

Pruebas estadísticas

TAREA 6

Alumno:

Joaquín Arturo Velarde Moreno

1. Introducción

Los objetivos de esta tarea son responder los cuestionamientos acerca del tema de pruebas estadísticas derivados de cuatro artículos diferentes dados por la Dra. Sara Verónica Rodríguez Sánchez y, además, aplicar una serie de pruebas estadísticas acerca del turismo internacional con el programa R [5].

2. Preguntas y respuestas

2.1. ¿Relación entre contraste de hipótesis y pruebas estadísticas?

Es el procedimiento para encontrar la veracidad de una hipótesis contrastándola con otra para lo cual es necesario realizar pruebas estadísticas y comprobar cuál de dichas pruebas es más factible.

2.2. ¿Qué indicaría rechazar la hipótesis nula?

Dicha hipótesis indicaría que nuestra prueba estadística nos dio un *valor-p* menor al nivel de significancia ($p < \alpha$), aceptando así la hipótesis alternativa, lo cual no significa necesariamente que la hipótesis alternativa sea correcta [2], pues podemos tener un error de tipo 1.

2.3. ¿Cómo se interpreta la salida de una prueba estadística?

La prueba estadística produce un número denominado *valor-p*, el cual tiene como límites 0 y 1; el *valor-p* es la probabilidad de obtener los datos bajo la hipótesis nula, esta se debe comparar con el nivel de significancia, y en caso de tener $p < \alpha$ rechazamos la hipótesis nula [2].

2.4. ¿Cómo seleccionar el alfa?

El valor alfa, también denominado nivel de significación, debe ser primeramente un valor de 0 a 1 y, generalmente, dicho valor se fija en 0.05, 0.01 o 0.001. La elección de alfa debe depender de cuán peligroso sea rechazar la hipótesis nula. Por ejemplo, en un estudio que se proponga demostrar los beneficios de un tratamiento médico, deberá tener un alfa bajo. Por otro lado, si tomamos en cuenta la apreciación de un producto, podemos ser más moderados [2]. Además, es importante tomar en cuenta que tan factible es realizar el experimento múltiples veces para no ser muy exigentes con él.

2.5. ¿Cuáles son los errores frecuentes de interpretación del valor p?

Debido a que la prueba estadística suele usar muestras de una población aleatoriamente escogida, puede ocasionar dos situaciones: que una hipótesis nula que es verdadera sea rechazada, denominado error tipo 1 o que una hipótesis nula falsa sea aceptada denominado error tipo 2.

2.6. ¿Qué es la potencia estadística y para qué sirve?

Es la capacidad de un experimento para conducir al rechazo de la hipótesis nula [2]. La potencia de un experimento aumenta con alfa, con lo preciso que son las mediciones y con el numero de repeticiones del experimento, equivale a 1 menos el riesgo de ser errónea cuando se acepta H_0 ($1 - \beta$), De este modo, mientras mayor sea la potencia, menor es el riesgo de equivocarse al aceptar H_0 .

2.7. Ejemplos de pruebas estadísticas paramétricas y no paramétricas.

1. Paramétricas:

- a) "t" de student,
- b) el coeficiente de correlación de Pearson,
- c) la regresión lineal,
- d) el análisis de varianza unidireccional (ANOVA Oneway),
- e) análisis de varianza factorial (ANOVA),
- f) análisis de covarianza (ANCOVA).

2. No paramétricas:

- a) chi squared test,
- b) coeficientes de correlación e independencia para tabulaciones cruzadas,
- c) coeficientes de correlación por rangos ordenados Spearman y Kendall,
- d) prueba de suma de rango Wilcoxon.

2.8. Resume la guía para encontrar la prueba estadística que buscas

Al escoger una prueba estadística debemos preguntarnos primeramente ¿cuál es nuestro objetivo? Asociar o comparar, aunque ambas establecen relaciones, la comparación evalúa estas relaciones entre uno o varios grupos. También hay analizar qué tipo de muestras tenemos, si es que son independientes o relacionadas, las relacionadas pueden ser tipo antes-después, como por ejemplo el estudio de pacientes donde se comparan antes y después de la aplicación de un tratamiento.

La siguiente pregunta que debemos contestar es ¿qué tipo de datos tenemos? ya sean variables cualitativas o cuantitativas, al tener estos debemos corroborar si nuestros datos cumplen o no con los supuestos (normalidad, homogeneidad, independencia), con esto podremos elegir entre pruebas paramétricas, no paramétricas y robustas [3].

Como paso final hay que analizar si podemos confiar en los resultados datos por nuestra prueba estadística *p-value* graficándolos para decidir.

2.9. ¿Cuáles son los supuestos para aplicar técnicas paramétricas?

Se requiere contar con la información de la distribución, además, tener eventos independientes, una muestra grande de números y seleccionar muestras al azar.

3. Pruebas estadísticas

A continuación veremos las mas comunes pruebas estadísticas [4] y su aplicación en R [5], junto con datos obtenidos de INEGI [1].

3.1. One sample t-Test.

Esta es una prueba estadística de tipo paramétricas y es usada para probar si la media de una muestra de una distribución normal puede ser un valor específico [4]. Como primer ejemplo utilizaremos el número de visitantes que ingresan al país por mes por la vía aérea (Figura 3.1), según los datos obtenidos del INEGI [1].

Lo expresaremos en R como:

```
t.test(VisitantesPorAvion, mu = 1540000)

data: VisitantesPorAvion
t = 0.076095, df = 11, p-value = 0.9407
alternative hypothesis: true mean is not equal to 1540000
```

En esta prueba estadística quisimos demostrar que la media de personas que entran al país por avión al mes es de 1,540,000 siendo esta nuestra hipótesis nula, sin embargo la prueba produjo un *p-value* mayor al nivel de significancia de 0.05, por lo que no rechazamos nuestra hipótesis nula.

3.2. Wilcoxon Signed Rank Test.

Esta prueba estadística es un método no paramétrico que evalúa la media de una muestra sin asumir que esta distribuida normalmente, este puede ser una alternativa al t-Test, especialmente cuando no se tiene información de la distribución en que sigue [4]. Usaremos como muestra el número de visitantes que ingresan al país por mes por la vía Terrestre el cual asumimos no tiene una distribución normal (Figura 3.3), según los datos obtenidos del INEGI [1]. Lo expresaremos en R como:

```
wilcox.test(VisitantesTerrestres, mu=640049, conf.int = TRUE)

data: VisitantesTerrestres
V = 40, p-value = 0.9697
alternative hypothesis: true location is not equal to 640049
```

En esta prueba estadística quisimos demostrar que la media de personas que entran al país por la vía terrestre al mes es de 640,049 siendo esta nuestra hipótesis nula, sin embargo la prueba produjo un *p-value* mayor al nivel de significancia de 0.05, por lo que no rechazamos nuestra hipótesis nula.

3.3. Two Sample t-Test and Wilcoxon Rank Sum Test.

Tanto t-Test como Wilcoxon rank pueden ser usados para comparar la media de 2 muestras, la diferencia como ya dijimos es que t-Test asume que la muestra sigue una distribución normal mientras que Wilcoxon rank no [4]. Usaremos nuestras 2 muestras anteriores que son el número de visitantes que ingresan al país por mes a través de la vía Terrestre y Aérea (Figura 3.3 y Figura 3.1), según los datos obtenidos del INEGI [1]. Lo expresaremos en R como:

```
wilcox.test(VisitantesTerrestres, VisitantesPorAvion, paired = TRUE)

data: VisitantesTerrestres and VisitantesPorAvion
V = 0, p-value = 0.0004883
alternative hypothesis: true location shift is not equal to 0
```

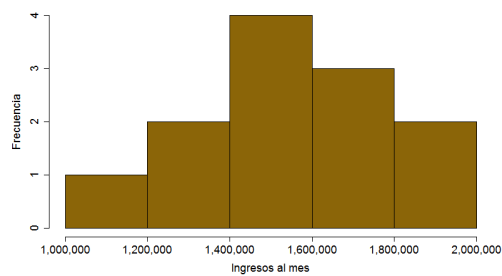
En esta prueba estadística quisimos demostrar nuestra hipótesis nula el cual es que la media de personas que entran al país por la vía terrestre al mes es la misma que el promedio de personas que entran por la vía aérea, sin embargo la prueba produjo un *p-value* menor al nivel de significancia de 0.05, por lo que rechazamos nuestra hipótesis nula.

3.4. Shapiro Test.

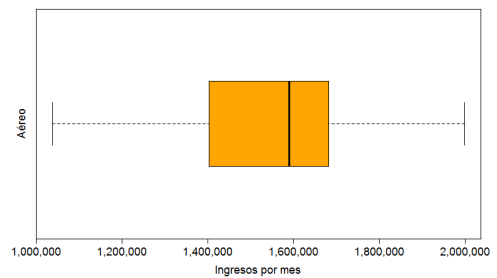
Esta prueba estadística evalúa si una muestra sigue una distribución normal [4]. Usaremos nuestra muestra de número de visitantes que ingresan al país por mes a través de la vía marítima (Figura 3.2), según los datos obtenidos del INEGI [1]. Lo expresaremos en R como:

```
shapiro.test(VisitantesMaritimos)
data: VisitantesMaritimos
W = 0.82518, p-value = 0.01838
```

En esta prueba estadística quisimos demostrar que nuestra población de personas que entran al país por la vía marítima al mes sigue una distribución normal, sin embargo la prueba produjo un *p-value* menor al nivel de significancia de 0.05 , por lo que rechazamos nuestra hipótesis nula.

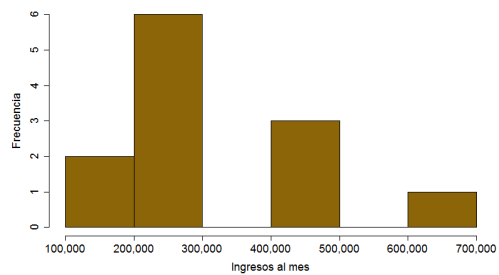


(a) Histograma que muestra la frecuencia de turistas ingresados al mes por la vía Aérea.

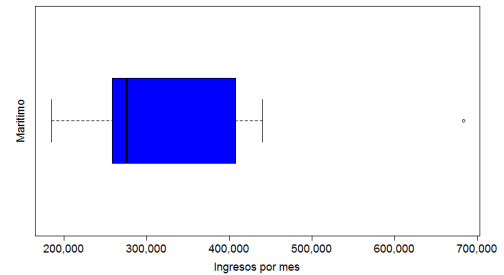


(b) Gráfico con información de la media de ingresos turísticos por la vía Aérea.

Figura 3.1: Gráficos que muestran el ingreso de turistas al país cada mes por medio de la vía Aérea.

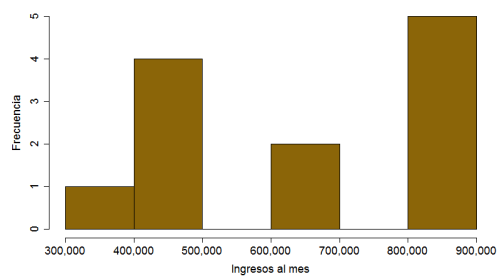


(a) Histograma que muestra la frecuencia de turistas ingresados al mes por la vía marítima.

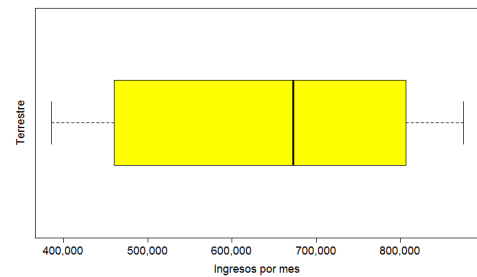


(b) Gráfico con información de la media de ingresos turísticos por la vía marítima.

Figura 3.2: Gráficos que muestran el ingreso de turistas al país cada mes por medio de la vía marítima.



(a) Histograma que muestra la frecuencia de turistas ingresados al mes por la vía terrestre.



(b) Gráfico con información de la media de ingresos turísticos por la vía terrestre.

Figura 3.3: Gráficos que muestran el ingreso de turistas al país cada mes por medio de la vía terrestre.

Referencias

- [1] INEGI. *Numero de visitantes internacionales que ingresaron al país por mes*. https://www.inegi.org.mx/contenidos/temas/economia/turismo/EVI_R_01.xlsx. 2020.
- [2] *Pruebas estadísticas*. Sitio en, <https://help.xlstat.com/s/article/que-es-una-prueba-estadistica?language=es>.
- [3] Rosana Ferrero. *Guía definitiva para encontrar la prueba estadística que buscas*. Sitio en, <https://www.maximaformacion.es/blog-dat/guia-para-encontrar-tu-prueba-estadistica/>.
- [4] Selva Prabhakaran. *Statistical Tests*. Sitio en, <http://r-statistics.co/Statistical-Tests-in-R>.
- [5] The R Foundation. *The R Project for Statistical Computing*. <https://www.r-project.org/>. 2019.