Ejercicio 6 - Estadística Computacional

Byron Caices Lima

2023-11-29

Para los siguientes enunciados es importante que identifiquen los parámetros y/o estadísticos de interés, planteen las pruebas de hipótesis, además de determinar cual es el test indicado a utilizar y por que. Además debe de comentar y concluir respecto de los resultados que obtenga de sus tests. El ejercicio 5 y 6 debe de utilizar test NO parámetricos.

Problema 1

Una empresa de tecnología quiere saber si la duración promedio de la batería de su nuevo modelo de teléfono móvil supera las 24 horas. Se sabe que la duración promedio de la batería de modelos anteriores es de 24 horas, y la desviación estándar poblacional es de 2 horas. Se selecciona una muestra aleatoria de 40 teléfonos del nuevo modelo y se registra la duración de la batería. La muestra arroja una duración media de 25 horas. Utilizando un nivel de significancia del 5 %, ¿existe suficiente evidencia para afirmar que la duración media de la batería del nuevo modelo es superior a las 24 horas? Realiza un z-test para responder esta pregunta. Para trabajar con una muestra, puede generarla con rnorm utilizando los parámetros indicados en el enunciado.

Solución:

1. **Identificamos parámetros de interés:**
   - **Parámetro Poblacional:** Duracion promedio de la bateria μ
   - **Valor Hipotético:** 24 horas μ_0
2. **Definir hipótesis:**

$$H_0 : \mu = 24$$
$$H_1 : \mu > 24$$

3. **Determinar y aplicar la prueba estadística apropiada:**
   - Z test dado que tenemos la desv. estándar y la muestra de 40

```
set.seed(123)
datos_muestra <- rnorm(n=40,mean=25,sd=2)

resultado <- z.test(datos_muestra,
                    mu = 24,
                    sigma.x=2,
                    alternative="greater")
print(resultado)
```

```
##
##  One-sample z-Test
##
## data:  datos_muestra
## z = 3.448, p-value = 0.0002823
## alternative hypothesis: true mean is greater than 24
## 95 percent confidence interval:
##  24.57022       NA
## sample estimates:
## mean of x
##  25.09037
```

4. **Conclusión:**

La conclusión es que la duración promedio de la batería mayor a 24 horas ya que $p = 0.0002823 < 0.05$ entonces se rechaza H_0.

- El valor p es un indicador de la probabilidad de que el resultado observado se deba al azar. Un valor p menor a 0.05 indica que es muy poco probable que el resultado observado se deba al azar.
- El valor Z indica la distancia entre la media de la muestra y la media poblacional hipotética. Un valor Z significativo indica que la media de la muestra está lejos de la media poblacional hipotética.

Problema 2

Un restaurante introduce un nuevo plato y desea saber si el tiempo de preparación promedio de este plato es diferente del tiempo estándar de 30 minutos. El restaurante no tiene datos previos sobre la variabilidad del tiempo de preparación de este plato.Se toma una muestra de 25 preparaciones del nuevo plato, obteniendo un tiempo promedio de preparación de 32 minutos. Utilizando un nivel de significancia del 5 %, ¿existe suficiente evidencia para afirmar que el tiempo de preparación promedio del nuevo plato es diferente de 30 minutos?.

Solución:

1. **Identificamos parámetros de interés:**
    - **Parámetro Poblacional:** Tiempo promedio de preparación del plato μ
    - **Valor Hipotético:** 30 minutos μ_0
2. **Definir hipótesis:**

$$H_0 : \mu = 30$$
$$H_1 : \mu \neq 30$$

3. **Determinar y aplicar la prueba estadística apropiada:**
    - Se realizará un T-test, debido a que la varianza poblacional no se conoce y el tamaño de la muestra es 25

```r
set.seed(123)
tiempo_preparacion = rnorm(n=25,mean=32,sd=5)

resultado = t.test(tiempo_preparacion,
                   mu = 30,
                   alternative="two.sided")
print(resultado)
```

```
##
##  One Sample t-test
##
## data:  tiempo_preparacion
## t = 1.9365, df = 24, p-value = 0.06467
## alternative hypothesis: true mean is not equal to 30
## 95 percent confidence interval:
##  29.87939 33.78731
## sample estimates:
## mean of x
##  31.83335
```

4. **Conclusión:**

Como t es igual a 1.9365 con 24 grados de libertad, y el valor p resultante es 0.06467. Estos valores indican que la diferencia entre la media de la muestra y la media poblacional teórica no es estadísticamente significativa, dado que el valor p supera el umbral de 0.05. Esto implica que no existe una base estadística suficiente para afirmar que la duración promedio para preparar el plato sea distinta de 30 minutos.

Problema 3

Una empresa de fabricación de neumáticos desea evaluar si el proceso de producción de neumáticos nuevos está bajo control en términos de variabilidad. La empresa tiene un estándar de calidad que establece que la variabilidad en el espesor de los neumáticos no debe exceder una varianza de 4 mm². Se toma una muestra aleatoria de 30 neumáticos y se mide el espesor de cada uno, resultando en una varianza muestral de 5 mm². Utilizando un nivel de significancia del 5 %, ¿existe suficiente evidencia para afirmar que la variabilidad en el proceso de producción es mayor que el estándar establecido?.

Solución:

1. **Identificamos parámetros de interés:**
   - **Parámetro Poblacional:** Varianza en el espesor de los neumáticos σ^2
   - **Valor Hipotético:** $\sigma^2 = 4mm$
2. **Definir hipótesis:**

$$H_0 : \sigma^2 = 4mm^2$$
$$H_1 : \sigma^2 > 4mm^2$$

3. **Determinar y aplicar la prueba estadística apropiada:**

    - Se usara la prueba χ^2 ya que la varianza poblacional es desconocida y el tamaño de la muestra es 30

```
# Datos de la muestra

n = 30 # tamaño de la muestra
varianza_muestral = 5 # varianza muestral en mm^2

varianza_h0 = 4 # varianza estándar en mm^2

chi_cuadrado_estadistico = (n - 1) * varianza_muestral / varianza_h0

grados_libertad = n - 1

p_valor = pchisq(chi_cuadrado_estadistico, grados_libertad, lower.tail = FALSE)

list(estadistico = chi_cuadrado_estadistico, p_valor = p_valor)
```

```
## $estadistico
## [1] 36.25
##
## $p_valor
## [1] 0.1663945
```

4. **Conclusión:**

El valor chi cuadrado es 36.25, lo que indica que la variabilidad observada en la muestra no se alinea con lo que se esperaría si reflejara exactamente la variabilidad de la población en general. El valor p de aproximadamente 0.167 indica que no hay suficiente evidencia para rechazar la hipótesis nula, lo que en este caso específico significa que no se puede concluir que haya una diferencia significativa en la variabilidad del grosor de dos neumáticos que se está comparando con un valor especificado en la hipótesis (4 mm^2).

Problema 4

Una cadena de cines está interesada en aumentar la venta de palomitas de maíz. Un estudio previo mostró que aproximadamente el 40 % de los clientes compran palomitas. Tras implementar una nueva estrategia de marketing, la cadena desea saber si el porcentaje ha aumentado. Después de la implementación de la nueva estrategia, se selecciona una muestra aleatoria de 200 clientes, y se encuentra que 96 de ellos compraron palomitas de maíz. Utilizando un nivel de significancia del 5 %, ¿existe suficiente evidencia para afirmar que el porcentaje de clientes que compran palomitas ha aumentado?

Solución:

1. **Identificamos parámetros de interés:**

    - **Parámetro Poblacional:** Proporción de clientes que compran palomitas ρ

- **Valor Hipotético:** $\rho_0 = 0.40$

2. **Definir hipótesis:**

$$H_0 : \rho = 0.40$$
$$H_1 : \rho > 0.40$$

3. **Determinar y aplicar la prueba estadística apropiada:**

- Se realizará Z-test para proporciones binomiales

```r
set.seed(123)

n_clientes = 200
x_clientes = 96
p0_clientes=0.40

resultado = prop.test(x_clientes, n_clientes, p=p0_clientes,alternative = "greater")
print(resultado)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  x_clientes out of n_clientes, null probability p0_clientes
## X-squared = 5.0052, df = 1, p-value = 0.01264
## alternative hypothesis: true p is greater than 0.4
## 95 percent confidence interval:
##  0.4200931 1.0000000
## sample estimates:
##    p
## 0.48
```

4. **Conclusión:**

Tomando en cuenta los resultados obtenidos, donde el p-valor es de 0.01264 y teniendo un nivel de significancia del 5%, procedemos a rechazar la hipótesis nula. Esto nos permite concluir que existe suficiente respaldo estadístico para afirmar que la proporción de clientes que adquieren palomitas es mayor al 40%.

Problema 5

Una compañía de suplementos alimenticios está investigando el efecto de un nuevo producto en la mejora del rendimiento atlético. Se lleva a cabo un estudio con un grupo de deportistas para evaluar el impacto del suplemento. Un grupo de 15 deportistas participa en un experimento donde se mide su rendimiento en una prueba de resistencia. Se toman mediciones antes de iniciar el consumo del suplemento y después de cuatro semanas de uso continuo. Los investigadores están interesados en determinar si el consumo del suplemento ha tenido un efecto significativo en el rendimiento de los deportistas. Analiza los datos para proporcionar una respuesta a esta pregunta.

Número Atleta	Rendimiento Antes	Rendimiento Después
1	20.5	22.0
2	18.7	19.1
3	21.3	21.8
4	19.5	20.0
5	22.1	23.5
6	17.8	18.2
7	20.0	20.7
8	23.4	25.1
9	21.5	22.3
10	18.0	28.5
11	19.2	19.7
12	22.6	23.0
13	17.9	18.3
14	21.7	22.2
15	20.3	21.5

Solución:

Aspectos Clave para la Aplicación del Test de Wilcoxon en el Ejemplo Propuesto

1. **Datos Emparejados:** La información indica que se evaluó el mismo grupo de atletas antes y tras la ingesta del suplemento, lo cual implica un diseño de medidas repetidas.
2. **Comparación de medidas realcionadas:** Existe una evaluación comparativa del desempeño pre y post intervención (ingesta del suplemento), lo que sugiere un análisis bivariado dentro del mismo conjunto de participantes.
3. **Ausencia de suposición de normalidad:** Si bien no se especifica, el test de Wilcoxon se presenta como una opción robusta ante la falta de distribución normal de las diferencias, situación típica en muestrarios de tamaño reducido o en datos no paramétricos.
4. **Parámetros de Interés:**
   - **Parámetro de la Población:** Cambio en la performance de los atletas antes y después de tomar el suplemento (µ)
   - **Valor Teórico:** µ0 = 0
5. **Establecimiento de Hipótesis (no paramétricas):**
   - H_0, lo que indica que no hay variación en el desempeño de los atletas con la ingesta del suplemento.
   - H_1, sugiriendo que el suplemento sí tuvo un efecto en el rendimiento de los atletas.
6. **Selección y Ejecución del Test Estadístico Adecuado:**
   - Se optará por el test de Wilcoxon debido a la naturaleza emparejada de los datos y la inexistencia de una distribución normal.

```r
# Datos proporcionados
rendimiento_antes <- c(20.5, 18.7, 21.3, 19.5, 22.1, 17.8, 20.0, 23.4, 21.5, 18.0, 19.2, 22.6, 17.9, 21
rendimiento_despues <- c(22.0, 19.1, 21.8, 20.0, 23.5, 18.2, 20.7, 25.1, 22.3, 28.5, 19.7, 23.0, 18.3, 

# Realizar el test de Wilcoxon
test_wilcoxon <- wilcox.test(rendimiento_antes, rendimiento_despues, paired = TRUE)
```

```
## Warning in wilcox.test.default(rendimiento_antes, rendimiento_despues, paired =
## TRUE): cannot compute exact p-value with ties
```

```
# Mostrar los resultados del test
test_wilcoxon
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  rendimiento_antes and rendimiento_despues
## V = 0, p-value = 0.0007051
## alternative hypothesis: true location shift is not equal to 0
```

7. **Conclusión:** Un indicador estadístico V próximo a cero denota una consistencia en las diferencias a favor de una condición sobre la otra. Dado que el p-valor resultó ser menor, se procede a descartar la hipótesis nula, lo cual nos permite inferir estadísticamente que el cambio en la performance de los atletas es significativo tras la ingesta del suplemento, proporcionando fundamento suficiente para rechazar la hipótesis de no efecto.

Problema 6

Un restaurante popular ha decidido modificar la receta de su plato estrella para hacerlo más saludable, cambiando algunos ingredientes para reducir el contenido de grasa. Para evaluar la aceptación de la nueva receta, el restaurante realiza una encuesta con 20 de sus clientes habituales que han probado tanto la versión original como la nueva del plato. Cada cliente indica si prefiere la versión original o la nueva. El restaurante necesita saber si la nueva receta es tan buena como la original según la percepción de los clientes. Examina los resultados de la encuesta para determinar si hay una preferencia clara por alguna de las dos versiones. Sugerencia : puede señalar en un arreglo la preferencia por un plato u otro, como un vector de valores binarios, es decir, c(1,0,1,1,0,. . .) y puede hacerlo aleatorio.

Solución:

1. **Datos Binarios o de Categorías:** La selección de los clientes se divide entre la alternativa tradicional o la actualizada del menú, lo que nos lleva a trabajar con variables categóricas de dos niveles (binarios).

2. **Prueba de Preferencias o Elecciones:** Adecuado para evaluar la elección entre dos opciones, el test de signos se convierte en el instrumento idóneo para discernir la contabilidad de las preferencias distribuidas en dos categorías distintas, tal como ocurre en la decisión entre dos variantes de un mismo plato

3. **No se Analiza la Magnitud de la Diferencia:** El test de signos prioriza determinar la tendencia en las elecciones más que el tamaño del cambio, siendo este enfoque pertinente para la naturaleza de los datos presentados.

4. **Parámetro de interés**

    - Parámetro poblacional preferencia de los clientes por la versión nueva del platillo v/s la original.

5. **Definir Hipótesis**

- H_0 : La nueva versión del platillo no modifica la preferencia de la clientela respecto a la original.
- H_1: Existe una preferencia distinguible por parte de la clientela hacia la versión renovada del platillo en detrimento de la original.

6. **Determinar y aplicar la prueba estadística apropiada**

   - El método de los signos será el utilizado, considerando la binariedad de los datos y el interés por contrastar la predilección de los clientes entre la versión innovadora del platillo y la clásica.

7. **Conclusión**

```r
preferencias = c(1,0,1,1,0,0,1,1,1,0,1,0,1,1,0,1,1,0,1,0)
# Realizar la prueba de signos
resultado_signos = SIGN.test(preferencias, md = 0.5)

# Mostrar resultados
print(resultado_signos)
```

```
##
##  One-sample Sign-Test
##
## data:  preferencias
## s = 12, p-value = 0.5034
## alternative hypothesis: true median is not equal to 0.5
## 95 percent confidence interval:
##  0 1
## sample estimates:
## median of x
##           1
##
## Achieved and Interpolated Confidence Intervals:
##
##                   Conf.Level L.E.pt U.E.pt
## Lower Achieved CI     0.8847      0      1
## Interpolated CI       0.9500      0      1
## Upper Achieved CI     0.9586      0      1
```

Basándonos en el resultado de la prueba de signos obtenida, con un p-valor de 0.5034 y un intervalo de confianza del 95% que incluye el 1, no encontramos evidencia estadística suficiente para rechazar la hipótesis nula H_0. Esto sugiere que, con respecto a la preferencia expresada por los 20 clientes encuestados, no hay una diferencia significativa en la aceptación de la nueva versión del plato en comparación con la versión original. En otras palabras, la nueva receta, diseñada para ser más saludable al reducir el contenido de grasa, parece tener una recepción comparable a la versión original del plato según la percepción de los clientes. El restaurante podría interpretar esto como una señal positiva de que la modificación de la receta no ha afectado negativamente la preferencia de sus clientes y que la nueva versión es tan buena como la original en términos de aceptación.