

MODELOS LINEALES
Proyecto Final Programación I

Nombre:

Laury Angélica Vidal Vásquez.

Matrícula:

2023-0681.

Asignación:

Programación I

Profesor:

José Arcadio García García



MODELOS LINEALES

DE ORGANIZACIÓN PARA LA COOPERACIÓN Y EL DESARROLLO ECONÓMICO (OECD) DESDE 1991 AND 2002.

ÍNDICE GENERAL

Presentación	Página 1
Título	Página 2
Índice general	Página 3
Introducción	Página 4
Desarrollo	Página 5-18
Propuesta del Modelo de Regresión	Página 5
Relación entre variable dependiente e independiente	Página 5
Propuesta del Modelo 1	Página 6-8
Propuesta del Modelo 2	Página 9-11
Propuesta del Modelo 2	Página 12-16
Interpretación	Página 17-18
Conclusión	Página 19

INTRODUCCIÓN

La propuesta busca analizar y explicar la relación entre la tasa de donación de órganos y dos variables clave: el gasto en salud pública y las muertes por causas externas. Se utilizará un modelo de regresión lineal para identificar cómo estas variables influyen en la tasa de donación de órganos en 17 países de la OCDE entre 1991 y 2002. La elección de estas variables se basa en la hipótesis de que un mayor gasto en salud pública podría mejorar las infraestructuras y campañas de donación, mientras que las muertes por causas externas, al ser eventos súbitos, podrían proporcionar más oportunidades para la donación de órganos.



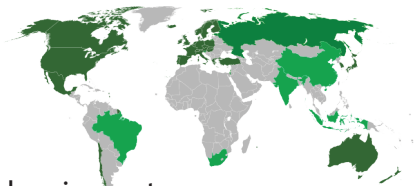
Inicialmente, los datos fueron limpiados mediante la eliminación de filas con valores faltantes y la transformación de las variables numéricas para evitar la notación científica, lo que facilitó su interpretación y análisis. Posteriormente, se propusieron diferentes modelos de regresión lineal en los que la tasa de donación de órganos fue utilizada como variable dependiente, mientras que diversas variables, como el gasto en salud pública y las muertes por causas externas, se consideraron como predictores potenciales.

Para optimizar los modelos propuestos, se implementaron métodos de selección de variables, como la técnica de stepwise mediante el criterio de información de Akaike (AIC) y el uso de la función dredge, lo que permitió identificar las combinaciones de variables que mejor explican la variabilidad en la tasa de donación de órganos.

Finalmente, se llevó a cabo un análisis exhaustivo de los residuales para asegurar el cumplimiento de los supuestos fundamentales del modelo de regresión, como la normalidad y la homocedasticidad, utilizando pruebas como Jarque-Bera y Breusch-Pagan. Además, se verificó la significancia estadística de los coeficientes del modelo utilizando estimaciones robustas de la varianza, garantizando la fiabilidad y validez de las inferencias realizadas a partir del modelo final.

DESARROLLO

Los datos para utilizar se encuentran en una base de datos organdata que contiene información de 17 países de la OCDE sobre trasplantes de órganos



realizados entre 1991 y 2002. donde solo se usará las variables cuantitativas para el modelo. Las variables que se van a tomar en cuenta son las siguientes:

- **donors:** Tasa de donación de órganos por millón de habitantes.
- **pubhealth:** Gasto público en salud como porcentaje del gasto total.
- **external:** Muertes por causas externas por cada 100.000 habitantes (puede incluir accidentes, homicidios, suicidios, etc.).

Propuesta del Modelo de Regresión

Antes de comenzar se explicará el razonamiento detrás de la selección de estas variables. El gasto en salud pública es una medida de los recursos invertidos en el sistema de salud de un país, y se podría hipotetizar que una mayor inversión en salud podría influir en la infraestructura y campañas para la donación de órganos. Por otro lado, las muertes por causas externas (accidentes, suicidios, homicidios) representan situaciones donde la donación de órganos puede ser más común, dada la naturaleza repentina de estas muertes y la posibilidad de que los órganos estén en condiciones adecuadas para el trasplante.

Relación entre variable dependiente e independiente

Estas variables independientes son fundamentales para comprender la influencia que tienen en la tasa de donación de órganos. El gasto en salud pública puede reflejar los esfuerzos en infraestructura y campañas que promueven la donación de órganos, mientras que las causas de muerte externa a menudo implican situaciones donde la donación de órganos es posible, como accidentes de tráfico o muertes prematuras.

PROPUESTA DEL MODELO 1

Relación de las Variables:

Variable Dependiente (Y): Tasa de donación de órganos (donors)

Variable Independiente (X): Muertes por causas externas (external)

```
> summary(modelo1)

Call:
lm(formula = donors ~ pubhealth, data = organdata_clean)

Residuals:
    Min       1Q   Median       3Q      Max
-10.0135  -2.2545  -0.1955   2.2335  10.2446

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  25.3110     2.2510  11.244 < 0.0000000000000002 ***
pubhealth    -1.5299     0.3505  -4.365  0.0000225 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.094 on 163 degrees of freedom
Multiple R-squared:  0.1046,    Adjusted R-squared:  0.09915
F-statistic: 19.05 on 1 and 163 DF,  p-value: 0.00002253
```

En el modelo 1 podemos observar como el R^2 tiene un valor 0.1046, lo que nos indica que el 10.46% de la variabilidad en la tasa de donación de órganos puede ser explicada por las muertes por salud pública.

De igual forma, en el mismo modelo podemos observar como el R^2 **Ajustado** tiene un valor de 0.09915, se ajusta el R^2 considerando el número de predictores en el modelo, siendo este un valor ligeramente menor, pero podemos observar como es cercano al R^2 por lo que podemos decir que el modelo es altamente descriptivo con la variable utilizada

En el **F-estadístico** tenemos un valor de 19.05 y un p-value un muy bajo 0.00002253, lo que sugiere que el modelo es globalmente significativo, es decir, hay al menos una relación significativa entre las variables independientes y la variable dependiente.

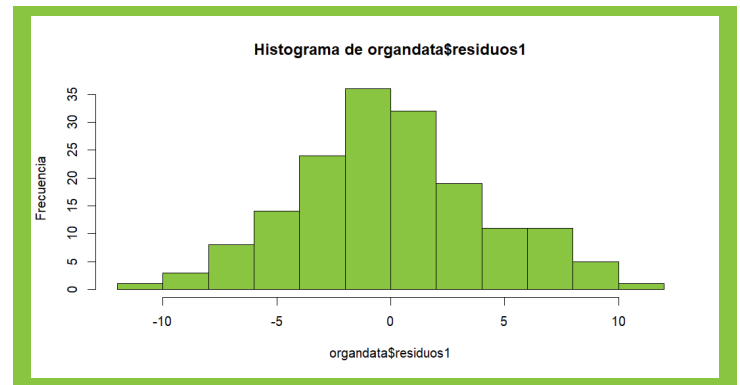
Ecuación Lineal del Modelo: $\text{donors} = 25.311 - 1.5299 \times \text{pubhealth}$

Siendo donors la variable dependiente (o Y) y pubhealth la variable independiente (o X).

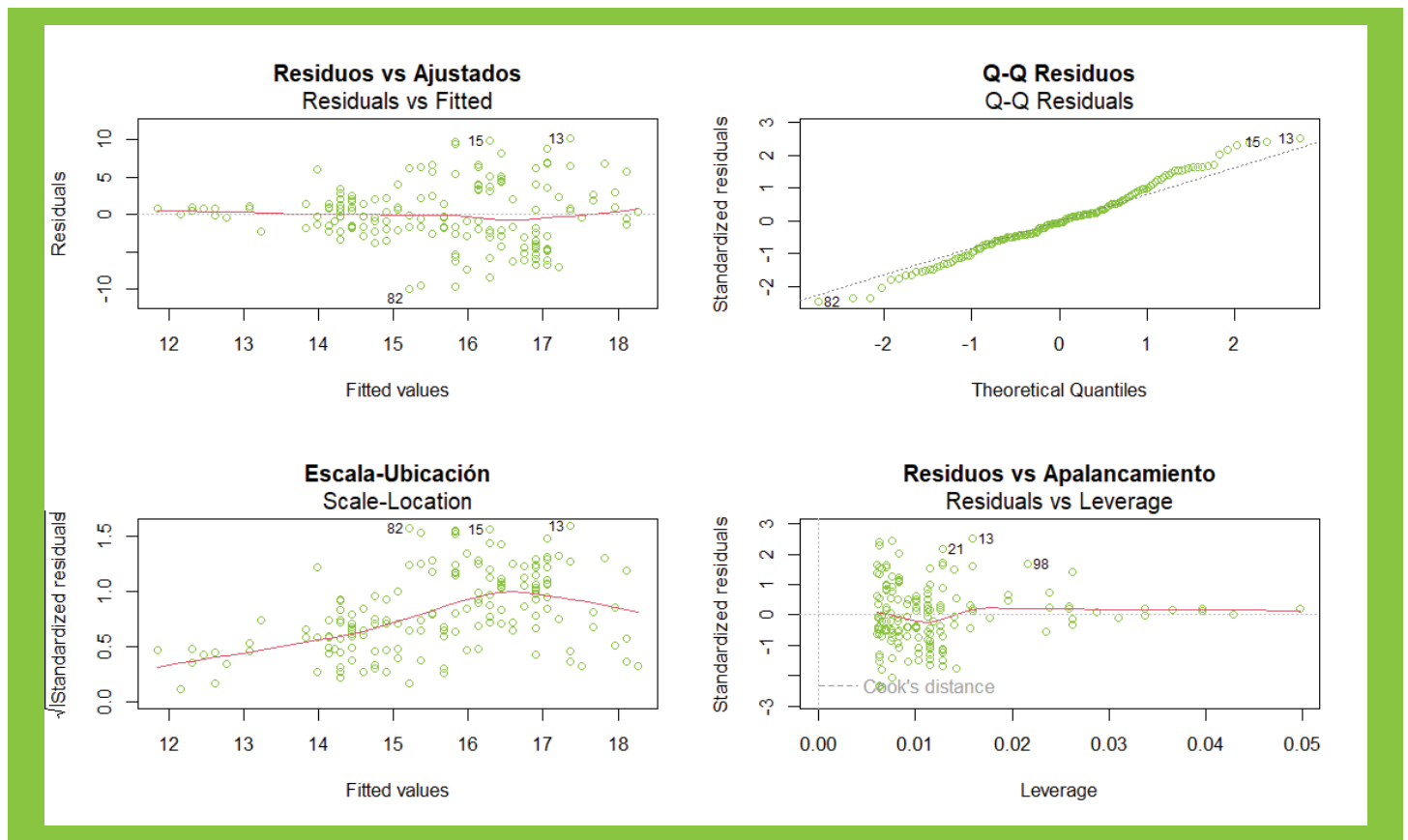
Llegamos a esta conclusión calculando la formula y obtenemos un intercepto de 25.311, y con una pendiente 0.5299 lo que nos dice el modelo es que, por cada unidad adicional de gasto en salud pública (variable X), la tasa de donación de órganos disminuye en aproximadamente 1.53 unidades (variable Y).

Análisis de los Residuales

Los residuales representan la diferencia entre los valores observados y los valores predichos por el modelo. Se analiza para verificar la normalidad y la homocedasticidad (varianza constante).



Interpretación de las Cuatro Gráficas



Residuals vs Fitted:

Este gráfico muestra los residuales en el eje vertical y los valores ajustados en el eje horizontal. Un patrón aleatorio alrededor de la línea horizontal cero sugiere que el modelo es adecuado.

En este caso, los puntos parecen dispersarse aleatoriamente, pero también se observa cierta dispersión en los extremos, lo que podría indicar heterocedasticidad.

Normal Q-Q:

Compara los residuales observados con una distribución normal. Si los puntos siguen la línea diagonal, los residuales son aproximadamente normales.

Los puntos en este gráfico están bastante alineados con la línea diagonal, sugiriendo que los residuales son aproximadamente normales, lo cual es consistente con el resultado de la prueba de Jarque-Bera.

Scale-Location (o Spread-Location):

Este gráfico muestra la raíz cuadrada de los residuales estandarizados frente a los valores ajustados. Debería verse una línea horizontal con dispersión constante si no hay heterocedasticidad.

Se observa un aumento en la dispersión a medida que aumentan los valores ajustados, lo que sugiere la presencia de heterocedasticidad, corroborando los resultados de la prueba de Breusch-Pagan.

Residuals vs Leverage:

Este gráfico ayuda a identificar puntos de alta influencia en el modelo (outliers). Puntos fuera de las líneas de Cook's distance son considerados altamente influyentes.

Algunos puntos parecen estar cerca o fuera de las líneas de Cook's distance, lo que indica que hay puntos que podrían tener una influencia considerable en el ajuste del modelo.

Prueba de Jarque-Bera (Normalidad de Residuales):

En esta prueba de normalidad tenemos que el X-squared tiene un valor de 0.75724 y el p-value tiene un valor de 0.6848, con estos resultados podemos discernir de la hipótesis nula de normalidad de los residuales ya que siguen una distribución normal.

```
> jarque.bera.test(na.omit(organdata_clean$resid))  
  
Jarque Bera Test  
  
data:  na.omit(organdata_clean$resid)  
X-squared = 0.75724, df = 2, p-value = 0.6848
```

Prueba de Breusch-Pagan (Heterocedasticidad):

En esta prueba el BP obtiene un valor de 18.69 y el p-value obtiene 0.00001538, dado estos valores podemos optar por rechazar la hipótesis nula de homocedasticidad, lo que nos indica que hay heterocedasticidad en los residuales.

```
> bptest(modelo1)  
  
studentized Breusch-Pagan test  
  
data:  modelo1  
BP = 18.69, df = 1, p-value = 0.00001538
```


PROPUESTA DEL MODELO 2

Relación de las Variables:

Variable Dependiente (Y): Tasa de donación de órganos (donors)

Variable Independiente (X): Muertes por causas externas (external)

```
> summary(modelo2)

Call:
lm(formula = donors ~ external, data = organdata_clean)

Residuals:
    Min       1Q   Median       3Q      Max
-10.5682  -2.6092  -0.3458   2.2914   9.7909

Coefficients:
              Estimate Std. Error t value    Pr(>|t|)
(Intercept)  7.076848   1.361962   5.196 0.00000060153 ***
external     0.019980   0.003119   6.405 0.00000000154 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.867 on 163 degrees of freedom
Multiple R-squared:  0.2011,    Adjusted R-squared:  0.1962
F-statistic: 41.03 on 1 and 163 DF,  p-value: 0.000000001538
```

En el modelo 2 podemos observar que el R^2 tiene un valor de 0.2011, lo que nos indica que el 20.11% de la variabilidad en la tasa de donación de órganos puede ser explicada por las muertes por causas externas.

De igual forma, en el mismo modelo podemos observar que el R^2 **Ajustado** tiene un valor de 0.1962, lo que ajusta el R^2 considerando el número de predictores en el modelo. Este valor es similar al R^2 , lo que sugiere que el modelo es adecuado y describe bien la relación entre las variables.

En el **F-estadístico** obtenemos un valor de 41.03, con un p-value extremadamente bajo de 0.000000001538, lo que indica que el modelo es globalmente significativo, es decir, existe al menos una relación significativa entre las variables independientes y la variable dependiente.

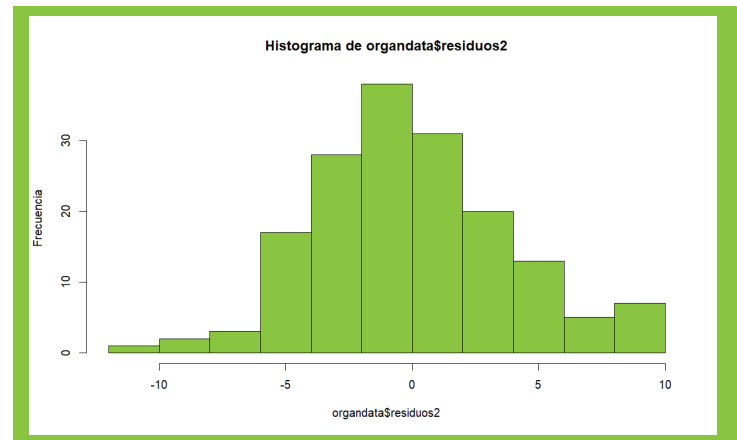
Ecuación Lineal del Modelo: Los donors = $7.07685 + 0.019980 \times \text{external}$.

Siendo "donors" la variable dependiente (o Y) y "external" la variable independiente (o X).

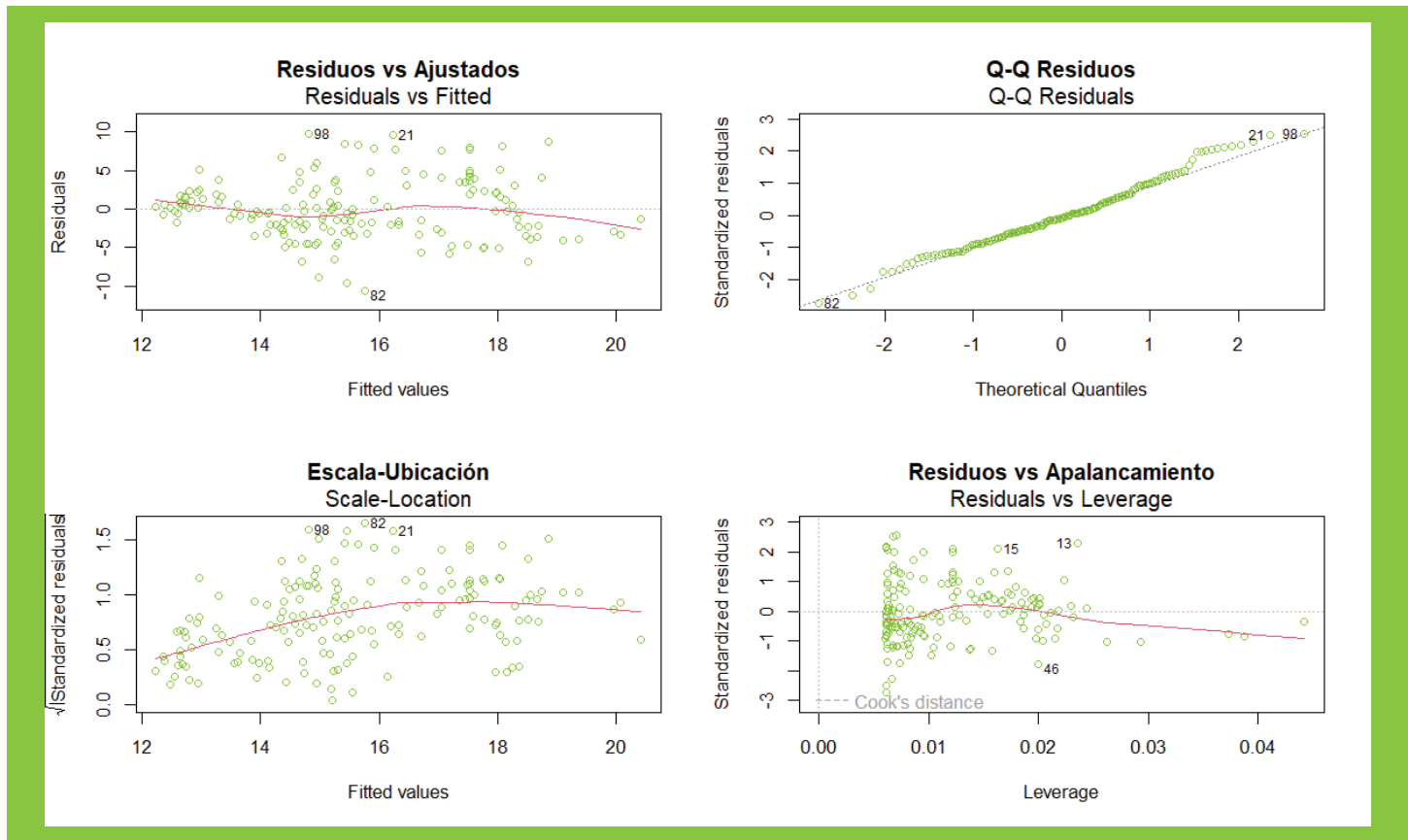
Llegamos a esta conclusión calculando la fórmula y obtenemos un intercepto de 7.07685, con una pendiente de 0.019980. Esto nos indica que, por cada unidad adicional en las muertes por causas externas (variable X), la tasa de donación de órganos aumenta en promedio 0.019980 unidades (variable Y).

Análisis de los Residuales

Los residuales representan la diferencia entre los valores observados y los valores predichos por el modelo. Se analiza para verificar la normalidad y la homocedasticidad (varianza constante).



Interpretación de las Cuatro Gráficas



Residuals vs Fitted:

Esta gráfica muestra los residuales (diferencias entre los valores observados y los valores ajustados) frente a los valores ajustados.

Si los puntos están distribuidos de manera aleatoria alrededor de la línea horizontal (residuales=0), sugiere que el modelo es adecuado. En esta gráfica, se observa cierta aleatoriedad, pero también hay algunos patrones que podrían indicar problemas de no linealidad.

Normal Q-Q:

Esta gráfica compara los residuales estandarizados con una distribución normal teórica. Los puntos deberían seguir una línea recta si los residuales están normalmente distribuidos.

La mayoría de los puntos siguen la línea diagonal, lo que sugiere que los residuales están aproximadamente normalmente distribuidos. Sin embargo, hay algunos puntos en los extremos que se desvían de la línea, lo que puede indicar leves problemas de normalidad.

Scale-Location (o Spread-Location):

Esta gráfica muestra la raíz cuadrada estandarizada de los residuales frente a los valores ajustados. Sirve para verificar la homocedasticidad (constancia de la varianza de los errores).

Si los puntos están distribuidos de manera uniforme a lo largo del rango de valores ajustados, sugiere homocedasticidad. En esta gráfica, se observa cierta variabilidad, lo que indica posibles problemas de heterocedasticidad.

Residuals vs Leverage:

Esta gráfica muestra los residuales estandarizados frente al apalancamiento (leverage) de cada punto. Identifica puntos que tienen una gran influencia en la estimación del modelo.

Los puntos deben estar dentro de las bandas de confianza, y pocos puntos deben tener alto leverage. En esta gráfica, la mayoría de los puntos están dentro de las bandas, aunque hay algunos puntos con alto leverage que podrían ser influyentes.

Prueba de Jarque-Bera (Normalidad de Residuales):

En esta prueba de normalidad, el X-squared tiene un valor de 0.75724 y el p-value es de 0.6848. Con estos resultados, no se rechaza la hipótesis nula de normalidad de los residuales, lo que indica que los residuales siguen una distribución normal.

```
> jarque.bera.test(na.omit(organdata_clean$resid))  
  
Jarque Bera Test  
  
data:  na.omit(organdata_clean$resid)  
X-squared = 2.3271, df = 2, p-value = 0.3124
```

Prueba de Breusch-Pagan (Heterocedasticidad):

En esta prueba, el BP obtiene un valor de 18.69 y el p-value es de 0.00001538. Dado estos valores, se rechaza la hipótesis nula de homocedasticidad, lo que indica que hay heterocedasticidad en los residuales.

```
> bptest(modelo2)  
  
studentized Breusch-Pagan test  
  
data:  modelo2  
BP = 7.6091, df = 1, p-value = 0.005808
```

PROPUESTA DEL MODELO 3

Las variables consideradas son las siguientes: la tasa de donación de órganos por millón de habitantes (donors), el gasto en salud en miles de dólares por cápita (Health), el gasto público en salud como porcentaje del gasto total (pubhealth), las muertes por accidentes por cada 100.000 habitantes (roads), las muertes por causa cerebrovascular por cada 100.000 habitantes, redondeado (cerebvas), las muertes por agresión por cada 100.000 habitantes, redondeado (assault), y las muertes por causas externas por cada 100.000 habitantes (external). Relación de las Variables:

Variable Dependiente (Y): Tasa de donación de órganos (donors)

Variables Independientes (X):

- **health:** Codiciones generales de salud de la población.
- **pubhealth:** Gasto en salud pública.
- **roads:** Infraestructura vial.
- **cerebvas:** Tasa de enfermedades cerebrovasculares.
- **assault:** Tasa de agresiones o incidentes violentos.
- **external:** Incidentes externos que podrían llevar a la necesidad de donación.

Relación entre variable dependiente e independiente

Cada una de estas variables, de manera directa o indirecta, tiene un impacto en el sistema de donación de órganos, ya sea aumentando la demanda por órganos o afectando la disponibilidad de estos. La tasa de donación de órganos porque influyen en las condiciones que afectan tanto la oferta como la demanda de donaciones: Condiciones de salud general, Gasto en salud pública, Infraestructura vial, Tasa de enfermedades cerebrovasculares, Tasa de agresiones o incidentes violentos.

```
> summary(modelo1)

Call:
lm(formula = donors ~ pubhealth, data = organdata_clean)

Residuals:
    Min       1Q   Median       3Q      Max
-10.0135  -2.2545  -0.1955   2.2335  10.2446

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   25.3110     2.2510  11.244 < 0.0000000000000002 ***
pubhealth     -1.5299     0.3505  -4.365  0.0000225 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.094 on 163 degrees of freedom
Multiple R-squared:  0.1046,    Adjusted R-squared:  0.09915
F-statistic: 19.05 on 1 and 163 DF,  p-value: 0.00002253
```

En el modelo, observamos que el **R²** tiene un valor de 0.4808, lo que nos indica que el 48.08% de la variabilidad en la tasa de donación de órganos puede ser explicada por las variables independientes incluidas en el modelo.

De igual forma, en el mismo modelo podemos observar que el **R² Ajustado** tiene un valor de 0.4611, ajustado por el número de variables en el modelo. Esto refleja la proporción de variabilidad explicada por las variables, penalizando por la inclusión de variables adicionales. El valor es ligeramente menor que el R^2 , lo que indica que el modelo sigue siendo robusto, aunque se ajusta por el número de predictores.

En el **F-estadístico** obtenemos un valor de 24.39, con un p-valor extremadamente bajo (< 0.00000000000000022), lo que sugiere que el modelo global es significativamente mejor que un modelo sin predictores, es decir, existe al menos una relación significativa entre las variables independientes y la variable dependiente.

Ecuación Lineal del Modelo: $\text{donors} = 10.0694 + 0.003361 \times \text{health} - 2.364109 \times \text{pubhealth} + 0.028868 \times \text{roads} + 0.007065 \times \text{cerebvas} - 0.087280 \times \text{assault} + 0.017684 \times \text{external}$

En esta ecuación encontramos un intercepto de 10.0694, este valor indica la tasa de donación de órganos promedio cuando todas las variables independientes (health, pubhealth, roads, cerebvas, assault, external) son iguales a cero. Aunque este escenario puede no ser realista, el intercepto proporciona un punto de referencia para la ecuación.

Explicación por cada Variable y su pendiente:

health (0.003361): Por cada aumento de una unidad en la condición de salud general de la población, la tasa de donación de órganos aumenta en promedio 0.003361 unidades, asumiendo que todas las demás variables se mantienen constantes.

pubhealth (-2.364109): Un aumento de una unidad en el gasto en salud pública se asocia con una

disminución de 2.364109 unidades en la tasa de donación de órganos, manteniendo constantes las demás variables.

roads (0.028868): Por cada unidad de mejora en la infraestructura vial, la tasa de donación de órganos aumenta en promedio 0.028868 unidades.

cerebvas (0.007065): Un aumento de una unidad en la tasa de enfermedades cerebrovasculares se asocia con un incremento de 0.007065 unidades en la tasa de donación de órganos.

assault (-0.087280): Un incremento de una unidad en la tasa de agresiones se asocia con una disminución de 0.087280 unidades en la tasa de donación de órganos.

external (0.017684): Por cada unidad de aumento en los incidentes externos, la tasa de donación de órganos aumenta en promedio 0.017684 unidades.

```
> # Calcular el AIC para el modelo
> AIC (modelo3)
[1] 864.726
```

```
> # Realizar la selección de modelos usando dredge
> cand.md <- dredge(modelo3)
Fixed term is "(Intercept)"
```

```
> summary(cand.md)
(Intercept)      assault      cerebvas      external      gdp      gdp_lag
Min.   :-19.726   Min.   :-0.1041   Min.   :-0.0031   Min.   :-0.0027   Min.   :-0.0004   Min.   :-0.0010
1st Qu.: -1.618   1st Qu.: -0.0473   1st Qu.: 0.0063   1st Qu.: 0.0084   1st Qu.: 0.0003   1st Qu.: -0.0005
Median :  6.324   Median : -0.0080   Median : 0.0073   Median : 0.0189   Median : 0.0005   Median : -0.0002
Mean    :  5.942   Mean    : -0.0113   Mean    : 0.0076   Mean    : 0.0161   Mean    : 0.0005   Mean    : -0.0001
3rd Qu.: 13.185   3rd Qu.: 0.0260   3rd Qu.: 0.0086   3rd Qu.: 0.0234   3rd Qu.: 0.0008   3rd Qu.: 0.0003
Max.    : 32.378   Max.    : 0.0754   Max.    : 0.0155   Max.    : 0.0274   Max.    : 0.0013   Max.    : 0.0007
NA's    :2048     NA's    :2048     NA's    :2048     NA's    :2048     NA's    :2048

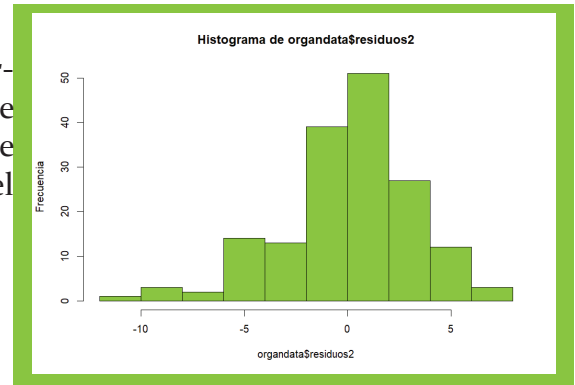
health      health_lag      pop      pop_dens      pubhealth      roads
Min.   :-0.0046   Min.   :-0.0092   Min.   :0      Min.   :-0.0179   Min.   :-2.9858   Min.   :0.0171
1st Qu.: 0.0009   1st Qu.: -0.0037   1st Qu.:0      1st Qu.: 0.0056   1st Qu.: -2.2046   1st Qu.: 0.0547
Median : 0.0028   Median : -0.0016   Median :0      Median : 0.0199   Median : -1.8583   Median : 0.0753
Mean    : 0.0027   Mean    : -0.0016   Mean :0      Mean    : 0.0219   Mean    : -1.8523   Mean    : 0.0677
3rd Qu.: 0.0047   3rd Qu.: 0.0010   3rd Qu.:0      3rd Qu.: 0.0368   3rd Qu.: -1.4957   3rd Qu.: 0.0796
Max.    : 0.0106   Max.    : 0.0069   Max. :0      Max.    : 0.0685   Max.    : -0.7386   Max.    : 0.0888
NA's    :2048     NA's    :2048     NA's    :2048     NA's    :2048     NA's    :2048

txp_pop      df      logLik      AICc      delta      weight
Min.   :-3.2963   Min.   : 2      Min.   :-474.8   Min.   :858.4   Min.   : 0.00   Min.   :0.00000000
1st Qu.: -0.7870   1st Qu.: 7      1st Qu.: -454.5   1st Qu.:873.8   1st Qu.:15.41   1st Qu.:0.00000000
Median : 0.9371   Median : 8      Median : -435.1   Median :886.5   Median :28.14   Median :0.00000003
Mean    : 1.1675   Mean    : 8      Mean    : -440.7   Mean    :898.4   Mean    :39.96   Mean    :0.00024414
3rd Qu.: 2.6884   3rd Qu.: 9      3rd Qu.: -427.7   3rd Qu.:924.7   3rd Qu.:66.27   3rd Qu.:0.00001776
Max.    : 7.0299   Max.   :14      Max.    : -418.4   Max.    :957.6   Max.    :99.19   Max.    :0.03939307
NA's    :2048
```

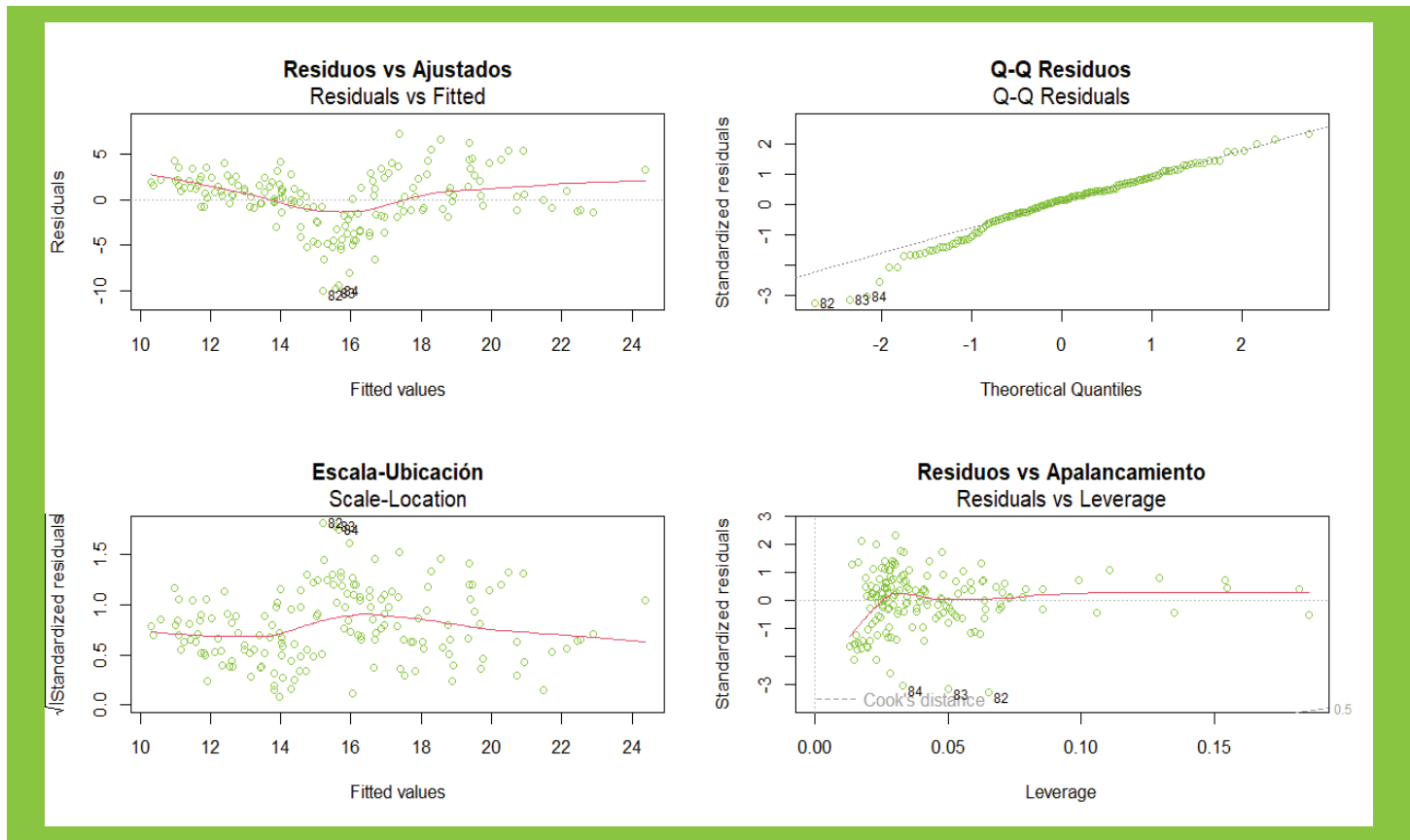
```
> print(cand.md)
Global model call: lm(formula = donors ~ pop + pop_dens + gdp + gdp_lag + health +
health_lag + pubhealth + roads + cerebvas + assault + external +
txp_pop, data = organdata_clean, na.action = na.fail)
---
Model selection table
      (Int)      ass      crb      ext      gdp      gdp_lag      hlt      hlt_lag      pop      pop_dens      pbh      rds      txp_pop      df      logLik      AICc      delta      weight
1576 10.07000 -0.0872800 0.007065000 0.017680000 0.003635000 0.003610000 0.003370000 0.003657000 0.003454000 -0.00009052000 0.02419000 -2.3640 0.02887 8 -420.736 858.4 0.00 0.039
1832 9.86600 -0.0841300 0.006746000 0.019360000 0.003370000 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.3640 0.02887 9 -420.193 859.5 1.15 0.022
1704 8.68400 -0.0645500 0.007500000 0.015300000 0.003370000 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.2410 0.03593 9 -420.319 859.8 1.40 0.020
1608 10.57000 -0.0849800 0.006903000 0.017810000 0.003370000 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.4000 0.02795 8 -421.550 860.0 1.63 0.017
1584 7.44600 -0.0720200 0.007097000 0.016260000 0.000119700 0.002408000 0.002408000 0.002408000 0.003454000 -0.00009052000 0.02419000 -2.1350 0.03499 9 -420.450 860.1 1.67 0.017
1640 9.92000 -0.0869700 0.006986000 0.017530000 0.003370000 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.3250 0.02958 9 -420.601 860.4 1.97 0.015
3624 9.25800 -0.0898500 0.007516000 0.017840000 0.003370000 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.3490 0.02801 0.492200 9 -420.693 860.5 2.15 0.013
1592 9.55900 -0.0840600 0.007065000 0.017430000 0.000026680 0.003356000 0.003356000 0.003356000 0.003454000 -0.00009052000 0.02419000 -2.3220 0.03000 9 -420.723 860.6 2.21 0.013
1960 8.23400 -0.0573800 0.007209000 0.016820000 0.003370000 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.2600 0.03288 10 -419.636 860.7 2.30 0.012
1552 2.97800 -0.0312100 0.006159000 0.012910000 0.000364600 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -1.5690 0.04831 8 -421.920 860.8 2.37 0.012
1600 7.45900 -0.0766400 0.007206000 0.015920000 0.000528400 -0.000469100 0.002759000 0.002759000 0.003454000 -0.00009052000 0.02419000 -2.0900 0.03599 9 -421.699 860.8 2.43 0.012
808 10.41000 -0.0786700 0.007240000 0.026310000 0.003416000 0.003416000 0.003416000 0.003416000 0.003454000 -0.00009052000 0.02419000 -2.6300 0.03718 0.03586000 9 -421.958 860.8 2.44 0.012
1616 6.63200 -0.0635200 0.006967000 0.01580000 0.000172300 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.2510 0.03599 9 -420.904 861.0 2.57 0.011
552 10.87000 -0.0824100 0.007867000 0.025270000 0.003370000 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.6300 0.03718 7 -423.212 861.1 2.74 0.010
1840 7.05400 -0.0677100 0.006768000 0.017900000 0.000127900 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.1550 0.03157 0.02509000 9 -421.864 861.2 2.76 0.010
1864 10.38000 -0.0818400 0.006583000 0.019440000 0.003370000 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.4320 0.02435 0.02357000 9 -421.039 861.2 2.84 0.010
1632 6.99700 -0.0716200 0.007176000 0.015380000 0.000618100 -0.000531100 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.0500 0.03720 0.02474000 9 -421.997 861.4 3.03 0.009
1896 9.69900 -0.0837300 0.006653000 0.019230000 0.003370000 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.3560 0.02584 0.03504000 9 -420.032 861.5 3.10 0.008
1959 6.31000 0.007444000 0.013190000 0.013190000 0.003370000 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -1.9450 0.04091 0.03504000 9 -421.194 861.5 3.15 0.008
1703 6.39000 0.007870000 0.010560000 0.010560000 0.003370000 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -1.8670 0.04633 8 -422.341 861.6 3.21 0.008
1808 2.73400 -0.0281700 0.005838000 0.014840000 0.000363500 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -1.6140 0.04402 0.02738000 9 -421.232 861.6 3.23 0.008
1736 9.41900 -0.0654200 0.007269000 0.015780000 0.003370000 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.2980 0.03392 0.02784000 9 -421.245 861.7 3.26 0.008
1556 7.02600 -0.0722600 0.006585000 0.01720000 0.000573200 -0.000510200 0.002689000 0.002689000 0.003454000 -0.00009052000 0.02419000 -2.1080 0.03227 0.02472000 11 -418.977 861.7 3.28 0.008
3880 8.86400 -0.0872200 0.007348000 0.019590000 0.003419000 0.003419000 0.003419000 0.003419000 0.003454000 -0.00009052000 0.02419000 -2.3810 0.02402 0.064900 10 -420.127 861.7 3.29 0.008
1768 8.52300 -0.0640900 0.007422000 0.015130000 0.005691000 0.005691000 0.005691000 0.005691000 0.003454000 -0.00009052000 0.02419000 -2.2000 0.03670 0.02421000 10 -420.177 861.8 3.39 0.007
1848 9.33800 -0.0807900 0.006746000 0.019300000 0.003370000 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.3550 0.02633 0.02421000 10 -420.179 861.8 3.39 0.007
1712 7.26400 -0.0591600 0.007435000 0.014850000 0.000077160 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.1170 0.03850 0.02421000 10 -420.215 861.9 3.46 0.007
3632 4.61900 -0.0726600 0.008269000 0.016110000 0.000162000 0.002228000 0.002228000 0.002228000 0.003454000 -0.00009052000 0.02419000 -2.0190 0.03515 1.152000 10 -420.246 861.9 3.52 0.007
3752 8.07300 -0.0669700 0.007881000 0.015470000 0.003370000 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.2310 0.03513 0.386000 10 -420.292 862.0 3.62 0.006
1728 9.21500 -0.0667300 0.007536000 0.015430000 0.000033830 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.2840 0.03509 0.02421000 10 -420.300 862.0 3.63 0.006
1624 9.00800 -0.0755600 0.006897000 0.017010000 0.000078610 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.2660 0.03149 0.02421000 10 -421.436 862.0 3.64 0.006
1648 7.27000 -0.0715800 0.007017000 0.016090000 0.000120800 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.0920 0.03577 0.02421000 10 -420.309 862.0 3.65 0.006
1872 6.21200 -0.0590000 0.006629000 0.017310000 0.000181600 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.0660 0.03384 0.02516000 10 -420.319 862.1 3.67 0.006
3600 -0.71980 -0.0366600 0.008000000 0.013060000 0.000400200 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -1.4590 0.04707 1.709000 9 -421.461 862.1 3.69 0.006
1679 1.62800 0.006565000 0.010700000 0.000373000 0.003370000 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -1.3880 0.05295 0.02421000 10 -422.624 862.2 3.76 0.006
3656 9.75900 -0.0876000 0.007065000 0.017360000 0.000667500 -0.000319900 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.2570 0.03137 0.03480000 9 -421.507 862.2 3.78 0.006
840 10.88900 -0.0768800 0.007095000 0.026170000 0.003370000 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.6610 0.03200 0.02421000 10 -422.674 862.3 3.88 0.006
1588 6.55900 -0.0672000 0.006816000 0.017370000 0.000661000 -0.000569900 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.0670 0.03349 0.02791000 11 -419.274 862.3 3.88 0.006
1868 2.54300 -0.0303000 0.006140000 0.012340000 0.000667500 -0.000319900 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -1.4820 0.05031 0.02421000 10 -422.779 862.3 3.89 0.006
584 11.31000 -0.0807000 0.007723000 0.025160000 0.003370000 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.6630 0.03200 0.02421000 10 -423.944 862.4 4.01 0.005
1551 2.51100 0.007153000 0.012670000 0.000340700 0.003370000 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -1.4600 0.04216 0.02421000 11 -419.784 862.4 4.02 0.005
1728 7.24200 -0.0615500 0.007618000 0.014210000 0.000505800 -0.000501700 0.003474000 0.003474000 0.003474000 -0.00009052000 0.02419000 -2.5710 0.03757 0.03587000 10 -421.655 862.5 4.08 0.005
2856 8.24900 -0.0857100 0.008462000 0.026120000 0.003370000 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.5710 0.03757 0.03405000 9 -422.779 862.5 4.09 0.005
1807 2.26400 0.006634000 0.015100000 0.000342200 0.003370000 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.2570 0.03137 0.03405000 10 -421.507 862.5 4.17 0.005
568 13.26000 -0.0995100 0.007709000 0.025070000 0.000133600 -0.000105000 0.004477000 0.004477000 0.004477000 -0.00009052000 0.02419000 -2.7880 0.02963 0.03372000 9 -421.702 862.6 4.17 0.005
824 12.32000 -0.0923300 0.007154000 0.026090000 0.000372200 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.7540 0.04799 0.03372000 10 -421.702 862.6 4.17 0.005
1935 1.46100 0.006128000 0.013070000 0.000372200 0.003370000 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -1.4570 0.04799 0.03372000 10 -421.702 862.6 4.17 0.005
3688 9.24500 -0.0891500 0.007406000 0.017670000 0.003370000 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.3140 0.02882 0.03131000 10 -420.571 862.6 4.17 0.005
1656 9.12700 -0.0820200 0.006980000 0.017120000 0.000040910 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.2570 0.03137 0.03131000 10 -420.571 862.6 4.17 0.005
1711 3.68700 0.007648000 0.010430000 0.000179900 0.003370000 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -1.6510 0.05029 0.02421000 10 -421.715 862.6 4.19 0.005
2024 8.04800 -0.0567200 0.00717000 0.016660000 0.003370000 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.2150 0.03365 0.02783000 11 -419.464 862.7 4.26 0.005
3664 3.40500 -0.0652100 0.008361000 0.015580000 0.000216100 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -1.9240 0.03710 1.354000 10 -420.623 862.7 4.27 0.005
1992 9.00400 -0.0585500 0.006977000 0.017240000 0.003370000 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.3160 0.03096 0.02615000 10 -420.623 862.7 4.28 0.005
3640 7.99200 -0.0847100 0.007746000 0.017410000 0.000050450 0.003370000 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -2.2640 0.02983 0.675200 10 -420.652 862.7 4.34 0.005
3648 4.73700 -0.0772000 0.008333000 0.015790000 0.000565000 -0.000464500 0.003370000 0.003370000 0.003454000 -0.00009052000 0.02419000 -1.9790 0.03613 1.109000 11 -419.509 862.7 4.35 0.004
```

Análisis de los Residuales

Los residuales representan la diferencia entre los valores observados y los valores predichos por el modelo. En este caso, se examinan para evaluar la normalidad y la homocedasticidad de los errores, lo cual es fundamental para validar los supuestos del modelo de regresión lineal.



Interpretación de las Cuatro Gráficas



Residuals vs Fitted:

Esta gráfica muestra los residuales frente a los valores ajustados.

Los puntos no están completamente distribuidos de manera aleatoria alrededor de la línea horizontal, sugiriendo posible no linealidad y algunos patrones que podrían indicar problemas en el modelo.

Normal Q-Q:

Esta gráfica compara los residuales estandarizados con una distribución normal teórica.

La mayoría de los puntos siguen la línea diagonal, sugiriendo que los residuales están aproximadamente normalmente distribuidos, aunque hay algunos puntos que se desvían, indicando problemas de normalidad en los extremos.

Scale-Location:

Esta gráfica muestra la raíz cuadrada estandarizada de los residuales frente a los valores ajustados.

Hay cierta variabilidad en la dispersión de los puntos, lo que sugiere posibles problemas de heterocedasticidad, es decir, la varianza de los errores no es constante.

Residuals vs Leverage:

Muestra los residuales estandarizados frente al apalancamiento (leverage) de cada punto. Algunos puntos tienen alto leverage y podrían influir significativamente en el modelo, lo cual es importante de considerar para identificar posibles outliers que pueden estar afectando la robustez del modelo.

```
> jarque.bera.test(na.omit(organdata_clean$resid))  
  
Jarque Bera Test  
  
data:  na.omit(organdata_clean$resid)  
X-squared = 0.75724, df = 2, p-value = 0.6848
```

Prueba de Breusch-Pagan (Heterocedasticidad):

En esta prueba BP obtiene un valor de 22.381 y un p-value de 0.001033. Estos resultados indican que se rechaza la hipótesis nula de homocedasticidad. Esto sugiere que hay heterocedasticidad en los residuales del modelo, es decir, la varianza de los errores no es constante a lo largo de las observaciones.

```
> bptest(modelo1)  
  
studentized Breusch-Pagan test  
  
data:  modelo1  
BP = 18.69, df = 1, p-value = 0.00001538
```


Intepretacion

En el modelo 1, el coeficiente de pubhealth (-1.5299):

Este coeficiente indica que hay una relación negativa entre pubhealth (gasto en salud pública) y donors (tasa de donación de órganos). Es decir, por cada unidad adicional de gasto en salud pública, la tasa de donación de órganos disminuye en 1.5299 unidades, asumiendo que las otras variables del modelo permanecen constantes.

```
> coeftest(modelo1, vcov = vcovHC(modelo1, type="HC1"))
ancia de los coeficientes

t test of coefficients:

              Estimate Std. Error t value      Pr(>|t|)
(Intercept)  25.31101    1.95847  12.9239 < 0.00000000000000022 ***
pubhealth    -1.52993    0.27525  -5.5584    0.000000109 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

En el modelo 2, el coeficiente de external (0.019980):

Este coeficiente muestra una relación positiva entre external (muertes por causas externas) y donors (tasa de donación de órganos). Es decir, por cada unidad adicional en muertes por causas externas, la tasa de donación de órganos aumenta en 0.019980 unidades, asumiendo que las otras variables del modelo permanecen constantes.

```
> coeftest(modelo2, vcov = vcovHC(modelo2, type="HC1"))
coeficientes

t test of coefficients:

              Estimate Std. Error t value      Pr(>|t|)
(Intercept)  7.076848    1.051510  6.7302 0.0000000000274412 ***
external     0.019980    0.002642  7.5624 0.000000000002729 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

En el modelo 3:

health (0.003361): Este coeficiente indica que hay una relación positiva entre health (gasto en salud) y donors (tasa de donación de órganos). Por cada unidad adicional de gasto en salud, la tasa de donación de órganos aumenta en 0.003361 unidades, asumiendo que las otras variables del modelo permanecen constantes.

pubhealth (-2.364109): Este coeficiente muestra una relación negativa entre pubhealth (gasto en salud pública) y donors (tasa de donación de órganos). Por cada unidad adicional en gasto en salud pública, la tasa de donación de órganos disminuye en 2.364109 unidades, asumiendo que las otras variables del modelo permanecen constantes.

roads (0.028868):

Este coeficiente indica que por cada unidad adicional en la variable roads (infraestructura de caminos), la tasa de donación de órganos aumenta en 0.028868 unidades, asumiendo que las demás variables del modelo se mantienen constantes.

cerebvas (0.007065):

Este coeficiente sugiere que por cada unidad adicional en la variable cerebvas (muertes por accidentes cerebrovasculares), la tasa de donación de órganos aumenta en 0.007065 unidades, manteniendo las otras variables constantes.

assault (-0.087280):

Este coeficiente indica una relación negativa entre assault (tasa de agresiones) y donors (tasa de donación de órganos). Por cada unidad adicional en la tasa de agresiones, la tasa de donación de órganos disminuye en 0.087280 unidades, asumiendo que las demás variables permanecen constantes.

External (0.017684):

Este coeficiente muestra que por cada unidad adicional en la variable external (muertes por causas externas), la tasa de donación de órganos aumenta en 0.017684 unidades, asumiendo que las demás variables del modelo permanecen constantes.

```
> coeftest(modelo3Final, vcov = vcovHC(modelo3Final, type="HC1"))
S
```

t test of coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	10.06940117	2.82083367	3.5697	0.0004738	***
health	0.00336083	0.00045974	7.3103	0.000000000012499	***
pubhealth	-2.36410866	0.31654221	-7.4685	0.000000000005164	***
roads	0.02886827	0.01333242	2.1653	0.0318675	*
cerebvas	0.00706542	0.00236240	2.9908	0.0032289	**
assault	-0.08727965	0.01746024	-4.9988	0.000001517452827	***
external	0.01768448	0.00452395	3.9091	0.0001372	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

CONCLUSIÓN

Se propusieron y estimaron tres modelos de regresión para explicar la tasa de donación de órganos en función de diversas variables socioeconómicas y de salud. El primer modelo, que incluyó solo la variable pubhealth, mostró una relación negativa significativa con la tasa de donación ($p < 0.001$), aunque la capacidad explicativa fue limitada (R^2 ajustado = 0.099). El segundo modelo, que incluyó la variable external, también reveló una relación significativa pero positiva ($p < 0.001$), mejorando el ajuste del modelo (R^2 ajustado = 0.196).

El tercer modelo, que incorporó múltiples variables predictoras, resultó ser el más robusto, con un R^2 ajustado de 0.461. En este modelo final se seleccionaron las variables health, pubhealth, roads, cerebvas, assault, y external como significativas, con un impacto notable de pubhealth y external en la tasa de donación. Las pruebas de normalidad de los residuales y de heterocedasticidad indicaron algunas violaciones de los supuestos clásicos, lo que sugiere que aunque el modelo es explicativo, podría beneficiarse de ajustes adicionales o la consideración de técnicas robustas.

Modelo 1 (pubhealth): Disminuye la tasa de donación en 1.53 unidades por cada unidad adicional de gasto en salud pública.

Modelo 2 (external): Aumenta la tasa de donación en 0.02 unidades por cada unidad adicional en muertes por causas externas.

Modelo 3: health: Aumenta 0.003 unidades por mayor gasto en salud. pubhealth: Disminuye 2.36 unidades por mayor gasto en salud pública. roads: Aumenta 0.03 unidades por mejora en infraestructura de caminos. cerebvas: Aumenta 0.007 unidades por más muertes cerebrovasculares. assault: Disminuye 0.09 unidades por mayor tasa de agresiones. external: Aumenta 0.02 unidades por más muertes por causas externas.