LABORATORIO PARA EL ÁNALISIS DE DATOS ECONÓMICOS Y FINANCIEROS

Examen domiciliario de R — 2025

Universidad Torcuato Di Tella

Barahona Pineda Blanca, Cruceño Karen, Yapur Najla

Ejercicio de Análisis de datos

Para la resolución de este ejercicio se utilizó el dataset "Spotify Tracks" que contiene 114 géneros distintos. La descripción de cada una de las variables del dataset se puede encontrar en **SpotifyTracks.csv** , en Kaggle.

1. Identifique los géneros musicales que representan la mayor popularidad en Spotify. En particular, calcule en promedio los 5 géneros más y menos escuchados. Realice un **gráfico de barras** para ambos resultados, indicando en el eje horizontal los géneros y en el eje vertical su popularidad promedio.

   Respuesta: Los cinco géneros más escuchados en Spotify son:

   Table 1: Géneros más populares

track_genre	**prom_popularity**
pop-film	59.3
k-pop	56.9
chill	53.7
sad	52.4
grunge	49.6

   Los cinco géneros menos escuchados:

   Table 2: Géneros menos populares

track_genre	**prom_popularity**
iranian	2.21
romance	3.24
latin	8.30
detroit-techno	11.2
chicago-house	12.3

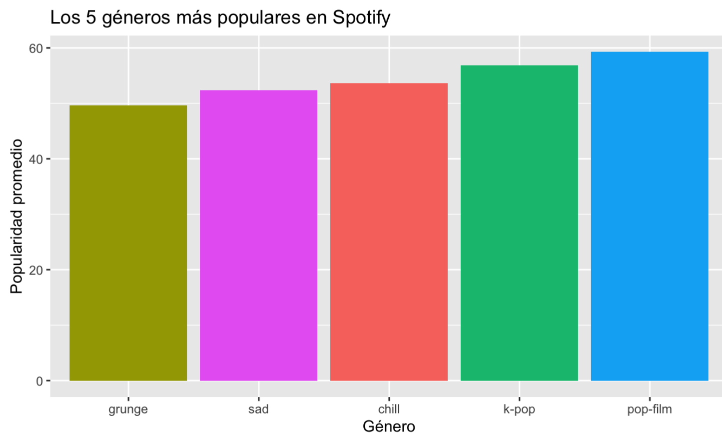

Figure 1: Los cinco géneros más escuchados

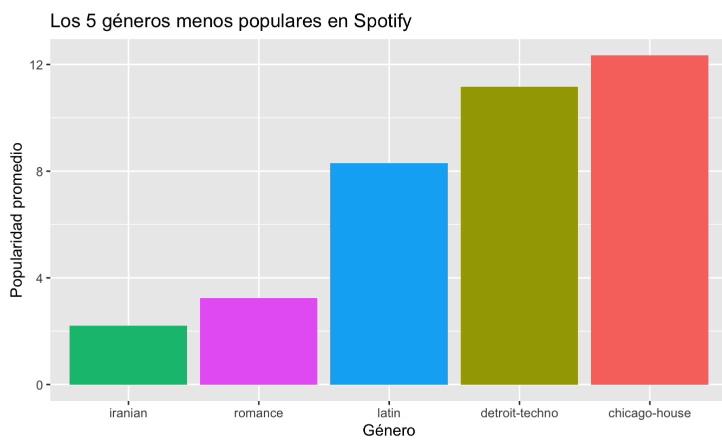

Figure 2: Los cinco géneros menos escuchados

2. Ahora intentaremos analizar si existe relación entre la capacidad de baile de una canción (*danceability*) y la energía de la misma (*energy*). Calcule el **coeficiente de correlación** entre ambas variables y grafíquelas en un **scatter plot** (diagrama de dispersión), indicando en el eje horizontal la variable "energy" y en el vertical "danceability". Interprete los resultados obtenidos.

   Respuesta: El coeficiente de correlación es r = 0.1343255. Esto nos indica que existe una correlación positiva, pero débil entre las variables *danceability* y la variable *energy*. En promedio, las canciones con mayor nivel de energía tienden a ser ligeramente más bailables. Este resultados lo podemos obervar en el scatter plot y la pendiente de la línea de regresión.

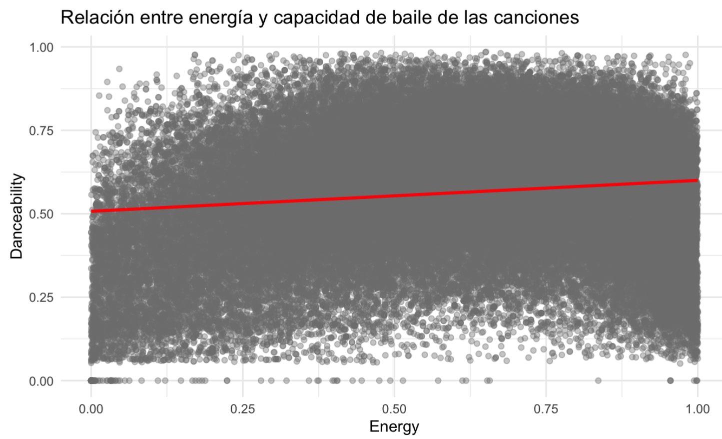

Figure 3: Scatter plor entre energy y danceability

3. El objetivo de este inciso es determinar si las canciones con contenido explícito tienden a ser más populares que las que no lo son.

   En el dataset "Spotify Tracks", se incluye la variable **"explicit"** la cual indica si la canción tiene letras explícitas (`TRUE` = tiene; `FALSE` = no tiene; o desconocido) y la variable **"popularity"** valuada entre 0 y 100, siendo 100 la más popular.

   - Calcule para cada categoría de "explicit", el **promedio** y la **desviación estándar** de la variable *popularity*.
   - Luego, realice un **boxplot** (diagrama de caja) que muestre la distribución de *popularity* según las categorías de *explicit*.
   - Interprete los resultados.

Respuesta: Utilizando los comandos `group by` y `summarise` obtenemos el promedio y la desviación estándar de la variable *popilarity*.

Table 3: Estadísticas de popularidad agrupadas por contenido explícito

explicit	**prom_popularity**	**sd_popularity**	**observaciones**
False	32.9	22.1	104253
True	36.5	24.3	9747

La canciones explîcitas presentan, en promedio, mayor popularidad de las que no tienden contenido explícito, estas tienden a ser ligeramente más escuchadas. Sin embargo, hay que tener en cuenta la cantidad de observciones en cada grupo. Observamos que hay 104253 canciones que no son explícitas y 9747 que sí lo son.

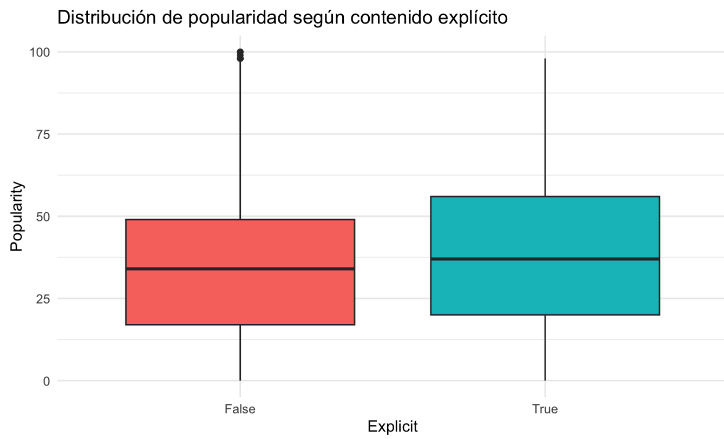

Figure 4: Análisis de la popularidad por contenido explícito

Observamos los resultados mencionados antes en el diagrama de caja. Las canciones explícitas poseen una mediana mas alta de las no explícitas. Por otro lado, se puede obervsar una distribución más dispersa en las canciones explícitas.

Ejercicio: Análisis econométrico con datos de `gapminder`

Parte 1: ingreso por persona

1. Grafica la evolución temporal de la variable `income per person` en Argentina. Comenta brevemente la tendencia observada.

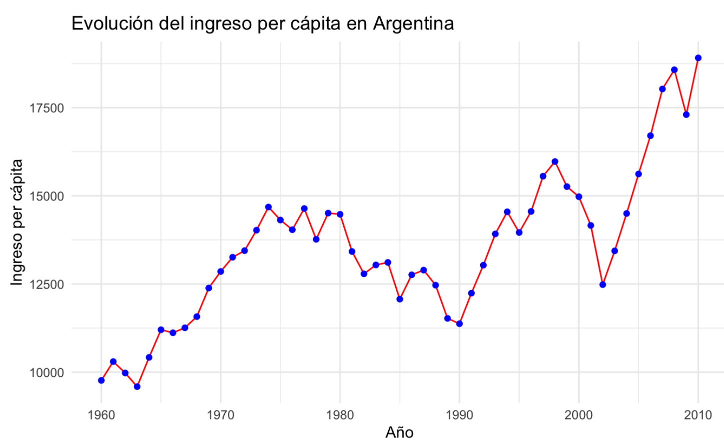

Figure 5: Evolución temporal de `income per person`

Respuesta: El gráfico muestra una economía cíclica característica de Argentina. Se observa una tendencia creciente marcada por una alta volatilidad. Es visible cierto patrón en la trayectoria que muestra una desaceleración en la economía o crisis posterior a un auge de crecimiento.

2. Para el entrenamiento, separar los últimos 10 años de Argentina para testear posteriormente. Estima regresiones de income per person sobre el tiempo t utilizando tres modelos:

   (a) Modelo lineal: $\text{income\_per\_person} = \beta_1 t + \beta_0 + \varepsilon$

   (b) Modelo polinómico de grado 2: $\text{income\_per\_person} = \beta_2 t^2 + \beta_1 t + \beta_0 + \varepsilon$

   (c) Modelo polinómico de grado 10: $\text{income\_per\_person} \sim \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_{10} t^{10} + \varepsilon$

   Representa gráficamente los tres ajustes junto con los datos observados y comenta las diferencias. Visualizar y cuantificar los resultados sobre el conjunto de testeo. ¿Qué observa?

   Respuesta: Utilizamos el comando `filter` para separar en conjunto de testeo y entrenamiento.

0.1 Estimaciones sobre el conjunto de entrenamiento

Regresiones de los tres modelos:

(a) Modelo lineal estimado es:

$$\widehat{\text{income\_per\_person}} = 92.95 \times \text{year} - 171079.77$$

Table 4: Coeficientes del Modelo de Regresión Lineal

	Estimate	**Std. Error**	**t value**	**Pr(> |t|)**	
(Intercept)	−171079.77	32497.28	−5.264	5.42e-06	***
year	92.95	16.41	5.663	1.52e-06	***

- Residual standard error: 1243 on 39 degrees of freedom
- Multiple R-squared: 0.4513
- Adjusted R-squared: 0.4372
- F-statistic: 32.07 on 1 and 39 DF, p-value: 1.521e-06

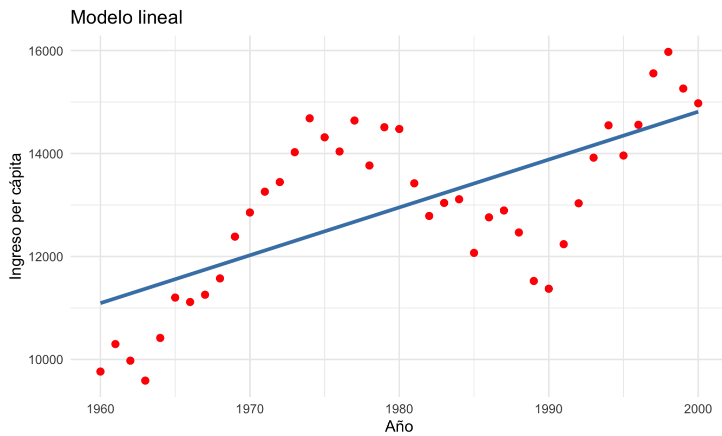

Figure 6: Modelo de regresión lineal para el ingreso per cápita

Interpretación: Observamos que el modelo lineal captura de manera optima la tendencia en la evolución del ingreso per cápita. El coeficiente estimado *year* nos indica el efecto marginal manteniendo el resto de variables contantes. Sin embargo, la medida de la bondad del ajuste de nuestro modelo nos sugiere que el modelo no tienen un alto poder explicativo, el R^2 es 0,45.

(b) Modelo polinómico de grado 2:

Table 5: Coeficientes del Modelo de Regresión Polinómico de Grado 2

	Estimate	**Std. Error**	**t value**	**Pr($>|t|$)**	
(Intercept)	12952.6	188.2	68.834	$<$ 2e-16	***
poly(year, 2)1	7041.8	1204.9	5.844	9.31e-07	***
poly(year, 2)2	-2266.0	1204.9	-1.881	0.0677	.

- Residual standard error: 1205 on 38 degrees of freedom
- Multiple R-squared: 0.498
- Adjusted R-squared: 0.4716
- F-statistic: 18.85 on 2 and 38 DF, p-value: 2.059e-06

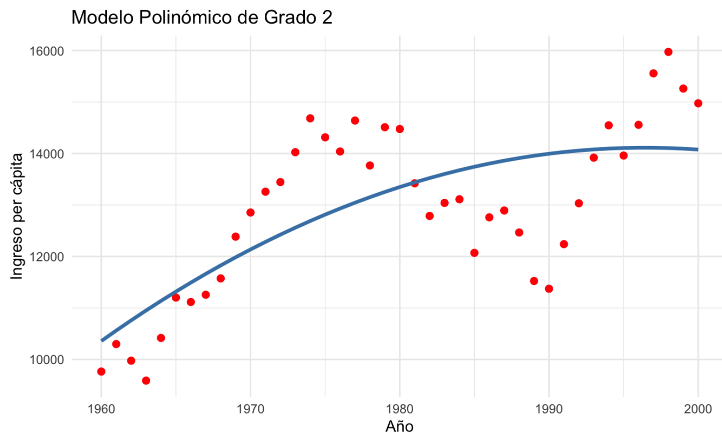

Figure 7: Modelo de regresión para el ingreso per cápita

Interpretación: El modelo polinómico de segundo grado difiere del lineal en la interpretacion de los coeficientes, ya no son lineales. Dada la concavidad, se puede ver que, a lo largo del período, el ingreso per cápita aumento, pero a una tasa decreciente. El R^2 ajustado si bien es levemente mayor al modelo lineal, nos sigue indicando que el modelo no ajusta perfectamente a los datos.

(c) Modelo polinómico de grado 10:

Table 6: Coeficientes del Modelo de Regresión Polinómico de Grado 10

	Estimate	**Std. Error**	**t value**	**Pr(> $\|t\|$)**	
(Intercept)	12952.610	71.663	180.743	< 2e-16	***
poly(year, 10)1	7041.821	458.869	15.346	9.55e-16	***
poly(year, 10)2	−2265.988	458.869	−4.938	2.77e-05	***
poly(year, 10)3	5427.527	458.869	11.828	8.00e-13	***
poly(year, 10)4	1973.707	458.869	4.301	0.000166	***
poly(year, 10)5	−3611.115	458.869	−7.870	8.79e-09	***
poly(year, 10)6	−1538.115	458.869	−3.352	0.002182	**
poly(year, 10)7	3.248	458.869	0.007	0.994400	
poly(year, 10)8	−131.186	458.869	−0.286	0.776927	
poly(year, 10)9	158.404	458.869	0.345	0.732349	
poly(year, 10)10	218.333	458.869	0.476	0.637659	

- Residual standard error: 458.9 on 30 degrees of freedom
- Multiple R-squared: 0.9425
- Adjusted R-squared: 0.9234
- F-statistic: 49.19 on 10 and 30 DF, p-value: 7.68e-16

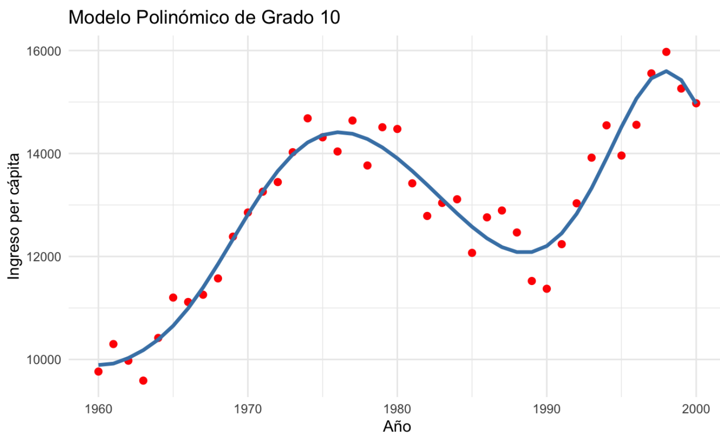

Figure 8: Modelo de regresión para el ingreso per cápita

Interpretación: El R^2 ajustado del modelo polinómico de grado 10 indica que el modelo se ajusta casi perfectamente a los datos. Sin embargo, el modelo presenta un problema de *overfitting*, es decir, se ajusta exactamente a nuestros datos. Por ende, estamos frente a un modelo ineficiente para hacer predicciones.

0.2 Testeo de los modelos

Calculamos el *RMSE* (Root Mean Square Error) para los tres modelos:

- Modelo lineal: el RMSE es 2012.47.
- Modelo polinómico de grado 2: el RMSE es 3114.47.
- Modelo polinómico de grado 10: el RMSE es 138836.8.

Al evaluar los tres modelos, se observa que el modelo lineal es el que obtiene el menor de predicción seguido del polinómico de segundo grado. En cambio, el modelo de grado 10, tal como habíamos mencionado anteriormente, arrojó un error extremandamente alto por lo que indica un sobreajuste.

Graficamos las predicciones sobre el conjunto de testeo:

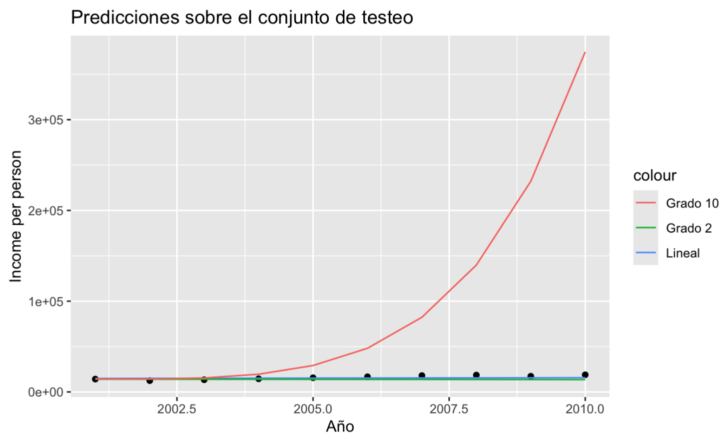

Figure 9: Predicciones sobre test

Conclusión: El modelo lineal y el polinómico de grado 2 son los que mejor se ajustan a los datos y los que poseen capacidad predictiva.

3. Selecciona cuatro países sudamericanos distintos de Argentina. Con ellos, arma:

   (a) Una matriz de correlaciones entre los ingresos (`income per person`) de los cinco países.

   (b) Una matriz de correlaciones entre las variaciones porcentuales anuales (crecimiento interanual, Y /Y) de dichos ingresos.

   Discute brevemente las similitudes y diferencias encontradas. Respuesta:

   (a) Utilizando los comandos `filter`, `select`, `pivot wider` y `cor`:

Table 7: Matriz de correlaciones entre los ingresos

	Argentina	**Brazil**	**Chile**	**Peru**	**Uruguay**
Argentina	1.000000	0.795111	0.765041	0.781651	0.829183
Brazil	0.795111	1.000000	0.771716	0.564863	0.871349
Chile	0.765041	0.771716	1.000000	0.548526	0.940793
Peru	0.781651	0.564863	0.548526	1.000000	0.577266
Uruguay	0.829183	0.871349	0.940793	0.577266	1.000000

   (b) Utilizando los comandos `mutate` y `cor`:

Table 8: Matriz de correlaciones entre variaciones porcuentuales anuales

	Argentina	**Brazil**	**Chile**	**Peru**	**Uruguay**
Argentina	1.00000000	0.27205223	0.16919886	0.35531638	0.5127562
Brazil	0.2720522	1.00000000	0.00612927	0.41054674	0.2779881
Chile	0.1691989	0.00612927	1.00000000	0.04320042	0.3655066
Peru	0.3553164	0.41054674	0.04320042	1.00000000	0.4126808
Uruguay	0.5127562	0.27798812	0.36550663	0.41268077	1.0000000

Al analizar la matriz de correlaciones entre los ingresos

Al analizar ambas matrices, se observa que tanto el nivel como la variación del ingreso de un país, si es alto, tiende a asociarse con un nivel alto en los demas países. La magnitud de la matriz de correlaciones entre los ingresos es significativamente alta, por lo que deja ver que el nivel de ingresos de los países se mueven en la misma medida. En cambio, en la matriz de correlaciones entre variaciones la magnitud es relativamente más baja que en el de ingresos. Por lo tanto, el crecimiento anual de los países es independiente entre ellos y no responde a shocks externos de manera significativa.

Parte 2: Esperanza de vida y género

5. Gráfica `life_expectancy` frente a `life_expectancy_female`. ¿Qué relación observas entre ambas variables?

   Respuesta: Filtrando los datos para el año 2000, graficamos `life_expectancy` frente a `life_expectancy_female`.

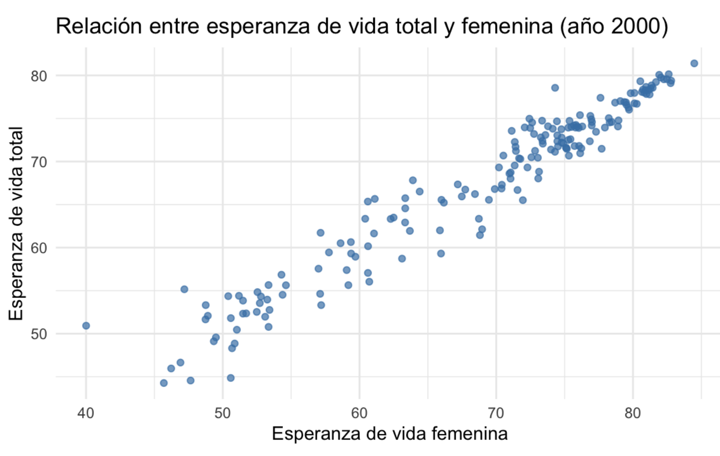

Figure 10: Esperanza de vida

Observamos una tendencia positiva y creciente entre ambas variables. Se puede apreciar que a mayor esperanza de vida, mayor será la esperanza de vida de las mujeres.

6. Estima una regresión simple de `life_expectancy` sobre `life_expectancy_female`. Analiza el valor del coeficiente estimado. Calcular R^2. ¿Qué observa? ¿Qué implica este resultado?

   Respuesta: Estimamos la regresión simple utilizando el comando `lm()`.

   Table 9: Coeficientes del modelo simple

	Estimate	**Std. Error**	**t value**	**Pr($>|t|$)**	
(Intercept)	7.91447	1.10302	7.175	$1.76e-11$	***
life_expectancy_female	0.86534	0.01585	54.583	$< 2e-16$	***

   Bondad de ajuste:

   - Múltiple R^2: 0.9424
   - Ajustado R^2: 0.9421

   El coeficiente estimado para `life_expectancy_female` es 0.86534, este es el efecto marginal de aumentar la esperanza de vida femenina sobre la esperanza de vida total manteniendo lo demás constante.

   Por otro lado, se observa un R^2 significativamente alto de 0.94. Este resultado nos indica que nuestro modelo estimado explica la variación de la esperanza de vida total casi en su totalidad.

7. Contrasta la hipótesis nula H_0: `life_expectancy_female` = `life_expectancy` contra la alternativa de que `life_expectancy_female` es mayor. ¿Cuál es la evidencia empírica al respecto?

   Respuesta:

   - $\mathbf{H_0}$: la media de la diferencia = 0 (las mujeres no viven más que la media total).
   - $\mathbf{H_1}$: la media de la diferencia > 0 (las mujeres viven más que la media total).

   Calculamos el `t_test`, obtenemos:

   - Estadístico $t = 6.754$
   - $df = 183$
   - p-valor $= 9.24e^{-11}$
   - Intervalo de confianza 95% : $(1.02, \infty)$
   - Media muestral de la diferencia = 1.34

Como $p - valor < 0,01$, hay evidencia suficiente para rechazar la hipótesis nula.

8. Estima una regresión múltiple de `life_expectancy` sobre `life_expectancy_female` e `income_per_person`. Interpreta los coeficientes de ambos regresores. Discute qué ocurre con la relación entre esperanza de vida y género al controlar por el nivel de ingresos. Calcular R^2 y comparar con la regresión simple. ¿Vale la pena incluir la variable `income_per_person`?

   Respuesta:

   Estimamos la regresión múltiple:

Table 10: Coeficientes del modelo múltiple

	Estimate	**Std. Error**	**t value**	**Pr(> |t|)**	
(Intercept)	$8.491e+00$	$1.233e+00$	6.886	$9.11e-11$	***
life_expectancy_female	$8.546e-01$	$1.888e-02$	45.277	$< 2e-16$	***
income_per_person	$1.162e-05$	$1.113e-05$	1.045	0.297	

- Error estándar residual: 2.29 sobre 181 grados de libertad
- R^2 múltiple: 0.9428
- R^2 ajustado: 0.9421
- Estadístico F: 1491 sobre 2 y 181 DF, p-valor: $< 2.2e^{-16}$

   El coeficiente estimado para `life_expectancy_female` es 0.8546, mientras que para `income_per_person` fue 0.0000116. Por ende, incluir la variable ingresos no cambia la magnitud, es decir, no aporta información relevante.

   El R^2 que se obtuvo en el modelo simple y múltiple fue 0.9424 y 0.9428, respectivamente. Dado que el R^2 es ligeramente mayor y el coeficiente estimado de los ingresos no es relevante, se puede concluir que no vale la pena incluir la variable `income_per_person` al modelo.

9. Excluyendo `life_expectancy_female` y las variables que por algún motivo considere conveniente excluir, elegir un modelo lineal de **máximo** tres covariables que explique `life_expectancy`. Explicar y justificar cada paso. **Aclaración:** No es necesario que esas tres covariables sean columnas originales del dataset. Pueden generarse una nueva columna a partir de las existentes.

   Respuesta: Excluimos `life_expectancy_female` para explicar `life_expectancy`. Se elegieron las siguientes variables para explicar la esperanza de vida: `child_mortality`, `income_per_person` y `children_per_woman`.

   Se eligieron esas variables, pues son relevantes a la hora de explicar la esperanza de vida de una economía. La tasa de mortalidad infantil nos indica la calidad del sistema sanitario, además, es una variable de la cual depende negativamente: si baja la mortalidad infantil, aumenta la esperanza de vida. Por otro lado, los ingresos también es una variable significativa, ya que, la calidad de vida de una persona es dependiente del nivel de ingresos de la misma. Por último, la cantidad de hijos por mujer tiene una interpretación similar a la mortalidad infantil, esta variable nos sugiere que en países con alta natalidad suele estar relacionada con un sistema sanitario ineficiente y, por lo tanto, una menor esperanza de vida.

Modelo lineal:

$$life\_expectancy = \beta_0 + \beta_1 \cdot income\_per\_person + \beta_2 \cdot child\_mortality + \beta_3 \cdot children\_per\_woman + \varepsilon$$

Utilizando el comando `lm()` para estimar el modelo obtuvimos los siguientes coeficientes:

Table 11: Coeficientes del Modelo de Regresión Lineal para life_expectancy

Variables	**Estimate**	**Std. Error**	t **value**	**Pr**($>\|t\|$)
(Intercept)	$7.548e+01$	$1.128e-01$	668.98	$< 2e-16$ ***
income_per_person	$5.989e-05$	$2.545e-06$	23.53	$< 2e-16$ ***
child_mortality	$-8.687e-02$	$8.099e-04$	-107.26	$< 2e-16$ ***
children_per_woman	$-9.667e-01$	$3.261e-02$	-29.64	$< 2e-16$ ***

*** $p < 0.001$

Ejercicio Simulación 3: Envido en el Truco

1. Genera una variable mazo que contenga todas las cartas del Truco (mazo español de 40 cartas: valores del 1 al 7 y del 10 al 12 en los cuatro palos: espadas, bastos, oros y copas).

2. Define una función sacar al azar(mazo) que simule la extracción aleatoria de tres cartas y devuelva tanto la mano obtenida como el mazo restante. Graficar un histograma del valor medio.

3. Implementa una función contar tantos(mano) que, dada una mano de tres cartas, calcule los tantos de envido siguiendo las reglas:

   • Si hay al menos dos cartas del mismo palo, se suman sus valores de envido (las cartas del 1 al 7 valen su número, y las del 10, 11 y 12 valen 0) y se agrega un bono de 20.

   • Si no hay dos del mismo palo, el tanto es el mayor valor de envido entre las tres cartas.

4. Escribe una función que, dada tu mano y el mazo restante, calcule la probabilidad exacta de que el oponente gane el envido en un partido de a dos. Considera que en caso de empate gana el jugador que es mano.

5. Simula la siguiente situación: estás jugando al Truco de seis jugadores (tres contra tres). Ya tienes tus tres cartas, y tus compañeros te hicieron seña de que poseen menos de 27 tantos cada uno. Calcula la probabilidad de que tu equipo gane el envido frente al equipo rival, teniendo en cuenta que:
   - El envido lo gana el jugador con el mayor puntaje individual.
   - En caso de empate, gana el equipo del jugador que sea mano
6. (Opcional) Simula muchas partidas y presenta un histograma del valor del máximo puntaje de envido alcanzado entre los tres jugadores del equipo rival. Analiza que tan frecuentes son los valores altos y como esto influye en la estrategia de tu equipo.

En este ejercicio estudiaremos, mediante simulaciones en R, el cálculo de tantos de envido y las probabilidades de ganar en distintas situaciones de juego del Truco.

A continuación, analizaremos los códigos de R que hemos utilizado para cada inciso e interpretaremos el significado de los resultados y gráficos.

Comenzando por el **Primer Inciso**, vamos a diseñar un código para que nos devuelva un total de 40 cartas dónde se encuentran los 4 palos y los valores de las cartas para jugar al Truco.

En este caso, expand.grid nos da todas las combinaciones entre los valores y los palos. Como pusimos la función head(mazo,10) nos aparecen las 10 cartas de espadas porque es el primer palo, pero si modificamos y ponemos 12, por ejemplo, aparece hasta 2 más de bastos.

En el **Segundo Inciso** definimos la función sacar_al_azar(), la cúal nos da 3 cartas aleatorias del mazo con sample() eligiendolas del mazo sin reemplazo y devuelve la mano y el mazo_restante eliminando esas filas. Luego, vamos a repetir este proceso 10.000 veces la extracción para estimar cómo se distribuye el valor promedio de una mano.

Como resultado, obtenemos el 1 de copas y 4 y 5 de bastos, con un mazo restante de 37 cartas.

Por otra parte, vemos nuestro histograma del valor promedio de muchas manos con la función hist(). La intuición detrás de esto es que cada punto del histograma es el promedio de los valores de 3 cartas tomadas al azar del mazo del Truco (valores 1–7, 10–12). Es decir, para valores medios de entre 4 y 7 vemos que la mayoría de las manos tienen cartas mezcladas, algunas bajas y otras medias, lo cuál conlleva a promedios intermedios. Mientras tanto si observamos a los extremos, para valores medios de 0 a 4 y de 8 a 12 es poco común que te salgan 3 cartas muy bajas o 3 muy altas.

En el **Tercer Inciso**, nos pide implementar una función contar tantos para las cúales cumple determinandas reglas, en este caso, si o si debemos usar el condicional If.

Teniendo en cuenta los datos obtenidos de 1 de copas, 5 de bastos y 4 de bastos obtenemos un número como resultado que es 29, ya que se repite el hecho

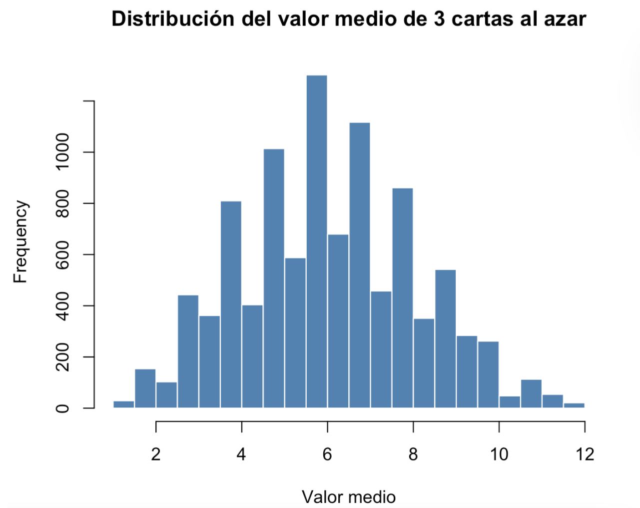

Figure 11: Distribución del valor medio de 3 cartas al azar

de que al menos tienen 2 cartas del mismo palo y ambas suman 29 (5+4+20 porque son del mismo palo).

Siguiendo por el **Cuarto Inciso**, vamos a calcular la probabilidad de que el oponente gane el envido en un partido de a dos. Como es de costumbre, primero ponemos `set.seed(123)` para que los resultados siempre se basen en la mismas cartas y no halla repeticiones. Luego de plantear las variables y funciones, llegamos a qué la probabilidad exacta de que el oponente gana el envido es 0.13 apróximadamente. Mientras que la probabilidad exacta de que gane el oponente que no es mano gane el envido es de 0.086 apróximadamente. Estos dos números son los resultados de la función `prob_oponente_gana_envido()`. ¿Qué quiere decir esto?

- 0.132175 (Significa que el oponente gana el envido con una probabilidad de 13,2 %) Cuando el oponente es mano, ya que en caso de empate gana él.
- 0.08648649 (Significa significa que el oponente gana el envido con una probabilidad de 8,6 %) Cuando "yo" soy mano ya que en caso de empate ganó yo.

Podemos asumir que el que es mano tiene una pequeña ventaja, porque gana los empates.

Si vamos al caso de nuestras 3 cartas que nos salieron en el segundo inciso con la función `sacar_al_azar()` es en su mayoría buena, ya que el rival solo gana el envido en el 13 % de los casos cuando tiene la ventaja del desempate, y en el 8 % cuando vos la tenés.

En el **Quinto inciso**, en base a nuestro escenario hipotético, tenemos 6 jugadores, en total 3 por equipo. Además, conocemos que "yo" tengo mis cartas y mis compañeros tienen menos de 27 puntos cada uno. Entonces, el envido se define por el mayor puntaje individual y en caso de que haya empate se define por el equipo que es mano. Por ende, nuestro objetivo es estimar la probabilidad de que nuestro equipo gane el envido.

¿Cómo lo calculamos? Empezamos simulando muchas partidas, 3 contra 3. En cada una, se comparan los tantos individuales de todos los jugadores. Dónde se quitan las 3 cartas del mazo. Se reparten 6 manos de 3 cartas (3 para nuestro equipo + 3 para el contrincante).

Adémas, por datos de enunciado sabemos que nuestro equipo tiene mayor o igual a 27 tantos, suponemos porque "me hicieron seña de menos de 27". Comparamos el máximo individual de tu equipo vs. el del contrincante y si hay empate gana el que es mano. Repetimos este proceso muchas veces con `(n_sim)` y devuelve la proporción de victorias de tu equipo.

Este es el resultado [1] 0.381, significa que la probabilidad estimada de que nuestro equipo gane el envido es de apróximadamente un 38,10 %.

Gracias a al parámetro `oponente_es_mano = TRUE`, con el resultado 0.381 indica que nuestro equipo gana solo el 38,10 % de las veces cuando el contrincante tiene la ventaja del desempate, es mano. Mientras que si el `oponente_es_mano = FALSE`, nuestro equipo gana el 38,10% de las veces aún así teniendo ventaje del desempate.

Por último, en el **Sexto Inciso**, vamos a simular y graficar. Cada simulación representa una partida distinta, con nuevas cartas. En cada una, se calcula el máximo tanto entre los 3 jugadores del equipo rival. A continuación, vamos a interpretar nuestros resultados con lo que nos muestra el histograma según cuán frecuentes son los valores altos de envido. Por ejemplo, cuántas veces alcanzan 30, 31, 32, etc.

Nuestros resultados son :

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
`max_rival`	2.00	25.00	27.00	25.94	29.00	33.00

Table 12: Resumen estadístico de `max_rival`.

La variable `max_rival` presenta un valor mínimo de 2 y un máximo de 33, con una media de 25.94 y una mediana de 27. El hecho de que la media y la mediana sean similares sugiere una distribución apróximadamente simétrica, sin presencia de valores atípicos extremos. El rango intercuartílico (IQR = 29 - 25 = 4) indica una dispersión moderada de los datos alrededor del centro. En general, los valores se concentran entre 25 y 29, mostrando que la mayoría de las observaciones se ubican cerca del promedio. El `summary()` nos permite visualizar rápido estos indicadores, calculando de manera automática los cuantiles, la media y los valores extremos de la variable.

Como podemos observar el histograma muestra solo la parte alta porque estamos graficando el máximo del equipo rival.

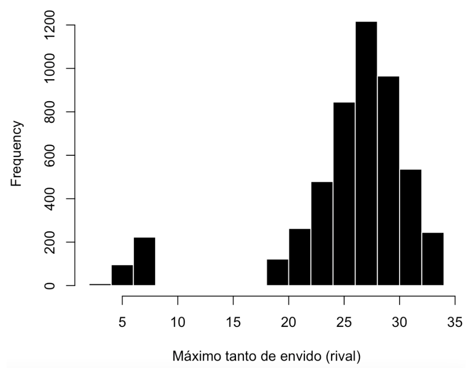

Figure 12: Distribución del máximo puntaje del rival

Los valores altos de envido (30–33) resultan poco frecuentes lo que confirma que manos muy altas no aparecen tan seguido, concentrándose la mayoría de los equipos rivales entre 25 y 29 puntos.

Por eso, si tus compañeros tienen menos de 27, y vos también estás cerca de ese rango, no conviene arriesgar un "quiero" si el rival canta envido con confianza, ya que la probabilidad de que ellos tengan una mano muy fuerte es baja pero decisiva.

En cambio, si tus tantos están por encima de 30, es una oportunidad buena para aceptar o incluso cantar.

Es importante aclarar que es normal y esperado que no haya valores entre 10 y 20 en el histograma del envido. Porque en el Truco, las cartas 10, 11 y 12 valen 0 tantos. Por eso, el valor del envido lo calculamos solo con las cartas 1 a 7 del mismo palo + 20 si hay pareja.