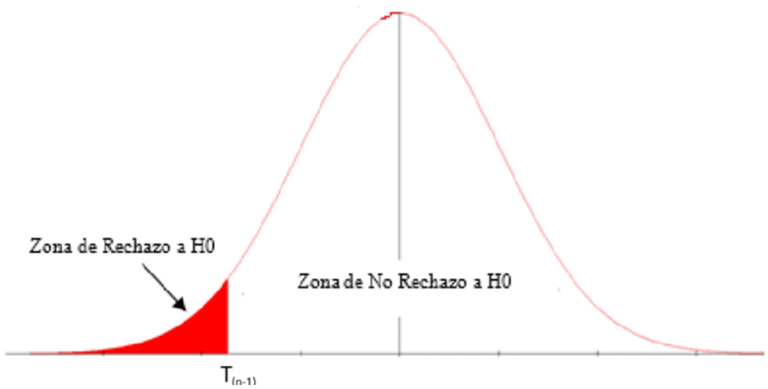

Figura 2.10: Prueba de hipótesis unilateral izquierda

rechaza si el valor crítico cae en la zona de rechazo, es decir, cuando el valor crítico es menor que el estadístico de prueba.

Caso 2. Prueba de hipótesis unilateral derecha. Con las hipótesis:

$$H_0 : \mu < \mu_0$$

$$H_1 : \mu \geq \mu_0$$

Y con un nivel de significancia de α, y obteniendo el etadístico de prueba T se tiene que la región de rechazo está dada por: Tener en cuenta que la hipótesis nula se rechaza si el valor crítico cae en la zona de rechazo, es decir, cuando el valor crítico es mayor que el estadístico de prueba.

Caso 3 Prueba de hipótesis bilateral. Con las hipótesis:

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

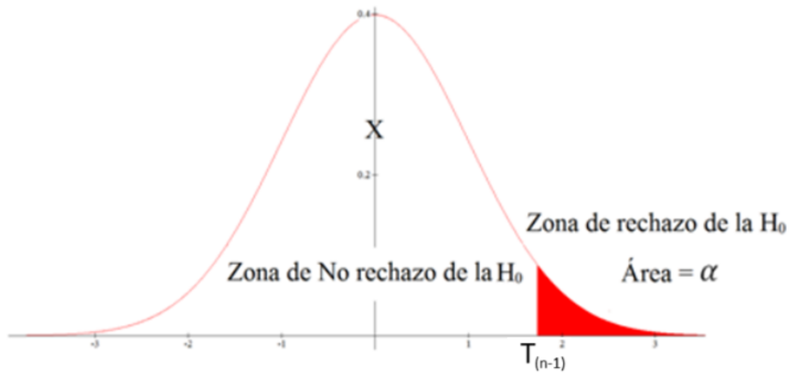

Figura 2.11: Prueba de hipótesis unilateral derecha

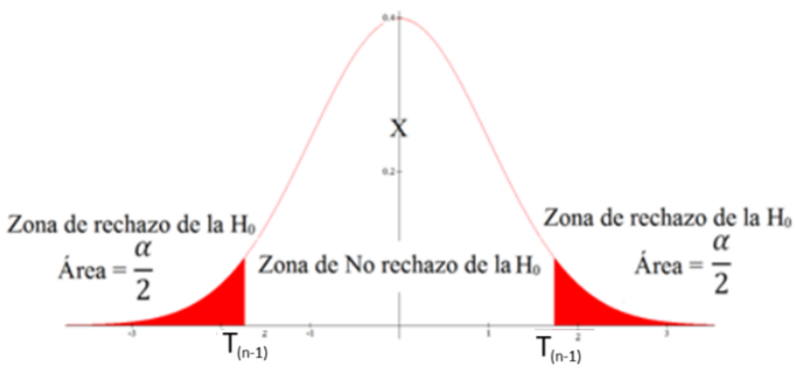

Figura 2.12: Prueba de hipótesis bilaterial

Y con un nivel de significancia de α, y obteniendo el etadístico de prueba T se tiene que la región de rechazo está dada por: Tener en cuenta que la hipótesis nula se rechaza si el valor crítico cae en la zona de rechazo, es decir, cuando el valor crítico es mayor al valor absoluto que el estadístico de prueba.

Ejemplo: El dueño de un café cree que la edad promedio de los jóvenes que entran a su establecimiento es diferente de 20 años. Para comprobarlo toma una muestra aleatoria de 22 clientes, obteniendo los siguientes datos:

30, 19, 18, 26, 24, 27, 27, 29, 21, 24, 20, 25, 18, 16, 20, 28, 22, 18, 23, 20, 23, 22.

Realizar una prueba de hipótesis usando un nivel de confianza del 95 %

- Planteamiento de la hipótesis:

$$H_0 : \mu = 20$$

$$H_1 : \mu \neq 20$$

- Nivel de significancia:

$$\alpha = 5\% = 0,05$$

- Estadístico de prueba:

  Para determinar el estadístico se debe tener el promedio y la desviación estándar muestral, de los datos brindados. Se tiene (aproximando a dos cifras decimales).

$$\bar{x} = 22.73, s = 3.95$$

```r
# Datos del problema
xbarra <- 22.73  # Media muestral
desvia <- 3.95   # Desviación estándar muestral
n <- 22          # Tamaño de la muestra
mu <- 20         # Media bajo la hipótesis nula

# Calcular la estadística t
est <- (xbarra - mu) / (desvia / sqrt(n))
```

Figura 2.13: Código para hallar el estadístico de prueba en R - T de Student.

  El valor estadístico es $T = 3.2417$

- Valor crítico:

```r
alpha <- 0.05   # Nivel de significancia
n <- 22         # Tamaño de la muestra
df <- n - 1     # Grados de libertad

# Valor crítico t para una prueba bilateral (alfa/2)
t_critico <- qt(1 - alpha / 2, df)
```

Figura 2.14: Código para determinar el valor crítico en prueba de hipótesis bilateral.

  El valor crítico (zona de rechazo) es $T_{1-\alpha} = \pm 2.0796$

- Decisión:

  Dado que el valor crítico es mayor que el valor estadístico entonces H_0 se rechaza, es decir, existe suficiente evidencia estadística para determinar que la creencia del dueño del café es cierta, puesto que con una confiablidad del 95% la edad promedio de los jóvenes que entran a su establecimiento es diferente de 20 años.

2.3 Prueba de hipótesis para proporciones

La prueba de hipótesis para proporciones es una técnica estadística utilizada para determinar si existe suficiente evidencia en una muestra para inferir que una proporción poblacional cumple con una afirmación específica. Cuando se realiza una prueba de hipótesis para una proporción, se tiene una proporción poblacional desconocida p, y se quiere probar una hipótesis sobre esa proporción. Por tal motivo el estadístico que surge para realizar inferencias sobre p, bajo estas condiciones es el siguiente:

$$Z = \frac{\widehat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

De manera similar que en los dos apartados anteriores, la hipótesis nula y alternativa surgen a partir de lo que se quiera probar.

Ejemplo un hospital ha desarrollado un nuevo tratamiento para la hipertensión y afirma que, con este tratamiento, al menos el 80% de los pacientes se recuperan o experimentan una mejora significativa en su salud. Un investigador quiere probar esta afirmación, y para ello realiza un estudio.

En el estudio, se incluyen 150 pacientes tratados con el nuevo medicamento, y de estos, 115 pacientes muestran una mejora significativa en su condición. El investigador quiere saber si la proporción de pacientes que se recuperan es realmente 80% (como afirma el hospital) o si hay suficiente evidencia para rechazar esta afirmación. Se realiza la prueba con un nivel de confianza del 95%.

Por lo que el planteamiento está dado por:

$$H_0 : p < 0.8$$

$$H_1 : p \geq 0.8$$

```
> # Definir los valores
> n <- 150   # Tamaño de la muestra
> x <- 115   # Número de pacientes que se recuperan
> p0 <- 0.80 # Proporción bajo la hipótesis nula
> alpha <- 0.05  # Nivel de significancia
>
> # Calcular la proporción muestral
> p_hat <- x / n
>
> # Calcular el error estándar de la proporción bajo la hipótesis nula
> se <- sqrt(p0 * (1 - p0) / n)
>
> # Calcular el estadístico de prueba Z
> Z <- (p_hat - p0) / se
>
> # Calcular los valores críticos de Z para una prueba bilateral (α = 0.05)
> z_critical_left <- qnorm(alpha / 2)  # Valor crítico inferior
> z_critical_right <- qnorm(1 - alpha / 2)  # Valor crítico superior
>
> # Mostrar resultados
> cat("Proporción muestral (p̂):", p_hat, "\n")
Proporción muestral (p̂): 0.7666667
> cat("Estadístico de prueba Z:", Z, "\n")
Estadístico de prueba Z: -1.020621
> cat("Valor crítico izquierdo (Z<sub>α/2</sub>):", z_critical_left, "\n")
Valor crítico izquierdo (Z<sub>α/2</sub>): -1.959964
> cat("Valor crítico derecho (Z<sub>1-α/2</sub>):", z_critical_right, "\n")
Valor crítico derecho (Z<sub>1-α/2</sub>): 1.959964
```

Figura 2.15: Prueba de hipótesis bilaterial

Teniendo en cuenta el valor y el estadístico de prueba se concluye que no se rechaza la hipótesis nula, significa que no hay evidencia estadística suficiente para afirmar que al menos el 80% de los pacientes se recuperan o experimentan una mejora significativa en su salud.

2.4 Prueba de hipótesis para las medias

Redactar una guía de laboratorio o clase afín a la formación del estudioso en LaTeX. Esta guía debe cumplir los siguientes mínimos:

La prueba de hipótesis para la diferencia de medias se utiliza cuando se desea comparar las medias de dos grupos (por ejemplo, dos tratamientos, dos poblaciones, dos métodos, etc.) para determinar si existe una diferencia significativa entre ellas.

Supuestos comunes en la prueba de la diferencia de medias:

- Las muestras son aleatorias.
- La variable sigue una distribución aproximadamente normal.
- Si las muestras son grandes (por ejemplo,$n \geq 30$), el Teorema Central del Límite asegura que la distribución muestral de la diferencia de medias se aproxima a una distribución normal.

Bibliografía

[1] Susan Miltón, J. (2201). Estadística para Biología y Ciencias de la Salud. 3ra Edición. McGRAW-HILL.

[2] Shahbaba. B. (2011). Biostatistics with R: An introduction to Statistics Through Biological Data. Springer: Baltimore.