Tarea 6
Pruebas estadísticas

Erick Cervantes Mendieta
Matrícula: 2032430

Modelos Probabilistas Aplicados

Octubre 2020

1. Actividad 1

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

En estadística, una **hipótesis** es una aseveración o afirmación acerca de una propiedad de una población [4]. Una **prueba de hipótesis** (o *prueba de significancia*) es un procedimiento estándar para probar una aseveración acerca de una propiedad de una población. Por otro lado, una **prueba estadística** es una forma de evaluar la evidencia que los datos proporcionan para *probar una hipótesis*. Esta hipótesis se denomina *hipótesis nula*, y suele denominarse H_0. Bajo H_0, los datos se generan mediante procesos aleatorios. En otras palabras, los procesos controlados (las manipulaciones experimentales, por ejemplo) no afectan a los datos. Normalmente, H_0 establece la igualdad (entre las medias, o entre las varianzas, o entre un coeficiente de correlación y cero, por ejemplo). H_0 normalmente se opone a una hipótesis denominada *hipótesis alternativa*, denominada H_1 o H_a. La mayoría de las veces, la hipótesis alternativa es aquella que el usuario querría demostrar. Implica establecer una diferencia (por ejemplo, diferencia entre medias). Si se desea realizar un estudio y se desea emplear una prueba de hipótesis para sustentar una aseveración, ésta debe redactarse de tal manera que se convierta en la hipótesis alternativa. Esto quiere decir que la aseveración debe expresarse utilizando sólo estos símbolos: $<$ o $>$ o $\neq$. No se puede utilizar una prueba de hipótesis para sustentar la aseveración de que algún parámetro es igual a algún valor en específico.

2. ¿Qué indicaría rechazar la hipótesis nula?

Rechazar la hipótesis nula H_0 implica que la hipótesis alternativa es *verdadera*. Es decir, no se tiene suficiente evidencia estadística que apoye a H_0 a ser verdadera.

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

Una prueba estadística produce un número denominado valor-p (cuyos límites son 0 y 1). El valor p es la probabilidad de obtener los datos o datos más extremos bajo la hipótesis nula. En términos más prácticos, el valor p debería compararse con alfa: si $p <$ alfa, rechazamos H_0 y aceptamos H_1 con un riesgo proporcional al valor p de ser errónea; si $p >$ alfa, no rechazamos H_0, pero esto no implica necesariamente que debamos aceptarla. Significa, bien que H_0 es verdadera, bien que H_0 es falsa, pero nuestro experimento y nuestra prueba estadística no han sido suficientemente *fuertes* para producir un valor p inferior a alfa.

4. ¿Cómo seleccionar el valor de alpha (α)?

Cuando diseñamos un estudio, debemos especificar un umbral de riesgo por encima del cual H_0 no debería ser rechazada. Este umbral se conoce como **nivel de significación alfa**, y debería estar entre 0 y 1. Valores bajos de alfa son más conservadores. La *elección* de alfa debería depender de cuán peligroso sea rechazar H_0 en el caso de que sea verdadera. Por ejemplo, en un estudio que se proponga demostrar los beneficios de un tratamiento médico, alfa debería ser bajo. Por otro lado, cuando revisamos los efectos de muchos atributos en la apreciación de un producto, alfa podría ser moderado. En la mayoría de los casos, alfa se fija en valores iguales a 0.05, 0.01 o 0.001. Al examinar la hipótesis alternativa, podemos determinar si la prueba es de cola derecha, de cola izquierda o de dos colas. La cola corresponderá a la región crítica que contiene los valores que entrarán en conflicto, de manera significativa, con la hipótesis nula. En la figura 2 se indica el signo de desigualdad de H_1 y su relación en la dirección de la región crítica [4].

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

El valor p no nos da una probabilidad de que la hipótesis nula sea cierta, solamente nos permite definir su aceptación o rechazo, confrontándolo con el valor de alfa establecido (posibilidad de cometer error tipo I). El valor p no nos informa la probabilidad de que la hipótesis alternativa sea cierta; de la misma manera que lo anterior, si el valor p es menor que el valor definido como valor de rechazo (generalmente < 0.05) significa que, a partir de los datos obtenidos, no podemos aceptar la hipótesis de no diferencias. Pero esto no significa que el valor p asigne una probabilidad de que la hipótesis alternativa sea cierta. El valor p con magnitudes muy bajas no es una medida de efecto de la variable estudiada, es decir, el valor p muy bajo no demuestra "mayor efecto" que un valor p cercano a 0.05[10]. Tampoco debe entenderse que la significación estadística de una determinada prueba es evidencia de efecto, ni lo contrario, no es evidencia de no efecto la falta de significación estadística ($p = 0.05$).

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

La potencia de una prueba de hipótesis es la probabilidad $(1-\beta)$ de rechazar una hipótesis nula falsa, que se calcula utilizando un nivel de significancia α particular y un valor específico del parámetro de población que representa una alternativa al valor asumido como verdadero en la hipótesis nula. Es decir, la potencia de una prueba de hipótesis es la probabilidad de sustentar una hipótesis alternativa que es verdadera. Se utiliza para describir la probabilidad de que una prueba identifique correctamente un efecto genuino, real. Dicho de una manera más sencilla, es la capacidad de distinguir la señal del ruido. La señal que buscamos es el impacto de un tratamiento sobre algún resultado que nos interesa.

7. Ejemplos de pruebas estadísticas paramétricas y no paramétricas.

Pruebas paramétricas: Si el investigador quiere comparar dos grupos con *variables cuantitativas continuas* y con distribución normal, (dicho de otra forma, comparación de promedios entre dos grupos), se puede elegir una **prueba t** (hay diferentes, la más conocida es la denominada **t de Student**). Tomando en cuenta lo descrito, esta prueba puede utilizarse en dos escenarios diferentes: en muestras relacionadas (un solo grupo antes y después) y en muestras independientes (comparación de dos grupos). Ahora bien, si lo que se desea es comparar más de tres grupos (comparación de tres o más promedios) se debe seleccionar una prueba denominada **análisis de varianza** o **ANOVA**. De esta última prueba se pueden distinguir al menos dos variantes: ANOVA de una vía, cuando se comparan los promedios de tres o más grupos independientes, y ANOVA de dos vías, que se emplea cuando se pretende comparar los promedios de muestras relacionadas medidas tres o más veces.

Pruebas no paramétricas: cuando se tienen *datos cuantitativos* que no siguen una distribución normal, se puede usar la **prueba de Wilcoxon**, la cual se utiliza para comparar un grupo antes y después, es decir, *muestras relacionadas*. Para la comparación de *grupos independientes* se debe emplear **U de Mann–Withney**. En el caso de tres o más grupos independientes se debe utilizar la **prueba de Kruskal–Wallis** (la cual es equivalente a ANOVA de una vía). La **prueba Friedman** es la que se recomienda cuando se comparan tres o más muestras relacionadas (equivalente a ANOVA de dos vías). Cuando se tienen *datos cualitativos* y se desea comparar dos *grupos independientes* se puede utilizar la **prueba de chi cuadrada**, pero si las *muestras son relacionadas* deberá emplearse la **prueba de McNemar**. En caso de comparar tres o más grupos independientes también se utiliza la **prueba chi cuadrada**; en caso de muestras relacionadas, la preuba **Q de Cochran**.

Se puede escoger entre aplicar algún método no paramétrico o un método paramétrico correspondiente, aunque la eficiencia de estos últimos es mejor en la mayoría de las veces que sus contrapartes, generalmente es mejor utilizar las pruebas paramétricas cuando sus supuestos requeridos se satisfacen.

8. Resume LA GUÍA para encontrar la prueba estadística que buscas.

En la figura 1 se presenta un resumen de la guía definitiva para encontrar la prueba estadística que buscas, en la cual se tratan los puntos más importantes a considerar en la implementación de pruebas estadísticas.

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

Un requisito indispensable para seleccionar una prueba paramétrica es la distribución de los datos; en este sentido, solamente se debe utilizar este tipo de prueba cuando los datos muestran una **distribución normal** (es decir, semejante a una curva de Gauss). Otros re-

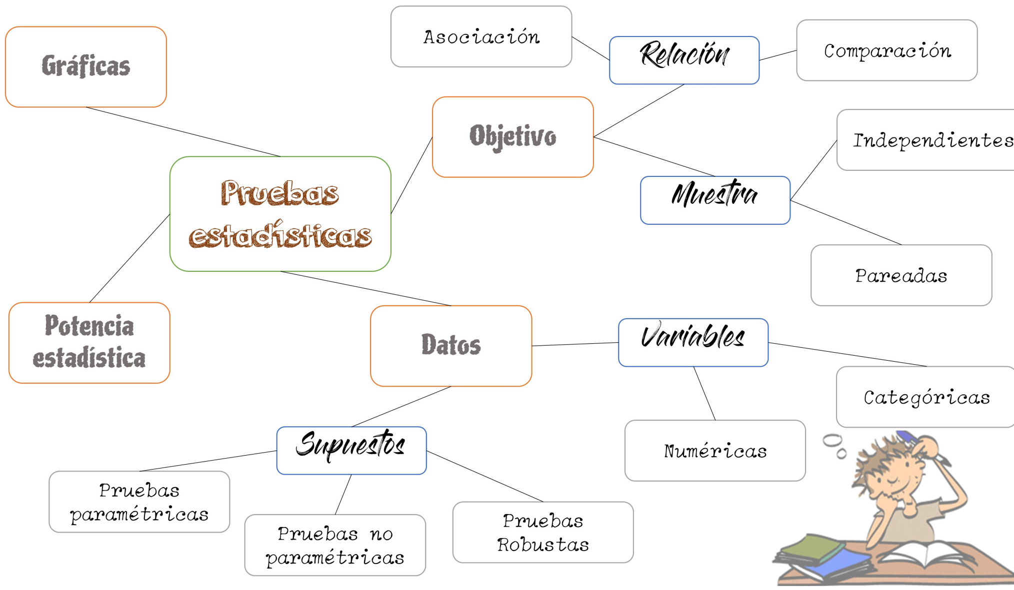

Figura 1: Resumen de la guía definitiva para encontrar la prueba estadística que buscas.

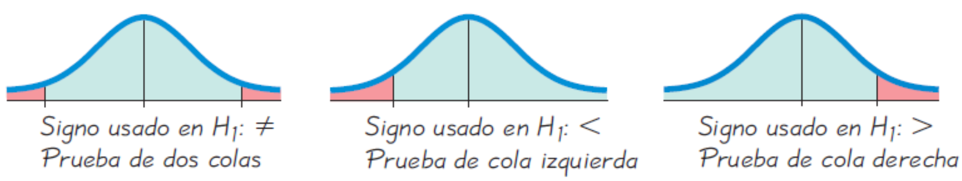

Figura 2: Pruebas de dos colas, cola izquierda, cola derecha y la relación que hay con el valor de α (sección roja).

quisitos son: que las observaciones deben ser independientes entre sí, las poblaciones deben tener la misma varianza, las variables deben haberse medido por lo menos en una escala de intervalo de manera que sea posible utilizar las operaciones aritméticas.

2. Actividad 2

El uso del método del valor p como auxiliar en la toma de decisiones es muy natural y casi todos los programas de cómputo que proporcionan el cálculo de pruebas de hipótesis ofrecen valores p, junto con valores del estadístico de prueba adecuado Walpole***. Sin embargo, se debe tener en cuenta que hay diferencias entre el entre el enfoque y la filosofía del método clásico de α fija, que tiene su momento más importante en la conclusión *rechazar* H_0 o *no rechazar* H_0 y el método del valor p. En este último no se determina una α fija y las conclusiones se obtienen con base en el tamaño del valor p, según la apreciación subjetiva del ingeniero o del científico. Los modernos programas de cómputo proporcionan valores p, es importante comprender ambos enfoques, por lo que a continuación se enlistan los pasos del dichos métodos.

Aproximación a la prueba de hipótesis con probabilidad fija del error tipo I:

1. Establezca las hipótesis nula y alternativa.
2. Elija un nivel de significancia α fijo.
3. Seleccione un estadístico de prueba adecuado y establezca la región crítica con base en α.
4. Rechace H_0 si el estadístico de prueba calculado está en la región crítica. De otra manera, no rechace H_0.
5. Saque conclusiones científicas y de ingeniería.

Prueba de significancia (método del valor p):

1. Establezca las hipótesis nula y alternativa.
2. Elija un estadístico de prueba adecuado.
3. Calcule el valor p con base en los valores calculados del estdístico de prueba.
4. Saque conclusiones con base en el valor p y los conocimientos del sistema científico.

A continuación se presentan diferentes bases de datos del sistema educativo nacional del INEGI [2] y algunas pruebas estadísticas implementadas en el entorno R (Versión 4.0.2) [3].

1. One Sample t-Test

La prueba t de una muestra se utiliza para estimar la media de procesos y compararla con un valor objetivo. Esta prueba se considera un procedimiento robusto debido a que es extremadamente sensible al supuesto de normalidad cuando la muestra es moderadamente grande. De acuerdo a la mayoría de los libros de texto de estadística, la prueba t de una muestra son apropiados para cualquier muestra con un tamaño de 30 o más.

Para ver cómo funciona esta prueba utilizamos los datos obtenidos del grado promedio de escolaridad de la población de 15 años y más en toda la República Mexicana durante el año 2015, los datos se encuetran en `Prueba1.txt`. El INEGI tiene como dato que en México, los habitantes de 15 años y más tienen **9.1** grados de escolaridad en promedio, lo que significa un poco más de la secundaria concluida. Esto nos permite plantear nuestras hipótesis como sigue:

H_0 : El promedio del grado de escolaridad de la población de 15 años y más en México es igual a 9.1.
H_1 : El promedio del grado de escolaridad de la población de 15 años y más en México no

es igual a 9.1.

Al realizar la prueba con la instrucción `t.test(x, mu)` obtenemos un valor de $p = 0.6638 > 0.05$, por lo que no hay evidencia estadística para rechazar H_0, es decir, el promedio del grado de escolaridad de la población de 15 años y más en México es igual a 9.1 como lo dice el INEGI.

2. Wilcoxon Signed Rank Test

En el ejemplo anterior se supuso que no habría problema al aplicar la prueba t con el supuesto de normalidad, debido a que se tenía una muestra mayor a 30 observaciones, en la figura 3 se puede apreciar que no del todo se ajustan nuestros datos a una distribución normal, por lo que procedemos a implementar una prueba de los rangos con signo de Wilcoxon con las mismas hípótesis planteadas, ya que esta prueba puede ser una alternativa a la prueba t, especialmente cuando no se supone que la muestra de datos siga una distribución normal, y en este caso como método paramétrico, se utilizó para probar si una estimación es diferente de su valor real.

H_0 : El promedio del grado de escolaridad de la población de 15 años y más en México es igual a 9.1.
H_1 : El promedio del grado de escolaridad de la población de 15 años y más en México no es igual a 9.1.

Al realizar la prueba con la instrucción `wilcox.test(x, mu, conf.int = TRUE)` obtenemos un valor de $p = 0.3526 > 0.05$, por lo que no hay evidencia estadística para rechazar H_0, es decir, el promedio del grado de escolaridad de la población de 15 años y más en México es igual a 9.1 como lo dice el INEGI, nótese que a esa misma conclusión se había llegado con la prueba t.

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

Tanto la prueba t como la prueba de rango de Wilcoxon se pueden utilizar para comparar la *media de dos muestras*. La diferencia es que la prueba t supone que las muestras que se prueban se extraen de una distribución normal, mientras que la prueba de suma de rangos de Wilcoxon no lo hace. Para nuestro ejemplo, tomamos la misma base de datos, solo que en este caso separamos por género (hombres y mujeres), así las hípótesis planteadas en este caso son:

H_0 : El promedio del grado de escolaridad de hombres es igual al de las mujeres (de 15 años y más en México).
H_1 : El promedio del grado de escolaridad de hombres no es igual al de las mujeres (de 15 años y más en México).

Al realizar la prueba con la instrucción `t.test(x, y)` obtenemos un valor de $p = 0.2494 > 0.05$, por lo que no hay suficiente evidencia estadística para rechazar H_0, es decir, el promedio del grado de escolaridad entre hombres y mujeres es el mismo en el país, esto

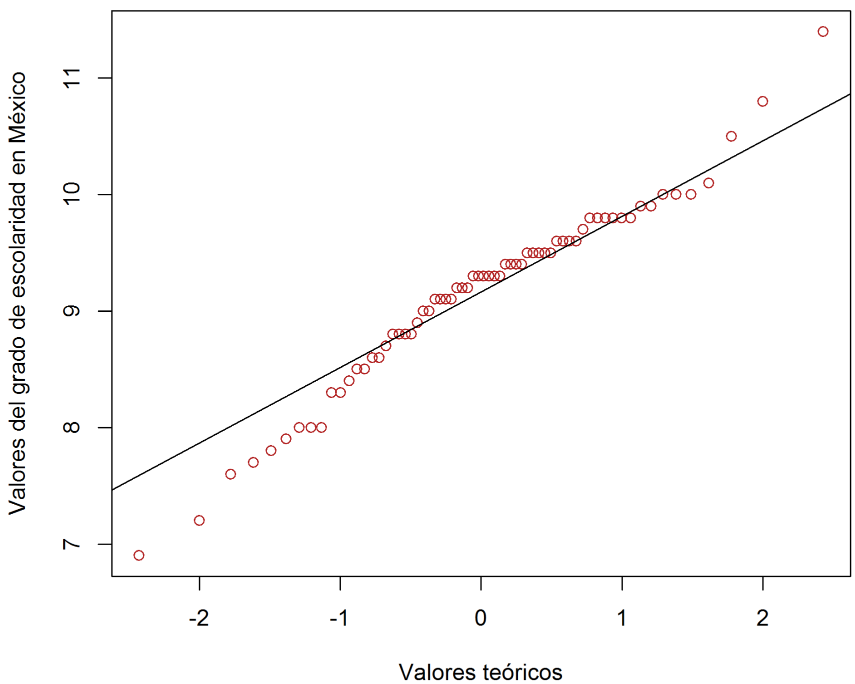

Figura 3: Gráfica QQ Normal para los datos del grado de escolaridad en el 2015 de la población en México.

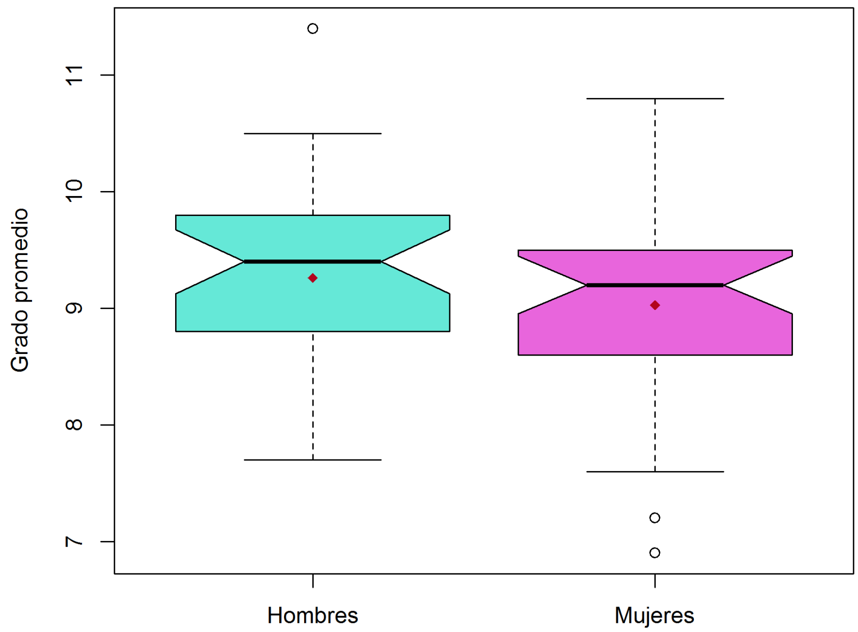

Figura 4: Gráfica de cajas para los datos del grado de escolaridad de hombres y mujeres.

suponiendo que no tenemos problema con que los datos sigan una distribución normal, ya que se tienen 33 observaciones. En la figura 4 se puede apreciar el gráfico de caja para las dos variables, de igual forma se pueden ver los extremos de los intervalos de confianza para las medias, así como la media muestral para cada variables, lo que nos hace pensar que efectivamente no parece haber una diferencia entre las medias.

Al realizar la prueba con la instrucción `wilcox.test(x, y)` obtenemos un valor de $p = 0.2749 > 0.05$, por lo que no hay suficiente evidencia estadística para rechazar H_0, es decir, concluimos igual que con la prueba t.

4. Shapiro Test

Esta prueba se utiliza para saber si una muestra sigue una distribución normal. La hipótesis nula aquí es que la muestra que se está probando tiene una distribución normal. Para nuestro ejemplo, utilizamos la base de datos del INEGI (`Prueba2.txt`), que hace referencia a la tasa de abandono escolar por entidad federativa según nivel educativo (nivel superior) en el ciclo escolar 2018/2019. Las hipótesis quedan planteadas como sigue:

H_0 : La tasa de abandono escolar en el nivel superior de los estados del país siguen una

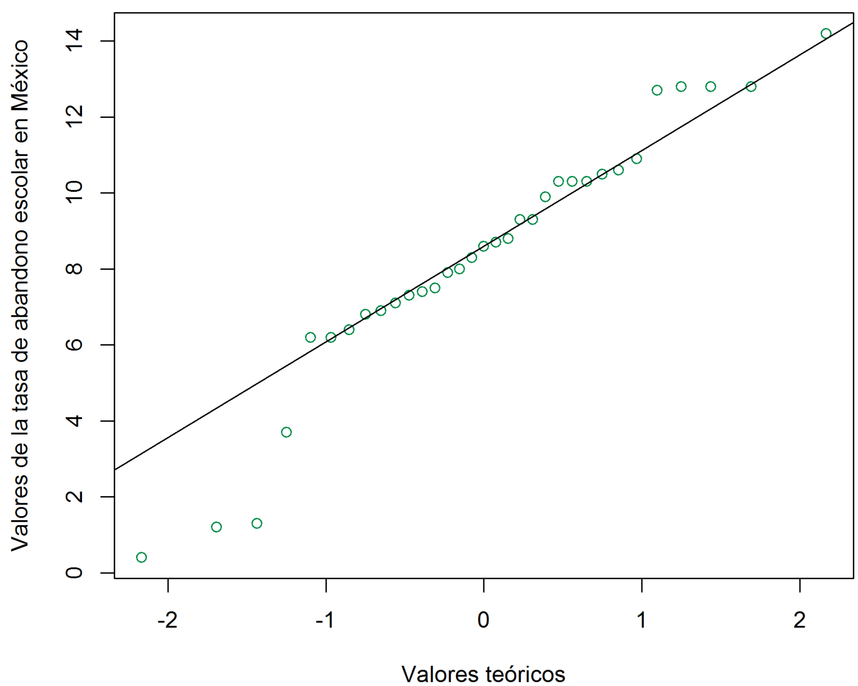

Figura 5: Gráfica QQ Normal para la tasa de abandono escolar en el nivel superior del país.

distribución normal.
H_1 : La tasa de abandono escolar en el nivel superior de los estados del país no siguen una distribución normal.

Al realizar la prueba con la instrucción `shapiro.test(x)` obtenemos un valor de $p = 0.077 > 0.05$, nótese que apenas el valor de p es ligeramente mayor, por lo que se tendría que utilizar algunas otras herramientas estadísticas para poder aceptar H_0, pero, siguiendo que $p > 0.005$ podemos concluir que hay un poco de evidencia estadística para no rechazar H_0, es decir, al parecer nuestros datos provienen de una distribución normal. En la figura 5 se puede apreciar un poco la falta de evidencia estadística, ya que al inicio y al final de la gráfica podemos visualizar como los datos se despegan de la recta, pero la mayoría de los puntos se sobre ponen a ella.

5. Kolmogorov and Smirnov Test

Esta prueba se utiliza para comprobar si dos muestras siguen la misma distribución. En este caso se consideró una base de datos del INEGI (`Prueba3.txt`), que hace referencia a la tasa de abandono escolar por entidad federativa según nivel educativo (nivel primaria y superior) en el ciclo escolar 2018/2019. Las hipótesis a considerar en este caso son:

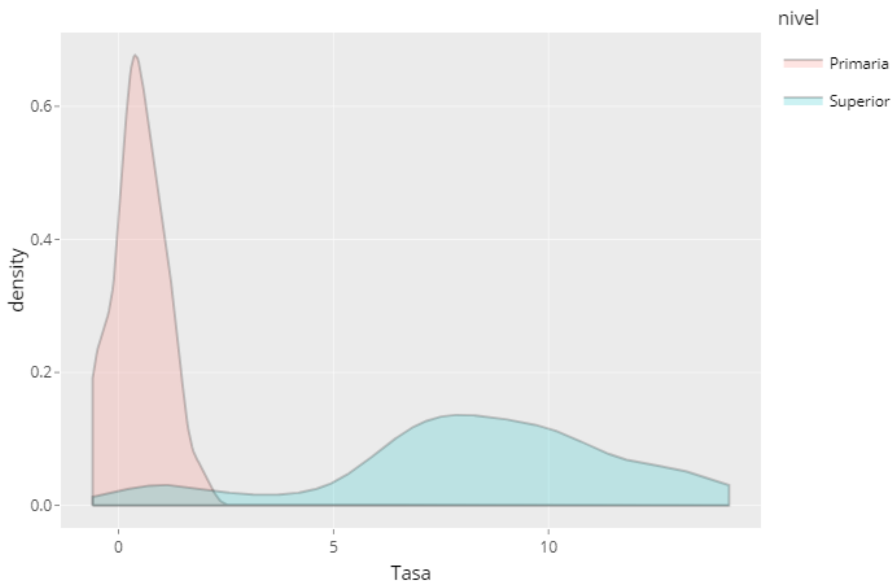

Figura 6: Gráfica QQ Normal para la tasa de abandono escolar en el nivel superior del país.

H_0 : La tasa de abandono escolar en el nivel primaria y superior siguen la misma distribuión.
H_1 : La tasa de abandono escolar en el nivel primaria y superior no siguen la misma distribuión.

Al realizar la prueba con la instrucción `ks.test(x, y)` obtenemos un valor de $p = 2.862 \times 10^{-12}$, lo cual es exageradamente menor a 0.05, por lo que hay suficiente evidencia estadística para rechazar H_0, por lo que se puede concluir que la distribución de la tasa de abandono escolar en el nivel primaria no es la misma que la del nivel superior. En la figura 6 podemos observar la gráfica de densidad de las variables para el nivel primaria y el nivel superior, y visualmente podemos observar que efectivamente no siguen una misma distribución.

6. Fisher's F–Test

La prueba F de Fisher se utiliza para verificar si dos muestras tienen la misma varianza. Para este caso se utilizo la base de datos respecto a la tasa de abandono escolar por entidad federativa según nivel educativo (nivel primaria y superior) en el ciclo escolar 2018/2019. Las hipótesis a considerar en este caso son:

H_0 : La tasa de abandono escolar en el nivel primaria y la del nivel superior tienen la misma varianza.
H_1 : La tasa de abandono escolar en el nivel primaria y la del nivel superior no tienen la misma varianza.

Para comprobar estas hipótesis utilizamos la instrucción `var.test(x, y)`, dicha prueba

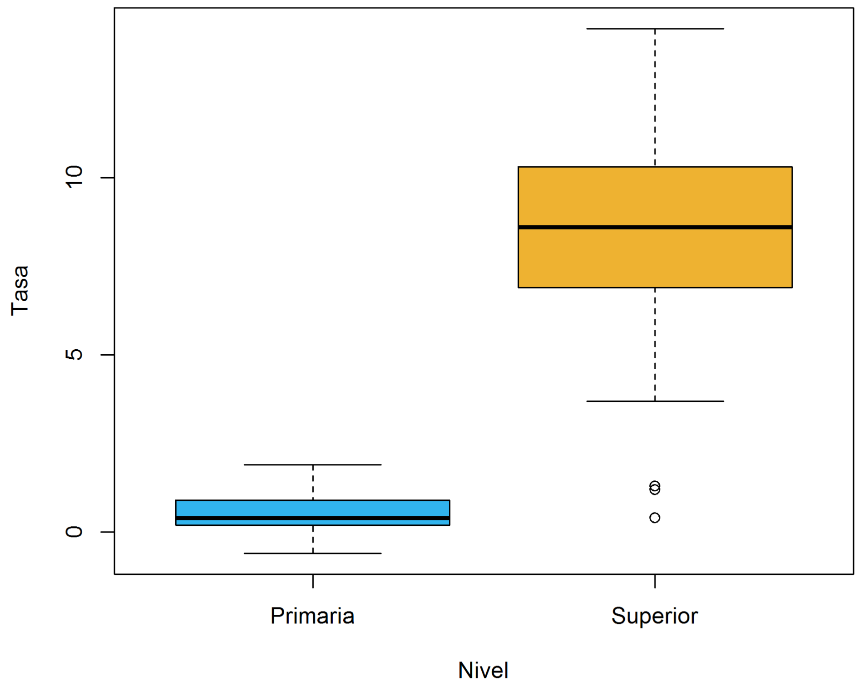

Figura 7: Gráfica de cajas para los datos de la tasa de abandono escolar en el nivel primaria y superior.

arroja un valor de $p = 2.804 \times 10^{-16}$, por lo que hay suficiente evidencia estadística para rechazar H_0, por lo que se puede concluir que la varianza en los datos de la tasa de abandono escolar en el nivel primaria es diferente al del nivel superior, por lo que habría que hacer otro análisis más profundo para conocer cual presenta más variación en sus datos. La figura 7 nos da la idea de que efectivamente nuestros datos no presentan igualdad de varianzas.

7. Chi Squared Test

La prueba de χ^2 implementada en el lenguaje R se puede utilizar para probar si dos variables categóricas son dependientes, mediante una tabla de contingencia. Esta prueba evalúa si existe una asociación significativa entre las categorías de las dos variables. En esta ocasión se utilizó una base de datos con las variables *Educación* (Maestros, Escuelas) y *Nivel escolar* (Preescolar, Primaria, Secundaria, Media_superior, Superior), en la cual se indica la cantidad de maestros y escuelas que hubo en todo el país durante el ciclo escolar 2018/2019, la tabla de contigencia se muestra en el cuadro 1, y la figura 8 se muestra una matriz gráfica de la tabla de contigencia, donde cada celda contiene un punto cuyo tamaño refleja la magnitud relativa del componente correspondiente. Las hipótesis a considerar en este caso son:

	Maestros	Escuelas
Preescolar	236,509	90,446
Primaria	572,104	96,508
Secundaria	406,084	39,967
Media_superior	418,893	21,010
Superior	414,408	5,535

Cuadro 1: Tabla de contigencia para la cantidad de maestros y escuelas que hay en los diferentes niveles educativos del país.

H_0 : La variable Educación es independiente de la variable Escuela.
H_1 : La variable Educación no es independiente de la variable Escuela.

Para comprobar estas hipótesis utilizamos la instrucción `chisq.test(tabla)`, dicha prueba arroja un valor de $p < 2.2 \times 10^{-16}$, por lo que hay suficiente evidencia estadística para rechazar H_0, es decir, la variable Educación y la variable Nivel escolar son dependientes, es decir, habrá menos o más maestros/escuelas dependiendo del nivel educativo.

8. Correlation

La prueba de correlación se utiliza para comprobar la relación lineal entre dos variables continuas y prueba si una de ellas es dependiente de la otra. Para esta prueba se consideró una base de datos del INEGI (`Prueba5.txt`), la cual contiene la cantidad de hombres y mujeres inscritos en el nivel superior en todos los estados del país durante el ciclo escolar 2018/2019, las hipótesis a considerar en este caso son:

H_0 : La cantidad de hombres (x) inscritos no esta relacionada con la cantidad de mujeres (y) inscritas en el nivel superior ($\rho = 0$).
H_1 : La cantidad de hombres inscritos esta relacionada con la cantidad de mujeres inscritas en el nivel superior ($\rho \neq 0$).

En la figura 9 se muestra el diagrama de dispersión de los datos, en dónde se puede observar que existe una relación lineal positiva entre las variables, esto se comprobó al realizar la prueba de correlación con la instrucción `cor.test(x, y)`, dicha prueba arrojó un valor de $p < 2.2 \times 10^{-16}$, por lo que hay evidencia estadística para rechazar H_0, es decir, existe una correlación significativa entre las variables.

Un contra de la prueba es que no nos dice qué tipo de relación lineal existe entre las variables, para esto utilizamos lo que se conoce como **coeficiente de correlación**, este se utiliza para medir la magnitud de la relación lineal entre dos variables, es decir, indica cuán fuerte o débil es una relación lineal. También se le conoce como r de Pearson, en honor a Karl Pearson. El rango de valores varia entre -1 y $+1$; los valores intermedios pueden interpretarse, de forma intuitiva [1] como se muestra en la figura 10. La implementación de la prueba de correlación en el ambiente R arroja también el valor del coeficiente de correlación,

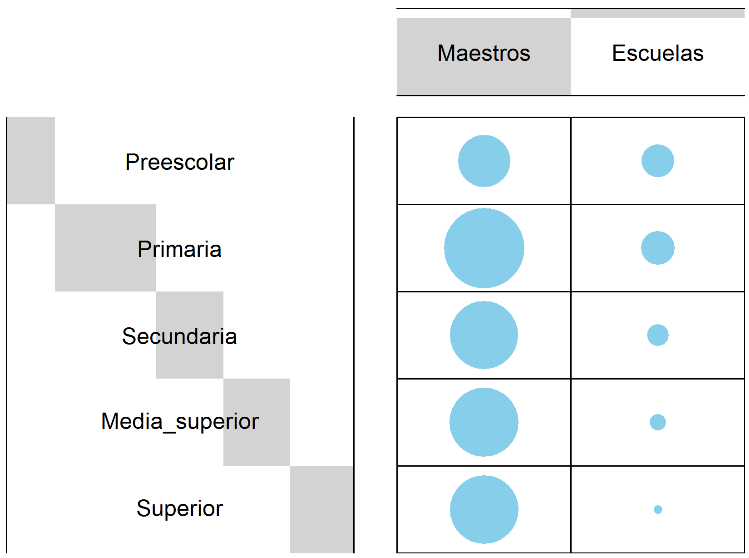

Figura 8: Tabla de contigencia para la cantidad de maestros y escuelas que hay en los diferentes niveles educativos del país.

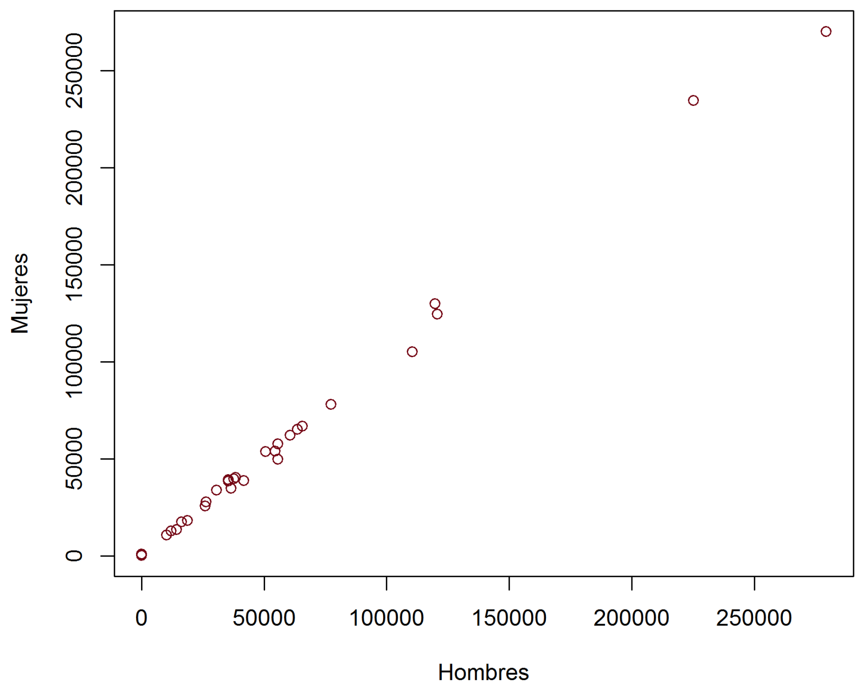

Figura 9: Diagrama de dispersión para la cantidad de hombres y mujeres inscritos(as) en el nivel superior.

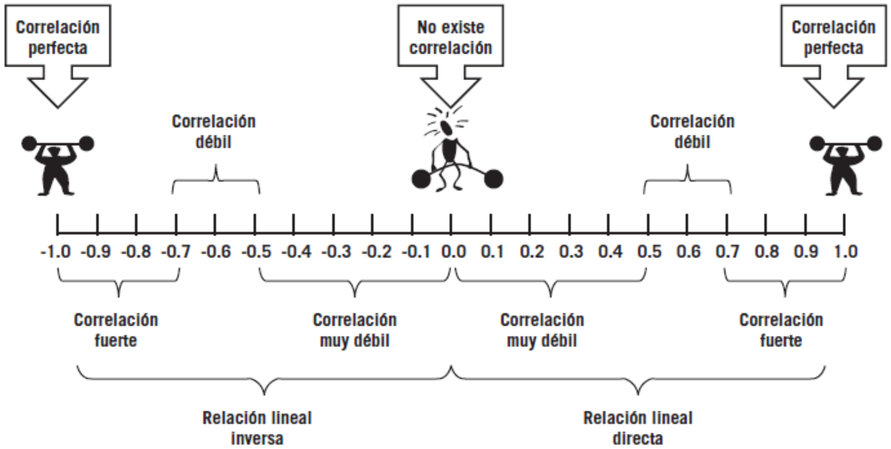

Figura 10: Interpretación del índice de correlación.

para este caso es $r = 0.9984378$, lo cual nos indica que la asociación lineal entre la cantidad de hombres inscritos y la cantidad de mujeres inscritas en el nivel superior es directa y fuerte, es decir, cuanto mayor es el número de hombres inscritos, mayor es el número de mujeres inscritas, y viceversa.

Referencias

[1] A. G. Banegas. *Probabilidad y Estadística, Enfoque por competencias.* McGraw Hill, México, 2012.

[2] I. N. de Estadística y Geografía. Características educativas de la población. https://www.inegi.org.mx/temas/educacion/default.html#Tabulados, 2020.

[3] R Core Team. R: A Language and Environment for Statistical Computing. https://www.R-project.org/, 2020.

[4] M. F. Triola. *Estadística.* Pearson Education, México, 2004.