

¡Estadístic@s en acción!

**INFERENCIA ESTADÍSTICA
2024**

Damaris ALANIZ - Yoana Antonella CUELLO - María Emilia
FUNES - Selva Agustina MAICA - Baltazar MALDONADO -
Giuliana RAITERI EBERLE - Ana Paula TORRES - Lucas
Gabriel VILLAMIL

Compiladora
Maria Paula DIESER



FACULTAD DE CIENCIAS
EXACTAS Y NATURALES

Universidad Nacional de La Pampa

Índice

Prólogo	2
Análisis estadístico de la población infantil del asentamiento “El Salitral”, años 2021 y 2022 <i>Yoana Antonella Cuello y María Emilia Funes</i>	4
La peligrosidad de ser fumador pasivo a comparación de no estar expuesto al humo nocivo, por sus niveles de cotinina <i>Lucas Gabriel Villamil</i>	13
¿Disminuye la cantidad de accidentes aplicando el límite de velocidad? <i>Damaris Alaniz y Baltasar Maldonado</i>	19
Efectos de la terapia hormonal sobre el colesterol en mujeres premenopáusicas <i>Ana Paula Torres, Giuliana Raiteri Eberle y Selva Agustina Maica</i>	25

Prólogo

Hace ya algunos años, como respuesta a las demandas de nuevas competencias profesionales por parte de una sociedad basada en la economía de servicios, y debido a las investigaciones sobre la manera de enseñar y aprender de las personas, se piensa en una educación eficaz en la medida que ésta sea capaz de desarrollar habilidades de alto nivel que ayuden a los estudiantes a aprender a lo largo de toda su vida, *i.e.* una educación capaz de ofrecer a los ciudadanos un conocimiento sólido y a la vez flexible que pueda dar respuestas ajustadas a las situaciones cambiantes que se presentan (Barberà, 2005). La enseñanza de la estadística no ha quedado ajena a esta perspectiva de transformación, siendo objeto de un marcado interés en la comunidad educativa, en donde se habla de construir una cultura estadística. Gal (2002) se refiere a dos componentes interrelacionados: la capacidad para interpretar y evaluar críticamente la información estadística, y la capacidad para discutir o comunicar sus opiniones respecto a tal información. Franklin *et al.* (2005) indican que la enseñanza de la estadística debe, como principal objetivo, ayudar a los estudiantes a aprender los elementos básicos del pensamiento estadístico como la importancia de los datos, la ubicuidad de la variabilidad, así como su cuantificación y explicación. Según Wild & Pfannkuch (1999), el razonamiento estadístico es esencial para el aprendizaje e incluye cinco componentes fundamentales: reconocer la necesidad de los datos, la transnumeración, percibir la variabilidad, razonar con modelos estadísticos e integrar la estadística al contexto.

Asimismo, es necesario diferenciar entre conocer y ser capaz de aplicar un conocimiento. La habilidad para aplicar los conocimientos estadísticos requiere no sólo de conocimientos técnicos (construir un gráfico o calcular un promedio), sino también conocimientos estratégicos (saber cuándo usar un concepto o gráfico dado). Los problemas y ejercicios presentes en libros de texto sólo se concentran en conocimientos técnicos. Sin embargo, el trabajo con datos reales demanda utilizar conocimientos estratégicos.

En consecuencia, suponemos que la mejor forma de seguir estas recomendaciones es introducir, en las clases de estadística, problemas con datos reales, sean propuestos por el profesor o diseñados por los estudiantes. En este tipo de experiencias, se procura reemplazar la introducción de conceptos y técnicas descontextualizadas, o aplicadas a problemas tipo, difíciles de encontrar en la vida real, por una actividad integral donde se presenten y desarrollen las diferentes fases de una investigación estadística. Por estos motivos, desde 2011, proponemos a los estudiantes de “Inferencia estadística”¹ de las carreras de Profesorado y Licenciatura en Matemática de la Facultad de Ciencias Exactas y Naturales (UNLPam), la realización de proyectos con datos reales desarrollando las diferentes etapas de una investigación estadística, con los objetivos no sólo de promover el aprendizaje de la estadística a partir del trabajo con datos reales, fomentando el uso de la tecnología digital como herramientas necesarias en la resolución de este tipo de problemas, sino además estimular una mayor reflexión, compromiso y proactividad.

En este documento reunimos los trabajos de los estudiantes del ciclo lectivo 2024, quienes abordaron temáticas vinculadas con los efectos de las restricciones en los límites de velocidad, los

¹ Antes “Estadística” y “Probabilidad y Estadística II”

efectos de la exposición al humo del cigarrillo en fumadores pasivos, y de la terapia hormonal sobre el colesterol en mujeres premenopáusicas, así como un estudio poblacional del asentamiento “El Salitral” en la ciudad de Santa Rosa.

La compiladora, septiembre de 2024

REFERENCIAS

- Barberá, E. (2005). La evaluación de competencias complejas: la práctica del portafolio. *Educere*, 31(9), pp 497-504.
- Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2005). *Guidelines for assessment and instruction in statistics education (GAISE) report: A Pre-K-12 curriculum framework*. Alexandria, VA: American Statistical Association. Disponible en: www.amstat.org/Education/gaise/
- Gal, I (2002). Adult's statistical literacy. Meanings, components, responsibilities. *International Statistical Review*, 70(1), 1-25.
- Wild, C. & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3), 223-265.

ANÁLISIS ESTADÍSTICO DE LA POBLACIÓN INFANTIL DEL ASENTAMIENTO “EL SALITRAL”, AÑOS 2021 Y 2022

Yoana Antonella Cuello, María Emilia Funes
yoicuello.17@gmail.com , maremiliafunes@gmail.com

Resumen

En este trabajo se realizó un análisis estadístico descriptivo e inferencial de la población de personas menores de 18 años residentes en el asentamiento “El salitral” de Santa Rosa, durante 2021 y 2022. Se contó con 89 observaciones, las cuales se clasificaron por grupo familiar, dando un total de 34 grupos. Se establecieron tres objetivos y para cada uno de ellos se definió una variable. Luego concluimos que los integrantes del asentamiento se caracterizan por ser principalmente familias con pocos hijos, con una alta proporción de menores de hasta doce años y una población estable entre 2021 y 2022.

Introducción

La presente investigación trata sobre la población del asentamiento de lo que fue el barrio El Salitral durante los años 2021 y 2022. El barrio El Salitral surgió al este de la laguna Don Tomás en Santa Rosa, entre las décadas de 1930 y 1970. Fue habitado por peones, tanto indígenas como no indígenas, en condiciones precarias de salubridad. Estaba dividido en dos sectores: uno con viviendas más consolidadas al sur, ubicado –como todo el barrio– al oeste de las vías entre las calles Uruguay y Roca y otro más humilde hacia el norte, que comenzaba a partir de las actuales Avenida San Martín Oeste y calle 9 de Julio, y se extendía hacia el norte. Los residentes trabajaban en actividades informales, mientras que los niños a menudo colaboraban económicamente. Un plan de erradicación impulsado en la década de 1960 culminó con la desaparición del barrio el 24 de marzo de 1976.

La Defensoría del Pueblo denomina “Nuevos Asentamientos Urbanos” (NAUs) como procesos informales de ocupación de terrenos, ya sean privados o públicos, por familias de escasos recursos. Por lo general, estos asentamientos suelen ubicarse en zonas con características topográficas desfavorables como terraplenes del ferrocarril, autopistas, fábricas abandonadas y basurales. El término “nuevos” destaca su aparición posterior, especialmente a partir de la crisis económica del 2001, cuando un aumento de la desocupación y precarización laboral obligó a que muchas personas se vieran imposibilitadas de cumplir con el pago de sus alquileres.

Como parte de nuestra investigación, nos parece necesario definir el concepto de familia numerosa. Para ello, extraemos del capítulo VII el artículo 20 de la ley de asignaciones familiares: *ARTÍCULO 20. La asignación por familia numerosa se abonará al agente que tenga a cargo por lo menos tres hijos menores de veintiún años o mayores de esa edad que estén incapacitados. [...].

Los objetivos de este trabajo son:

- Realizar una descripción de las características de las personas menores de 18 años del asentamiento El Salitral.
- Analizar cómo se distribuyen las edades de los menores de edad asentados en el relleno sanitario.

- Determinar si la cantidad de niños o jóvenes por familia sigue una distribución de Poisson.
- Analizar cómo varía la población de menores de 18 años del relleno sanitario de Santa Rosa entre los años 2021 y 2022.

Metodología

Los datos fueron proporcionados por colaboradores del merendero “El Amanecer”, ubicado en el centro del asentamiento, a partir de relevamientos que realizaron en los años 2021 y 2022. Éstos constituyeron un total de 89 niños y adolescentes. La información obtenida fue plasmada en una hoja de cálculo para su posterior análisis y consideramos las siguientes variables:

F_i ="Cantidad de menores de 18 años en la familia i ", $i=1,2,\dots,34$, de tipo cuantitativa discreta

E =Edad de un menor de 18 años (registrada en el año 2022), en años, de tipo cuantitativa continua

R_1 =Residencia en asentamiento en el año 2021 , de tipo cualitativa nominal, que se categoriza (sí, no)

R_2 =Residencia en asentamiento en el año 2022 , de tipo cualitativa nominal, que se categoriza (sí, no)

Es importante aclarar que sólo se registra información de menores de 18 años (debimos adaptar nuestra investigación en base a lo proporcionado).

Comenzamos realizando un análisis descriptivo e inferencial para conocer si la distribución que moldea a la variable E es normal. Para esto, calculamos las medidas resúmenes y realizamos distintas pruebas gráficas, Boxplot, histograma y QQ-Plot. Posteriormente analizamos si la variable cumplía con las reglas empíricas. A partir del análisis transformamos la variable.

Luego, realizamos un análisis descriptivo de F_i mediante una tabla de frecuencias, Boxplot y un gráfico de barras. Para el análisis inferencial, elaboramos una prueba de bondad de ajuste Chi-cuadrado (Dieser, 2023), con un nivel de significancia del 10%, utilizando el método de máxima verosimilitud para estimar sus parámetros.

Por último, describimos las variables R_1 y R_2 mediante una tabla de contingencia, y trabajamos con la prueba de McNemar ya que teníamos datos binomiales pareados.

El análisis se realizó a partir de los resultados numéricos y gráficos del software R Commander (versión 4.3.1) (R Core Team, 2023) y GeoGebra (Versión Classic 5.0).

Resultados y Discusión

A continuación se detallan los análisis descriptivos e inferenciales de las variables mencionadas anteriormente.

Variable a trabajar:

E =Edad de un menor de 18 años (registrada en el año 2022)

En la siguiente tabla se encuentran las medidas resumen, y para analizarlas mejor se elaboró un Box-plot:

n	$\bar{x}$	s	Varianza	Q_1	Mediana	Q_2
89	9.97	3.75	14.08	7	10	12

Tabla 1: Medidas de resumen para la variable en interés.

El cuadro presenta un resumen estadístico para la variable E , basado en una muestra de 89 observaciones. La media (estimada a partir de $\bar{x}$) es 9.97 y su desvío (estimado por s) es 3.75. Además, los cuartiles uno, dos y tres son, respectivamente, 7, 10, y 12.

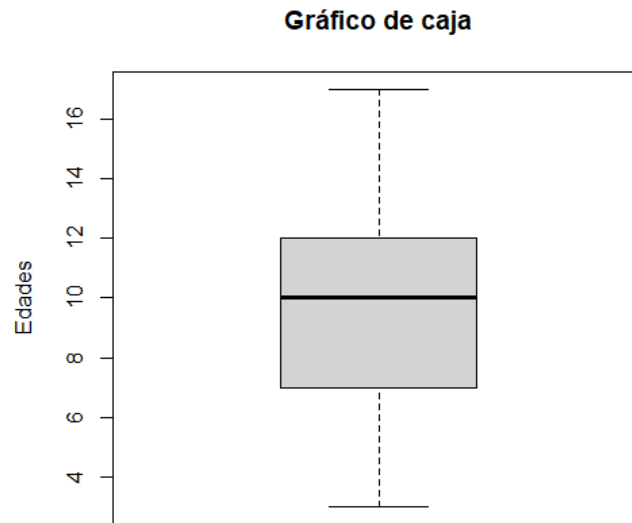


Figura 1: Box-plot de E .

Como se puede observar, la mediana se desvía un poco hacia los valores superiores, por lo que no hay simetría, sino que notamos que hay una asimetría negativa. Además los datos se desvían equitativamente hacia los valores superiores e inferiores y no hay puntos extremos.

Para saber si la distribución de E es normal, decidimos realizar un diagnóstico de normalidad utilizando el software R .

En el siguiente gráfico se observa el histograma generado y la curva normal superpuesta obtenida a partir de la media y la varianza muestrales.

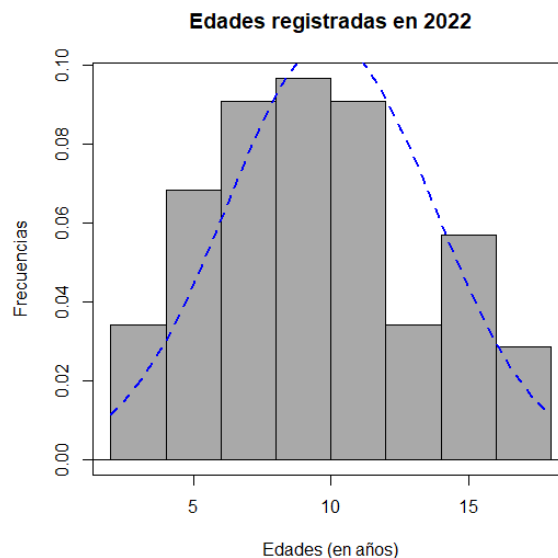


Figura 2: Histograma de E .

Nuestro análisis del histograma es que se asemeja a una distribución normal, pero está sesgada a la derecha. Para tener una mayor evidencia con la cual decidir, hicimos un gráfico qq-plot. Éste nos muestra que la mayor parte de los datos se aproximan a una recta, pero hay varios que se desvían de ella, sobre todo cuando las edades son mayores o iguales a 14.

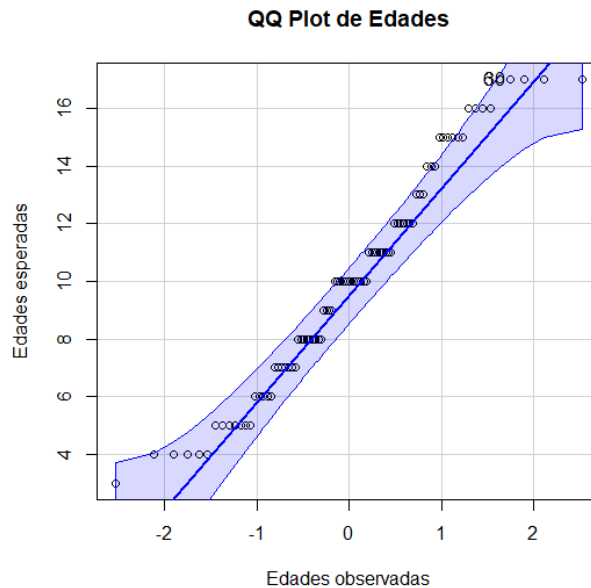


Figura 3: QQ-Plot de los datos de las Edades.

Utilizando el software *GeoGebra*, hicimos una tabla de frecuencias con el objetivo de analizar si la variable sigue la regla empírica.

Intervalo	Cuenta
$-1.23 - 2.5$	0
$2.5 - 6.24$	18
$6.24 - 13.7$	53
$13.7 - 17.43$	18
$17.43 - 21.16$	0

Tabla 2: Tabla de frecuencias para la regla empírica de E .

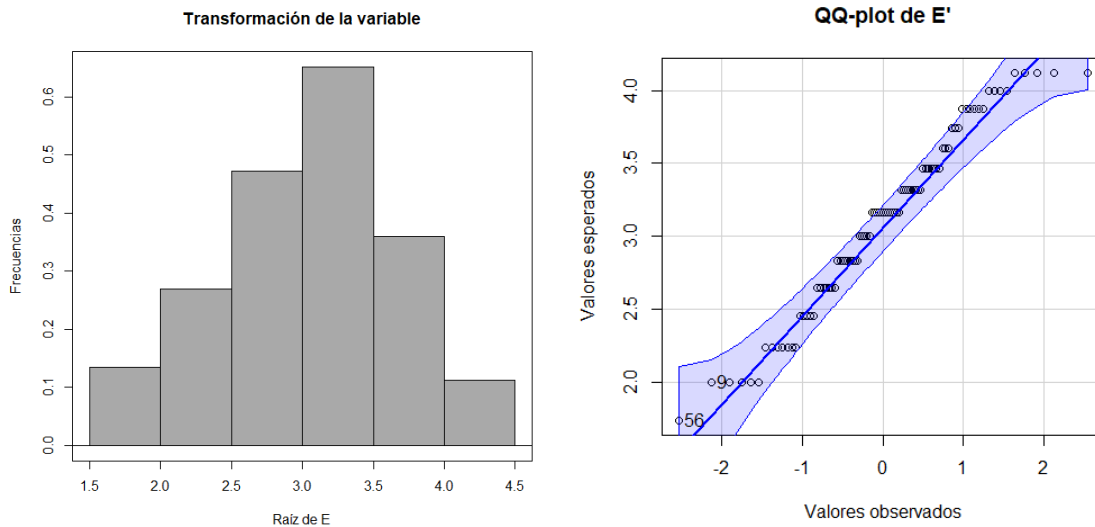
Los resultados fueron:

- En el intervalo $(\bar{x} - s; \bar{x} + s) = (6.24; 13.7)$ se encuentra el 47.17% de los datos;
- En el intervalo $(\bar{x} - 2s; \bar{x} + 2s) = (2.5; 17.43)$ se encuentra el 100% de los datos
- En el intervalo $(\bar{x} - 3s; \bar{x} + 3s) = (-1.23; 21.16)$ se encuentra el 100% de los datos

Esto nos confirma que la variable no sigue una distribución normal, dado que lo esperado es que aproximadamente el 68% de los datos se encuentre en el primer intervalo, el 95% en el segundo y el 99% en el tercer intervalo.

Luego de la realización y análisis de los gráficos y de la regla empírica, concluimos que la variable no sigue una distribución normal. Para lograr que siga esta distribución, transformamos la variable que, al ser sesgada a la derecha, nos permite calcular una nueva definida como $E' = \sqrt{E}$.

A continuación adjuntamos el histograma y el qq-plot generados por el software R, donde es posible observar que la distribución de la nueva variable se aproxima a una normal:



Figuras 4 (izquierda): Histograma. **Figura 5 (derecha):** QQ-plot de E' .

Variable a trabajar:

$$F_i = \text{"Cantidad de menores de 18 años en la familia } i\text{"}, i = 1, 2, \dots, 34$$

Analizando los datos observamos que la F_i toma valores enteros positivos, con lo cual es de tipo cuantitativa discreta. Además la muestra es independiente ya que la presencia de un menor en una familia no afecta a la cantidad de menores de otra.

Por medio de una tabla de frecuencias hicimos un recuento de la cantidad de menores de 18 años que hay por familia, y esto fue lo que obtuvimos:

<i>cant menores</i>	Frecuencias	Porcentaje
1	8	23.53
2	12	35.29
3	5	14.71
4	5	14.71
5	2	5.88
6	2	5.88
Suma total	34	100

Tabla 3: Tabla de frecuencias para F_i .

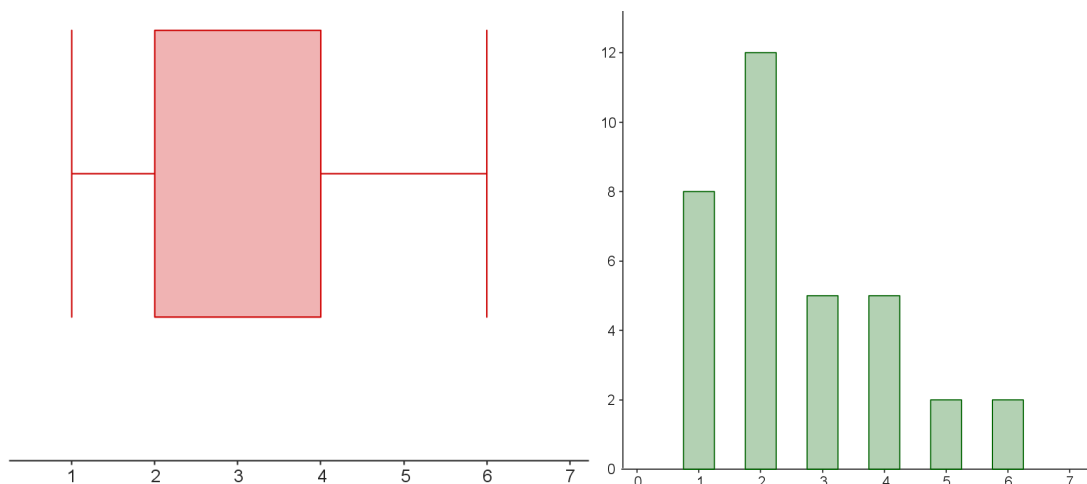
En ella se observa que las familias que se consideran numerosas representan el 41.18%, mientras que aquellas con uno o dos hijos/as constituyen el 58.82%.

A partir de los datos completamos la siguiente tabla de las medidas resumen:

n	$\bar{x}$	s	Varianza	$Q1$	Mediana	$Q3$
34	2.62	1.46	2.12	2	2	4

Tabla 4: Tabla de medidas de resumen para F_i .

El cuadro presenta un resumen estadístico para la variable F , basado en una muestra de 34 observaciones. La media (estimada a partir de $\bar{x}$) es 2.62 y su desvío (estimado por s) es 1.46. Además, los cuartiles uno, dos y tres son, respectivamente, 2, 2, y 4. Luego, elaboramos un diagrama de caja y un gráfico de barras.



Figuras 6 (izquierda): Diagrama de caja. **Figura 7 (derecha):** Gráfico de barras de F_i .

Tanto en el Box-plot como en el gráfico de barras no se observa simetría. En la figura 6 encontramos que los datos son más variables en los valores superiores que en los inferiores. Sin embargo, no hay puntos extremos. La figura 7 da cuenta de que en $F = 2$ la cantidad de datos es mucho mayor con respecto a los demás valores.

Teniendo en cuenta las características mencionadas, creemos que la variable sigue una distribución de Poisson, ya que nos parece que es ideal para describir la ocurrencia de eventos en un intervalo fijo de tiempo o espacio con una tasa constante. Además de que ésta modela eventos discretos e independientes como, en este caso, la cantidad de menores en una familia.

Para probar si la variable sigue una distribución de Poisson vamos a hacer un análisis de bondad de ajuste, el cual se puede realizar ya que contamos con una muestra suficientemente grande: esto asegura que las frecuencias esperadas en cada categoría sean lo suficientemente grandes para cumplir con la aproximación asintótica.

Planteamos las hipótesis:

$$H_0: F \sim P(\lambda) \quad \text{vs} \quad H_a: F \not\sim P(\lambda).$$

Estimamos a partir de GG la media, utilizando a $\bar{x}$ como estimador, el cual fue estimado por el método de máxima verosimilitud, ya que nos servirá como parámetro λ de la distribución, recordando que en una distribución de Poisson la esperanza y la varianza son iguales.

$$\hat{\lambda} = \bar{x} = \frac{89}{34} \cong 2.62$$

Realizamos una tabla de contingencia:

Clases	Frecuencias Observadas	Frecuencias Esperadas
1	8	8.96
2	12	8.50
3	5	7.42
4	5	4.86
5	2	2.55
6	2	1.71

Tabla 5: Tabla de Frecuencias observadas y esperadas de F_i

Ya que las frecuencias esperadas son muy pequeñas, se corre el riesgo de que el test no tenga suficiente poder. Como las frecuencias esperadas tienen que ser mayores o iguales a 5 entonces la tabla nos queda:

Clases	Frecuencias Observadas	Frecuencias Esperadas
1	8	8.96
2	12	8.50
3	5	7.42
4 o más	9	9.12

Tabla 6: Tabla de Frecuencias observadas y esperadas de F_i con clases agrupadas

Con estos datos, realizamos un test de bondad de ajuste en Geogebra, y obtuvimos el estadístico de prueba: $\chi^2 = 2.33$

Además, los grados de libertad son $glib = k - m - 1 = 4 - 1 - 1 = 2$, dado que tenemos 4 clases y un solo parámetro estimado. Luego, calculamos el valor p:

$$P(\chi^2_{(2)} \geq 2.33) = 0.31$$

En conclusión, tanto para un nivel de significancia del 5%, como para 1% y 10%, $p \geq \alpha$. Por lo tanto, no hay suficiente evidencia para rechazar la hipótesis nula, y podemos afirmar que $F \sim P(2.62)$.

Variables a trabajar:

$R_1 = \text{Residencia en asentamiento en el año 2021}$

$R_2 = \text{Residencia en asentamiento en el año 2022}$

Definiendo la variable $R = R_2 - R_1$, realizamos una tabla de contingencias para describir los movimientos poblacionales a analizar. Observemos que esto es posible porque tenemos dos muestras de datos apareados.

AÑOS	2022		
2021		si	no
	si	43	21
	no	25	0

Tabla 7: Tabla de contingencias para R .

Luego, establecimos las siguientes hipótesis:

$$H_0: R_2 - R_1 \leq 0 \quad \text{vs} \quad H_a: R_2 - R_1 > 0$$

Realizamos una prueba de McNemar en R commander para evaluar si existe un cambio significativo en proporciones entre dos condiciones (en nuestro caso, el cambio poblacional del asentamiento entre 2021 y 2022). Elegimos esta prueba porque contamos con datos emparejados binomiales en una tabla de contingencia 2x2. Obtuvimos lo siguiente:

χ^2	Glib	Valor p
0.19565	1	0.6583

Tabla 8: Resultados de la prueba McNemar para R .

En conclusión, tanto para un nivel de significancia del 5%, como para el 1% y el 10%, $p \geq \alpha$. Esto indica que la probabilidad de equivocarnos al rechazar la hipótesis nula sería muy alta. Por lo tanto, no hay suficiente evidencia estadística para rechazar la hipótesis nula, lo que implica que no podemos descartar que $R_2 - R_1 \leq 0$ sea consistente con los datos observados.

Conclusiones

Luego del análisis descriptivo e inferencial, se concluye que, con respecto al primer objetivo, las edades de los menores residentes en el asentamiento se distribuyen asimétricamente, determinando que hay una mayor concentración de niños y niñas de hasta doce años.

Con respecto al segundo objetivo, observamos que a menor cantidad de hijos, mayor es la cantidad de familias con ese número de menores, con lo cual la cantidad de familias numerosas presentes en los años registrados constituye menos de la mitad.

En cuanto al tercer objetivo, concluimos que la población trabajada no presentó un crecimiento entre los años 2021 y 2022.

Podemos concluir que los integrantes del asentamiento se caracterizan por ser principalmente familias con pocos hijos, con una alta proporción de menores de hasta doce años y una población estable entre 2021 y 2022.

Este trabajo fue el primer paso de una investigación que puede ser más profunda y abarcativa. Todavía restan preguntas por responder, las cuales invitan a seguir indagando: ¿Actualmente sigue habiendo una mayor proporción de niños y niñas de hasta doce años? Con respecto a las familias numerosas, ¿hubo un aumento o disminución? En cuanto a la escolaridad, ¿los jóvenes que residen en el lugar llegan a tener un estudio secundario completo? Si no es así, sería interesante investigar por qué no lo logran.

Referencias

- Argentina.gob.ar. (s. f.). Argentina.gob.ar. Recuperado el 30 de octubre de 2024, de <https://www.argentina.gob.ar/normativa/provincial/ley-9290-123456789-0abc-defg-020-0000svorpyel/actualizacion>
- Rodríguez, M. F. (2009). Notas sobre los conceptos de los “Nuevos Asentamientos Urbanos” (NAUs) en la ciudad de Buenos Aires. PAMPA. Extraído del 30 de octubre de 2024, de <https://doi.org/10.14409/PAMPA.V1I5.3170>
- Tarquini, C. S. (s. f.). *El Salitral: un barrio central en la historia de Santa Rosa*. Gob.Ar. Extraído del 30 de octubre de 2024 de https://historia.santarosa.gob.ar/wp-content/uploads/2021/12/pdfs/notas_breves/nota-claudia-salomc3b3n-tarquini.pdf
- Dieser, MP. (2022). Inferencia estadística. Nota de Clases . [material no publicado]. Facultad de Ciencias Exactas y Naturales. Universidad Nacional de La Pampa. [Junio de 2022].
- GeoGebra (2022). GeoGebra para enseñar y aprender Matemáticas. [software]. Disponible en: <https://www.geogebra.org/>
- Szczepanek, A. (10 de septiembre). *McNemar's Test Calculator - Contingency table*. Omni Calculator. Recuperado el 30 de octubre de 2024, de <https://www.omnicalculator.com/statistics/mcnemars-test>

LA PELIGROSIDAD DE SER FUMADOR PASIVO A COMPARACIÓN DE NO ESTAR EXPUESTO AL HUMO NOCIVO, POR SUS NIVELES DE COTININA

Lucas Gabriel Villamil
lucasgvillamil@gmail.com

Resumen

La cotinina es el principal metabolito de la nicotina, y por lo tanto, medirlo es un muy buen método para controlar el consumo y exposición ambiental al humo de tabaco. A través de este informe se dará un análisis de la presencia de este biomarcador en el organismo de dos poblaciones: fumadores pasivos y personas no expuestas al humo de tabaco ambiental. Se realizó un estudio descriptivo e inferencial para determinar la peligrosidad de ser un fumador pasivo al comparar la cantidad de muestras de cotinina superiores a 9,2 ng/ml (valor del límite de detección que permiten diferenciar entre los fumadores y los no fumadores) en los grupos previamente mencionados.

Introducción

El consumo de tabaco es la principal causa de muerte evitable en el mundo y es responsable de más de 8 millones de muertes al año, es una problemática que azota a la población desde tiempos remotos y que, a pesar de que en la actualidad hay más información y mayor concienciación sobre los múltiples problemas que ocasiona, se calcula que el tabaquismo se cobra del 10% al 12% de las muertes por enfermedades cardiovasculares a nivel mundial. Según la Organización Panamericana de la Salud (OPS/OMS) (2018).

Existen los fumadores activos que son los que consumen cigarrillos por elección propia, fumadores pasivos que son las víctimas del humo nocivo. Según el sitio web Salud180 (2013) el humo que inhalan los fumadores pasivos contiene hasta tres veces más nicotina y alquitrán que el que aspira el fumador, y unas cinco veces más monóxido de carbono.

Con todo esto en mente es importante destacar que el tabaco no solamente causa enfermedades cardiovasculares, ni tampoco es como se cree que afecta “recién en unos años”, el cigarrillo puede causar cáncer de: vejiga, sangre (leucemia mielógena aguda), cuello uterino, colon, recto, esófago, riñón, pelvis renal, hígado, pulmones, bronquios y tráquea, boca y garganta, páncreas, estómago, laringe y eso solo hablando de cáncer. Según afirma el sitio web CDC (2023). A corto plazo puede generar arrugas prematuras en la zona del labio superior, alrededor de los ojos (patas de gallo), barbilla y mejillas, coloración grisácea de la piel que constituye el llamado “rostro del fumador”, manchas en los dientes, infecciones, caries dentales, manchas amarillentas en uñas y dedos, además de aumentar la posibilidad de engriparse y de reducir el potencial deportivo, según el ministerio de sanidad del gobierno de España (2020)

Para poder marcar y seguir la recuperación de una persona a la adicción al tabaco se ha demostrado que la cotinina es un buen biomarcador para estudiar el consumo de tabaco y la exposición al humo ambiental del tabaco. La cotinina es el principal metabolito de la nicotina y puede medirse en el pelo y en diferentes fluidos corporales, como la saliva, la orina y el suero. La vida media de la

nicotina en los fluidos corporales es de dos a tres horas antes de metabolizarse a cotinina, que es mucho más estable y tiene una vida media de 15 a 17hs. La cotinina es un marcador específico del tabaco, fácilmente detectable, y mantiene una razón constante con otros productos del tabaco. Aunque el mejor indicador de la dosis absorbida de nicotina es la concentración de cotinina en la sangre, ésta puede estimarse razonablemente bien a partir de los valores de cotinina en la saliva y la orina (Sociedad Española de Salud Pública y Administración Sanitaria (2009)).

Esta sociedad realizó un estudio midiendo varios parámetros y tomando en cuenta la regularidad del vicio entre los sujetos de estudio, también consideró el género de los mismos para establecer que el punto de corte óptimo que discrimina entre fumadores (diarios y ocasionales) y no fumadores fue de 9,2ng/ml.



Imagen 1: persona fumando y el gran humo tóxico que suelta

Podemos coincidir en que todos estamos o estuvimos en presencia del humo del tabaco, tanto en el hogar, trabajo, vía pública o en algún bar o lugar de recreación, según un estudio realizado por la Cámara Argentina de especialidades médicas en 2018 sólo en nuestro país se encuentran 9 millones de fumadores, eso nos da un porcentaje de 22,2% de fumadores en nuestro suelo. Siendo una cifra tan alta es muy probable que todos tengamos contacto con este humo más de una vez al día. Por lo tanto mi objeto de estudio es medir la peligrosidad de ser un fumador pasivo, tomando los datos dados y comparándolos con 9,2 ng/ml para saber que tasa de fumadores pasivos cumple con dicho punto de corte, y así probar que ser un fumador pasivo o de segunda mano no te exenta en absoluto de los problemas derivados por el tabaco.

Años atrás se había hecho un trabajo bien logrado sobre este tema pero tomando un punto de corte diferente, mucho más bajo, mi trabajo difiere en la fuente para situar dicho punto, los estudios usados para sustentar todo mi análisis son hechos por la Sociedad Española de Salud Pública y Administración Sanitaria, un ente respetable y que en el informe enunció múltiples referencias avalando lo que se muestra. Tal vez quede como trabajo próximo para algunos años en adelante revisar el punto de corte tomado en este informe y corroborar su certeza, o comparar la proporción hallada al final con la de existente en un futuro para saber qué tanto cambió todo con el tiempo.

Metodología

La tercera Encuesta Nacional de Salud y Nutrición (NHANES III), perteneciente al Department of Health and Human Services de EEUU, entre 1988 y 1994, registró información a nivel nacional del

estado de salud y nutrición de la población estadounidense. Con estos datos se logró registrar los niveles de cotinina sérica (en ng/ml) en fumadores activos (FUMADORES), en sujetos que no fuman pero están expuestos al humo de tabaco ambiental (fumadores pasivos) en su casa o trabajo (NFHTA) y sujetos que no fuman y que no se exponen al humo de tabaco ambiental (NFSHTA).

En base al anterior estudio a esas poblaciones se seleccionaron una serie de datos, basándose en tres muestras aleatorias simples independientes: 40 individuos FUMADORES, 40 sujetos NFHTA y 40 individuos NFSHTA. En todos los casos se registró la cantidad de cotinina (en ng/ml) que se hallaba en su organismo. Yo me voy a fijar más bien en los 40 sujetos NFHTA y los 40 sujetos NFSHTA por mi particular objetivo de estudio.

También debo notar que la Sociedad Española de Salud Pública y Administración Sanitaria en 2009 realizó un elaborado informe recaudando datos de otros estudios para poder fijar el punto de corte acerca de cómo distinguir un fumador de uno que no sobre el que apoyaré mi análisis.

Resultados

Siendo que mi objetivo es probar la peligrosidad del humo del tabaco, y el riesgo que corre todo fumador pasivo, deberé encontrar la forma de comparar los conjuntos de datos dados con 9,2 ng/ml que como bien se explicó en la introducción es un punto de corte bueno para dividir fumadores de no fumadores. Entonces:

X = nivel de cotinina superior a 9,2 ng/ml en un no fumador y no expuesto al humo

Y = nivel de cotinina superior a 9,2 ng/ml en un fumador pasivo

Voy a calcular algunas características de los grupos que a mi me interesa estudiar

Medidas	Cotinina en no fumadores (ng/ml)	Cotinina en fumadores pasivos (ng/ml)
Cantidad de datos	40	40
Mínimo	0	0
Máximo	309	551
Primer cuartil	0	1
Mediana	0	1,5
Tercer cuartil	0	32
Media	16,35	60,57
Desviación estándar	62,53	138,08

Tabla 1: medidas de resumen y otras de las poblaciones estudiadas

Para mayor visualización sobre nuestros grupos de estudio elabore un diagrama de caja de ambos:

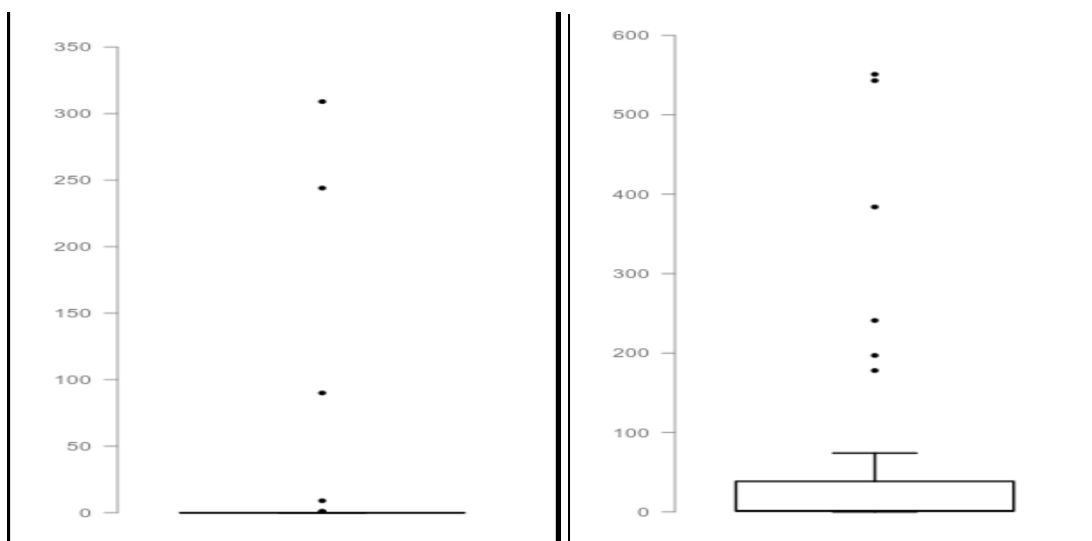


Gráfico 1:Box plot de cotinina de NFSHTA

Gráfico 2: Box plot de cotinina de NFHTA

Teniendo en cuenta todo esto, se buscó la cantidad de muestras de cotinina que superan a 9,2 ng/ml en ambos sectores, con el grupo de fumadores pasivos se encontró que trece muestras cumplieron dicha condición por la cual se tiene una proporción de 13/40, lo que sería un porcentaje de 32,5%. Casi un tercio de los fumadores pasivos de esta población analizada se lo puede considerar fumador activo, pero para poder decir si este porcentaje es alto o bajo lo compararé con el porcentaje de muestras superiores a 9,2 ng/ml en el grupo de no expuestos al humo. Este recuento arrojó que de 40 muestras solo 3 lograron superar el punto de corte en la población NFSHTA, es decir un porcentaje de 7,5%.

Con la prueba de homogeneidad veremos claramente cómo es su distribución bajo estos parámetros, con la cual tenemos por hipótesis nula es que los grupos de fumadores pasivos y no fumadores están distribuidos equitativamente respecto a los niveles de cotinina, y nuestra hipótesis alterna es que no lo están.

	<9.2 ng/ml	≥9.2 ng/ml	Total
Fumadores pasivos	27 (32)	13 (8)	40
No fumadores	37 (32)	3 (8)	40
Total	64	16	80

Tabla 2; tabla de contingencia, valores menores o mayor igual a 9,2 ng/ml

A través de la expresión “ $f_{ei} = (\text{total fila } i \times \text{total columna } i) / \text{total}$ ” obtenemos los valores esperados y los colocamos en rojo, mientras que en negro aparecen los reales. Luego buscamos nuestro estadístico de prueba con la expresión:

$$\chi_p^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

Imagen 2: fórmula para encontrar estadístico de prueba

Esto nos da como resultado 7,8125, teniendo $\alpha=0,05$ con $gl=(c-1)(r-1)=(1)(1)=1$. Buscando en la tabla de chi-cuadrado cuál es nuestro punto crítico obtenemos 3,841. Con lo que se concluye que

nuestra diferencia es significativa, por lo que se debe rechazar la hipótesis nula, dando paso a dar la alternativa como cierta, es decir los grupos analizados no están distribuidos homogéneamente respecto a los niveles de cotinina, con $\alpha=0,05$.

En base a los datos recaudados se confeccionó un nuevo gráfico que las comparo:

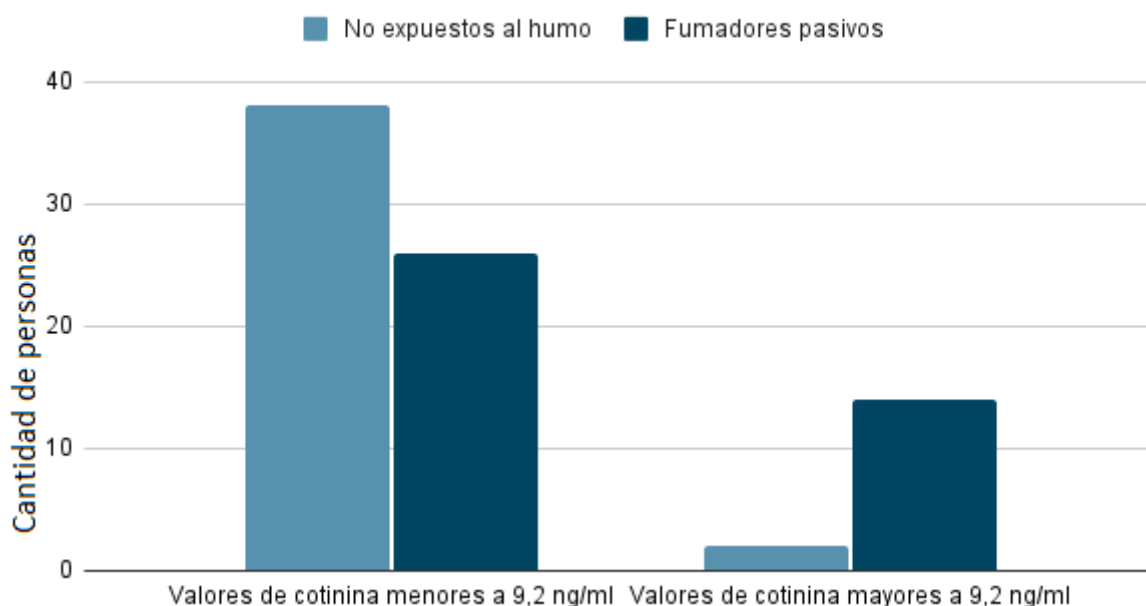


Grafico 3: grafico de barras comparativo de las variables presentadas

Conclusión

El objetivo de este informe fue concientizar y poder a establecer la peligrosidad de estar cerca del humo nocivo de un cigarro, para esto tuve que comparar 40 muestras de fumadores pasivos y 40 de no fumadores con 9,2 ng/ml, también conté con diferentes gráficos y análisis para argumentar mis pasos. Luego de usar las herramientas de la inferencia estadística logre establecer la proporción buscada como 4,333. Es interesante dejar para cerrar que no es necesario ser un fumador activo para morir por alguna de todas las enfermedades enumeradas al principio, y que cada vez que compartimos un ambiente lleno de humo nos exponemos 4 veces más a padecer alguno de los problemas que este vicio origina que las personas con un entorno limpio de tabaco.

Yo fui el primero en concienciarme al hacer este trabajo, y también es un buen empujón para dejar de fumar.

Referencias

CDC (24 de febrero de 2016), National Health and Nutrition Examination Survey. [en línea]. Disponible en: <http://www.cdc.gov/nchs/nhanes/index.htm>

Scielo (2009) sociedad Española de Salud Pública y Administración Sanitaria. Disponible en https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S0213-91112009000600003

OPS (2018) Organización Panamericana de la Salud (OPS/OMS)

<https://www.paho.org/es/temas/enfermedades-cardiovasculares>

CDC (2023) Cancer <https://www.cdc.gov/cancer/es/risk-factors/tobacco.html>

Plan nacional sobre Drogas (2020) Ministerio de sanidad del gobierno de españa
<https://pnsd.sanidad.gob.es/ciudadanos/informacion/tabaco/menuTabaco/efectos.htm>
Argentina.gob.ar (2018) Cámara Argentina de especialidades médicas
<https://www.argentina.gob.ar/salud/consumo-de-tabaco/preguntas-frecuentes#:~:text=Esta%20most%20que%20el%2029,observ%C3%B3%20en%20todas%20las%20provincias.>

¿DISMINUYE LA CANTIDAD DE ACCIDENTES APLICANDO EL LÍMITE DE VELOCIDAD?

Damaris Alaniz - Baltasar Maldonado

damaris3838@gmail.com - baltasarmaldonadospital@gmail.com

Resumen

La falta de una normativa que implique establecer el límite de velocidad puede incrementar el riesgo de accidentes y poner en peligro la seguridad vial. Es por ello que es importante realizar la evaluación de las políticas de tránsito para implementar medidas efectivas que contribuyan a la reducción de accidentes. Este trabajo tiene como objetivo estudiar la influencia de la aplicación o no de un límite de velocidad en relación a la cantidad de accidentes que ocurren diariamente. Para esto, se tomaron los datos de la cantidad de accidentes por día que ocurrieron en Suecia durante algunos días de los años 1961 y 1962. Se analizó qué distribución estadística puede modelar la cantidad de accidentes tanto si se aplica un límite de velocidad como cuando no. Y si existe una disminución en la cantidad de accidentes por día con la aplicación del límite de velocidad. Se concluyó que al aplicar el límite de velocidad existe una disminución en la cantidad de accidentes por día comparado cuando no se aplica este límite.

Introducción

El límite de velocidad es una norma que aplican los distintos gobiernos con el fin de reducir los accidentes de tránsito en las calles y rutas del país. Tiene una gran importancia para la seguridad vial y su incumplimiento supone una de las mayores causas de accidentes. En este contexto, en los años 1961 y 1962, en Suecia se aplicó ciertos días un límite de velocidad de, aproximadamente, 90 o 100 km/h en autopistas y autovías. (Svensson, University of Stockholm, 1981).

Los objetivos del presente trabajo son: (1) Estudiar qué tipo de distribución siguen los datos de la cantidad de accidentes cuando se aplica el límite de velocidad y cuando no; (2) Determinar si la aplicación de este límite disminuye la cantidad de accidentes por día.

Metodología

Trabajamos con los datos publicados por Svensson en el año 1981, donde se registraron los datos de 92 días en cada uno de los años 1961 y 1962 con la información de si se había aplicado o no el límite de velocidad y cuántos accidentes que resultaron con personas heridas ocurrieron ese día. Los datos fueron importados a una hoja de cálculo para facilitar su análisis.

Definimos las variables en estudio como:

X =Cantidad de accidentes por día cuando se aplicó el límite de velocidad.

Y =Cantidad de accidentes por día cuando no se aplicó el límite de velocidad.

Ambas variables cuantitativas discretas. Realizamos un análisis descriptivo de estas variables a través de un boxplot y calculando las medidas de resumen correspondientes. También un análisis inferencial de los datos, utilizando bondad de ajuste para probar qué distribución sigue la variable y un test de hipótesis, para probar la diferencia de las medias de la cantidad de accidentes por día

cuando se aplica o no el límite de velocidad, además una prueba Fisher para elegir el estadístico de prueba adecuado. Utilizamos estimación puntual a través del método de máxima verosimilitud. Para este análisis utilizamos el software estadístico R a través de su interfaz gráfica R Commander y el software Geogebra a través de la vista de Cálculos de probabilidad.

Resultados y discusión

Comenzamos realizando el análisis descriptivo de nuestras variables X e Y calculamos las medidas de resumen de una muestra aleatoria simple de estas variables utilizando GeoGebra, el cual nos arrojó los siguientes resultados:

Variables	X	Y
Número de días	69	115
Media	18,913	23,1304
Desvío	7,4749	9,158
Mínimo	7	8
Primer cuartil	13,5	17
Mediana	17	21
Tercer cuartil	23,5	29
Máximo	42	49

Tabla 1: Tabla con las medidas de resumen de las variables X e Y.

Luego, realizamos los gráficos de estas variables para tener una forma visual de resumir estos datos y analizar qué distribución pueden llegar a tener.

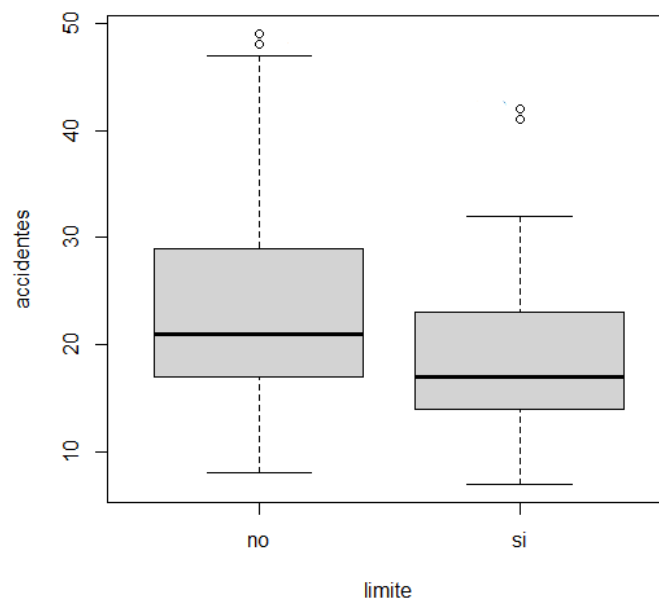


Imagen 1: Boxplots de la cantidad de accidentes por día respecto a la aplicación o no del límite de velocidad.

Estos resultados, el boxplot y número de las medias muestrales, nos llevan a intuir que la cantidad de accidentes en los días en los que se aplicó el límite de velocidad es menor que en los días en que no se aplicó el límite.

Por la definición de las variables, se podría pensar que las variables X e Y siguen una distribución Poisson de parámetro λ_X y λ_Y respectivamente. Por esto, realizamos el test de bondad de ajuste para analizar si las variables X e Y siguen esta distribución. Para esto, anteriormente estimamos el parámetro λ a través del método de máxima verosimilitud. Utilizamos este método de estimación ya que sabemos que arroja buenos resultados para el test de bondad de ajuste. Además, sabemos que cumple con las propiedades de ser insesgado y consistente.

Así, obtenemos que, tanto para X como para Y, el parámetro estimado λ es igual a la media muestral. Esto es,

$$\lambda_X = \bar{x} = 18,913 \quad \text{y} \quad \lambda_Y = \bar{y} = 23,1304$$

A partir de este parámetro, realizamos el test de bondad de ajuste de cada variable. Para la variable X, las hipótesis que fueron puestas a prueba son:

$$H_0: X \sim P(18,913) \quad \text{vs.} \quad H_a: X \text{ no tiene esta distribución}$$

Obtenemos la siguiente tabla de frecuencias observadas y esperadas:

Clase	{4;...;13}	{14;...;16}	{17;...;21}	{22;...;54}
Frecuencia observada	17	14	14	24
Frecuencia esperada	7,01	13,6	29,91	18,48

Tabla 2: Tabla de frecuencias observadas y esperadas de la variable $X \sim P(\lambda_X)$.

De esta forma, definimos el estadístico chi-cuadrado para la variable X de la siguiente forma:

$$T = \sum_{i=1}^4 \frac{(o_i - e_i)^2}{e_i} \sim \chi^2(2)$$

Así, encontramos la siguiente región de rechazo, fijando $\alpha = 0,05$

$$RR = \{T \geq \chi^2(2)_{\alpha}\} = \{T \geq \chi^2(2)_{0,05}\} = \{T \geq 5,9915\}$$

Calculando con los datos de nuestra muestra, obtenemos

$$\chi^2 = 24,3604 \in RR$$

Por lo tanto, tenemos evidencia al 5% para rechazar la hipótesis nula. Es decir, X no tiene una distribución Poisson de parámetro 18,913.

Para el caso de la variable Y, las hipótesis que fueron puestas a prueba son

$$H_0: Y \sim P(23,1304) \quad \text{vs.} \quad H_a: Y \text{ no tiene esta distribución}$$

Obteniendo la siguiente tabla de frecuencias observadas y esperadas

Clase	{4;...;15}	{16;...;21}	{22;...;27}	{28;...;52}
Frecuencia observada	24	38	21	32
Frecuencia esperada	5,67	37,93	50,72	20,68

Tabla 3: Tabla de frecuencias observadas y esperadas de la variable $Y \sim P(\lambda_Y)$.

Realizando el mismo procedimiento que para la variable X, obtenemos que la región de rechazo es

$$RR = \{T \geq \chi^2(2)_{\alpha}\} = \{T \geq \chi^2(2)_{0,05}\} = \{T \geq 5,9914\}$$

Y, calculando el valor del estadístico con los datos de nuestra muestra, tenemos

$$\chi^2 = 82,8687 \in RR$$

Por lo tanto, nuevamente, tenemos evidencia al 5% para rechazar la hipótesis nula. Esto es, Y no sigue una distribución Poisson de parámetro 23,1304.

A su vez, como se puede ver en la **Tabla 1**, para la variable X e Y, sabemos que la media y la varianza de las variables no son iguales. Y esto es una característica importante de las variables con distribución Poisson. Por lo tanto, como tenemos un caso de sobredispersión de los datos de la muestra (la varianza es mucho más alta que la media), la mejor forma para modelar nuestra variable es a través de una distribución Binomial Negativa. (López Vásquez, Universidad de Valparaíso, 2016, p. 37)

Ahora, realizamos el test de bondad de ajuste para analizar si las variables X e Y siguen una distribución Binomial Negativa.

La distribución Binomial Negativa cuenta con dos parámetros: r y p, donde r debe ser un número natural y p un valor entre 0 y 1. Para estimar estos parámetros, no podemos utilizar el método de máxima verosimilitud. Por lo tanto, se estima a partir de las características de esta distribución, de forma que

$$E(X) = \frac{r}{p} \quad y \quad V(X) = r \frac{(1-p)}{p^2}$$

Estimamos la esperanza de la variable X con la media muestral y la varianza de la variable con la varianza muestral. Así obtenemos, para la variable X:

$$\hat{p} \simeq 0,253 \quad y \quad \hat{r} \simeq 5$$

Obtenemos la siguiente tabla de frecuencia:

Clase	[5;11]	[12;13]	[14;15]	[16;18]	[19;20]	[21;23]	[24;29]	[30;∞)
Frecuencia observada	9	8	8	12	5	10	13	4
Frecuencia esperada	8,24	6,26	7,58	11,72	7,17	9,01	11,30	7,63

Tabla 4: Tabla de frecuencias observadas y esperadas de la variable $X \sim BN(5; 0,253)$

Habiendo estimado estos parámetros, ponemos a prueba las siguientes hipótesis para realizar el test de bondad de ajuste

$$H_0: X \sim BN(5; 0,253) \quad vs. \quad H_a: X \text{ tiene otra distribución.}$$

Luego, definimos nuevamente el estadístico chi-cuadrado como

$$T = \sum_{i=1}^8 \frac{(o_i - e_i)^2}{e_i} \sim \chi^2(5)$$

Encontrando que la región de rechazo, fijando $\alpha = 0,05$, es

$$RR = \{T \geq \chi^2(5)_{\alpha}\} = \{T \geq \chi^2(5)_{0,05}\} = \{T \geq 11,0705\}$$

Mientras que si calculamos el valor del estadístico con los datos de nuestra muestra de X, obtenemos que

$$\chi^2 = 3.3122 \notin RR$$

Por lo que podemos decir que no existe evidencia al 5% para rechazar la hipótesis nula. Entonces, $X \sim BN(5; 0,253)$

Del mismo modo, realizamos la prueba de bondad de ajuste para la variable Y. Considerando la media muestral y la varianza muestral estimamos los parámetros r y p tal que:

$$\hat{p} \simeq 0,216 \quad \text{y} \quad \hat{r} \simeq 5$$

Obtenemos la siguiente tabla de frecuencia para la variable Y:

Clases	Frecuencia observada o	Frecuencia esperada e	$\frac{(o-e)^2}{e}$
[5;10]	6	5,12	0,14958242
[11;12]	4	5,89	0,60398612
[13;15]	14	12,89	0,09465969
16	4	5,08	0,22832671
17	5	5,31	0,01773321
18	9	5,44	2,32855703
19	5	5,48	0,04274816
[20;23]	20	20,90	0,03921147
24	5	4,70	0,01906065
25	4	4,42	0,04034392
26	3	3,93	0,21872402
[27;28]	6	7,34	0,24600401
[29;30]	6	6,15	0,00352748
[31;33]	7	7,17	0,00397737
[34;38]	8	7,58	0,02304243
[39;inf)	9	7,59	0,26224553

Tabla 5: Tabla de frecuencias observadas y esperadas de la variable $Y \sim BN(5; 0,216)$

Por lo tanto, las hipótesis que se pusieron a prueba son

$$H_0: Y \sim BN(5; 0,216) \quad \text{vs.} \quad H_a: Y \text{ tiene otra distribución.}$$

Siguiendo el mismo procedimiento que para la variable X pero con la muestra de Y, obtenemos que la región de rechazo, fijando nuevamente $\alpha = 0,05$, es

$$RR = \{T \geq \chi^2(13)_{\alpha}\} = \{T \geq \chi^2(13)_{0,05}\} = \{T \geq 22,362\}$$

Mientras que si calculamos el valor del estadístico con nuestra muestra, utilizando los datos de la **Tabla 5**, obtenemos que

$$\chi^2 = 4,3217 \notin RR$$

Por lo tanto, como este valor no pertenece a la región de rechazo, podemos decir que no hay evidencia al 5% para rechazar la hipótesis nula. Esto es, $Y \sim BN(5; 0,216)$.

Por otra parte, buscamos determinar si es efectiva la aplicación del límite o no. Para este análisis, utilizamos R Commander, realizando los test de hipótesis correspondientes.

Para esto, en primera instancia, debemos analizar si las varianzas de las variables X e Y son iguales. A partir de esto, ponemos a prueba las siguientes hipótesis:

$$H_0: \frac{\sigma_Y}{\sigma_X} = 1 \quad vs. \quad H_a: \frac{\sigma_Y}{\sigma_X} \neq 1$$

Para ello se considera el estadístico de prueba

$$T = \frac{\frac{S_X^2}{\sigma_X^2}}{\frac{S_Y^2}{\sigma_Y^2}} \sim F(68; 114)$$

Fijando $\alpha = 0,05$, se tiene la región de rechazo

$$RR = \left\{ \frac{S_X^2}{S_Y^2} > F(68; 114)_{0,05} \right\}$$

Obtenemos que el valor p es 0,06972, por lo tanto, no hay evidencia suficiente al 5% para rechazar la hipótesis nula. Entonces, podemos suponer que las varianzas son iguales.

Con este resultado, para poder responder a dicho objetivo, planteamos las siguientes hipótesis

$$H_0: \mu_Y - \mu_X \leq 0 \quad vs. \quad H_a: \mu_Y - \mu_X > 0$$

Para poner a prueba las hipótesis el pivote adecuado es

$$\frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{S_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \sim t_{182}$$

Resulta que el valor p es de 0,0007286. Entonces, tenemos evidencia suficiente al 5% para rechazar la hipótesis nula. Por lo tanto, podemos decir que la media de la variable Y es mayor que la media de la variable X. Además, hallamos que el intervalo de confianza para la diferencia de medias es

$$IC(\mu_Y - \mu_X: 95\%) = [1,643104; 6,791679]$$

Conclusiones

Luego de analizar los resultados obtenidos con los datos registrados entre 1961 y 1962 en Suecia, podemos concluir que la cantidad de accidentes por día, tanto si se aplicó o no el límite de velocidad, tiene una distribución Binomial Negativa.

Además, podemos decir que la aplicación del límite de velocidad generó una baja significativa en la cantidad de accidentes por día. Es decir, podemos pensar que establecer un límite de velocidad previene, aproximadamente, entre 2 y 6 accidentes por día.

Por lo tanto, la aplicación de esta norma fue positiva para disminuir los accidentes y es sugerible la aplicación de un límite de velocidad en las carreteras alrededor del mundo.

Referencias

Svensson, A. (1981). On the goodness-of-fit test for the multiplicative Poisson model. *Annals of Statistics*, 9, 697-704.

https://www.jstor.org/stable/2240840?read-now=1&oauth_data=eyJlbWFpbCI6ImRhbWFyaXMzODM4QGdtYWlsLmNvbSIsImIuc3RpdHV0aW9uSWRzIjpbXSsicHJvdmlkZXIiOiJnb29nbGUifQ&seq=3

GeoGebra (2022). GeoGebra para enseñar y aprender Matemáticas. [software]. <https://www.geogebra.org/>

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. [software]. <https://www.R-project.org/>

EFFECTOS DE LA TERAPIA HORMONAL SOBRE EL COLESTEROL EN MUJERES PREMENOPÁUSICAS

Ana Paula Torres - Giuliana Raiteri Eberle - Selva Agustina Maica

pauelytorres009@gmail.com - giulianaraiterieberle@gmail.com - agustmaica03@gmail.com

Resumen

Este estudio analiza los niveles de colesterol en mujeres premenopáusicas de 30 a 53 años sometidas a histerectomía transabdominal y salpingo-ooforectomía unilateral (SOU), evaluando el efecto de la terapia de reemplazo hormonal con estrógenos en el control del colesterol. Tras un análisis estadístico, los niveles de colesterol no mostraron un incremento significativo dos meses después de la cirugía. Esto sugiere que el tratamiento hormonal contribuye a mantener los niveles de colesterol en un rango saludable.

Introducción

La histerectomía es una intervención quirúrgica que implica la extracción del útero y, en algunos casos, de uno o ambos ovarios y trompas de Falopio. Cuando esta operación incluye la extirpación bilateral de los ovarios, las mujeres experimentan una menopausia inducida debido a la interrupción en la producción de estrógenos, una hormona clave en la regulación de los niveles de lípidos en sangre (Instituto Nacional del Cáncer, s.f).

Por otro lado, la histerectomía total es una cirugía para extirpar el útero y el cuello uterino. En ocasiones también se extirpan los ovarios y las trompas de Falopio. La histerectomía total se puede hacer a través de la vagina (sin hacer cortes en el abdomen) o mediante una incisión (corte) en el abdomen (Instituto Nacional del Cáncer, s.f).

La salpingo - ooforectomía unilateral (SOU) es un procedimiento quirúrgico en el cual se extirpan uno de los ovarios y una de las trompas de Falopio. Este tipo de cirugía puede realizarse por diversas razones médicas, como para tratar cánceres ginecológicos, quistes ováricos grandes, endometriosis severa o para prevenir problemas futuros en el ovario o la trompa afectados (Instituto Nacional del Cáncer, s.f).

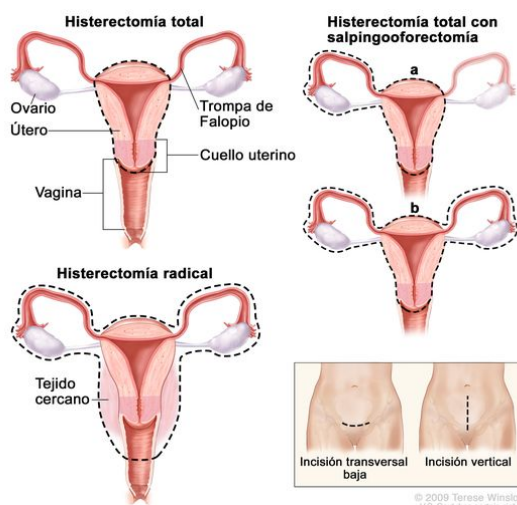


Figura 1. Histerectomías extraída de Instituto Nacional de Cáncer

Históricamente, el estrógeno equino ha sido utilizado en la terapia de reemplazo hormonal (TRH) para tratar síntomas de la menopausia en mujeres. Sin embargo, su uso ha disminuido debido a preocupaciones sobre posibles efectos secundarios y el desarrollo de alternativas más seguras y eficaces, como estrógenos sintéticos o bio-idénticos. Este refiere a una forma de estrógeno que se obtiene a partir de la orina de yeguas preñadas. El mismo contiene principalmente estrona y equilina, que son formas naturales de estrógeno presentes en los caballos.

En este contexto, la regulación del colesterol cobra especial relevancia, ya que los cambios hormonales asociados a la menopausia pueden alterar significativamente sus niveles. El colesterol es una sustancia grasa natural presente en todas las células del cuerpo humano necesaria para el normal funcionamiento del organismo. La mayor parte del colesterol se produce en el hígado, aunque también se obtiene a través de algunos alimentos.

Definamos su función:

- Interviene en la formación de ácidos biliares, vitales para la digestión de las grasas.
- Los rayos solares lo transforman en vitamina D para proteger la piel de agentes químicos y evitar la deshidratación.
- A partir de él se forman ciertas hormonas, como las sexuales y las tiroideas.

Los niveles normales de colesterol total son:

- Normal: menos de 200 mg/dl
- Normal-alto: entre 200 y 240 mg/dl. Se considera hipercolesterolemia a los niveles de colesterol total superiores a 200 mg/dl.
- Alto: por encima de 240 mg/dl. (CardioPatagonia, s.f)

La sangre conduce el colesterol desde el intestino o el hígado hasta los órganos que lo necesitan y lo hace uniéndose a partículas llamadas lipoproteínas. Según CardioPatagonia (s.f), existen dos tipos de lipoproteínas:

- De baja densidad (LDL): se encargan de transportar nuevo colesterol desde el hígado a todas las células de nuestro organismo.
- De alta densidad (HDL): recogen el colesterol no utilizado y lo devuelve al hígado para su almacenamiento o excreción al exterior a través de la bilis.

Según esta interacción podemos hablar de dos tipos de colesterol:

- Colesterol malo: el colesterol al unirse a la partícula LDL se deposita en la pared de las arterias y forma las placas de ateroma.
- Colesterol bueno: el colesterol al unirse a la partícula HDL transporta el exceso de colesterol de nuevo al hígado para que sea destruido. (CardioPatagonia, s.f)

Teniendo en cuenta la investigación llevada a cabo en el año 2019 (Guittlein, Martel & Miranda, 2019) sobre los niveles de colesterol de 22 mujeres premenopáusicas sometidas a histerectomía y SOB, antes y a los dos meses después de la cirugía, en cual se determinó que el tratamiento de sustitución de estrógeno no fué favorable, el **objetivo** de nuestra investigación es evaluar cómo los niveles de colesterol total varían en mujeres sometidas a histerectomía unilateral (SOU) con o sin extracción de ovarios y trompas, y analizar el impacto de la terapia de sustitución hormonal con estrógeno en la regulación del colesterol total considerando antes y dos meses después de la cirugía (CT0 y CT2).

Metodología

Para llevar a cabo la investigación, se considerará un muestreo aleatorio simple tomado por Kim et al. (1996). Siguiendo nuestro objetivo de investigar las variaciones en los niveles de lipoproteínas Lp(a) debido a la pérdida de hormonas sexuales femeninas en mujeres que experimentan menopausia quirúrgica y que están bajo terapia de reemplazo de estrógenos, se llevaron a cabo mediciones del colesterol en sangre (mg/dl) en diferentes momentos, previo, a dos meses y a los cuatro meses de la operación; nos enfocaremos en el momento cero, es decir, antes de la cirugía, y dos meses posterior a ésta.

La unidad experimental es una mujer premenopáusica entre 30 y 53 años de edad, que se ha sometido a una histerectomía transabdominal. La población de unidades son las mujeres premenopáusicas entre 30 y 53 años de edad que se sometieron a la histerectomía transabdominal, mientras que 44 mujeres seleccionadas de ésta población, corresponden a la muestra de unidades.

La muestra estadística que utilizamos, se basa en 22 mujeres premenopáusicas, que tenían histerectomía y salpingo - ooforectomía unilateral (SOU), este grupo de mujeres recibieron 0,625 mg de estrógeno equino durante dos meses después de la cirugía. A las 22 mujeres se le realizó una medición sobre los niveles de colesterol total en sangre (mg/dl), estas mediciones fueron realizadas antes de la cirugía, a los dos meses de dicha cirugía y a los cuatro meses después de la misma. Teniendo en cuenta el objetivo planteado, la variable de estudio se define como:

$X = \text{nivel de colesterol total en sangre (mg/dl), de una mujer premenopáusica, entre 30 y 53 años, con la cirugía de histerectomía transabdominal realizada.}$

Dicha variable, se clasifica como cuantitativa continua, pues resulta de un proceso de medición.

Partiendo del conjunto de datos previos, nos enfocaremos en niveles de colesterol en sangre entre de la cirugía y dos meses posterior a esta, si se realizaron una histerectomía y salpingo - ooforectomía unilateral (SOU).

Debido a la muestra estadística considerada, se seleccionan dos momentos:

$X_0 = \text{nivel de colesterol total en sangre (mg/dl) previo a la cirugía.}$

$X_2 = \text{nivel de colesterol total en sangre (mg/dl) dos meses posterior a la cirugía.}$

Para el análisis descriptivo, se realizaron tablas de distribución de frecuencias observadas de las variables mencionadas, con sus respectivos histogramas. Según lo estudiado, se consideró que los niveles de colesterol total en sangre se puede modelar con una distribución normal, de parámetros μ y σ^2 . Para poder verificar esta afirmación se realizó un análisis descriptivo mediante los gráficos QQ-plot. Para saber si el tratamiento fue efectivo, es decir, si la sustitución de estrógeno equilibra el nivel de colesterol total en sangre(mg/dl) para las mujeres premenopáusicas con histerectomía unilateral SOU, se realizó un test de hipótesis, para la diferencia de medias de poblaciones normales para muestras aleatorias dependientes (antes y después). En este proyecto, se utilizó el software R (versión 4.4.1) y Rcmdr (versión 2.9-2) para construir gráficos, tablas de frecuencias y hacer cálculos auxiliares.

Resultados y discusión

Análisis descriptivo

A partir de los datos obtenidos, por el método de Sturges (Dieser, 2022), se calcularon las clases de los niveles de colesterol en los momentos (X_0 y X_2), se realizaron gráficos (Histograma), tablas de frecuencias correspondientes y se presentaron medidas de resúmenes.

X_0 (SOU)	
Niveles de Colesterol	Cantidad de mujeres
[120;143)	5
[143;166)	5
[166;189)	5
[189;212)	5
[212;235)	2
[235;258]	0
TOTAL	22

Tabla 1. Tabla de distribución de frecuencias del nivel de colesterol total en sangre X_0 (SOU)

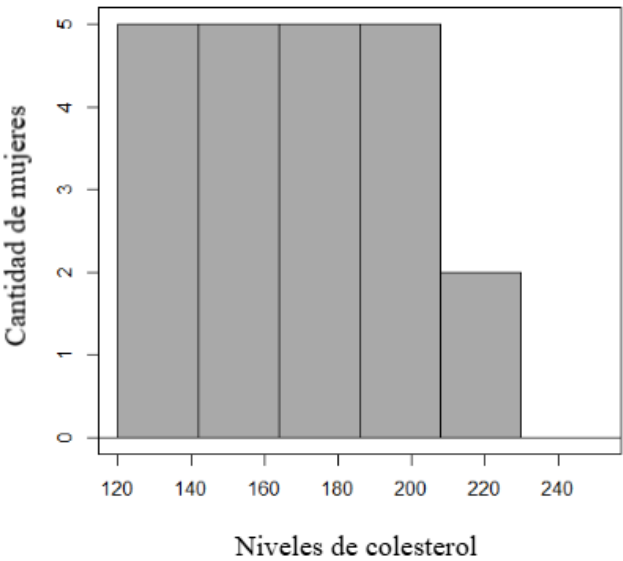


Figura 2. Histograma del nivel de colesterol total en sangre X_0 (SOU)

$X_2(\text{SOU})$	
Niveles de Colesterol	Cantidad de mujeres
[120;143)	4
[143;166)	3
[166;189)	9
[189;212)	3
[212;235)	0
[235;258]	3
TOTAL	22

Tabla 2. Tabla de distribución de frecuencias del nivel de colesterol total en sangre X_2 (SOU)

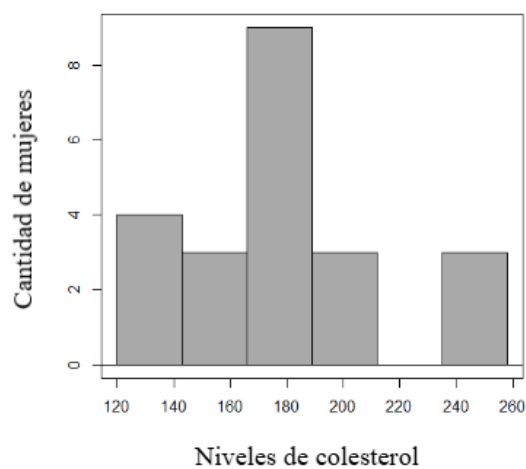


Figura 3. Histograma del nivel de colesterol total en sangre X_2 (SOU)

Nivel de colesterol	SOU		
	$\bar{X}$ Media	S Desvío	CV Coef. de variabilidad
X_0	172,77	30,56	0,18
X_2	178,36	35,43	0,20

Tabla 3. Tabla de medidas de resumen de X_0 (SOU) y X_2 (SOU)

Se observa en la **Tabla 1** y **Figura 2** la frecuencia de colesterol total en sangre de las 22 mujeres premenopáusicas previo a la intervención, es igual en los intervalos de 120 a 212 mg/dl mientras que, respecto de la **Tabla 2** y **Figura 3**, a los dos meses posteriores a la cirugía, aumenta significativamente la frecuencia entre 166 y 189 mg/dl. En relación a los registros de referencia de colesterol total en sangre mg/dl, podemos afirmar que los datos referidos a la **Figura 2** están en un nivel normal, mientras que los que refieren al **Figura 3** si bien predomina un nivel normal, también aumentan los niveles de colesterol en algunas mujeres.

Por otro lado, en cuanto a las medidas de resumen, se obtuvo la media como medida de tendencia central, el desvío y CV como medidas de dispersión. Como se puede ver, los coeficientes de variación obtenidos están entre (0,10; 0,33] es decir, son aceptables.

Además, comparando los niveles de colesterol de las 22 mujeres, en promedio sus niveles son menores antes que dos meses después de la cirugía, al igual que sus valores de dispersión.

Como representación gráfica consideramos elaborar para cada momento un QQ-plot, y así afirmar que si los datos provienen de una distribución normal se espera que tales puntos se acomoden aproximadamente alrededor de la recta identidad (Dieser, 2022).

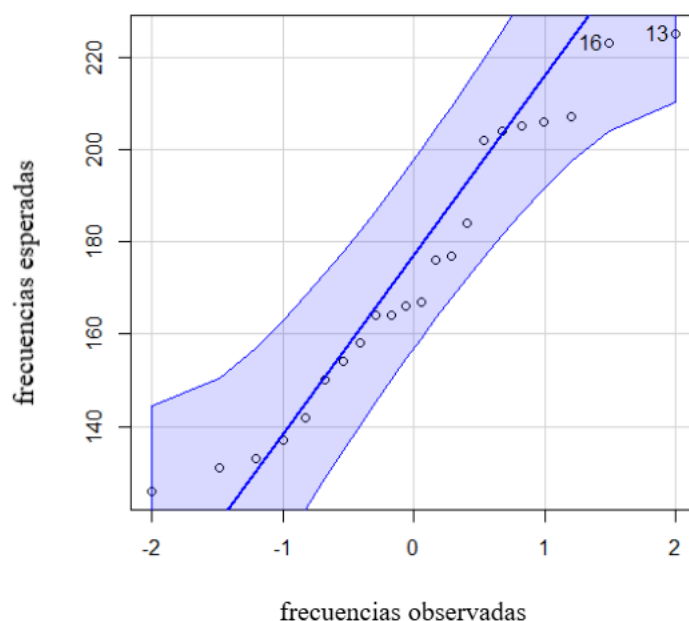


Figura 4. QQ-plot sobre los niveles de colesterol total en sangre X_0

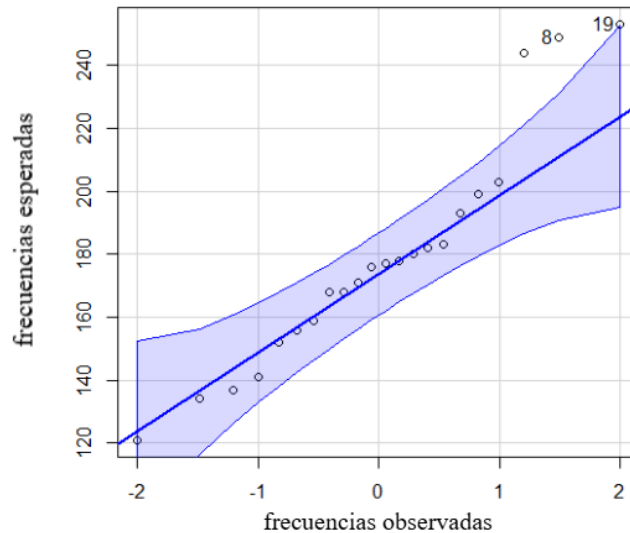


Figura 5. QQ-plot sobre los niveles de colesterol total en sangre X_2

Como se puede apreciar, en la **Figura 4** y **Figura 5**, los datos se ajustan adecuadamente a una recta, lo que sugiere que los momentos X_0 y X_2 pueden modelarse con una distribución normal.

Análisis inferencial

Para verificar si los niveles de colesterol aumentan a los dos meses después de la cirugía en mujeres premenopáusicas con SOU tratadas con sustitución de estrógenos, se realiza una prueba de hipótesis por el método clásico. Esto implica formular una hipótesis nula (H_0) que establece que no hay aumento en los niveles de colesterol tras el tratamiento, y una hipótesis alternativa (H_a) que sugiere un incremento.

Dado que tenemos muestras dependientes, pues la variable X depende de dos momentos, consideramos una nueva variable D denominada diferencia.

Sea $D = X_0 - X_2$ con $D_1; D_2; \dots; D_{22}$ m.a.s de $D \sim N(\mu_0 - \mu_2; \sigma^2)$. De esta manera tenemos:

$\bar{X}_D = -5,59$ y $S_D = 21,79$ en donde $\bar{X}_D$ denota la media muestral y S_D la desviación estándar.

Luego, planteamos las siguientes hipótesis para poner a prueba:

$$H_0: \mu_0 - \mu_2 \geq 0 \quad vs \quad H_a: \mu_0 - \mu_2 < 0$$

Sea el estadístico o pivote adecuado $T = \frac{\bar{X}_D - (\mu_0 - \mu_2)}{\frac{S_D}{\sqrt{n}}} \sim t(n - 1)$

Luego, se fijó un nivel de significancia $\alpha = 0,05$ y con los cálculos convenientes, se obtuvo la siguiente región de rechazo:

$$RR = \left\{ \bar{X}_D < -t(n - 1)_{\alpha} \frac{S_D}{\sqrt{n}} \right\} = \left\{ \bar{X}_D < -t(22 - 1)_{0,05} \frac{21,79}{\sqrt{22}} \right\} = \left\{ \bar{X}_D < -1,72074 \frac{21,79}{\sqrt{22}} \right\}$$

Por lo tanto, $RR = \left\{ \bar{X}_D < -7,99 \right\}$ y como $\bar{X}_D = -5,59 \notin RR$, es decir, no hay suficientes evidencias con un nivel de significancia del 5% para rechazar la H_0 . Esto equivale a decir que en

promedio, los niveles de colesterol después de dos meses de la cirugía, no son mayores con respecto a los niveles de colesterol obtenidos antes de la cirugía.

Observando los valores de la muestra tomada por Kim et al. (1996) en el momento cero, es decir, previo a la cirugía podemos observar que el 68% tiene un nivel normal de colesterol normal en sangre y un 32% tiene un colesterol normal alto; dos meses después de la cirugía el 82% de la muestra tiene un nivel de colesterol normal, mientras que un 4,5% tiene un colesterol normal-alto y un 13,5% tiene un colesterol alto.

A partir de la **Tabla 3**, podemos ver que el nivel de colesterol en sangre de las mujeres con cirugía SOU, en promedio es de 172,77 mg/dl en el momento cero, y 178,36 mg/dl dos meses después.

Para cada momento se elaboró un QQ-plot, y así corroborar o no la hipótesis de que la variable de estudio modelan una distribución normal, la cual efectivamente se pudo constatar según las gráficas obtenidas, ya que se ajustan a la recta identidad. Además, no se pudo realizar un test de bondad de ajuste, por tener una muestra pequeña; es decir, menor a 30.

Por último, al realizar el test de hipótesis concluimos que el tratamiento de sustitución de estrógeno para las mujeres premenopáusicas con histerectomía y SOU fué efectivo, ya que se obtuvo, en promedio, que los niveles de colesterol después de dos meses de la cirugía no son mayores con respecto a los niveles previos a la cirugía.

Conclusión

Luego de los análisis mencionados con anterioridad, y basándonos en la muestra recolectada por Kim et al. (1996) sobre mujeres premenopáusicas entre 30 y 53 años que se sometieron a una histerectomía transabdominal y salpingooforectomía unilateral (SOU), concluimos que el tratamiento mostró resultados favorables ya que se observó que, en promedio, los niveles de colesterol en sangre no aumentan a los dos meses de la cirugía.

Para futuro informe, sería interesante estudiar las diferencias entre las medidas de colesterol previo a la cirugía y cuatro meses después, diferenciando las que tuvieron SOU y SOB, para ver cual tuvo resultados más benéficos. Sería conveniente tener una muestra aleatoria simple más grande y así poder obtener conclusiones más significativas.

Referencias

- Kim, C., Ryu, W.S., Kwak, J.W., Park, C.T., & Ryoo, U.H. (1996). Changes in Lp(a) Lipoprotein and Lipid Levels Meter Cessation of Female Sex Hormone Production and Estrogen Replacement Therapy. *Archives of Internal Medicine*, 156, 500-504.
- NIH - Instituto Nacional del Cáncer (s.f) Diccionario de cáncer del NCI - “*Histerectomía total*”. Recuperado de <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/histerectomia-total>
- OpenAI. (2024). *Respuesta generada en ChatGPT* [Asistencia en línea]. Recuperado el [21 de Octubre de 2024] <https://chat.openai.com/>
- CardioPatagonia (s.f) Colesterol y riesgo cardiovascular - “*Cuáles son los niveles normales de colesterol, cuál es el bueno y cuál es el malo, cómo prevenirlo*”. Recuperado de <https://cardiopatagonia.com.ar/pacientes/factores-de-riesgo-cardiovascular/colesterol/>
- Guittlein, Martel, Miranda. (2019). “*Cambios en los niveles de colesterol en la premenopausia*” ReDi, Repositorio Digital, Universidad Nacional de La Pampa - Facultad de Ciencias

Exactas y Naturales. Recuperado de
<http://redi.exactas.unlpam.edu.ar/xmlui/handle/2013/325>

Dieser, M.P. (2022). “*Inferencia Estadística: Notas de clase*”. Material no publicado. Facultad de Ciencias Exactas y Naturales, Universidad Nacional de La Pampa. Santa Rosa.