

¡Estadístic@s en acción!

ESTADÍSTICA

**PROBABILIDAD Y ESTADÍSTICA II
2019**

Santiago ACOSTA - Ana Karen BLANCO - Cintia Vanera ELENA -
Viviana Vanesa GUITTLEIN - Milagros María Elena LECITRA -
Yanina Belén LICK - Sergio Nicolás MAIDANA - Fanny Ruth MAIER -
Tamara Macarena MARTEL - Giuliana OLIVERA -
Lourdes María Belén MIRANDA - Silvina Maribel PONKOSKY -
Deborah RAMBURGER - Ximena Sofía SANCHEZ - Fabián Gabriel SEIA
- Talía Micaela SCHMIDT

Compiladoras

Maria Paula DIESER - Lorena Veronica CAVERO



FACULTAD DE CIENCIAS
EXACTAS Y NATURALES

Universidad Nacional de La Pampa

Índice

Prólogo	2
Peso de los bebés recién nacidos con respecto a la edad de la madre <i>Ana Karen Blanco - Cintia Vanesa Elena - Talia Micaela Schmidt</i>	4
Identificación de los cangrejos <i>Leptograpsus Variegatus</i> según el ancho del caparazón <i>Santiago Acosta - Ximena Sofía Sanchez</i>	14
¿Peso ideal o peso real? Análisis estadístico del peso de los gatos <i>Deborah Ramburger - Fanny Maier - Giuliana Olivera</i>	20
Calculando el radio de la Tierra con un método milenario <i>Sergio Nicolas Maidana - Silvina Maribel Ponkosky - Fabian Gabriel Seia</i>	30
Alcohol en dibujos animados <i>Milagros Lecitra - Yanina Belén Lick</i>	39
Cambios en los niveles de colesterol en la premenopausia <i>Viviana Vanesa Guittlein - Tamara Macarena Martel - Lourdes María Belén Miranda</i>	49

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 nuestros estudiantes de las asignaturas “Estadística” y “Probabilidad y Estadística II” de las carreras de Profesorado y Licenciatura en Matemática de la Facultad de Ciencias Exactas y Naturales (UNLPam), respectivamente, 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 las TIC 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 nuestros estudiantes del Ciclo Lectivo 2019, quienes abordaron temáticas muy diversas. Un grupo abordó la problemática de la presencia de escenas de alcohol en dibujos animados desde el punto de vista legal, basándose en la clasificación propuesta por “Motion Picture Association of America!” en el año 1922. Otro grupo analizó la influencia de la edad de la madre sobre el peso de los bebés recién nacidos. En tanto que otro se centró en analizar la influencia del ancho del caparazón para diferenciar el tipo de especie (azul o naranja) en los cangrejos *Leptograpsus Variegatus*. Un cuarto grupo buscó corroborar o refutar el peso medio ideal considerado por la página oficial de una marca de alimento balanceado para gatos y analizar el vínculo entre el peso

corporal de los éstos y su respectivo sexo. Un quinto grupo de estudio estimó el radio de la Tierra analizando datos recolectados por distintas instituciones en Argentina, siguiendo el método de Eratóstenes. Finalmente, otro grupo de estudio se abocó a investigar la influencia de sustitución de estrógeno para mantener el nivel de colesterol total en sangre en pacientes con premenopausia.

Las profes, diciembre de 2019

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.



PESO DE LOS BEBÉS RECIÉN NACIDOS CON RESPECTO A LA EDAD DE LA MADRE

Ana Karen Blanco- Cintia Vanesa Elena - Talia Micaela Schmidt

karenblanco0504@gmail.com - vanee2422@gmail.com - taliarusa@gmail.com

Resumen

En el presente trabajo, se pretende analizar la influencia de la edad de la madre sobre el peso del bebé recién nacido. Para ello, se consideraron los datos obtenidos de 120 madres que se atendieron en el Baystate Medical Center de la ciudad de Springfield, Massachusetts, durante 1986; de las cuales, 51 de ellas quedaron embarazadas en su etapa adolescente y 69 en su etapa adulta.

Introducción

Uno de los factores que influyen en el peso que tiene el bebé al nacer es la edad de la madre: según los obstetras, la edad óptima para la reproducción se sitúa entre los 20 y los 30 años. Tanto los hijos de adolescentes como los de madres maduras tienden a ser de menor peso que los nacidos de madres en la edad considerada ideal.

Hoy se considera normal un peso de entre 3,400-3,500 kilos en los niños y unos 200 gramos menos en las niñas. Por debajo de los 2,500 kilos y por encima de los 4,500 se habla de bebés de bajo peso (microsomáticos) y con sobrepeso (macrosomáticos) (Lleó, n.d.).

Teniendo esto en consideración, nos interesa investigar si hay diferencia entre los pesos de los bebés recién nacidos de madres adolescentes y los pesos de los bebés recién nacidos de madres adultas.

Metodología

El estudio se realizó con una muestra aleatoria simple inicial de 189 madres, con datos recolectados en el *Baystate Medical Center* de la ciudad de Springfield, Massachusetts durante 1986, tales como la edad de la madre al momento de quedar embarazada y el peso del bebé al nacer, en gramos (Hosmer & Lemeshow, 1989). Para nuestra investigación consideramos una muestra de 120 madres, 51 de las cuales quedaron embarazadas en su etapa adolescente y 69 quedaron embarazadas en su etapa adulta; dejando de lado 69 que quedaron embarazadas en su etapa joven.

Tomamos las siguientes variables de estudio:

Edad2: variable cualitativa ordinal, clasificada según la Organización Mundial de la Salud (Wikipedia, 2019) en:

-Adolescente: de 10 a 19 años.

-Joven: de 20 a 24 años.

-Adulta: de 25 años en adelante.

Peso.bebe: variable cuantitativa continua, indica el peso del bebé al nacer, en gramos.

Primero se efectuó un análisis descriptivo sobre el peso de los bebés recién nacidos, los cuales fueron clasificados de acuerdo a la edad de la madre considerándose las siguientes variables: “Peso en gramos de un bebé recién nacido, con madre adolescente, atendido en el Baystate Medical Center de la ciudad de Springfield, Massachusetts, durante 1986” y “Peso en gramos de un bebé recién nacido, con madre

adulta, atendido en el Baystate Medical Center de la ciudad de Springfield, Massachusetts, durante 1986”.

Luego, se realizó un análisis inferencial. Se utilizó el software R a través de su interfaz gráfica R Commander (versión 3.6.0). (R Core Team, 2019).

Para llevar a cabo el objetivo, primero se utilizó una prueba chi cuadrado de bondad de ajuste con un nivel del 5% para poner a prueba la normalidad de las variables analizadas.

Además se realizó un test de hipótesis para la diferencia de medias de poblaciones independientes, con un nivel del 5%.

Resultados y Discusiones

A continuación se muestran los resultados obtenidos a partir del análisis descriptivo e inferencial realizado sobre las muestras correspondientes a cada una de las variables de estudio.

Consideramos las variables:

X= Peso en gramos de un bebé recién nacido, con madre adolescente, atendido en el Baystate Medical Center de la ciudad de Springfield, Massachusetts, durante 1986.

Y= Peso en gramos de un bebé recién nacido, con madre adulta, atendido en el Baystate Medical Center de la ciudad de Springfield, Massachusetts, durante 1986.

De las medidas de resumen obtenidas con Rcmdr, se puede observar:

	mean	sd	cv	0%	25%	50%	75%	100%	peso.bebe:n
adolescente	2973.529	612.4235	0.2059585	1885	2480.5	2920	3487	4238	51
adulta	2946.130	840.4972	0.2852885	709	2466.0	2948	3600	4990	69

Tabla 1: media, desvío, coeficiente de variación, cuartiles de las variables X e Y

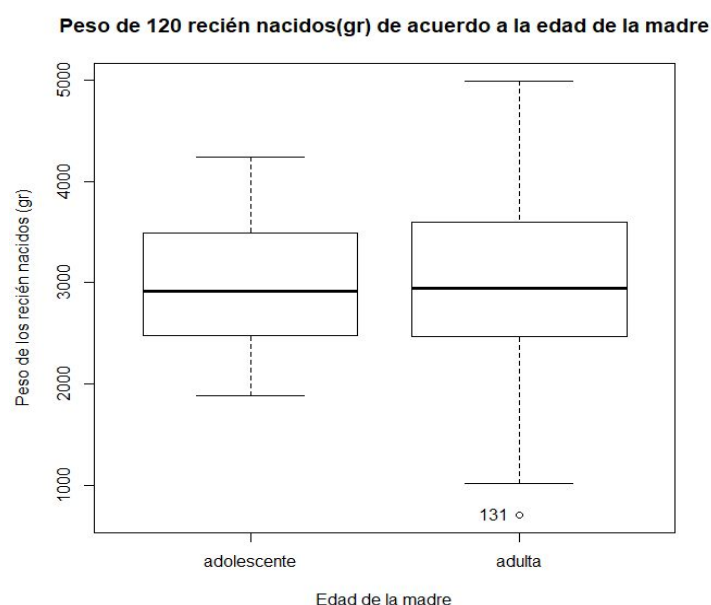


Imagen 1: boxplot de los pesos en gramos de los bebés recién nacidos según la edad de la madre

Observando la Imagen 1, podemos ver que estos diagramas de caja y brazos en paralelo, permiten comparar la distribución de los pesos de los bebés recién nacidos de madres adolescentes y el peso de los bebés recién nacidos de madres adultas. De esta manera, observamos que los pesos de los bebés de las madres adolescentes están más concentrados, ya que los datos se encuentran más agrupados, es decir, se encuentran más cercanos unos de otros. En el caso de los pesos de los bebés recién nacidos de madres adultas, se puede observar que hay casos más atípicos (por ejemplo, el dato en la posición 131) por ser que los brazos tienen mayor longitud. Además, se puede observar que las medianas en ambos casos están cercanas a los 3000 gramos.

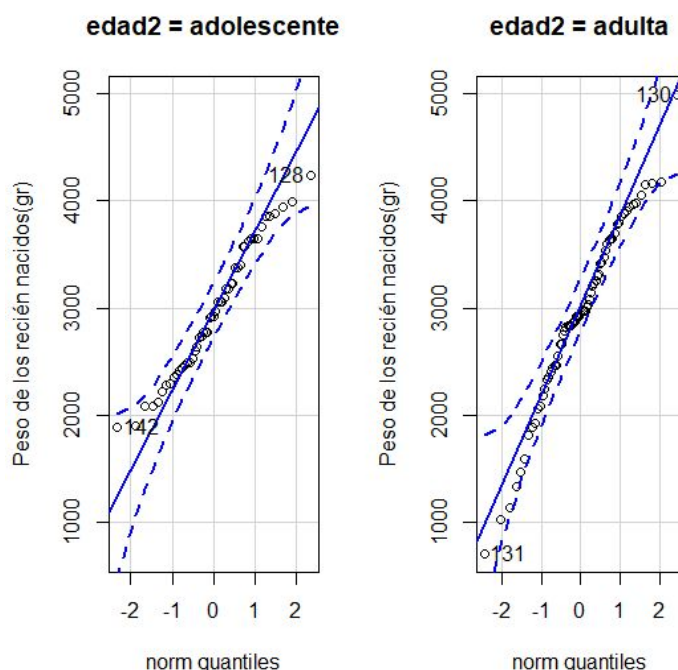


Imagen 2: QQ-plot de los pesos en gramos de los bebés recién nacidos según la edad de la madre

En los gráficos de la Imagen 2, se puede observar que los puntos (conformados por los cuantiles de las muestras y los cuantiles teóricos de la distribución normal estándar), se encuentran distribuidos próximos a la línea recta, por lo que aparenta tener una distribución normal.

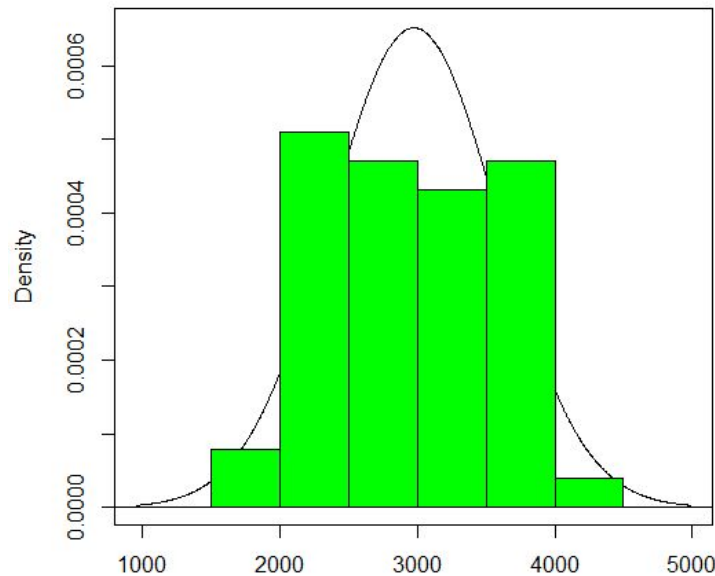


Imagen 3: histograma de los pesos en gramos de los bebés recién nacidos con madre adolescente

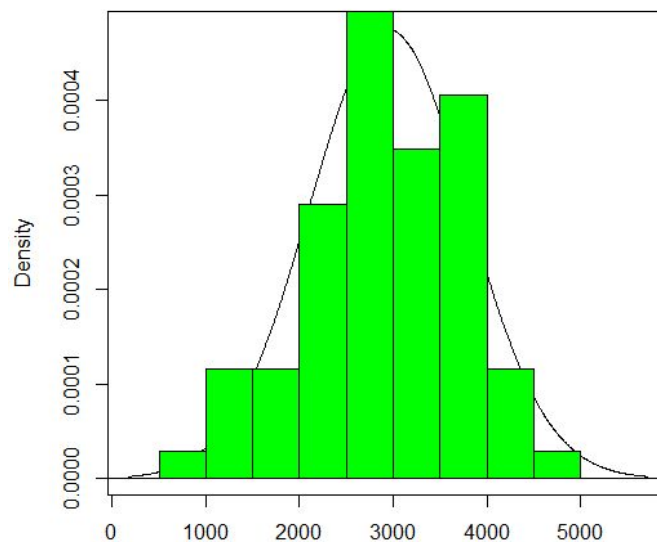


Imagen 4: histograma de los pesos en gramos de los bebés recién nacidos con madre adulta

De analizar las Imágenes 3 y 4, podemos ver que al ajustar la curva de una normal ambas aparentan seguir una distribución normal con los parámetros estimados, para la imagen 3: $\mu = 2973,529$ y $\sigma^2 = 612,4235$; para el caso de la imagen 4: $\mu = 2946,13$ y $\sigma^2 = 840,4972$.

Luego del análisis realizado sobre las imágenes 1, 2, 3 y 4, vamos a probar si las variables que nombramos anteriormente siguen, como suponemos, una distribución normal con los parámetros propuestos a continuación:

Como la media y el desvío poblacional son desconocidos, vamos a considerar como estimadores para estos parámetros:

$$\hat{\mu} = \bar{X} \quad ; \quad \hat{\sigma}^2 = s^2$$

Planteamos la prueba de bondad de ajuste (Dieser, 2019), para decidir si los conjuntos de datos se ajustan a la distribución normal con los parámetros estimados.

Vamos a poner a prueba las siguientes hipótesis:

$$H_0 : X \sim N(\mu; \sigma^2) \quad \text{vs} \quad H_a : X \text{ no sigue la distribución propuesta}$$

Donde utilizamos μ y σ^2 estimados por:

$$\hat{\mu} = 2973,529 \quad ; \quad \hat{\sigma}^2 = 612,4235^2$$

Se realizó la siguiente tabla para analizar las frecuencias observadas y esperadas, de los pesos en gramos de los bebés recién nacidos con madre adolescente.

C_i	Clases	Frecuencias observadas (o_i)	Probabilidades	Frecuencias esperadas (e_i)
i=1	$(-\infty ; 2000]$	2	0,0560	2,856
i=2	$(2000 ; 2500]$	13	0,1637	8,3487
i=3	$(2500 ; 3000]$	12	0,2975	15,1725
i=4	$(3000 ; 3500]$	11	0,2878	14,6778
i=5	$(3500 ; 4000]$	12	0,1481	7,5531
i=6	$(4000 ; \infty)$	1	0,0469	2,3919
	Total	51	1	51

Tabla 2: distribución de frecuencias de los pesos en gramos de los bebés recién nacidos con madre adolescente

Como hay clases en las cuales las frecuencias esperadas resultaron menores a 5 se deben agrupar:

C_i	Clases	Frecuencias observadas (o_i)	Probabilidades	Frecuencias esperadas (e_i)	$\frac{(o_i - e_i)^2}{e_i}$
i=1	$(-\infty ; 2500]$	15	0,2197	11,2047	1,2856
i=2	$(2500 ; 3000]$	12	0,2975	15,1725	0,6634
i=3	$(3000 ; 3500]$	11	0,2878	14,6778	0,9215
i=4	$(3500 ; \infty)$	13	0,195	9,945	0,9385
	Total	51	1	51	3.809

Tabla 3: distribución de frecuencias y desvíos cuadráticos de los pesos en gramos de los bebés recién nacidos con madre adolescente

Luego:

$$t = \sum_{i=1}^4 \frac{(o_i - e_i)^2}{e_i} = 1,2856 + 0,6634 + 0,9215 + 0,9385, \text{ por lo tanto, } t = 3,809$$

Vamos a calcular la región de rechazo con un nivel de confianza de $\alpha = 0,05$

Además tengamos en cuenta para el grado de libertad de la Chi-Cuadrado, que en la tabla 3 tenemos 4 clases, y estimamos 2 parámetros, por lo que nos queda de 1 grado de libertad.

$$RR = \{T \geq \chi^2(4 - 2 - 1)_{0.05}\}$$

$$RR = \{T \geq \chi^2(1)_{0.05}\}$$

$$RR = \{T \geq 3,84\}$$

Como obtuvimos que $t = 3,809$, no pertenece a la región de rechazo, es decir que no tenemos suficiente evidencia para rechazar H_0 , por lo tanto podemos afirmar a un nivel del 5% de confianza que

$$X \sim N(2973,529 ; 612,4235^2)$$

Ahora veamos si la variable Y tiene distribución con parámetros estimados por la media y el desvío muestral.

Vamos a poner a prueba las siguientes hipótesis:

$$H_0 : Y \sim N(\mu ; \sigma^2) \text{ vs } H_a : Y \text{ no sigue la distribución propuesta}$$

Donde utilizamos μ y σ^2 estimados por:

$$\hat{\mu} = 2946,130 ; \hat{\sigma}^2 = 840,4972^2$$

Se realizó la siguiente tabla para analizar las frecuencias observadas y esperadas, de los pesos en gramos de los bebés recién nacidos con madres adultas.

C_i	Clases	Frecuencias observadas (o_i)	Probabilidades	Frecuencias esperadas (e_i)
i=1	$(-\infty ; 1000]$	1	0,0103	0,7107
i=2	$(1000 ; 1500]$	4	0,0324	2,2356
i=3	$(1500 ; 2000]$	4	0,0875	6,0375
i=4	$(2000 ; 2500]$	10	0,1676	11,5644
i=5	$(2500 ; 3000]$	19	0,2278	15,7182
i=6	$(3000 ; 3500]$	12	0,2194	15,1386
i=7	$(3500 ; 4000]$	14	0,1501	10,3569
i=8	$(4000 ; 4500]$	4	0,0727	5,0163

i=9	(4500 ; ∞)	1	0,0322	2,2218
	Total	69	1	69

Tabla 4: distribución de frecuencias de los pesos en gramos de los bebés recién nacidos con madre adulta

Como hay clases en las cuales las frecuencias esperadas resultaron menores a 5 se deben agrupar:

C_i	Clases	Frecuencias observadas (o_i)	Probabilidades	Frecuencias esperadas (e_i)	$\frac{(o_i - e_i)^2}{e_i}$
i=1	($-\infty$; 2000]	9	0,1302	8,9838	0.00003
i=2	(2000 ; 2500]	10	0,1676	11,5644	0,2116
i=3	(2500 ; 3000]	19	0,2278	15,7182	0,6852
i=4	(3000 ; 3500]	12	0,2194	15,1386	0,6507
i=5	(3500 ; 4000]	14	0,1501	10,3569	1,2815
i=6	(4000 ; ∞)	5	0,1049	7,2381	0,6920
	Total	69	1	69	3.5210

Tabla 5: distribución de frecuencias y desvíos cuadráticos de los pesos en gramos de los bebés recién nacidos con madre adulta

Luego, calculando t resulta:

$$t = \sum_{i=1}^6 \frac{(o_i - e_i)^2}{e_i} = 0,00003 + 0,2116 + 0,6852 + 0,6507 + 1,2815 + 0,6920$$

$$t = 3,5210$$

Vamos a calcular la región de rechazo con un nivel de confianza de $\alpha = 0,05$

Además tengamos en cuenta para el grado de libertad de la Chi-Cuadrado, que en la tabla 5 tenemos 6 clases, y estimamos 2 parámetros, por lo que nos queda de 3 grado de libertad.

$$RR = \{T \geq \chi^2(6 - 2 - 1)_{0,05}\}$$

$$RR = \{T \geq \chi^2(3)_{0,05}\}$$

$$RR = \{T \geq 7,815\}$$

Como obtuvimos que $t = 3,5210$, no pertenece a la región de rechazo, es decir que no tenemos suficiente evidencia para rechazar H_0 , por lo tanto podemos afirmar a un nivel del 5% de confianza que $Y \sim N(2946, 130 ; 840, 4972^2)$.

Vamos a testear si el peso de un bebé recién nacido con madre adolescente difiere, en media, del peso de un bebé recién nacido con madre adulta.

$$H_0 : \mu_X - \mu_Y = 0 \quad \text{vs} \quad H_a : \mu_X - \mu_Y \neq 0 \quad (1)$$

Como no conocemos las varianzas poblacionales, primero analizaremos si X e Y tienen varianzas iguales o distintas realizando un TH para el cociente de varianzas (Dieser, 2019) de poblaciones normales.

$$H_0 : \frac{\sigma_2^2}{\sigma_1^2} \leq 1 \quad \text{vs} \quad H_a : \frac{\sigma_2^2}{\sigma_1^2} > 1$$

teniendo en cuenta que $\hat{\sigma}_2^2 > \hat{\sigma}_1^2$.

$$RR = \left\{ \frac{s_2^2}{s_1^2} > F(68, 50)_{0,05} \right\}$$

$$RR = \left\{ \frac{s_2^2}{s_1^2} > 1,56 \right\}$$

Como $\frac{s_2^2}{s_1^2} = 1,88$, pertenece a la región de rechazo, tenemos suficiente evidencia a un nivel del 0,05 para rechazar H_0 , por lo tanto, las varianzas σ_1^2 y σ_2^2 son distintas.

Como rechazamos H_0 , vamos a calcular un intervalo de confianza para el cociente de varianzas de estas poblaciones (Dieser, 2019).

$$IC \left(\frac{\sigma_2^2}{\sigma_1^2} : 95\% \right) = \left[\frac{840,4972^2}{612,4235^2 F(68; 50)_{0,025}}; \frac{840,4972^2}{612,4235^2 F(68; 50)_{0,975}} \right]$$

$$IC \left(\frac{\sigma_2^2}{\sigma_1^2} : 95\% \right) = \left[\frac{840,4972^2}{612,4235^2 * 1,70}; \frac{840,4972^2}{612,4235^2 * 0,60} \right]$$

$$IC \left(\frac{\sigma_2^2}{\sigma_1^2} : 95\% \right) = [1, 108 ; 3, 139]$$

Como las varianzas son distintas (Dieser, 2019), para testear (1), vamos a construir un TH para la diferencias de medias de poblaciones normales, con varianzas distintas:

$$\eta = \frac{\left(\frac{612,4235^2}{51} + \frac{840,4972^2}{69} \right)^2}{\frac{\left(\frac{612,4235^2}{51} \right)^2}{50} + \frac{\left(\frac{840,4972^2}{69} \right)^2}{68}}$$

$$\eta = 117,98$$

Redondeando al entero más bajo: $\eta = 117$

$$RR = \left\{ |\bar{X} - \bar{Y}| > t(117)_{0,025} \sqrt{\frac{612,4235^2}{51} + \frac{840,4972^2}{69}} \right\}$$

Calculando $\pm t(117)_{0,025} = \pm 1,98$

$$RR = \left\{ |\bar{X} - \bar{Y}| > 262,62 \right\}$$

Como $|\bar{x} - \bar{y}| = 27,4$, no pertenece a la región de rechazo. Por lo tanto no hay suficiente evidencia para rechazar H_0 a un nivel del 0,05. Entonces las medias poblacionales no son distintas.

Conclusiones

Con el análisis realizado sobre la muestra de los pesos de 120 bebés recién nacidos en el Baytaste Medical Center de la ciudad de Springfield, de los cuales según la clasificación utilizada en las edades,

51 tienen madres adolescentes y 69 adultas; podemos concluir que en base a estos datos la edad extrema de las madres caracterizadas como adolescentes y adultas no influye en el peso de los bebés.

Lo que se puede plantear para investigar a futuro, es analizar si los pesos de los bebés recién nacidos de madres adolescentes o madres adultas, difieren del peso de los bebés recién nacidos con madres que se encuentran en la etapa considerada “ideal” para ser madre, es decir, las que según la clasificación de la OMS (Organización Mundial de la Salud) serían las jóvenes.

Referencias

- Dieser, M.P. (2019). *Estadística para futuros matemáticos. Notas de Clases*. [material no publicado]. Facultad de Ciencias Exactas y Naturales. Universidad Nacional de La Pampa.
- Hosmer, D., y Lemeshow, S. (1989). *Regresión logística aplicada*. Wiley series en probabilidad y estadística matemática. Nueva York: Wiley.
- Lleó, V. (n.d.). *De qué depende el peso del bebé al nacer*. [blog]. Recuperado de <https://www.guiadelnino.com/embarazo/desarrollo-fetal/de-que-depende-el-peso-del-bebe-al-nacer> [10 de noviembre de 2019]
- R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- StatCrunch (2012). *Estudio del bajo peso al nacer*. [online] Disponible en <http://www.statcrunch.com/5.0/shareddata.php?keywords=weighing> [5 de Noviembre de 2019].
- Wikipedia (2019). *Juventud*. Recuperado de <https://es.wikipedia.org/wiki/Juventud> [7 de noviembre de 2019]

IDENTIFICACIÓN DE LOS CANGREJOS *LEPTOGRAPSUS VARIEGATUS* SEGÚN EL ANCHO DEL CAPARAZÓN

Santiago Acosta - Ximena Sofía Sanchez

santiago_a_03@hotmail.com; ximenasofiasanchez@gmail.com

Resumen

En este estudio nos centramos en ver si, a través de una variable es posible identificar a un cangrejo *Leptograpsus variegatus* de Australia Occidental de dos especies, naranja y azul; realizamos un análisis descriptivo e inferencial para ver si el ancho del caparazón efectivamente diferenciaba a las especies.

Introducción

Leptograpsus variegatus, también conocido como **jaiva corredora**, es una especie de cangrejo marino de la familia grapsidae, muy abundante en el litoral del archipiélago Juan Fernández (Wikipedia, 16 de Sep de 2019). Es un conjunto de islas ubicado en el Pacífico Sur, a más de 670 km de América del Sur (Wikipedia, 11 de Nov de 2019).

Mahon (1973), utilizando estudios ecológicos que ampliaron los de Shield (1959), y un análisis genético basado en un estudio electroforético, estableció la distinción específica de los cangrejos de roca de las formas azul y naranja del género *Leptograpsus* que se producen en las costas de Australia.

La conclusión de que, las dos formas de color son en realidad dos especies separadas ha hecho necesario examinar para determinar si esto también ocurre para las dos especies de Australia Occidental.

Ya que los cangrejos que se han conservado químicamente pierden su color distintivo, en un intento por resolver este problema de identificación, Campbell & Mahon (1974) realizaron un estudio morfológico de las especies de Australia Occidental. Se ha utilizado un enfoque **multivariante** para estudiar la variación morfológica en las especies de formas azules y naranjas de cangrejo de roca del género *Leptograpsus*. Bajo el supuesto de que las especies hermanas son casi indistinguibles en apariencia externa. La identificación de los individuos por medidas simples como las proporciones es poco probable que tenga éxito en algunas especies (Davidson, 1967, et al.).

Notaron un contraste entre el ancho del caparazón en relación con el ancho del lóbulo frontal y la profundidad del cuerpo; la especie de forma azul tiene un ancho de caparazón relativo mayor que la especie naranja.

Nuestro objetivo se basa en ver si tomando una variable existe una diferencia significativa del ancho del caparazón de los cangrejos, que nos permita identificar las especies. Tomando como muestra 100 cangrejos de la especie azul y 100 cangrejos de la especie naranja. También nos proponemos ver si los intervalos del ancho de cada especie se diferencian, para identificar una u otra especie.

Metodología

Para este estudio, nos basaremos en un muestreo aleatorio simple realizado por Campbell & Mahon (1974), quienes realizaron un estudio sobre cinco mediciones morfológicas de los cangrejos de dos colores diferentes y de ambos sexos de las especies *Leptograpsus variegatus*. Los datos fueron recogidos en Fremantle, Australia Occidental.

El conjunto original cuenta con 200 datos y 7 variables. A cada cangrejo muestreado se le determinó: la especie (A: azul, N: naranja), el sexo, el tamaño del lóbulo frontal (mm), la anchura trasera (mm), la

longitud y el ancho del caparazón (mm), y la profundidad del cuerpo (mm).

Se tomaron en nuestro trabajo dos de las siete variables dadas, ya que buscamos ver si una característica morfológica de los cangrejos nos permite diferenciar la especie.

Entonces trabajamos con las variables: **especie** (A: azul, N: naranja) de tipo cualitativa nominal; y **ancho del caparazón** (mm) de tipo cuantitativa continua.

Se desarrolló un análisis descriptivo sobre las variables involucradas y se aplicaron técnicas inferenciales utilizando el software estadístico R, (versión 3.6.1) y su interfaz RCommander (versión 2.5-3).

Para llevar a cabo el objetivo, primero se utilizó una prueba de bondad de ajuste sobre la variable ancho del caparazón en cada especie con un nivel de significación del 5% para poner a prueba su normalidad.

Luego se realizó un test de hipótesis para la diferencia de medias sobre el conjunto de los anchos del caparazón de las especies azul y naranja, junto con un test de cociente de varianzas ya que eran desconocidas. Para identificar si existe o no una diferencia significativa entre el ancho del caparazón de cada especie.

Además buscamos los intervalos de confianza para la media de cada especie con un nivel de confianza del 95%, para que al momento de medir podamos identificar de qué especie se trata.

Resultados

Teniendo en cuenta el conjunto de datos utilizado, tuvimos en cuenta dos variables

X= ancho del caparazón de la especie azul (mm)

Y= ancho del caparazón de la especie naranja (mm)

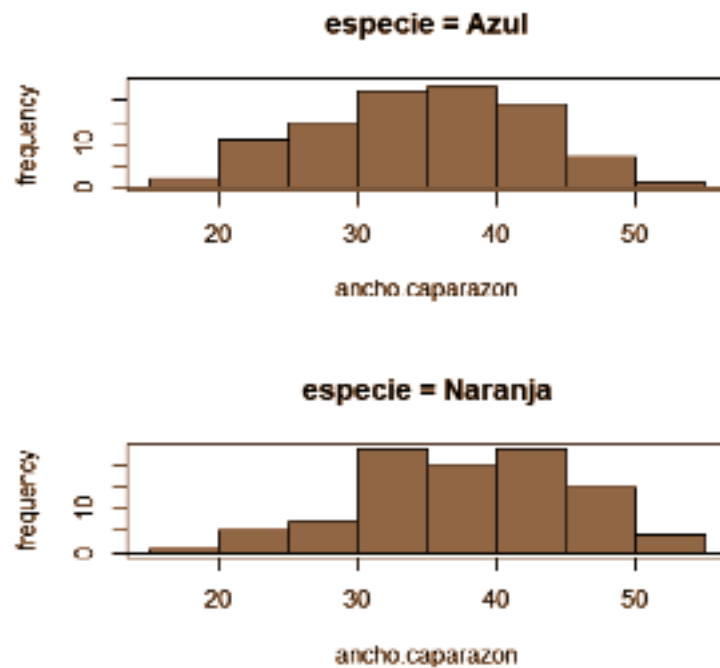


Figura 1: Gráficos de frecuencias del ancho del caparazón en mm.

En la Figura 1 podemos observar que la mayor frecuencia de especie azul ronda entre los 35 a 40 mm y la especie naranja tiene su mayor frecuencia entre los 30 a 35 mm y entre los 40 a 45 mm.

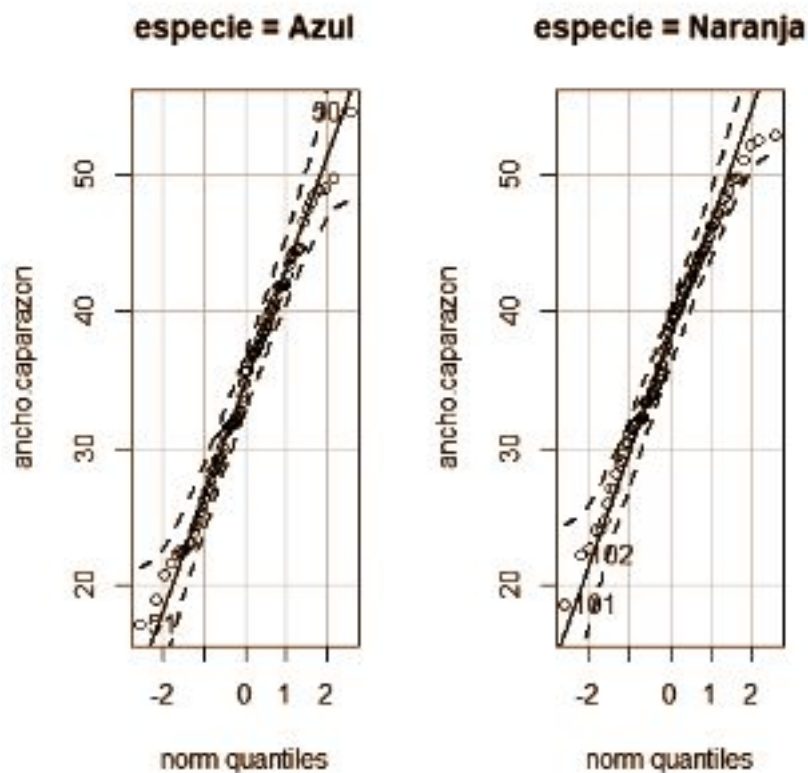


Figura 2: QQ-plot del ancho del caparazón de cada especie.

En ambos QQ-plot los puntos se acumulan, en su gran mayoría, alrededor de una recta, lo que nos da un indicio de que los datos provienen de una distribución normal.

Se realizó un test de bondad de ajuste chi-cuadrado en ambas variables para probar la normalidad de estas, utilizando media y varianza estimadas a partir de la muestra, siendo el estimador del primer momento y máxima verosimilitud de la media muestral $\mu = \bar{X}$ el cuál es insesgado, eficiente y suficiente; y de la varianza muestral $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ el cuál es asintóticamente insesgado, pero podemos obtener otro estimador insesgado para σ^2 , de la siguiente manera $\frac{n}{n-1} \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = s^2$.

El resultado que se obtuvo en ambos test de bondad de ajuste fue que no se rechazaba la hipótesis nula, por lo tanto teníamos suficiente evidencia para decir que ambas variables pueden modelarse bien con una distribución normal.

Estas variables se modelan:

$X \sim N(\mu_1; \sigma_1^2)$ siendo $X_1; \dots; X_{100}$ una m.a.s. de X

$Y \sim N(\mu_2; \sigma_2^2)$ siendo $Y_1; \dots; Y_{100}$ una m.a.s. de Y

Como no conocemos los parámetros poblacionales los estimaremos a partir de la muestra, utilizando Rcommander.

Variables	Media muestral	Desvío muestral
X	34.717	7.866038
Y	38.112	7.540922

Se comparan las medias del ancho del caparazón de las especies. Para esto, ya que las varianzas son desconocidas, debemos analizar si son iguales o no con un test para el cociente de varianzas de poblaciones normales para muestras aleatorias independientes (F de Fisher (99,99) grados de libertad) con un Rango de Rechazo $\left(\left\{ \frac{s_1^2}{s_2^2} > 1.394 \right\} \right)$, luego $\frac{s_1^2}{s_2^2} = 1.08$ no pertenece a la Región de Rechazo no se pudo rechazar la hipótesis, luego las varianzas son iguales.

Luego se realizó un TH para la diferencia de medias donde nuestras hipótesis fueron:

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{vs} \quad H_a : \mu_1 - \mu_2 \neq 0$$

donde $\hat{\mu}_1 = 34.717$ y $\hat{\mu}_2 = 38.112$, el estadístico tiene distribución t de Student con 198 grados de libertad, con una probabilidad de 0.05 y el desvío ponderado es 7.7, la región de rechazo es $RR = \left\{ \bar{X} - \bar{Y} < -2.15 \right\} \cup \left\{ \bar{X} - \bar{Y} > 2.15 \right\}$, con $\bar{x} - \bar{y} = -3.395$ por lo cual podemos rechazar H_0 , es decir las medias son distintas. Por lo tanto se construyó el IC correspondiente,

$IC(\mu_1 - \mu_2; 95\%) = [-5.543866; -1.246134]$, donde el intervalo no contiene al cero, es decir ratificamos nuestra conclusión anterior. Aún más, en media el ancho del caparazón de la especie naranja es mayor al de la especie azul.

Como un objetivo derivado se realizaron los Intervalos de confianza con un nivel del 95% para las medias de cada variable, por RCommander obtuvimos :

$IC(\mu_1 : 95\%) = (33.15621; 36.27779)$

$IC(\mu_2 : 95\%) = (36.61572; 39.60828)$

Podemos decir que el 95% de esos intervalos contendrán respectivamente a los parámetros , es decir, estamos 95% seguros que las medias poblacionales de cada especie están dentro de esos intervalos.

Discusión

En el trabajo anteriormente realizado por Campbell & Mahon (1974) en el cual se realizó un análisis multivariado concluyeron que el ancho del caparazón en relación con la profundidad del cuerpo y el lóbulo frontal de la especie azul era mayor al de la especie naranja.

En nuestro trabajo realizamos un análisis univariado y concluimos todo lo contrario al trabajo anteriormente realizado, es decir, el ancho del caparazón de la especie naranja es mayor al de la especie azul.

Una suposición de esta diferencia puede deberse que al no considerar más variables en conjunto, se puede perder información de la muestra, es decir, las variables pueden guardar alguna relación que, al nosotros tomar solamente dos. no estamos teniendo en cuenta.

Conclusiones

Luego de este análisis se concluye a un nivel de confianza del 95% que, de las muestras tomadas en Australia Occidental, en media el ancho del caparazón de la especie naranja es mayor al ancho del caparazón de la especie azul y, que a través de los intervalos de confianza nos permitimos identificar cada especie.

Referencias

- Campbell, N.A. & Mahon, R.J (1974). *A Multivariate Study of Variation in two species of rock crab of the genus Leptograpsus*. Australian Journal of Zoology, 22, 417-425
- Davidson, G., et al. (1967). *The Anopheles gambiae complex*. In 'Genetics of Insect Vectors of Disease'. (Eds J. W. Wright and R. Pal.) (Elsevier: Amsterdam.)
- Shield, J. W. (1959). *Population studies in the littoral at Rottnest Island*. In 'Rottnest Island: The Rottnest Biological Station and Recent Scientific Research'. (Eds E. P. Hodgkin and K. Sheard.) J. R. Soc. West. Aust. 42, 65-95
- Wikipedia (11 de Nov de 2019). *Archipiélago Juan Fernández*, (nd). Disponible en: https://es.wikipedia.org/wiki/Archipi%C3%A9lago_Juan_Fern%C3%A1ndez. [29 de Nov de 2019]
- Wikipedia (16 de Sep de 2019). *Leptograpsus variegatus*, (nd). Disponible en: https://es.wikipedia.org/wiki/Leptograpsus_variegatus . [16 de Nov de 2019]



¿PESO IDEAL O PESO REAL?

Análisis estadístico del peso de los gatos

Deborah Ramburger-Fanny Maier-Giuliana Olivera

deborahramburger@gmail.com fannymaier7@gmail.com oliveragiuliana1@gmail.com

Resumen

La delgadez extrema y el sobrepeso son frecuentes en los gatos, debido a una alimentación no adecuada. Los datos recolectados consistieron en el peso corporal en kilogramos de 144 gatos, de los cuales 47 son hembras y 97 son machos. El objetivo de este trabajo fue corroborar o refutar el peso medio ideal considerado por la página oficial de Purina (2019) de (3.6 kg; 4.5 kg); y analizar el vínculo entre el peso corporal de los gatos y su respectivo sexo. Obteniendo así un intervalo de confianza del peso medio corporal de [2.6180;2.8292], con lo que pudimos refutar una de las hipótesis planteadas; además para analizar si existía un vínculo entre el peso corporal y su respectivo sexo realizamos el test de independencia, obteniendo así un “alto” grado de asociación entre las variables de interés.

Introducción

Según la página oficial de Purina (2019), la preocupación por nuestros gatos domésticos adultos lleva a preguntarnos en ocasiones si se encuentran bien de salud, además de las revisiones en el veterinario y de los cuidados que les procesamos a diario, es importante controlar su alimentación para que tengan el peso adecuado.

Conocer el peso ideal de un gato es tan complicado como saber cuál es el de un humano. El problema del sobrepeso o de la delgadez extrema del felino no es ajeno, son trastornos nutricionales frecuentes en los gatos. Principalmente, el sobrepeso se debe a que el gato se ha convertido en una mascota de ciudad, lo que provoca cambios en su estilo de vida y sus necesidades energéticas. Si se suma además, que muchos de ellos están castrados y que tienen un acceso limitado (o incluso, nulo) al exterior, estos cambios son aún más notorios. Pero también influye la relación con sus dueños, ya que suele confundirse "atención" con "alimentación" (Royal Canin, 2019).



Figura 1: Gato doméstico adulto.

Para desarrollar un análisis conveniente debemos conocer que en los gatos se considera una presencia de sobrepeso (o de delgadez) cuando su peso corporal supera (o está por debajo) en más (o menos) del 10% su “peso corporal ideal” (Lund et al., 2005); tomando como peso ideal correspondiente a 3,6 kilogramos y 4,5 kilogramos, según la página oficial de Purina (2019).

Además, Lund et al. (2005) especifica que el sexo es un factor predisponente en el peso de los gatos, con lo que veremos más adelante si realmente es así.

Se debe remarcar que nuestro análisis consistirá únicamente en el peso corporal en kilogramos de los gatos (hembras y machos) para respondernos las siguientes preguntas:

“¿El peso medio de los gatos difiere del ideal?”

“¿Existe relación entre el peso de los gatos y su respectivo sexo?”

Teniendo en cuenta las siguientes hipótesis: “El peso medio de los gatos difiere del ideal” y “Existe

relación entre el peso de los gatos y su respectivo sexo”; para esto, nuestros objetivos son: corroborar o refutar el peso medio ideal considerado (3.6 kg; 4.5 kg) y analizar el vínculo entre el peso corporal de los gatos y su respectivo sexo.

Metodología

Para este estudio nos basaremos en un muestreo aleatorio simple realizado por R. A. Fisher con el respaldo de la Universidad de Cambridge en el año 1947 (**Figura 2**).



Figura 2: Lugar de recolección de datos, Universidad de Cambridge (Inglaterra).

Los datos recolectados consistieron en el peso corporal en kilogramos y el peso del corazón en gramos, ambas variables cuantitativas continuas de 144 gatos. De los cuales 47 son hembras y 97 son machos, aquí destacamos la variable “sexo” clasificada en cualitativa nominal.

Para contestar nuestros interrogantes consideramos sólo dos de las variables originales, por un lado, Q = peso corporal (en kg) de los gatos; y por otro lado, S = sexo de los gatos (F-M).

Acorde a los datos disponibles, realizamos un análisis estadístico para determinar si el peso medio de los gatos difiere del ideal y, si existe relación entre el peso de los gatos y su respectivo sexo. Esta posible relación se analizó mediante una comparación de gráficos (**Figura 4**) establecidos uno junto al otro y respetando las escalas. Y además, para darle una fundamentación matemática se utilizó un test de independencia para el cual se transformó la variable “ Q ” en una variable cualitativa ordinal, considerando “bajo”, cuando los valores que tome la variable se encuentren por debajo o igual a 2.48; “moderado”, cuando los valores estén entre (2.48;2.96] y “alto” cuando los valores superen a 2.96; así, se cuenta con dos variables cualitativas con clases mutuamente excluyentes.

Para ello, se realizó inferencia estadística a través del software estadístico R 3.6.1 y de su interfaz gráfica R Commander 2.2-0, Microsoft Excel 2016 y Geogebra Classic 6, interpretamos los valores resultantes de la aplicación de las técnicas descriptivas y posteriormente extrajimos conclusiones y tomamos decisiones sobre la población.

Se supuso a la variable “ Q ” cuantitativa continua con una distribución normal, comprobado mediante técnicas descriptivas gráficas, como se ilustrará en la **Figura 3** y pruebas chi cuadrado de bondad de

ajuste.

Entonces, tenemos $Q_1, Q_2, \dots, Q_{144}$ una m.a.s de $Q \sim N(\mu, \sigma^2)$, por lo nombrado anteriormente. Los parámetros estimados fueron seleccionados en base a cualidades razonables tales como ser insesgado, eficiente, consistente y suficiente (Dieser y Cavero, 2019). Por consiguiente, la media poblacional se estimó mediante la media muestral de “Q” y el desvío estándar poblacional mediante el desvío estándar muestral de “Q”. Así, $\hat{\mu} = \bar{q}$ y $\hat{\sigma} = s$.

Realizando los cálculos de la media y desvío estándar muestrales mediante R (Versión 3.1.0), se pusieron a prueba las siguientes hipótesis:

$$H_0 : Q \sim N(2.723611; 0.4853066^2) \quad \text{vs.} \quad H_a : Q \text{ no sigue la distribución propuesta}$$

Para ello se consideró el estadístico de prueba $T = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i}$, con o_i : frecuencias observadas, e_i : frecuencias esperadas y k: cantidad de clases.

En nuestro caso, según la distribución de frecuencias realizadas con el conjunto de datos muestrales ingresados en R (Versión 3.1.0) se obtuvieron 7 clases. Además, utilizamos $m=2$, donde m representa la cantidad de parámetros que ha sido necesario estimar a partir de la muestra.

Primer interrogante

Para responder a la primera pregunta planteada: “¿El peso medio de los gatos difiere del ideal?” consideramos cada unidad observacional independiente de la otra, puesto que un gato muestreado no depende del otro.

Luego de verificar la distribución de la variable Q se trabajó con este supuesto para la realización de un test de hipótesis para la media de una población normal para comprobar o refutar el peso medio ideal señalado por la página oficial de Purina (2019)

$$H_0 : \mu = 4.05 \text{ kg} \quad \text{vs.} \quad H_a : \mu \neq 4.05 \text{ kg}$$

A partir del estadístico de prueba propuesto por Dieser (2019): $T = \frac{\bar{Q} - \mu}{\frac{s}{\sqrt{n}}} \sim t(n-1)$, con n=tamaño muestral, en nuestro caso, $n=144$.

Luego, fijando $\alpha = 0.01$ tenemos $0.01 = P(T > t(143)_{\alpha} | \mu = 4.05)$ y la Región de Rechazo (RR) nos permitió rechazar la hipótesis nula.

En consecuencia, se realizó un Intervalo de Confianza (IC) para establecer valores factibles para el peso medio de interés. definido el pivote $T = \frac{\bar{Q} - \mu}{\frac{s}{\sqrt{n}}}$ (Dieser, 2019, pág. 88) con $T \sim t(n-1)_{\alpha}$, en nuestro caso $T \sim t(143)_{0.01}$.

Segundo interrogante

Para el segundo interrogante “¿Existe relación entre el peso de los gatos y su respectivo sexo?”

Analizamos la existencia de alguna posible relación entre dos características diferentes como son el “sexo” y el “peso” de la población.

Suponiendo que las variables “S” y “Q” cualitativa ordinal no guardaban entre sí ningún tipo de relación, nos interesó conocer el grado de dependencia existente de modo que se pueda prever el resultado o comportamiento de una de ellas cuando conocemos la realización de la otra, basándonos en nuestra muestra.

Para poner a prueba la relación de las variables se realizó un test de independencia para el cual se desarrollaron tres tablas de doble entrada (**Tabla 2, Tabla 3 y Tabla 4**) y se utilizó la siguiente región de rechazo $RR = \{T \geq X^2((r-1)(c-1))_{\alpha}\}$ (Dieser, 2019, pág. 135) con r: cantidad de parámetros de filas y

c: cantidad de parámetros de columnas de la **Tabla 1**. Luego, se cuantificó el grado de dependencia, es decir, el grado de asociación entre el peso de los gatos y el sexo.

Resultados

A continuación, se muestran los resultados obtenidos en el análisis de cada una de las variables para responder a las preguntas planteadas.

En la **Figura 3** se presenta el histograma del peso (en kg) de los gatos sobre la curva representativa de la función de densidad de una variable con distribución normal de parámetros $\mu = 2.723611$ y $\sigma = 0.4853066$.

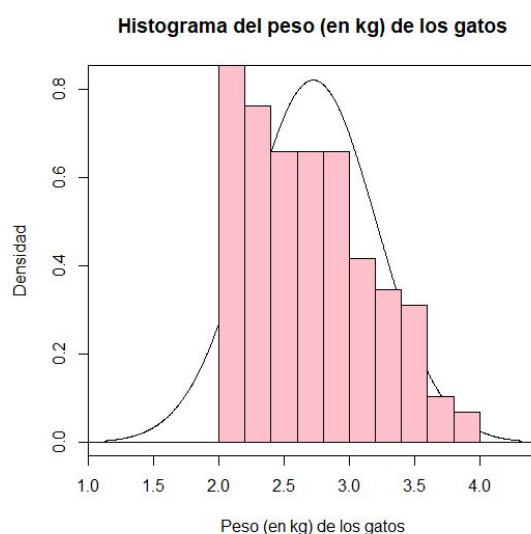


Figura 3: Histograma del peso (en gr) del peso de los gatos.

La figura anterior nos brindó una primera aproximación a la idea intuitiva con la que comenzamos. Al obtener una gráfica semejante a una campana, nuestra suposición es válida.

Para la realización del test de bondad de ajuste se tuvo en cuenta la **Tabla 1**, que consiste en el peso (en kg) de los gatos agrupados por clases basadas en la fórmula de Sturges (Dieser, 2019, pág. 8) con las frecuencias observadas y esperadas en cada una. Y también, los desvíos cuadráticos obtenidos con estas

dos tal que $T = \sum_{i=1}^7 \frac{(o_i - e_i)^2}{e_i} \sim X^2(4)$.

Clases	Frecuencias observadas	Frecuencias esperadas	Desvíos cuadráticos
≤ 2.24	29	22,9392	1,6013
$(2.24; 2.48]$	22	21,3552	0,0195
$(2.48; 2.72]$	31	27,2448	0,5176
$(2.72; 2.96]$	15	27,3456	5,5736
$(2.96; 3.20]$	17	21,6	0,9796
$(3.20; 3.44]$	16	13,4352	0,4896
$3.44 \leq$	14	10,08	1,5307
Total	144	144	10,7119

Tabla 1: Tabla de frecuencias observadas, esperadas y desvíos cuadráticos del peso (en kg) de los gatos.

Fijada una probabilidad $\alpha = 0.01$ de error de tipo I, tenemos una región de rechazo $RR = \{T \geq \chi^2(7 - 2 - 1)_{0.01}\} = \{T \geq 18.46683\}$.

Por lo tanto, sumando los desvíos cuadráticos relativos de la cuarta columna de la **Tabla 1** se tiene que $t = 10.7119 \notin RR$, luego no hubo suficiente evidencia al nivel de confianza de 1% para suponer que los datos no provienen de una distribución normal de parámetros estimados 2.723611 y 0.4853066^2 . En consecuencia, afirmamos que $Q \sim N(2.723611; 0.4853066^2)$.

Como contamos con una muestra aleatoria correspondiente a una población normal, construimos un Test de Hipótesis bilateral para la media de la población normal basados en el método clásico. Según la página oficial de Purina (2019) se afirma que el peso ideal de un gato varía entre 3.6 y 4.5 kg con lo que vamos a considerar como peso ideal el valor medio de ambos pesos, es decir $\frac{3.6 \text{ kg} + 4.5 \text{ kg}}{2} = 4.05 \text{ kg}$. Por lo tanto, tenemos:

$$H_0 : \mu = 4.05 \text{ kg}$$

vs .

$$H_a : \mu \neq 4.05 \text{ kg}$$

Entonces reemplazando en la Región de Rechazo (RR) obtuvimos:

$$RR = \{\bar{Q} < 4.05 - t(143)_{\frac{0.01}{2}} \frac{0.4853066}{\sqrt{144}}\} \cup \{\bar{Q} > 4.05 + t(143)_{\frac{0.01}{2}} \frac{0.4853066}{\sqrt{144}}\}$$

Con $t(143)_{\frac{0.01}{2}} = 2.610647$ tenemos:

$$RR = \{\bar{Q} < 3.9444\} \cup \{\bar{Q} > 4.1556\}$$

Por lo tanto, dado que $\bar{q} = 2.723611 \notin RR$ existe suficiente evidencia para rechazar H_0 con un nivel de confianza de 1%, es decir que el peso medio de los gatos no se encuentra entre los valores establecidos por la página oficial de Purina (2019).

En consecuencia, para nuestro caso:

$$IC(\mu : 99\%) = 2.723611 \pm t(143)_{\frac{0.01}{2}} \frac{0.4853066}{\sqrt{144}}$$

$$IC(\mu : 99\%) = [2.6180; 2.8292]$$

Estableciendo así, que el peso medio de los gatos ronda entre 2.6180 y 2.8292, con un coeficiente de confianza de 99%.

En consecuencia, obtenemos nuevos parámetros para el peso medio ideal de los gatos.

Existe una técnica descriptiva gráfica, donde se tienen dos variables (una cualitativa y otra cuantitativa) que permite comparar la relación, en nuestro caso el peso (en kg) de los gatos y el sexo (F-M). Tal como se muestra en la **Figura 4**.

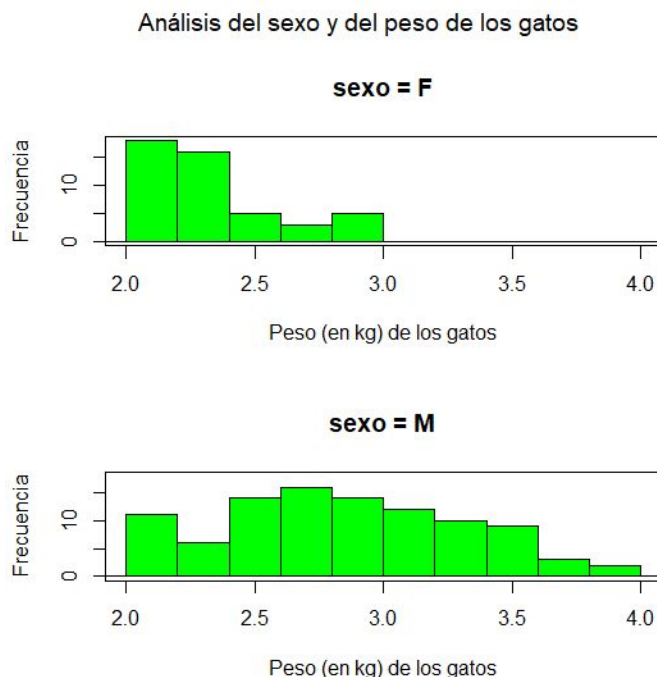


Figura 4: Histogramas del peso (en kg) de los gatos agrupados por sexo (F-M).

Los gráficos que se mostraron anteriormente nos permite ver que el peso (en kg) de las hembras rondan entre los dos kilogramos y los tres kilogramos, mientras que el peso (en kg) de los machos tiene un mínimo de dos kilogramos y un máximo de cuatro kilogramos. Entonces, intuitivamente podemos señalar que existe relación entre el peso (en kg) de los gatos y el sexo (F-M).

Ahora, para fundamentar matemáticamente la noción intuitiva hemos transformado el peso (en kg) de los gatos en una variable cualitativa ordinal según las categorías mutuamente excluyentes: Bajo: ≤ 2.48 ; Moderado: $(2.48; 2.96]$, Alto: > 2.96 .

Como muestra la **Tabla 2** hemos construido una tabla de contingencia en la que las cantidades o_{ij} representan las frecuencias observadas para la clasificación (i, j) , es decir número de casos en los que se observa la categoría i de la variable Q y la categoría j de la variable S.

Clases	Femenino	Masculino	Totales
Bajo	34	17	51
Moderado	11	35	46
Alto	2	45	47
Total	47	97	144

Tabla 2: Tabla de contingencia según el peso de los gatos y el sexo

El estadístico tiene distribución aproximadamente chi-cuadrado con $(3 - 1)(2 - 1)$ grados de libertad, que surge de considerar: $\sum_{i=1}^3 p_{i.} = 1$ por lo que se pierden $(3 - 1)$ grados de libertad al estimar esta

cantidad de parámetros de fila. Análogamente, $\sum_{j=1}^2 p_j = 1$ se pierden $(2 - 1)$ grados de libertad al estimar esta cantidad de parámetros de columna.

Con nuestra muestra obtuvimos la **Tabla 3** en la que se obtuvieron las frecuencias esperadas del peso de los gatos y el tipo de sexo. (Dieser, 2019, pág. 29)

Clases	Femenino	Masculino	Totales
Bajo	16,6458	34,3542	51
Moderado	15,0139	30,9861	46
Alto	15,3403	31,6597	47
Total	47	97	144

Tabla 3: Frecuencias esperadas según el peso de los gatos y el tipo sexo.

A partir de la **Tabla 2** y **Tabla 3**, se obtuvo la **Tabla 4** que contiene los desvíos cuadráticos relativos según el peso de los gatos y el tipo de sexo: $X^2 = \sum_{i=1}^3 \sum_{j=1}^2 \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$.

Clases	Femenino	Masculino	Totales
Bajo	18,0927	8,7666	26,8593
Moderado	1,0731	0,5199	1,593
Alto	11,6011	5,6211	17,2222
Total	30,7669	14,9076	45,6745

Tabla 4: Tabla de desvíos cuadráticos relativos según el peso de los gatos y el sexo.

Así, fijando el $\alpha = 0.01$ se obtuvo la región de rechazo dada por

$$RR = \{T \geq X^2((3 - 1)(2 - 1))_{0.01}\}$$

$$RR = \{T \geq 13.81551\}$$

Y, como sumando los desvíos cuadráticos relativos de la cuarta columna de la **Tabla 4** se obtuvo $t = 45.6745 \in RR$, tal que hay suficiente evidencia para suponer que las variables consideradas no son independientes.

Para cuantificar la dependencia calculamos el coeficiente de contingencia con nuestros datos:

$$C = \sqrt{\frac{t}{t+n}} = \sqrt{\frac{45.6745}{45.6745 + 144}} \approx 0.4907$$

cuya corrección arroja

$$C^* = \frac{C}{\sqrt{\frac{s-1}{s}}} = \frac{0.4907}{\sqrt{\frac{2-1}{2}}} = \frac{0.4907}{\sqrt{\frac{1}{2}}} \approx 0.6940,$$

Por lo tanto, el grado de asociación entre las variables de interés es “alto” y se comprueba la idea intuitiva del comienzo.

Discusión

Los datos de la sección anterior permiten hacer dos comparaciones. Por un lado, se podría contrastar el estudio previo de la página oficial de Purina (2019) sobre el peso medio ideal de los gatos; y por otro lado, frente a la relación de los gatos y el sexo, hay suficiente evidencia como para concordar con la investigación preliminar encontrada de Lund (2005).

La primera comparación nos permitiría rechazar el peso medio ideal propuesto por la página oficial de Purina (2019). Desde una primera discusión sospechábamos que el peso medio “ideal” no correspondía con el peso de los gatos muestreados; pues, según la página oficial los gatos de la muestra podrían considerarse no saludables, ya que estarían dentro de “nuevos parámetros ideales”. Cabe aclarar que el año del estudio de Purina es actual mientras que los datos que tomamos para nuestro análisis son de 1947 entonces si se hubiera realizado una comparación en el mismo año podríamos llegar a un acuerdo.

Con respecto a la relación del peso (en kg) de los gatos y el sexo, no se notaba claramente la dependencia entre ambas variables de interés, pero mediante los análisis estadísticos ya descriptos, al igual que Lund (2005), encontramos que podría ser cierto que el sexo es un factor que influye según el sexo de los gatos, por esto para la segunda comparación hubo concordancia.

Quedaría para una posible investigación conocer cuál de los dos sexos genera en los gatos domésticos adultos un mayor peso, si esta fuera la idea recomendaríamos realizar un Test de diferencia de medias.

Conclusiones

El objetivo de este estudio consistió en probar si los pesos medios de los gatos muestreados coinciden con el peso ideal propuesto por la página oficial de Purina (2019), y si existía relación entre el peso de los gatos y el sexo.

Luego de distintos análisis realizados se pudo concluir, con un nivel de confianza del 1%, que el peso medio de un gato ronda entre 2.6180 kg y 2.8292 kg, el cual no se encuentra entre los valores establecidos por el sitio web consultado. Además para el segundo cuestionamiento se demostró que las variables consideradas no son independientes, y cuyo grado de asociación entre ellas es “alto”: de 0.6940.

En general, podemos ver como el estereotipo de gato creado por las marcas de alimento generan en la población una excesiva alimentación en los animales que lleva a mayor consumo para las empresas pero perjudica a los felinos.

La principal causa de que el gato no esté en su peso ideal es la alimentación, pues se debe vigilar cuántas veces al día come y las porciones que se le da, y por supuesto no saturar de golosinas o en el peor de los casos darle restos de tu comida.

Referencias

- Dieser, M. P. (2019). *Estadística para futuros matemáticos*. La Pampa, Argentina: Universidad Nacional de La Pampa, Facultad de Ciencias Exactas y Naturales.
- Dieser, M. P. & Cavero, L. (2019). *Estimación puntual y propiedades*. La Pampa, Argentina: Universidad Nacional de La Pampa, Facultad de Ciencias Exactas y Naturales.
- Fisher, R. (1947). The analysis of covariance method for the relation between a part and the whole. Inglaterra: *Biometrics*, 3, pp. 65-68
- Hohenwarter, M. (2008). GeoGebra (Versión 6.0.518.0) [software].
- Ihaka, R.; Gentleman, R. (1993). R (Versión 3.1.0) [software].
- Lund E.M. , Armstrong P.J., Kirk C.A., et al. *Prevalence and risk factors for obesity in adult cats from private US veterinary practices*. Intern J Appl Res Vet Med 2005; 3: 88-96.
- Pibot, P., Biourge, V. & Elliott, D. (2008). *Enciclopedia de la Nutrición Clínica Felina*. Buenos Aires: Royal Canin.
- Purina (2019). *Conoce la tabla de peso ideal para gatos*. Disponible en <https://www.purina.es/proplan/consejos/tabla-peso-ideal-gatos> [18/11/19]
- Royal Canin (2019). *Riesgo para la salud de los gatos con sobrepeso y obesos*. Disponible en <https://www.royal-canin.com.ar/> [18/11/19]
- Simonyi, C. (1981). Microsoft Excel (Versión 2016) [software].

CALCULANDO EL RADIO DE LA TIERRA CON UN MÉTODO MILENARIO

Maidana, Sergio Nicolas - Ponkosky, Silvina Maribel- Seia, Fabian Gabriel

nicolasmaidana43@gmail.com-sponkosky@gmail.com-fabian_seia@hotmail.com

Resumen

En el año 240 aC, Eratóstenes, matemático, astrónomo, poeta, geógrafo y filósofo griego, se convirtió en el primer hombre en medir la circunferencia de la Tierra. Motivado por la curiosidad científica, por su espíritu sin límites y por un documento que encontró en la biblioteca de Alejandría, donde se aseguraba que en la ciudad de Siena (al sureste de Alejandría) una estaca clavada en el suelo no daba sombra a mediodía en el solsticio de verano, decide comenzar una serie de mediciones, con elementos de la época, que lo llevan a establecer datos medianamente precisos sobre la circunferencia de la Tierra.

En el siguiente trabajo se estimará el radio de la Tierra analizando descriptiva e inferencialmente 203 datos recolectados de la Argentina, utilizando el método de Eratóstenes.

Introducción

Consultando varias fuentes (Wikipedia, 2019a; Wikipedia, 2019b; Moreno V., 2019; Collazos D., 2018) se destacan los siguientes hechos de Eratóstenes.

Eratóstenes: (griego antiguo Ἐρατοσθένης) Nació en Cirene, 276 a. C. (hoy Shahhat, Libia) y falleció en Alejandría, 194 a. C.

Era hijo de Aglaos. Estudió en Alejandría y durante algún tiempo en Atenas. Fue discípulo de Aristón de Quíos, de Lisánias de Cirene y del poeta Calímaco y también gran amigo de Arquímedes.

Vivió en Atenas hasta que fue llamado a Alejandría (245 a.C.) para educar a los hijos de Ptolomeo III Evergetes. Para el año de 236 antes de Cristo, el faraón le pidió que esté a cargo de la Biblioteca de Alejandría, cosa que realizó a lo largo de toda su vida, teniendo la suerte de vivir hasta el día de su destrucción.

Poseía una gran variedad de conocimientos y aptitudes para el estudio. Matemático, astrónomo, poeta, geógrafo y filósofo, su apellido fue Pentathlos, nombre que se reservaba al atleta vencedor en las cinco competiciones de los Juegos Olímpicos. Se afirma que también era conocido como el segundo Platón y diversos autores dicen que se le daba el sobrenombre de Beta, por la segunda letra del alfabeto griego, porque ocupó el segundo lugar en todas las ramas de la ciencia que cultivó.

Pese a la multiplicidad de sus aportes, es muy poco lo que conocemos de su pensamiento debido a que gran parte de éste se hallaba almacenado en la Biblioteca de Alejandría, la cual fue destruida por un incendio. Aun así, sabemos que su poética se divide en dos publicaciones: Hermes, un poema que trata sobre astronomía y geografía; y Erigone, que recibió los elogios del mismo Longino. Se sabe, pese a las atribuciones apócrifas, que su primer trabajo sobre filosofía fue Platonikos, en donde estudia a Platón desde las matemáticas; que escribió Cronografía, uno de los primeros documentos históricos que se han realizado y que contenía la fecha de varios sucesos importantes; que analizó la obra de Cratino y de Aristófanes en su tratado titulado Sobre la antigua comedia ática, hablando del valor de las

declamaciones, los vestuarios y las decoraciones; y que hizo un estudio y una biografía de Homero, la cual no sobrevivió hasta la actualidad.

En su honor se ha nombrado un monte submarino ubicado en el mar Mediterráneo, un cráter de la Luna, un periodo dentro de la escala del tiempo geológico de nuestro satélite, y un asteroide proveniente del cinturón de asteroides de nuestro Sistema Solar.

Estando en la Biblioteca de Alejandría, encontró un informe de observaciones sobre Siena, ciudad situada a unos 800 km. al sur de Alejandría, en el que se decía que el día del solsticio¹ de verano (21 de junio) al mediodía, los objetos (como por ejemplo, los obeliscos) no producían sombra y en el fondo de los pozos de agua podía verse la luz del sol reflejándose. Eratóstenes observó que, en Alejandría, el mismo día y a la misma hora no se producía este mismo fenómeno, la proyección de los rayos solares generaban una sombra en el piso. Asumió de manera correcta que el Sol se encontraba a gran distancia y que sus rayos, al alcanzar la Tierra, lo hacían en forma paralela. Esto corroboraba su idea de que la superficie de la Tierra era curva, de haber sido plana, no se hubiese producido esta diferencia entre las dos ciudades. Acto seguido midió en Alejandría el ángulo que formaba los rayos del sol con la vertical de una torre, si tomamos los rayos del sol y la propagación de la luz hacia el fondo del pozo hasta el centro de la Tierra, dicho ángulo es el mismo del que se forma en el centro de la Tierra, ya que los ángulos alternos internos entre paralelas son de igual medida (ver **Figura 1**). Este ángulo resultó ser de 7° 12' (7,2°) que unido al hecho conocido de que la distancia entre las dos ciudades era de 5.000 estadios, dieron como conclusión que la circunferencia de la Tierra puede calcularse de la siguiente manera:

$$\frac{\text{Distancia de Siena a Alejandria}}{\text{Circunferencia de la Tierra}} = \frac{7,2^\circ}{360^\circ (\text{giro terrestre})}$$

Por lo tanto la circunferencia de la Tierra es

$$\frac{360^\circ \cdot 5000 \text{ estadios}}{7,2^\circ} = \text{Circunferencia de la Tierra}$$

$$250000 \text{ estadios} = \text{Circunferencia de la Tierra}$$

Aunque no se tienen datos exactos, se estima que un estadio equivale entre 158 y 160 metros. Por lo tanto, 250000 estadios son aproximadamente:

$$\frac{250000 \cdot 160}{1000} = 40000 \text{ km.}$$

Esto equivale a un radio de 6366 km. o 6286 si tomamos los 158m, contra los 6371 km. que son los admitidos hoy en día.

¹ Proviene del latín *sol* ("Sol") y *sistere* ("permanecer quieto") es cuando el semieje de un planeta, ya sea en el hemisferio norte o en el sur, está más inclinado hacia la estrella de su órbita. El Sol alcanza su posición más alta en el cielo.

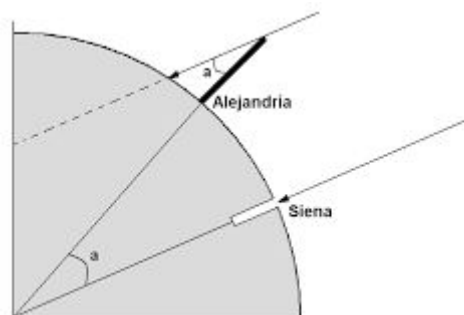


Figura 1: Ángulo “a” entre Alejandria y Siena, equivalente al calculado por Eratóstenes en Alejandria.

Nos proponemos siguiendo el método de Eratóstenes estimar el radio de la Tierra con mediciones otorgadas por uno de los encargados del “Proyecto Eratóstenes”², y analizar si difiere con mediciones científicas y contrastarlo con los resultados obtenidos por Irribarra P. y Uribe W. (2015).

Metodología

Para el desarrollo de este trabajo nos pusimos en contacto el Dr. Guillermo Mattei, dpto. de Física de la FCEyN UBA, para recabar información de la actualización de datos de mediciones realizadas con el fin de obtener el radio de la Tierra utilizando el método de Eratóstenes, ya que los otorgados por la cátedra de Estadística eran del 2012. El Dr. nos proporciona la información de 579 mediciones de diferentes ciudades del mundo, de las cuales seleccionamos los datos de 225 mediciones realizadas en distintas ciudades Argentinas. El criterio que se estableció para esta selección fue considerar que si el proyecto Eratóstenes puede utilizarse como una propuesta didáctica para estudiantes de nivel secundario, validar la información o recolectar nuevas mediciones, se facilitaría la comunicación con instituciones nacionales que con organismos internacionales.

Primeramente con la información brindada (Longitud y Latitud en unidades decimales, Ciudad, País, longitud del gnomon³, longitud de la sombra, horario y fecha donde se realizaron dichas mediciones) calculamos ángulo formado por el gnomon con los rayos del sol y las distancia de cada una de las ciudades al Ecuador, ya que para determinar el radio terrestres se necesitan dos ubicaciones, utilizando la página web <https://www.tutiempo.net/calcular-distancias.html>. En la Figura 2 se muestra un ejemplo de medición de la distancia calculada desde unos de los lugares que se hace el experimento (34°32'38.4" S - 58°26'27.6" W) al Ecuador (0 - 58°26'27.6" W).

² El Proyecto Eratóstenes es una iniciativa del Depto. de Física de la Facultad de Ciencias Exactas y Naturales de la UBA (ARG), del Laboratorio Pierre Auger, UTN Regional Mendoza (ARG) y de la Asociación Física Argentina.

³ Objeto alargado cuya sombra se proyectaba sobre una escala graduada para medir el paso del tiempo.

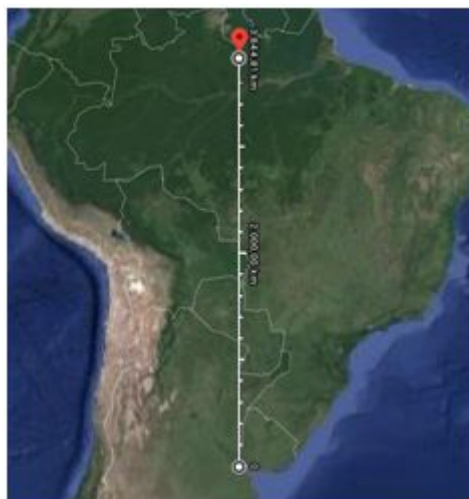


Figura 2: Distancia de CABA (ARG) al Ecuador sobre el mismo meridiano.

Para determinar el radio de la Tierra, con las mediciones de cada ciudad, se calcula el ángulo formado por el gnomon y los rayos del sol, se volcaron los resultados en una plantilla de Excel y seguidamente con este conjunto de datos comenzamos a trabajar con el software Estadístico R (x64 3.5.1- Paquete Rcmdr 2.3-5).

Con este software realizamos un gráfico de boxplot para visualizar la existencia o no de valores extremo.

Nuestra variable de interés es $X = \text{“valor calculado para el radio de la Tierra, desde una ciudad Argentina al centro del planeta (en kilómetros)”}$. Se lleva a cabo un análisis descriptivo para determinar la distribución de nuestra variable de estudio utilizando un boxplot, un diagrama de puntos, regla empírica y un histograma, tabla con medidas de posición y dispersión. Como técnicas de análisis inferencial realizaremos un Test de Hipótesis y una prueba chi-cuadrado de bondad de ajuste.

Resultados

Con el software R realizamos un gráfico de boxplot, dado de que nuestra variable de análisis es cuantitativa continua, los 225 datos correspondientes al valor calculado del radio de la Tierra para visualizar valores atípicos. Decidimos descartar dichos valores porque consideramos que los mismos podrían tener errores de medición, errores de cálculos y/o tipeo, logrando así un total de 203 mediciones Nacionales (Figura 3).

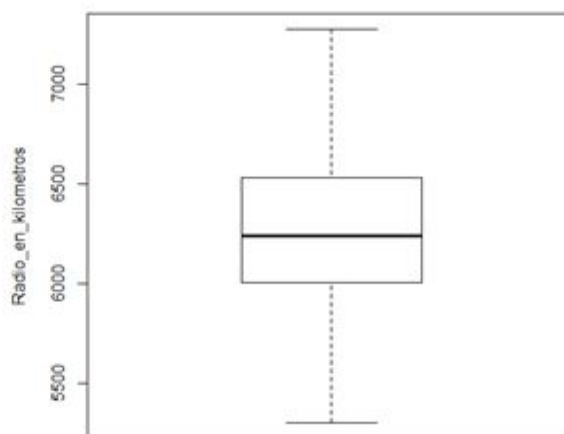


Figura 3: Boxplot de los 203 radios de la Tierra calculados en Km.

A partir del análisis numérico de nuestra variable de estudio se exponen en la siguiente tabla los siguientes resultados:

Media	Desvío	Coefficiente de variación	Cuantiles					Cantidad de datos
6255,182	367,3652	0,05873	0%	25%	50% (mediana)	75%	100%	203
			5300,275	6004,814	6236,571	6529,03	7272,228	

Tabla 1: Medidas de posición y de dispersión.

Continuando con el análisis del radio de la Tierra se realizó un diagrama de puntos en el software R obteniendo el siguiente resultado:

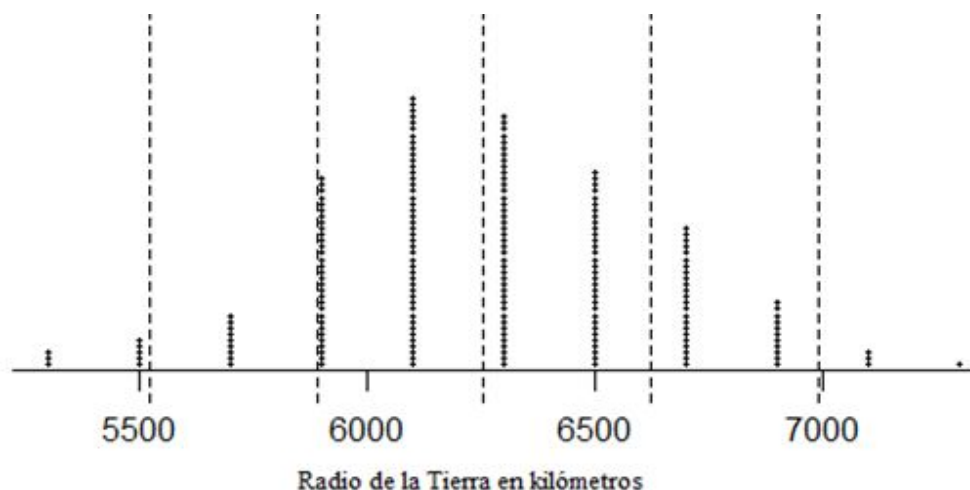


Figura 4: Diagrama de puntos de la variable analizada

Basados en los resultados obtenidos anteriormente se puede observar que la variable presenta cierta simetría, por lo cual se utilizó la regla empírica como diagnóstico de normalidad arrojando los siguientes resultados:

Intervalos		Cantidad de datos
$(\bar{x} \pm s)$	(5587,8168 ; 6625,472)	68.27%
$(\bar{x} \pm 2s)$	(5520,4516 ; 6989,9124)	95.45%
$(\bar{x} \pm 3s)$	(5153,0864 ; 7357,2776)	99.73%

Tabla 2: Intervalos de regla empírica

Analizando los resultados obtenidos podemos decir que la variable sigue una distribución normal de parámetros $\hat{\mu} = \bar{x}$ y $\hat{\sigma}^2 = s^2$ estimados a través del método de máxima verosimilitud asegurando que son insesgados, suficientes y de mínima varianza.

Realizamos, con R, un histograma al que se le ajusta una curva de densidad:

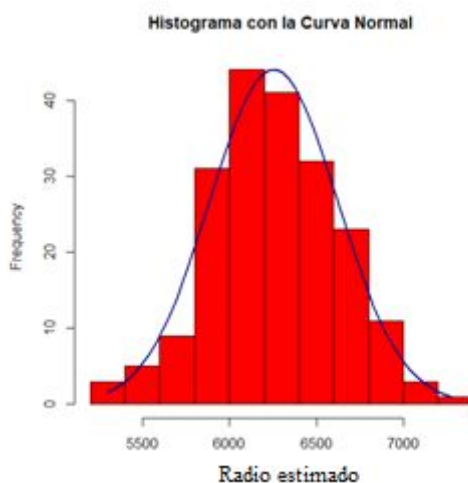


Figura 5: Histograma de las 203 mediciones con la función de densidad añadida

Por lo tanto podemos suponer que la variable sigue una distribución normal, conjetura (hipótesis nula) a la que le realizamos un test chi-cuadrado como prueba de bondad de ajuste con un nivel de significación $\alpha = 0,05$, arrojando un estadístico de prueba $T = 3,408$, que al compararlo con un nivel de referencia $\chi^2_{(6)} = 12,591$, no hay evidencia suficiente al 5% para rechazar nuestra hipótesis nula.

Al saber que nuestra variable de estudio presenta una distribución normal y desconocer la varianza, realizaremos un Test de Hipótesis para la media con varianza desconocida para una población normal.

Donde nuestra hipótesis nula $H_0 : \mu = 6371$ contra la hipótesis alternativa $H_a : \mu \neq 6371$. A continuación R nos informa que bajo un nivel de confianza con un $\alpha = 0,05$ nos arroja un valor del estadístico (t-student) $t = 4,4919$, con 202 grados de libertad y un valor de $p = 0,00001187$ ($p - value$). El resultado del Test de Hipótesis para la media de una población normal con varianza desconocida nos arroja un valor menor que α , por lo que tenemos suficiente evidencia para rechazar la hipótesis nula ($\mu = 6371$), es decir, la media es distinta de 6371 km.

Por lo tanto obtenemos el siguiente intervalo de confianza del 95% para la media:

$$IC(\mu : 95\%) = (6204,341 ; 6306,022)$$

Luego teniendo en cuenta el intervalo de confianza, el radio oscila entre 6204,341 km y 6306,022 km para la muestra analizada.

Calculamos la diferencia porcentual entre el valor estimado y el aceptado científicamente:

$$\frac{(|6255,182 - 6371|)}{6371} \cdot 100\% = 1,8179\%$$

Calculamos la diferencia porcentual entre el valor estimado y el calculado por Irribarra P. y Uribe W. (2015):

$$\frac{(|6255,182 - 6472,069|)}{6472,069} \cdot 100\% = 3,3511\%$$

Discusión

Analizando Tabla 1, Tabla 2, Figura 3, Figura 4, Figura 5 y los resultados del test chi-cuadrado de bondad de ajuste no hay evidencia suficiente al 5% para rechazar nuestra hipótesis nula, es decir, la variable posee una distribución normal de media 6255,182 y varianza de 134957,1902. Comparando con los resultados del informe de Irribarra P. y Uribe W. (2015) se obtiene la misma distribución para la misma variable de estudio pero con media 6472,069 y varianza 691495,3598, sobre los 85 datos del radio de la Tierra obtenidos por pares de instituciones inscriptas en el *Proyecto Eratóstenes* en el año 2012.

El resultado del Test de Hipótesis para la media de una población normal con varianza desconocida nos arroja un valor menor que α , por lo que tenemos suficiente evidencia para rechazar la hipótesis nula ($\mu = 6371$), es decir, la media es distinta de 6371 km, mientras que en el análisis del 2015 no tuvieron suficiente evidencia para rechazar la hipótesis nula.

Luego teniendo en cuenta el intervalo de confianza, el radio oscila entre 6204,341 km y 6306,022 km para la muestra analizada. Sin embargo en el análisis de Irribarra P. y Uribe W. (2015) el radio terrestre varía entre 6292,705 y 6651,432 km.

Observando que no se obtuvo mediante este análisis estadístico de los datos, la suficiente evidencia para asegurar que el radio medio de la Tierra es 6371 km, calculamos la diferencia entre el valor estimado y el aceptado científicamente siendo este menor al 2% y inferior al 3,5% al contrastarlo con el obtenido

por Irribarra P. y Uribe W. (2015). Esta diferencia podría deberse a la metodología e instrumentos utilizados al momento de recolectar los datos.

Consideramos que el radio medio de la Tierra aceptado científicamente es el correcto, ya que posee una metodología más rigurosa para medirlo, mientras que en los datos que nos proporcionaron desde el departamento de Física de la UBA, pueden presentar inconvenientes como por ejemplo:

- El gnomon podría no estar perfectamente perpendicular a la superficie terrestre.
- El instrumento de medición no fue el mismo para todos los datos.
- La diferencia horaria al momento de realizar el experimento.
- Que la superficie donde se estableció el instrumento de medición no fuese realmente plana, lo cual la proyección de la sombra modifica su longitud.
- La precisión con la cual medimos la distancia al Ecuador.

Conclusiones

El objetivo del trabajo fue estimar el radio de la Tierra con las mediciones realizadas en distintos puntos de Argentina siguiendo el método de Eratóstenes. Sabiendo que la diferencia porcentual de cálculo entre el valor estimado y el valor aceptado en la actualidad es inferior al 2%, a pesar que los datos trabajados no nos otorgaron suficiente evidencia para afirmar que el radio de la Tierra es de 6371 km, podemos considerar que por lo bajo que fue esta diferencia, el método de Eratóstenes es un buen estimador de la longitud del radio de la Tierra.

Como el error obtenido es ínfimo este método se puede implementar como propuesta didáctica para estudiantes de nivel secundario.

Referencias

- Collazos, D. (2018). *Eratóstenes*. Recuperado el 13 de noviembre de 2019 de <https://historia-biografia.com/eratostenes/>
- Departamento de Física, UBA. (2019). *El proyecto Eratóstenes*. Disponible en <https://www.df.uba.ar/es/difusion/proyecto-eratostenes1/informacion-conctacto-eratostenes>
- Irribarra, P. y Uribe, W. (2015). ¿Cuánto mide el radio de la Tierra? MP. Dieser y LV. Caverio (Comps.) *¡Estadístic@s en acción!*. Recuperado el 19 de noviembre de 2019. Disponible en <http://redi.exactas.unlpam.edu.ar/xmlui/handle/2013/166>
- Moreno, V. (2019). *Eratóstenes*. Recuperado el 13 de noviembre de 2019 de <https://www.buscabiografias.com/biografia/verDetalle/7352/Eratostenes>
- R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [software] Disponible en <https://www.R-project.org/>



Tu Tiempo Network, S.L. (2019). *Calcula la distancia entre dos puntos de la Tierra*. Consultado el 19 de noviembre del 2019. Disponible en: <https://www.tutiempo.net/calcular-distancias.html>

Wikipedia (13 de noviembre de 2019a). *Eratóstenes*. Recuperado el 19 de noviembre de 2019 de <https://es.wikipedia.org/wiki/Eratóstenes>

Wikipedia (31 de octubre de 2019b). *Tierra*. Recuperado el 19 de noviembre de 2019 de <https://es.wikipedia.org/wiki/Tierra>



ALCOHOL EN DIBUJOS ANIMADOS

Milagros Lecitra; Yanina Belén Lick

milagroslecitra@outlook.com; yanibelenlick97@gmail.com

Resumen

En este proyecto, se encuentra una investigación que tiene como objetivo corroborar si la mayoría de las películas de dibujos animados (estrenadas desde 1922 hasta 1997) cumplen con la clasificación G propuesta por 'Motion Picture Association of America' en el año 1922 la cual, establece que dichas películas son aptas para cualquier edad y que, entre otras cosas, no hay presencia de alcohol. Se analizaron 49 largometrajes dirigidos a niños, producidos en Estados Unidos por distintas compañías y que fueron estrenados luego de 1922. Se realiza un análisis descriptivo e inferencial, concluyendo que la mayoría de las películas no cumplen con la clasificación establecida, que su comportamiento no depende de la compañía que produjo la película y que el tiempo de aparición del alcohol es mayor en el período nuevo hollywood.

Introducción

Las bebidas alcohólicas son aquellas bebidas que contienen etanol (alcohol etílico) en su composición. Atendiendo a la elaboración se pueden distinguir entre ellas las bebidas producidas por 'fermentación alcohólica' (vino, cerveza, sidra, entre otras) y las producidas por 'destilación' a partir de un proceso de fermentación previo (como brandy, whisky, tequila, ron, vodka, entre otras). La cantidad de alcohol de una bebida se mide por el volumen de alcohol que contenga.

El 'etanol', presente en las bebidas alcohólicas, es el depresor más utilizado en el mundo y su consumo es adictivo. Según la dosis y la frecuencia con que se consuma, el etanol puede causar coma etílico, pérdida de conocimiento, parálisis respiratoria o incluso la muerte (Wikipedia, 2019a).

El consumo de alcohol es un factor causal de más de 200 enfermedades y trastornos. Está asociado con el riesgo de desarrollar problemas de salud tales como trastornos mentales y comportamentales, incluido el alcoholismo, importantes enfermedades no transmisibles tales como la cirrosis hepática, algunos tipos de cáncer y enfermedades cardiovasculares, así como traumatismos derivados de la violencia y los accidentes de tránsito (Organización Mundial de la Salud, 2019).

Podemos distinguir que dentro de los diferentes grupos de consumidores, los que resultan ser los más vulnerables son los adolescentes. El cerebro del adolescente se encuentra en constante desarrollo y crecimiento; durante la adolescencia el joven va a adquirir todas las capacidades de razonamiento, planificación, procesamiento de la información, capacidad discursiva y todas las propiedades del pensamiento abstracto. El consumo de alcohol durante esta etapa de adquisición y desarrollo de estas habilidades puede estancar el proceso y dar como resultado un adulto que presenta importantes carencias en el pensamiento maduro.

La Organización Mundial de la Salud recomienda restringir la publicidad de alcohol y los patrocinios de las compañías productoras y hasta algunos países tienen prohibiciones sobre ciertos medios y para ciertas bebidas.

Hoy en día, las caricaturas en la televisión y ahora en las películas representan para los niños un gran atractivo de audiencia influyendo en el desarrollo de diferentes patrones de comportamiento, por tal motivo es que surgió el “Sistema de clasificación por edades”, el cual permite clasificar las películas con respecto a la idoneidad para el público en términos de temas como: sexo, violencia, abuso de sustancias, blasfemia u otros tipos de contenido para adultos.

Las clasificaciones generales tienen recomendaciones de edad con capacidad de asesoría o restricción, y con frecuencia se dan en lugar de la censura.

Esta clasificación varía según cada país, en EE.UU., la clasificación de una película que determina la edad mínima de su audiencia es decidida por la ‘Motion Picture Association of America’ (MPAA) una asociación industrial fundada en 1922. Esta asociación se compone de 13 padres de familia que ven cada película y acuerdan en una clasificación. Aunque esta clasificación sólo es válida únicamente en Estados Unidos, sirve como una guía para el resto del mundo. En ella se menciona: G, PG, PG13, R, PT y NC-17. Con respecto a las películas animadas, nos centraremos en:

*G: todas las edades admitidas. No hay desnudos ni referencias sexuales, no hay drogas, alcohol y/o tabaco estos dos últimos sí pueden aparecer, pero muy poco frecuente, violencia mínima, muertes no violentas ni explícitas o ausencia de ella y lenguaje cortés, pero tal vez haya un poco de sangre por heridas corrientes y comunes como cortadas de papel. La mayoría son películas animadas, la minoría son películas reales y solo una es del género bélico. (Wikipedia, 2019c)

Por lo que en este trabajo analizaremos el uso de alcohol en películas animadas, teniendo en cuenta que en años anteriores se ha realizado un trabajo similar pero en relación al uso de tabaco en las películas animadas (Burgués et al., 2017).

Por tal motivo, nos resulta interesante complementar la información dada desde otro ángulo.

Luego, teniendo en cuenta la clasificación ‘G’ dada por la ‘Motion Picture Association of America’ (MPAA), nuestros objetivos son los siguientes:

*Analizar y determinar si la mayoría de las películas de dibujos animados estrenadas desde 1922 (año en que se fundó MPAA) cumplen con la normativa de clasificación G.

*Analizar y determinar si existe algún vínculo entre las compañías que produjeron las películas muestreadas y el cumplimiento o no de la clasificación.

*Analizar el tiempo de aparición de alcohol en las películas animadas según la época a la que pertenece.

Metodología

Para este estudio, nos basaremos en un muestreo aleatorio simple realizado por el profesor Goldstein Adam y dos asistentes, quienes recolectaron películas aptas para niños. Estas fueron estrenadas entre los años 1921 y 1997 por diferentes compañías (Walt Disney Co, MGM/United Artist, Warner Brothers Studios, Universal Studios y 20th Century Fox).

Las películas que fueron seleccionadas para su revisión han tenido una duración de al menos 60 minutos. El conjunto original de datos consta de 50 datos y 5 variables: nombre de la película, compañía a la que pertenece, duración total, tiempo total de aparición de tabaco y tiempo total de aparición de alcohol.

Agregaremos 3 nuevas variables a nuestro conjunto de datos, la primera hace referencia al ‘Año en el que fue estrenada la película en Estados Unidos’, la cual fue ingresada manualmente.

Definimos la variable X, la cual detalla el cumplimiento de la clasificación G en películas propuesta por

MPAA.

Como nuestro análisis se centrará en la aparición del ‘alcohol en las películas animadas’, eliminamos la variable que hace referencia al ‘tabaco (en segundos)’ de nuestro conjunto de datos.

Por último, definimos la variable que clasifica a las películas según el año de estreno, teniendo en cuenta que la industria cinematográfica estadounidense ha tenido un profundo efecto en el cine alrededor del mundo desde inicios del siglo XX. Su historia se divide en ocasiones en cuatro períodos principales: la era muda, el cine clásico de Hollywood (desde 1917 a 1960), el nuevo Hollywood (mediados de los 60’ hasta los 80’), y el período contemporáneo (a partir de 1980). (Wikipedia, 2019b)

Solo tomaremos los últimos tres períodos ya que, contienen a las películas muestreadas.

Como solo nos interesan aquellas películas que cumplan con la clasificación G establecida a partir del año 1922, eliminaremos de nuestro conjunto de datos a la película “Winnie the Pooh” ya que se estrenó antes de dicha fecha. Por lo tanto, nuestra ‘**muestra de unidades**’ se reduce a 49 películas de dibujos animados, y la ‘**unidad observacional**’ es una de estas películas.

Las variables de estudio son definidas del siguiente modo:

Variables de tipo ‘Cualitativa nominal’:

X= “Cumplimiento de la clasificación G de películas establecida por el MPAA en 1922 en Estados Unidos”

$$X = \begin{cases} 1 & \text{si cumple con la clasificación G} \\ 0 & \text{no cumple con la clasificación G} \end{cases}$$

La cual se distribuye $X \sim Bi(1, p)$ ya que se realizan ensayos de Bernoulli independientes (el contenido de una película es independiente del de la de otra).

Y= “Compañía a la que pertenece la película”

$\text{Im}(Y) = \{\text{Walt Disney Co, MGM/United Artist, Warner Brothers Studios, Universal Studios y 20th Century Fox}\}$

E= “Período al que pertenece la película de dibujos animados”

$\text{Im}(E) = \{\text{clásico, nuevo Hollywood, contemporáneo}\}$

Z= “Año en que se estrenó la película de dibujos animados en Estados Unidos”

$\text{Im}(Z) = [1922; 1997]$

Variables de tipo: ‘Cuantitativa continua’:

W= “Tiempo (en segundos) en que aparece el alcohol en la película de dibujos animados”

Se utilizó el software estadístico R (versión 3.5.3 y 3.6.0) a través de su interfaz gráfica R Commander (versión 2.5-3 respectivamente) para dar respuesta a nuestros objetivos.

Resultados

Analizando los datos obtenidos luego de la depuración, intentaremos dar respuesta a nuestros objetivos planteados al inicio de nuestro trabajo:

A partir del gráfico de barra de la **Figura 1**, se puede observar que de las 49 películas estrenadas a partir del año 1922, sólo 24 de ellas cumplieron con la clasificación G, es decir, no hay referencia a bebidas

alcohólicas en dicho film. Las 25 restantes siguieron mostrando escenas de uso de bebidas alcohólicas.

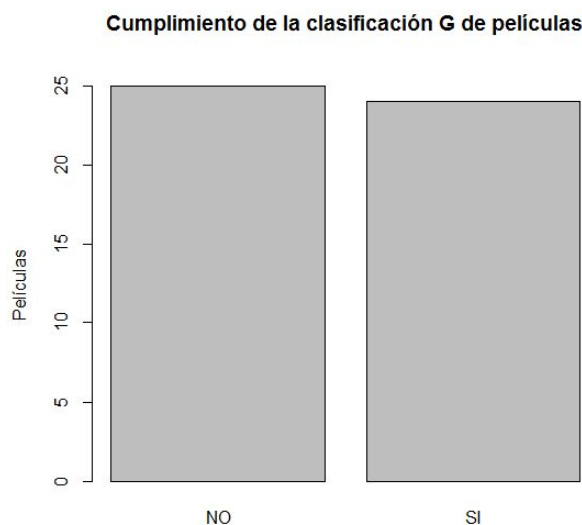


Figura 1: Gráfico de barras según el cumplimiento de la clasificación G de películas

Anteriormente consideramos que la variable X ="Cumplimiento de la clasificación G de películas establecida por el MPAA en 1922 en Estados Unidos" sigue una distribución binomial con un solo ensayo (Bernoulli), dado que a cada película se le analiza si cumple o no con la clasificación G. Esta variable puede tomar solo dos valores (1 si la película cumple con la clasificación G; 0 si no la cumple) y una probabilidad p , la cual sería la probabilidad de que una película cumpla con la clasificación.

Ahora veremos si es cierto que la mayoría de las películas animadas cumplen con la clasificación G, por lo que realizaremos un 'test de hipótesis' con nivel asintótico (debido a que el tamaño de nuestra muestra es lo suficientemente grande ($n \geq 30$)) donde nuestra hipótesis nula es que la minoría o la mitad de las películas cumplen con la clasificación "G" establecida en el año 1922 ($H_0 : p \leq 0.5$) y nuestra hipótesis alternativa, que la mayoría de las películas cumplen dicha clasificación ($H_a : p > 0.5$). Entonces queremos probar lo siguiente:

$$H_0 : p \leq 0.5 \text{ vs. } H_a : p > 0.5$$

Fijando un nivel de significancia del 5%, es decir $\alpha = 0.05$, el software utilizado arroja un valor $p=0.8864$, y como $\alpha < p$, no tenemos suficiente evidencia, a un nivel de confianza del 5%, de rechazar H_0 . Por lo que no es cierto que la mayoría de las películas cumplen con la clasificación G.

Ahora bien, hemos probado que la mayoría de las películas no cumplen con la clasificación G. Nos preguntamos si existe algún vínculo entre el cumplimiento de la clasificación y la compañía que la produjo. Por lo que, primero, analizamos gráficamente el cumplimiento de la clasificación según cada compañía:

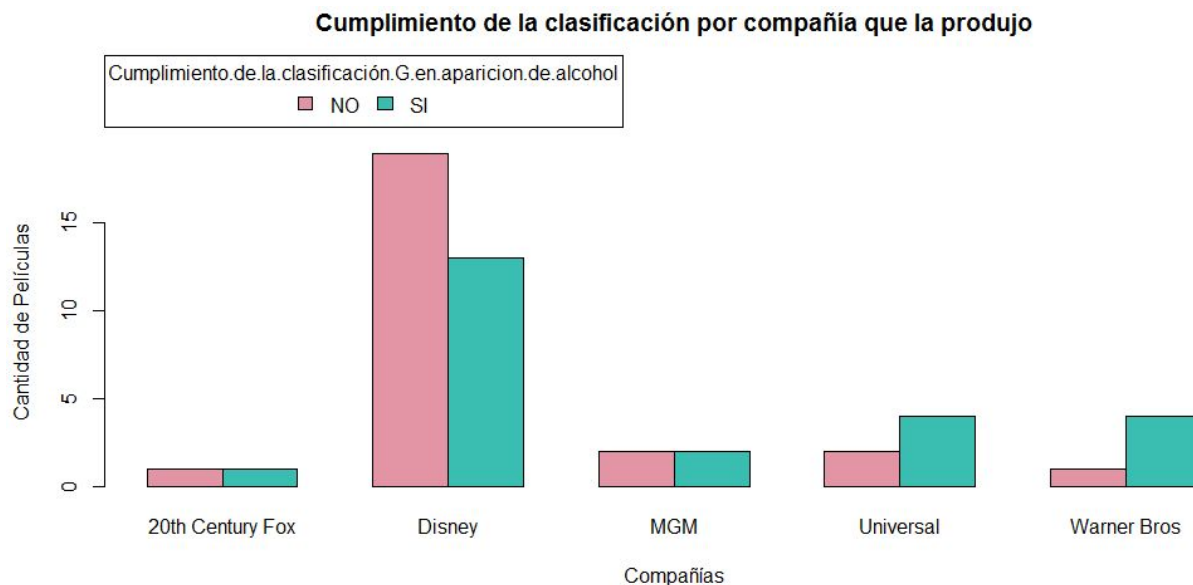


Figura 2: Diagrama de barra entre la compañía que produjo la película y el cumplimiento de la clasificación G

Podemos observar que tanto en ‘20th Century Fox’ como ‘MGM’ el porcentaje de películas que cumplen y no con la ley son iguales, mientras que en ‘Universal’ y ‘Warner Bros’ las películas que cumplen la ley son mayoría, aunque el único caso que registra lo contrario a este último son las películas de ‘Disney’.

Para poder concluir si existe o no relación entre el cumplimiento de la clasificación y la compañía que la produjo, realizaremos un ‘Test de independencia’, por lo cual planteamos lo siguiente:

Sean las variables **X**= “Cumplimiento de la clasificación G de películas establecida por el MPAA en 1922 en Estados Unidos” e **Y**= “Compañía a la que pertenece la película”

$$H_0 : p_{ij} = p_{i.} \cdot p_{.j} \text{ con } i \in \{1, 2, 3, 4, 5\} \text{ y } j \in \{1, 2\} \text{ vs. } H_a : \text{alguna igualdad no se cumple}$$

Compañía	Frecuencias Observadas		Frecuencias Esperadas	
	Comple	No cumple	Cumple	No cumple
Disney	13	19	16	16
MGM	2	2	2	2
Universal	4	2	3	3
Warner Bros	4	1	2	3
20 th Century Fox	1	1	1	1

Tabla N°1: Valores observados y esperados según el tipo de compañía y el cumplimiento o no de la clasificación G

Como ocurre que hay valores esperados menores a 5 ,agrupamos, obteniendo lo siguiente:

Compañía	Frecuencias Observadas		Frecuencias Esperadas		Desvíos Cuadráticos	
	Cumple	No cumple	Cumple	No cumple	Cumple	No cumple
Disney	13	19	16	16	0.5625	0.5626
MGM-Univers al-Warner Bros-20 th Century Fox	11	6	8	9	1.125	1

Tabla N°2: Valores observados, esperados y desvíos cuadrático según la compañía y el cumplimiento o no de la clasificación G

Considerando el estadístico de prueba:

$$T = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \sim X^2((2-1).(2-1))$$

Fijamos un nivel de confianza $\alpha = 0.05$, tenemos:

$$RR = \{T \geq X^2(1)_{0.05}\} = \{T \geq 3.841459\}$$

Dados los datos muestrales obtenidos en la **Tabla N°2**, el valor calculado $t = 3.25$ no pertenece a la región de rechazo, por lo que no hay suficiente evidencia, a un nivel de confianza del 5%, de rechazar H_0 . Es decir, que las variables relacionadas son independientes.

Por lo que el cumplimiento de la clasificación G es independiente del tipo de compañía que produjo la película de dibujos animados.

En este caso se observa que el valor calculado se encuentra muy cercano al límite inferior de la región de rechazo, por lo que para asegurarnos podríamos analizarlo utilizando otro valor de significancia:

*Si fijamos un $\alpha = 0.10$, tenemos:

$$RR = \{T \geq X^2(1)_{0.10}\} = \{T \geq 2.705543\}$$

Por lo que habiendo considerando un nivel de confianza mayor, en este caso el valor calculado $t=3.25$ pertenece a la región de rechazo. Por lo que hay suficiente evidencia, a un nivel de significancia del 10%, de rechazar H_0 . Es decir que las variables involucradas no son independientes. Por lo que el cumplimiento de la clasificación G depende de la compañía que produjo la película.

Analizando el grado de dependencia de las variables a través del 'coeficiente de contingencia' tenemos:

$$C = \sqrt{\frac{t}{t+n}} = \sqrt{\frac{3.25}{3.25+49}} = 0.25$$

Y su corrección:

$$C^* = \frac{C}{\sqrt{\frac{s-1}{s}}} = \frac{0.25}{\sqrt{0.5}} = 0.50$$

Por lo que el grado de asociación entre las variables involucradas es moderado.

Ese resultado lo comprobamos utilizando Rcmdr, al usar la herramienta de 'Tabla de doble entrada' la cual trabaja con los datos puros. Nos encontramos con la dificultad de que Rcmdr no agrupa en el caso de que las frecuencias esperadas sean menores a 5.

Ante esta situación optamos por utilizar la herramienta de Rcmdr ‘Introducir una tabla de doble entrada’, considerando los datos ya agrupados. Obtuvimos un valor del estadístico chi-cuadrado estimado igual a $X^2 = 2.5763$ con 1 grados de libertad y un valor $p = 0.1085$. Como habíamos fijado un nivel de confianza del 5%, tenemos que $\alpha < p$, por lo que no hay suficiente evidencia a un nivel de confianza del 5% de rechazar H_0 . Por lo que las variables son independientes.

Es decir, que el tiempo de duración del alcohol en películas animadas es independiente de la compañía que la produjo.

Si consideramos el segundo caso analizado, ocurre que el valor $\alpha = 0.10$ se encuentra muy cercano al valor $p = 0.1085$. Por lo que habiéndolo probado anteriormente a través del método clásico podemos concluir que las variables no son independientes.

Habiendo mencionado los períodos establecidos sobre la industria cinematográfica estadounidense desde inicios del siglo XX, siendo los mismos: cine clásico de Hollywood (desde 1917 a 1960), el nuevo Hollywood (mediados de los 60’ hasta los 80’), y el contemporáneo (a partir de 1980). Según los datos de nuestra muestra realizamos el siguiente Boxplot:

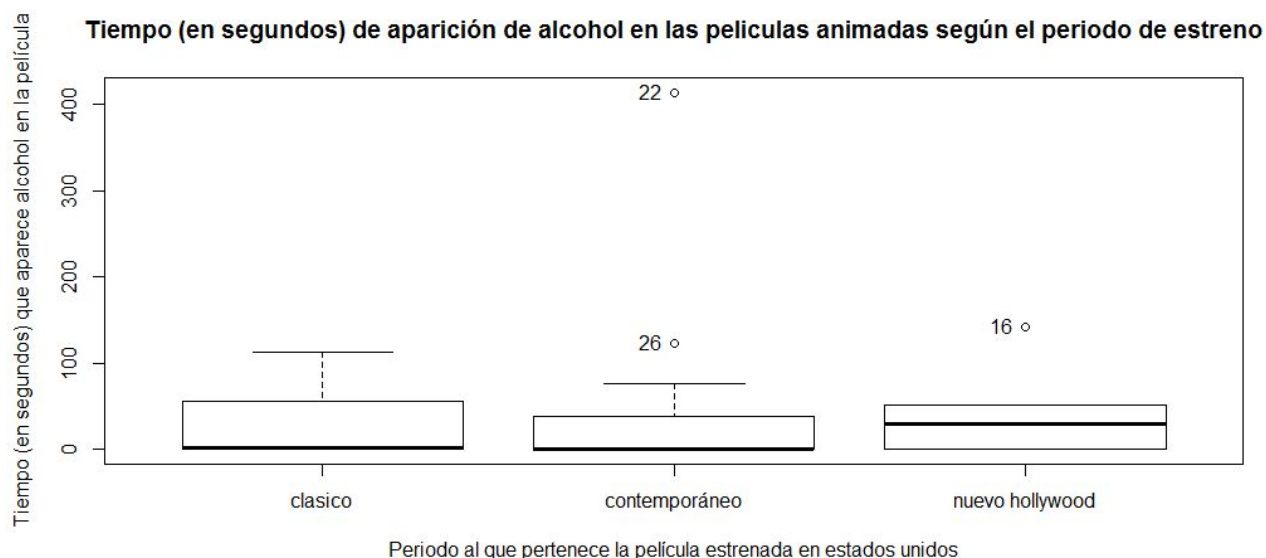


Figura 3: Boxplot del tiempo (en seg.) de aparición de alcohol y periodo de estreno de la película animada

Utilizando la herramienta de ‘resúmenes numéricos’ de Rcmdr, complementamos los datos obtenidos en la **Figura 3**:

	mean	sd	IQR	cv	0%	25%	50%	75%	100%	Alcohol.Use..sec.:n
clasico	27.16667	41.80655	44.75	1.538892	0	0	2.0	44.75	113	12
contemporáneo	33.70968	77.53201	38.50	2.299993	0	0	0.0	38.50	414	31
nuevo hollywood	42.00000	53.10744	43.00	1.264463	0	5	29.5	48.00	142	6

Figura 4: Resúmenes Numéricos entre la variable de Tiempo (en seg.) de aparición de alcohol y el Período

Analizando las **Figuras 3 y 4**, podemos comparar la distribución del tiempo (en segundos) que aparece el alcohol en la película animada según el período de estreno de la misma en Estados Unidos.

*Podemos ver que las películas estrenadas durante el ‘periodo clásico’, el tiempo medio (en seg.) de aparición del alcohol es de 27 segundos. La duración de aparición se encuentra comprendido entre 0 y 113 segundos. Además de poder observarse que el diagrama es asimétrico positivo, con una alta

disgregación de valores por sobre su mediana de 2 segundos.

*En el período ‘Nuevo Hollywood’, el tiempo medio de aparición de alcohol es de 42 segundos. La duración de su aparición oscila entre 0 y 48 segundos. Se observan datos atípicos en relación a la película 16. Además de poder observar que el diagrama es asimétrico negativo, con mayor concentración de los datos por sobre la mediana de 29.5 segundos.

*Por último, durante el período ‘Contemporáneo’, el tiempo medio de aparición del alcohol es de 34 segundos. La duración de su aparición se encuentra entre 0 y 38.5 segundos, además de observarse datos atípicos correspondiente a las películas 22 y 26. Se observa que el diagrama es asimétrico positivo, con una alta dispersión de los datos por sobre la mediana de 0 segundos.

Discusión

Teniendo en cuenta el trabajo realizado sobre el ‘Tabaco en dibujos animados’ (Burgués, otros, 2017), coincidiendo con los 2 primeros objetivos que hemos planteado en nuestro trabajo, llegamos a las mismas conclusiones en cuanto al no cumplimiento tanto de la ley que prohíbe el uso de tabaco en dibujos animados como de la clasificación G del MPAA. Además de llegar a la conclusión de que la aparición tanto de alcohol como de tabaco es independiente de la compañía que la produjo.

Por lo que nuestro análisis sobre el uso del alcohol en películas animadas complementa el realizado sobre el uso del tabaco, ya que podemos comparar en relación a los objetivos comunes de las diferentes problemáticas. Además de tener en cuenta, que ambos trabajos se desprenden de los datos dados por el profesor Goldstein Adam, el cual analiza ambas situaciones desde varias perspectivas. En nuestro caso aportamos los resultados a dicho trabajo desde nuestro planteo del problema.

Conclusiones

Según la clasificación G de películas dada por la MPAA la cual, menciona entre otras cosas la prohibición del alcohol, concluimos según nuestro análisis lo siguiente:

*Realizando un TH de nivel asintótico podemos afirmar con un nivel de significancia del 5%, que la mayoría de las películas animadas no cumplen con la clasificación G.

*Realizando un ‘Test de independencia’ con nivel asintótico podemos afirmar con un nivel de significancia del 5%, que la aparición de alcohol en las películas animadas es independiente de la compañía que la produjo.

Como el valor calculado ‘t’ está muy cercano al límite inferior de la región de rechazo, si consideramos un nivel de significancia mayor, se concluye que las variables involucradas no son independientes. En cuyo caso el grado de relación entre las variables es moderado.

*Concluimos que el tiempo de aparición de alcohol es mayor en el período nuevo hollywood que en los demás. Además, ningún período presenta una distribución simétrica de los datos respecto de su mediana.

Creemos que sería necesario ampliar la investigación sobre el tema, recolectando información sobre películas animadas estrenadas luego del año 1997 hasta la actualidad, lo cual permitirá contrastar si dicha clasificación (aún vigente) se cumple o no.

Además de que este trabajo complementa el análisis desarrollado por otros compañeros referido a el uso de ‘Tabaco en los dibujos animados’, por lo que resultaría interesante poder analizar y comparar si existe alguna relación entre la aparición del tabaco y el alcohol en las películas de dibujos animados estrenadas en EE.UU, como así también poder argumentar si la aparición de alguno de ellos es mayor o

no en las películas animadas de EE.UU, entre otros interrogantes que podrían surgir durante su desarrollo.

Referencias

- Burgués, I.; Calderón, J. & Esteban, E. (2017) Tabaco en dibujos animados. En MP Dieser & LV Cavero. (Comps.) *¡Estadístic@s en acción! 2017*. (pp. 4-10) FCEyN – UNLPam. Disponible en <http://redi.exactas.unlpam.edu.ar/xmlui/handle/2013/255>
- Dieser, MP. (2019). *Estadística para futuros matemáticos: Notas de clase*. [Material no publicado]. FCEyN-UNLPam. Santa Rosa (LP)
- Organización Mundial de la Salud (2018), *Alcohol*. Recuperado de: <https://www.who.int/es/news-room/fact-sheets/detail/alcohol>. [10/08/19]
- R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. [software]. Disponible en: <https://www.R-project.org/>.
- Wikipedia (2019a). *Bebida alcohólica*. Recuperado de: https://es.wikipedia.org/wiki/Bebida_alcoh%C3%B3lica. [09/08/19]
- Wikipedia (2019b). *Cine de los Estados Unidos*. Recuperado de: https://es.wikipedia.org/wiki/Cine_de_los_Estados_Unidos. [10/08/19]
- Wikipedia (2019c). *Sistema de calificaciones de la Asociación Cinematográfica de Estados Unidos*. Recuperado de: https://es.wikipedia.org/wiki/Sistema_de_calificaciones_de_la_Asociaci%C3%B3n_Cinematogr%C3%A1fica_de_Estados_Unidos. [10/08/19]

CAMBIOS EN LOS NIVELES DE COLESTEROL EN LA PREMENOPAUSIA

Viviana Vanesa Guittlein - Tamara Macarena Martel - Lourdes María Belén Miranda
marteltamara4@gmail.com - lourdesmariabelen@gmail.com - vanesaguittlein@gmail.com

Resumen

El siguiente proyecto, pretende contribuir a la investigación realizada por Kim, C. y colaboradores, sobre mujeres donde, se le sustrajeron el útero y a su vez un ovario y una trompa de falopio o ambos ovarios y ambas trompas de falopio. El objetivo de este documento está basado en la investigación sobre los niveles de colesterol de 44 mujeres que poseen las cirugías antes mencionadas. Las mujeres que tienen histerectomía y a su vez, le sustrajeron ambos ovarios y ambas trompas de falopio entran en proceso de menopausia, por lo tanto los niveles lipoproteicos LP (a), tienden a aumentar por no producir naturalmente la hormona sexual femenina (estrógeno), por lo tanto se les realizó un tratamiento de sustitución de estrógeno, para poder mantener dentro de los rangos normales los niveles de colesterol total en sangre (mg/dl). Se realiza un análisis descriptivo e inferencial, sobre los niveles de colesterol de las mujeres de nuestra muestra, determinando que el tratamiento de sustitución de estrógeno no fué favorable.

Introducción

El colesterol, es una sustancia grasa natural presente en todas las células del cuerpo humano, necesaria para el normal funcionamiento del organismo. Pero en cuerpos con niveles altos de colesterol total, puede aumentar la probabilidad de padecer una enfermedad cardíaca, accidente cerebrovascular y otros problemas (Fundación Española del corazón, 2019; Medlineplus, 2019).

El colesterol total, es la cantidad total de colesterol en la sangre. Incluye ambos tipos, el colesterol de lipoproteína de baja densidad (LDL) y el colesterol de lipoproteína de alta densidad (HDL) (Medlineplus, 2019).

El nivel de colesterol total saludables o normales(deseable), es menor a 200 mg/dl, y los niveles de colesterol total normal-alto(alto-sospechoso), se encuentran entre 200 y 239 mg/dl. Y los niveles elevados de colesterol total en sangre se encuentran por encima de 240 mg/dl (Fundación Española del corazón, 2019).

En el caso de mujeres que están atravesando la menopausia, se producen alteraciones en el patrón lipoproteico, relacionado con el descenso de los estrógenos: disminuye el HDL (colesterol bueno) y aumentan el colesterol total y el LDL (colesterol malo). Los médicos recomiendan la terapia hormonal sustitutiva (estrógenos y progesterona) o la terapia hormonal de estrógenos. Ambos tratamientos, reducen las molestias habituales de la menopausia y previene la osteoporosis. Además, en ocasiones, también pueden elevar ligeramente el HDL (colesterol bueno) y reducen el colesterol total (Fundación Española del corazón, 2019).

En esta investigación, se estudia los niveles de colesterol total en mujeres premenopáusicas, que son aquellas donde sus ovarios dejan de funcionar y se detiene la menstruación antes de los 40 años. Esto puede causar problemas de reproducción y síntomas de menopausia. Hay dos tipos de menopausia prematura: primaria y secundaria. Tener menopausia prematura primaria, significa que los ovarios no funcionan de manera normal, porque se extirpan mediante cirugía o por efecto de algunos tratamientos de cáncer, y determinadas enfermedades o afecciones genéticas. En la menopausia prematura secundaria, los ovarios son normales pero tienen un problema para recibir las señales hormonales que

vienen del encéfalo. En ocasiones, algunas mujeres con menopausia prematura tienen períodos menstruales y pueden tener hijos. También se llama insuficiencia del ovario, insuficiencia ovárica y menopausia precoz (Instituto Nacional del Cáncer, nd).

Como se menciona anteriormente, en el caso de mujeres que tuvieron que ser intervenidas quirúrgicamente, y así entrar en la etapa de menopausia prematura primaria. Por ejemplo, al extirpar el útero (histerectomía) provoca el cese de la menstruación. La extirpación de un ovario junto con la trompa de Falopio se denomina salpingooforectomía o salpingooforectomía unilateral (SOU). Cuando se extirpan ambos ovarios y ambas trompas de Falopio, se usa el término salpingooforectomía bilateral (SOB), en éste último caso los síntomas de la menopausia comienzan de inmediato, sin importar la edad. En muchos casos, la extirpación quirúrgica de los ovarios se realiza simultáneamente con una histerectomía” (Wikipedia, 2019).

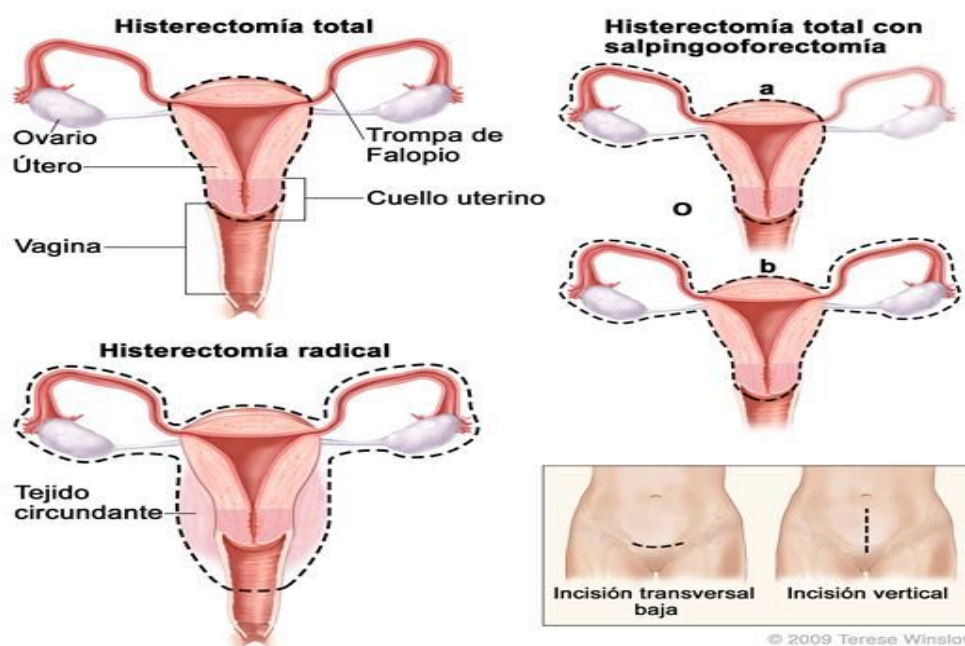


Figura 1: Diferentes Histerectomías. Ilustración tomada del sitio web Instituto Nacional del Cáncer.

Por tal motivo, es necesario que mujeres con esos cuadros sean controladas, mediante un análisis de sangre para determinar los niveles de colesterol total en sangre. Y en el caso que sea necesario, administrarles el tratamiento o la terapia hormonal necesaria.

Consideramos la siguiente tabla como referencia sobre los valores de nivel de colesterol total en sangre:

Colesterol Total		
Interpretación del Colesterol en sangre	mg/dl	mmol/l
Deseado	Menor de 200	Menor de 5.17
Alto. Sospechoso	200-239	5.17 - 6.18
Elevado desde	Mayor de 240	Mayor o igual a 6.20

Tabla 1: Referencia de nivel de colesterol total en sangre. Tomada del sitio web SH-SCI.org.

El propósito de este proyecto, es analizar los niveles de colesterol de aquellas mujeres que fueron sometidas a histerectomía y (SOB), con el objetivo de poder conocer si, con el tratamiento de sustitución de estrógenos, el nivel de colesterol total en sangre aumentó o no aumentó. Para poder realizar este análisis, consideramos importante estudiar los niveles de colesterol antes de la cirugía y a los dos meses después de la cirugía.

Metodología

Para este estudio, nos basaremos en un muestreo aleatorio simple, realizado por Kim et al. (1996), con el propósito de investigar los cambios en serie en los niveles lipoproteicos Lp(a) con la pérdida de hormonas sexuales femeninas en mujeres con menopausia quirúrgica y terapia de sustitución de estrógenos, a las cuales se les realizó mediciones de nivel de colesterol en sangre (mg/dl), en diferentes etapas.

La unidad experimental, resulta una mujer premenopáusica entre los 30 y 53 años de edad, que se sometió a histerectomía transabdominal. La población de unidades son las mujeres premenopáusicas entre 30 y 53 años de edad que se sometieron a histerectomía transabdominal, mientras que 44 mujeres seleccionadas, 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), y 22 mujeres premenopáusicas que tenían histerectomía y salpingo-ooforectomía bilateral (SOB), este último grupo de mujeres, recibieron 0,625mg de estrógeno equino durante dos meses después de la cirugía. A las 44 mujeres se les 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 4 meses después de la misma.

Teniendo en cuenta el objetivo planteado, la variable de estudio se define de la siguiente manera:

$X = \text{"Nivel de colesterol total en sangre (mg/dl), de las mujeres premenopáusicas entre 30 y 53 años de edad con histerectomía transabdominal"}$.

Como la variable X , es resultado de un proceso de medición, entonces es una variable cuantitativa continua (Dieser, 2019a).

En relación a la muestra estadística, consideramos las siguientes variables como:

$CT_0 = \text{"Nivel de colesterol total en sangre (mg/dl), antes de la cirugía"}$.

$CT_2 = \text{"Nivel de colesterol total en sangre (mg/dl), 2 meses después de la cirugía"}$.

Cirugía = "Tipo de cirugía realizada" (esta variable es cualitativa, nominal).

De acuerdo al conjunto de datos ofrecidos, y el objetivo a realizar, solo tomaremos los niveles de colesterol total en sangre antes de la cirugía y dos meses después de la cirugía, es decir CT_0 y CT_2 , de aquellas mujeres que tienen histerectomía y SOB.

Para el análisis descriptivo, realizamos tablas de distribución de frecuencias observadas de las variables mencionadas anteriormente, con sus respectivos histogramas.

Según lo estudiado, las medidas de tipo morfológicos y fisiológicos, tales como la talla, presión arterial, respuesta a un fármaco, son bien modeladas con la distribución normal (Dieser, 2019b). Entonces, consideramos que los niveles de colesterol total en sangre se puede modelar con una distribución normal, de parámetro μ y σ^2 . Para poder verificar esta afirmación realizamos un análisis descriptivo de 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 y SOB, se realizó

un test de hipótesis, para la diferencia de medias de poblaciones normales para muestras aleatorias dependientes (antes y después).

En este proyecto, utilizamos el software R (versión 3.5.3) y Rcmdr (versión 2.5-3) para construir gráficos, tablas de frecuencias y hacer cálculos auxiliares.

Resultados

Análisis descriptivo:

En principio, realizaremos gráficos (Histogramas) y tablas de frecuencia de los niveles de colesterol antes de la cirugía (CT0) y después de la cirugía (CT2) de las 22 mujeres premenopáusicas con histerectomía y SOB, además analizaremos de manera descriptiva los datos obtenidos.

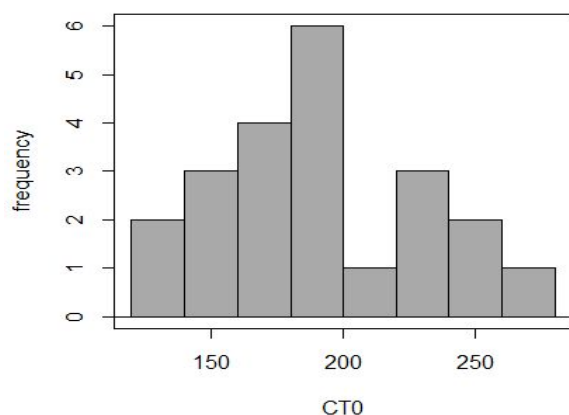


Figura 2: Histograma del nivel de colesterol total en sangre CT0, de 22 mujeres con cirugía SOB.

CT0 (SOB)		
Clases	frecuencia	porcentajes
(120,140]	2	9.09
(140,160]	3	13.64
(160,180]	4	18.18
(180,200]	6	27.27
(200,220]	1	4.55
(220,240]	3	13.64
(240,260]	2	9.09
(260,280]	1	4.55
Total	22	100.01

Tabla 2: tabla de distribución frecuencia del nivel de colesterol total en sangre CT0, de 22 mujeres con cirugía SOB

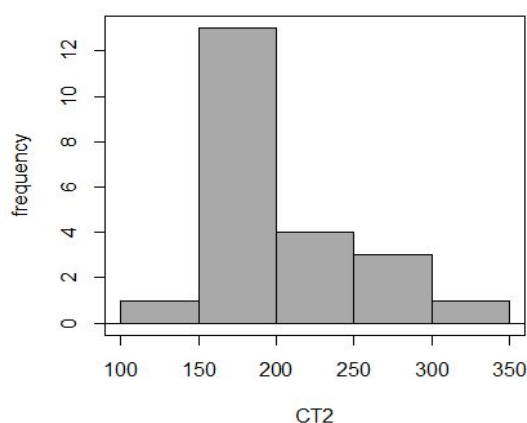


Figura 3: Histograma del nivel de colesterol total en sangre CT2, de 22 mujeres con cirugía SOB.

CT2 (SOB)		
Clases	frecuencia	porcentajes
(100,150]	1	4.55
(150,200]	13	59.09
(200,250]	4	18.18
(250,300]	3	13.64
(300,350]	1	4.55
Total	22	100.01

Tabla 3: tabla de distribución de frecuencia de los niveles de colesterol total en sangre CT2.

En el caso del nivel de colesterol antes de la operación, observando la **Figura 2** y **Tabla 2**, para las mujeres que fueron sometidas a la cirugía SOB, podemos ver que los niveles de colesterol total en sangre (mg/dl), se encuentran con mayor frecuencia entre 180-200 mg/dl y luego de dos meses transcurrida dicha cirugía, observando la **Figura 3** y **Tabla 3**, los niveles de colesterol total en sangre (mg/dl), se encuentran con mayor frecuencia entre los valores de 150-200 mg. Lo cual podemos decir que esos valores de nivel de colesterol total en sangre se encuentran en mayoría sobre el límite de los valores normales.

También, analizamos entre qué valores se encuentran, en promedio, los niveles de colesterol de las 22 mujeres que fueron sometidas a histerectomía y SOB, donde calculamos las medidas de tendencia central (Media), y medidas de dispersión (Desvío, CV):

Nivel de colesterol	SOB		
	Media	Desvío	CV
CT0	190.05	35.99	0.189
CT2	208.45	41.87	0.201

Tabla 4: Medidas de resumen de 22 mujeres con cirugía SOB

Como se puede observar en la **Tabla 4**, los coeficiente de variación está entre (0.10;0.33], indica que la dispersión en los datos es aceptable. Además, podemos decir, que antes de la cirugía, el nivel de colesterol en sangre de las mujeres en promedio es más bajo que después de la cirugía y que su dispersión es mucho mayor.

A partir de representaciones gráficas útiles podemos tener una idea bastante certera del comportamiento normal de las variables de estudio. Para ello utilizamos los gráficos de probabilidad normal QQ-plot, donde se espera que los datos observados se acomoden aproximadamente alrededor de una recta (recta identidad) (Dieser, 2019a).

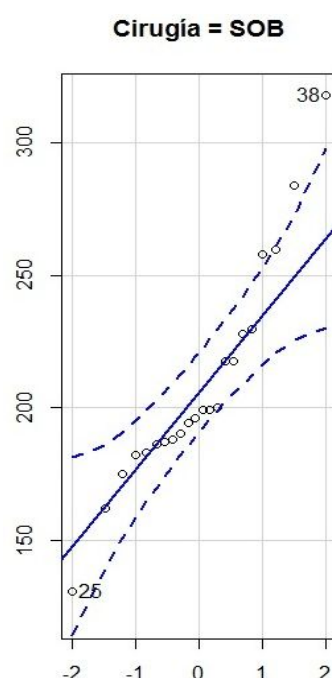
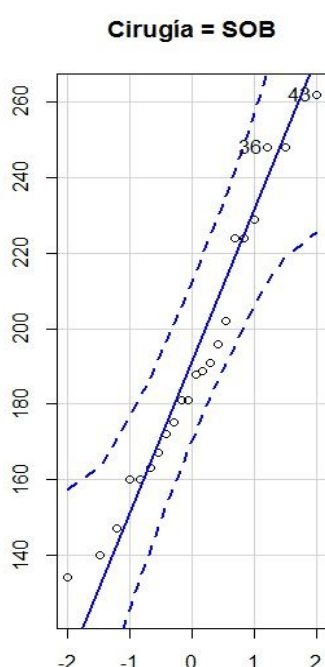


Figura 4: QQ-plot sobre los niveles de colesterol total en sangre CT0.

Figura 5: QQ-plot sobre los niveles de colesterol total en sangre CT2.

Cómo se puede observar, en ambas figuras (**Figura 4** y **Figura 5**) los datos se ajustan bastante bien a la línea recta, podemos decir que las variables (CT0 y CT2) se pueden modelar con una distribución normal. El tamaño de la muestra es pequeño, por eso detectamos salidas significativas de la distribución en los extremos.

Análisis inferencial:

Para determinar si el nivel de colesterol después de dos meses de la cirugía es mayor con respecto a los obtenidos antes de la cirugía, luego de haberles realizado el tratamiento de sustitución de estrógeno para aquellas mujeres premenopáusicas con SOB, realizamos un test de hipótesis por el método clásico.

Dado que tenemos muestras dependientes, consideramos una nueva variable diferencia $D = CT0 - CT2$ con $D_1; D_2; \dots; D_{22}$ m.a.s. de $D \sim N(\mu_1 - \mu_2; \sigma^2)$. Luego tenemos $\bar{d} = -18.409091$ y $s_D = 24.74$.

Para poner a prueba la hipótesis que planteamos antes:

$$H_0 : \mu_1 - \mu_2 \geq 0 \quad \text{vs} \quad H_a : \mu_1 - \mu_2 < 0$$

Un pivote adecuado es: $T = \frac{\bar{D} - (\mu_1 - \mu_2)}{\frac{s_D}{\sqrt{22}}} \sim t(22 - 1)$

Con un nivel de significación de $\alpha = 0.05$, realizando los cálculos convenientes se obtiene la región de rechazo:

$$\text{Bajo } H_0, RR = \left\{ \bar{D} < t(21)_{0.05} \frac{s_D}{\sqrt{22}} \right\} = \left\{ \bar{D} < -1.720743 \frac{24.74}{\sqrt{22}} \right\} = \left\{ \bar{D} < -9.076 \right\}$$

Por lo tanto, $\bar{d} = -18.4 \in RR$ hay suficiente evidencia para rechazar H_0 , es decir, que en promedio los niveles de colesterol después de 2 meses de la cirugía son mayores con respecto a los obtenidos antes de la cirugía.

Como se rechazó H_0 , se construyó un I.C: IC $(\mu_1 - \mu_2; 95\%) = \bar{D} \pm t(21)_{0.05} \frac{s_D}{\sqrt{22}} = [-29.3; -7.5]$, esto significa que la diferencia de los niveles de colesterol de las mujeres con SOB puede variar desde -18.41 ± 10.91 aproximadamente.

Discusión

Realizando el análisis descriptivo, al observar las figuras (**Figura 2 y 3**) y las tablas de distribución de frecuencias (**Tabla 2 y 3**), notamos que, un 32% aproximadamente de las mujeres con cirugía (SOB) superan el rango deseado de los niveles de colesterol en sangre antes de la cirugía y un 36% aproximadamente, a dos meses de dicha cirugía. Esto intuitivamente nos llevó a pensar que el tratamiento de sustitución de estrógeno realizado no fue satisfactorio para las 22 mujeres de la muestra, o que el tiempo de la realización del tratamiento fue insuficiente. Para verificar esta suposición, realizamos un test de hipótesis, teniendo en cuenta los niveles medios de colesterol en sangre antes de la cirugía y al terminar el tratamiento (2 meses después), considerando que sus niveles habían aumentado al finalizar con el tratamiento.

En la **Tabla 4**, llegamos a que el nivel de colesterol en sangre de las mujeres con cirugía (SOB) de la muestra en promedio es 190.05mg/dl antes de la cirugía, y luego, a los dos meses de ella, había aumentado a 208.45mg/dl en promedio con respecto al resultado obtenido antes la cirugía, lo que comprueba nuestra hipótesis.

Consideramos que nuestras variables de estudio se modelan con una distribución normal, al realizar un análisis descriptivo para nuestra variable de estudio, basado en el análisis de histograma y QQ-plot, ya que no pudimos realizar un test de bondad de ajuste por tener una muestra pequeña. Finalmente se puso a prueba la hipótesis planteada, donde se determinó que el tratamiento de sustitución de estrógeno para las mujeres premenopáusicas con histerectomía y SOB no fue efectivo.

Para saber por qué no dio el resultado esperado, el tratamiento en las mujeres de la muestra habría que investigar otras características de ellas, sobre sus hábitos de vida por ejemplo. Como una alimentación adecuada, la práctica de ejercicio y la supresión de hábitos tóxicos, porque si esto no estaría pasando serían el motivo del desorden en los niveles de colesterol.

Conclusiones

Luego de realizar las evaluaciones correspondientes y utilizando los métodos ya conocidos y mencionados con anterioridad, y en base a la evidencia aportada por la muestra, se concluye que, a las mujeres con cirugía quirúrgica y salpingo-ooforectomía bilateral(SOB), no les resultó positivo el tratamiento. Ya que pudimos analizar que el nivel de colesterol en sangre aumentó.

Como futuro análisis sería importante considerar una muestra aleatoria simple mucho mayor para poder obtener datos más precisos. También sería interesante evaluar los niveles de colesterol para estas mujeres luego de haber transcurridos unos meses, e investigar sobre sus hábitos de vida con el fin de diagnosticar el motivo que lleve al aumento en los niveles de colesterol.

Referencias:

Dieser, M.P. (2019a). *Estadística para futuros matemáticos: Notas de clase*. Material no publicado. Facultad de Ciencias Exactas y Naturales, Universidad Nacional de La Pampa. Santa Rosa.

Dieser, M.P. (2019b). *Probabilidad para futuros matemáticos: Notas de clase*. Material no publicado. Facultad de Ciencias Exactas y Naturales, Universidad Nacional de La Pampa. Santa Rosa.

Fundación española del corazón (2019). *Colesterol y riesgo cardiovascular*. Disponible en:

<https://fundaciondelcorazon.com/prevencion/riesgo-cardiovascular/colesterol.html> [7 de noviembre de 2019].

Instituto Nacional del Cáncer (nd). *Menopausia prematura*. Disponible en: <https://www.cancer.gov/espanol/publicaciones/diccionario/def/menopausia-prematura> [7 de noviembre de 2019].

Instituto nacional del Cáncer (nd). *Salpingooferectomía bilateral*. Disponible en:



<https://www.cancer.gov/espanol/publicaciones/diccionario/def/salpingooforectomia-bilateral>
[7 de noviembre de 2019].

Medlineplus (2019). *Niveles de colesterol en la sangre*. Disponible en:
<https://medlineplus.gov/spanish/ency/article/000403.htm> [6 de noviembre de 2019].

SH-SCI.org (2018). *Colesterol bueno y malo*. Disponible en:
<https://sh-sci.org/colesterol-bueno-y-malo/> [13 de noviembre de 2019].

Wikipedia (2019). *Menopausia*. Disponible en: <https://es.wikipedia.org/wiki/Menopausia> [6 de noviembre de 2019].