

# Análisis Cuantitativo de Sentencias

Jose Pina-Sánchez

25 January, 2026

(*Esta es la parte práctica del mini-curso ‘Introducción al Estudio Cuantitativo de Sentencing.’ Las diapositivas y ficheros de datos asociados con este ejercicio están disponibles aquí: <https://jmpinasanchez.github.io/sentencias.html>*)

## Introducción

En este taller vamos a practicar el análisis de datos de sentencias utilizando R. Antes de comenzar, tenemos que asegurarnos de haber descargado e instalado R y Rstudio, en ese orden.

La pregunta de investigación que vamos a explorar es: ¿Existe discriminación por género en la ‘Crown Court’ de Inglaterra y Gales?<sup>1</sup>

De manera progresiva vamos a ver como: i) preparar los datos para el análisis; ii) analizar los datos de manera exploratoria a través de tablas, gráficos y estadísticos descriptivos (uni y bivariados); iii) estimar e interpretar modelos de regresión; iv) considerar los pros y contras en el uso de variables de control; v) implementar tests de robustez.

Vamos a utilizar una muestra de 642 penas de robo con violencia (‘robbery’) impuestas en la ‘Crown Court’ de Inglaterra y Gales entre 2016 y 2018. Estos datos son recogidos por parte del ‘Sentencing Council for England and Wales’ a través de una encuesta a jueces. Para llevar a cabo tenemos que descargar el fichero de datos ‘Main dataset’ de la página web del Sentencing Council, aquí.

En esa misma página también podemos encontrar el cuestionario (‘Data collection form’) que se utilizó para la encuesta, y el archivo de metadata (‘Metadata’), donde se indica el contenido de cada variable. Las ‘sentencing guidelines’ utilizadas para decidir la sentencia para casos de robo de violencia se pueden consultar [aquí]: (<https://sentencingcouncil.org.uk/guidelines/robbery-street-and-less-sophisticated-commercial/?source=7511>).

## Preparamos el marco de trabajo

A lo largo de la práctica, vamos a utilizar cuatro comandos que no forman parte de la base de comandos disponibles en R: **tabyl** para generar tablas uni y bivariadas, **vif** para estimar el ‘variance inflation factor’, **sjPlot** para visualizar modelos, y **sensemakr** para llevar a cabo análisis de sensibilidad. Para poder utilizar estos comandos primero tenemos que descargar y activar las librerías donde se encuentran.

Copiad y pegad las siguientes líneas de código en la consola de R (abajo a la izquierda):

---

<sup>1</sup>Esta pregunta se ha investigado en el pasado pero sólo para el caso de agresiones, delitos de drogas, y robos con allanamiento de morada.

```
install.packages(c("janitor", "car", "sjPlot", "sensemakr"))
```

```
#Indicamos las librerias que vamos a utilizar.  
library(janitor) #Para utilizar tabyl()  
library(car) #Para utilizar vif()  
library(sjPlot) #Para utilizar sjPlot()  
library(sensemakr) #Para utilizar sensemakr()
```

Procedemos a importar el fichero de datos que vamos a analizar ('main-dataset-robbery-data.csv'). Para importar los datos utilizamos el comando `read.csv`, pero primero tenemos que indicar donde hemos guardado el fichero, para lo cual utilizamos el comando `setwd`.

```
#Importamos los datos.  
setwd("C:/Users/JPS/OneDrive - University of Leeds/Talks/Salamanca/Ejercicios/Robbery")  
#Acordaos de utilizar la dirección donde hayais guardado el fichero.  
data = read.csv("main-dataset-robbery-data.csv", header = TRUE)
```

En la ventana arriba a la derecha podemos ver el nuevo objeto que hemos creado `data`. Se trata de una base de datos con 642 observaciones y 73 variables. Al clickear sobre el nombre de este objeto podemos ver su contenido en la ventana de arriba a la izquierda. El número de variables mostrado al visualizar la base de datos de esta manera está sin embargo limitado a 50. Si queremos ver el nombre de todas las variables disponibles en la base de datos, podemos utilizar el comando `names`.

```
names(data)
```

```
## [1] "unique_id"                      "age_band"  
## [3] "gender"                          "principal_offence"  
## [5] "conspiracy"                     "location"  
## [7] "culp_useweapon"                  "culp_prodblade"  
## [9] "culp_prodfirearm"                "culp_prodoother"  
## [11] "culp_threaten"                  "culp_sigforce"  
## [13] "culp_minforce"                 "culp_leadrole"  
## [15] "culp_sigrole"                   "culp_sophisticated"  
## [17] "culp_littleplanning"            "harm_physical_serious"  
## [19] "harm_physical_minimal"          "harm_physical_neither"  
## [21] "harm_psych_serious"             "harm_psych_minimal"  
## [23] "harm_psych_neither"              "harm_detrim_serious"  
## [25] "harm_detrim_minimal"             "harm_detrim_neither"  
## [27] "harm_value_high"                "harm_value_low"  
## [29] "harm_value_neither"              "harm_soiling"  
## [31] "harm_limited"                   "culp_category"  
## [33] "harm_category"                  "starting_point_years"  
## [35] "agg_precons"                    "agg_precons_amount"  
## [37] "agg_bail"                        "agg_victimvulnerability"  
## [39] "agg_preventing"                 "agg_prolonged"  
## [41] "agg_restraint"                  "agg_location"  
## [43] "agg_timing"                     "agg_concealidentity"  
## [45] "agg_underinfluence"              "agg_concealevidence"  
## [47] "agg_communityimpact"             "agg_failuretocomply"  
## [49] "agg_licence"                     "agg_highvalue"  
## [51] "agg_significantplanning"        "agg_leading"  
## [53] "agg_child"                       "agg_victimcompelled"
```

```

## [55] "agg_otherfactors"           "mit_convictions"
## [57] "mit_remorse"                "mit_character"
## [59] "mit_disorder"                "mit_age"
## [61] "mit_steps"                   "mit_littleplanning"
## [63] "mit_otherfactors"            "GP_entered"
## [65] "GP_ind_police"              "GP_firstopportunity"
## [67] "GP_when"                     "GP_reduction"
## [69] "GP_percentagereduction"     "outcome"
## [71] "outcome_custody_years"       "outcome_suspension_years"
## [73] "totality_adjustment"

```

Tenemos variables describiendo al condenado (*gender*, *age\_band*), el tipo de robo (*principal\_offence*, *consiparity*, *location*), las características del caso que el juez tuvo en cuenta organizadas en factores de culpabilidad (*culp\_*), daño (*harm\_*), otros agravantes (*agg\_*) y mitigantes (*mit\_*), consideraciones por conformidad (*GP\_*), y la sentencia impuesta (*outcome\_*).<sup>2</sup>

## Análisis exploratorio

Las dos variables principales para analizar nuestra pregunta de investigación son el género del condenado y la sentencia que recibió. Vamos a echar un primer vistazo para valorar cómo están codificadas. Dado que la gran mayoría de casos de robo con violencia reciben pena de prisión, y por cuestiones de conveniencia técnica descritas en la ponencia previa, nos enfocamos en la duración de las penas de prisión (*outcome\_custody\_years*) como nuestra medida de severidad y variable dependiente.

```

#Tablas de frecuencia.
tabyl(data$gender)

##      data$gender   n percent
##      Female    55  0.0857
##      Male     585  0.9112
##  Unknown/Missing    2  0.0031

tabyl(data$outcome_custody_years)

```

Respecto al género del condenado, la mayor parte (91.1%) son hombres. Esto refleja la realidad criminológica del tipo de delitos con un componente de violencia. A nivel práctico también plantea un problema de precisión, ya que la muestra solo contiene a 55 mujeres. Una de las implicaciones es que no vamos a poder utilizar muchas variables de control en nuestros modelos antes de que nuestras estimaciones se hagan demasiado inciertas. Existe un ‘trade-off’ entre exactitud (ausencia de sesgo) y precisión (ausencia de ruido).

También vemos que existen 2 casos clasificados como *Unknown/Missing*. Para simplificar el análisis procedemos a reclasificar estos casos como verdaderamente ausentes, *NA*. Si no hacemos esto R los trataría como una categoría de género más. Para esta recodificación utilizamos el comando *ifelse*.

```

#Indicando datos ausentes.
data$gender = ifelse(data$gender=="Unknown/Missing", NA, data$gender)
tabyl(data$gender)

```

---

<sup>2</sup>El nivel de detalle que ofrecen estas bases de datos del ‘Sentencing Council’ es altísimo y además están disponibles en abierto, lo que nos ha permitido avanzar considerablemente nuestro conocimiento de la práctica de ‘sentencing’ en la jurisdicción de Inglaterra y Gales. Algo similar en España sería muy útil.

```

##  data$gender    n percent valid_percent
##    Female    55   0.0857        0.086
##    Male     585   0.9112        0.914
##    <NA>      2   0.0031        NA

```

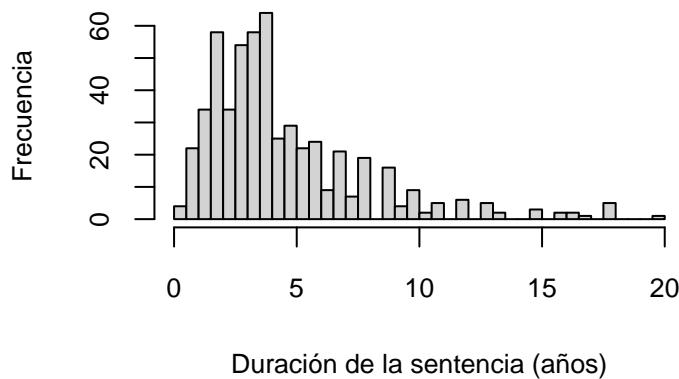
Al repetir la tabla de frecuencia vemos cómo la recodificación se llevó a cabo tal como queríamos. Este tipo de comprobaciones a la hora de transformar variables son importantísimas para evitar errores en nuestros resultados.

Si habéis ejecutado la tabla de frecuencias para la variable *outcome\_custody\_years*, os habréis dado cuenta que esta no es muy buena estrategia para describir su contenido. Esto es porque se trata de una variable continua, la cual se describe mejor a través de gráficas y estadísticos descriptivos. Para ello vamos a utilizar los comandos **hist** y **summary**.

```

#Histograma.
hist(data$outcome_custody_years, breaks = 30,
     main = "", xlab = "Duración de la sentencia (años)", ylab = "Frecuencia",
     cex.lab = 0.85, cex.axis = 0.8)

```



```

#Estadísticos descriptivos.
summary(data$outcome_custody_years)

```

```

##    Min. 1st Qu. Median    Mean 3rd Qu.    Max.    NA's
##    0.33    2.50    3.75    4.63    6.00   20.00    95

```

La pena media son 4.6 años, el mínimo 0.3 y el máximo 20. También observamos 95 datos ausentes. La mayoría de estos se refieren a sentencias donde no se dictaminó una pena de prisión, pero no todos, 13 son genuinamente datos ausentes; como podemos comprobar al ejecutar una tabla de frecuencias para el tipo de pena (*outcome*).

```
tabyl(data$outcome)
```

```

##          data$outcome    n percent

```

```

##           Community order    8  0.0125
##           Determinate 501  0.7804
##           Extended   52  0.0810
##           Life      7  0.0109
##           Other     4  0.0062
## Suspended sentence order  57  0.0888
## Unknown/Missing       13  0.0202

```

Procedemos a recodificar la duración de las sentencias en una nueva variable (*length*) que refleje solo 13 valores ausentes, clasificando el resto como penas diferentes de prisión, las cuales codificamos como 0s.

Pregunta: ¿Por qué es importante codificar las penas diferentes de prisión como 0s en lugar de dejarlos como datos ausentes?

Esto lo hacemos en tres pasos: i) creamos una nueva variable (*length*) que copie *outcome\_custody\_years*, ii) indicamos los valores que R debe entender como ausentes (i.e. donde no se recogió el resultado de la sentencia), y iii) recodificamos el resto de valores ausentes en *outcome\_custody\_years* como 0s.

*#Creando una nueva variable para la duración de las sentencias.*

```

data$length = data$outcome_custody_years
data$length[data$outcome == "Unknown/Missing"] = NA
data$length[is.na(data$length) == TRUE] = 0
summary(data$length)

```

```

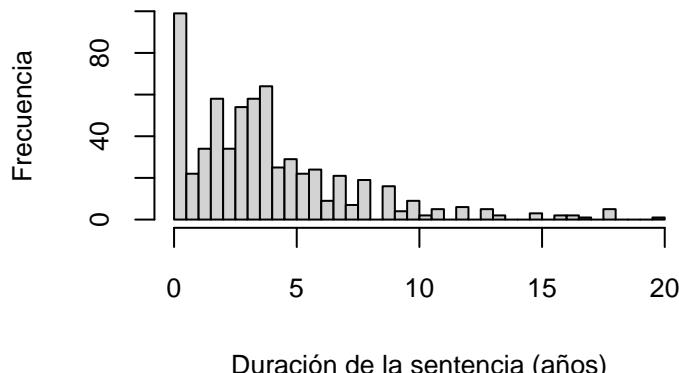
##      Min. 1st Qu. Median      Mean 3rd Qu.      Max.
##      0.00    1.67    3.33    3.94    5.33   20.00

```

```

hist(data$length, breaks = 30,
      main = "", xlab = "Duración de la sentencia (años)", ylab = "Frecuencia",
      cex.lab = 0.85, cex.axis = 0.8)

```



Hemos recuperado 83 casos (95-12) lo cual aumenta el tamaño de la muestra considerablemente, y con ello la precisión de nuestras estimaciones, pero también evitamos posibles sesgos de selección que surgen al

descartar sentencias que no conllevan prisión.<sup>3</sup> La media de la duración de penas de prisión se ha reducido considerablemente, y la distribución de las sentencias se aleja de una normal. Esta es una cuestión importante al que volveremos al final del ejercicio.

Llegado a este punto estamos preparados para comenzar el análisis de la asociación entre el género del condenado y la severidad de la sentencia. Dado que la primera variable es binaria y la segunda continua recomiendo llevar a cabo una comparación de medias. Esto lo podemos hacer a la vez que testamos la significatividad estadística de esa diferencia mediante el comando `t.test`.

```
#T-test para observar el diferencial de genero en la duración de sentencias.
t.test(data$length ~ data$gender)
```

```
##
##  Welch Two Sample t-test
##
## data: data$length by data$gender
## t = -3, df = 76, p-value = 0.002
## alternative hypothesis: true difference in means between group Female and group Male is not equal to
## 95 percent confidence interval:
## -1.91 -0.45
## sample estimates:
## mean in group Female   mean in group Male
##                      2.9                      4.0
```

Convictos de robo con violencia que son hombres reciben de media más de un año de prisión que las mujeres. Y a pesar de haber utilizado una muestra relativamente pequeña, la diferencia de medias es estadísticamente significativa ( $p\text{-value} = 0.002$ ).

Pregunta: ¿Podemos tomar este resultado como evidencia de discriminación contra los hombres?

Vamos a explorar si existen características del caso que los jueces deban tener en cuenta relacionadas con el género del condenado. Ya que el resto de variables en la base de datos son discretas, la mejor opción para hacer eso es utilizar tablas de contingencia. Debajo muestro como se pueden conseguir y expresar sus valores en porcentajes combinando los comandos `adorn_percentages` y `tabyl`.

```
#Tabla de contingencia para valorar el diferencial de género en ciertos factores
#legales.
adorn_percentages(tabyl(data, agg_child, gender), "col")
```

```
## agg_child Female  Male NA_
##      0  0.964 0.942   1
##      1  0.036 0.058   0
```

A menudo las variables categóricas binarias se codifican con 1 para indicar que el factor en consideración está presente en el caso y 0 para indicar lo contrario. Este es el caso para `agg_child` y la mayoría de agravantes y mitigantes codificados en nuestra base de datos.

Podemos ver como los hombres tienden a cometer robos con violencia donde el agravante por presencia de un menor es más frecuente, lo cual podría explicar parte del diferencial de género en la severidad de las sentencias.

Copiad y pegad el código en la consola (ventana abajo-izquierda) para explorar diferencias de genero en otras características del caso que el juez debería tener en cuenta. Solo hace falta reemplazar `agg_child` por el nombre de la variable que queráis explorar.

---

<sup>3</sup>Este es un problema muy común en la literatura de ‘sentencing’ (Bushway et al., 2007; Pina-Sánchez & Gosling, 2020).

Por razones de tiempo vamos a dejar el análisis exploratorio aquí, sin embargo es importante remarcar que antes de lanzarse a estimar modelos de regresión es imprescindible familiarizarse primero con las variables que van a ser utilizadas. En nuestro caso habría que explorar todas las variables que consideremos relevantes (i.e., las que vayamos a utilizar en nuestro modelo). Como mínimo deberíamos utilizar tablas de frecuencia para todas las variables categóricas y estadísticos descriptivos para las continuas.

## Modelos de regresión

Para estimar la asociación del género del condenado con la severidad de las sentencias controlando por características del caso, vamos a utilizar modelos de regresión. En particular, vamos a empezar mediante un modelo lineal simple, donde solo incluimos un coeficiente para la variable *gender* y otro para la constante.

Este modelo se suele representar matemáticamente de la siguiente manera:  $Y = \beta_0 + \beta_1 X + \epsilon$ .

- $Y$  es la variable dependiente, en nuestro caso representa la duración de las sentencias,
- $X$  es la variable independiente de interés, el género del condenado, donde las mujeres se representan con 0s y los hombres con 1s,
- $\beta_0$ , la constante, indica el valor de la variable dependiente ( $Y$ ) cuando las variables independientes ( $X$ ) adoptan el valor 0, en nuestro caso la duración media para mujeres,
- $\beta_1$ , la pendiente, indica el salto en la variable dependiente ( $Y$ ) cuando la variable independiente ( $X$ ) sube un valor, en nuestro caso la diferencia en la duración de las sentencias entre hombres y mujeres,
- $\epsilon$ , los residuos, recogen la variabilidad en la variable dependiente ( $Y$ ) que no puede ser explicada mediante la variabilidad en la/s independiente/s ( $X$ ), en nuestro caso la diferencia en la duración de sentencias que es debida a cualquier otro factor diferente del género del condenado.

Para estimar el modelo vamos a utilizar el comando `lm` y `summary` para visualizar los resultados.

```
#Modelo lineal simple.
simple = lm(length ~ gender, data = data)
summary(simple)

##
## Call:
## lm(formula = length ~ gender, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max 
## -4.049 -2.299 -0.719  1.281 15.951 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  2.870     0.467   6.15  1.4e-09 ***
## genderMale   1.179     0.488   2.41   0.016 *  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.5 on 638 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.00906,    Adjusted R-squared:  0.0075 
## F-statistic: 5.83 on 1 and 638 DF,  p-value: 0.016
```

Dada nuestra pregunta de investigación, debemos poner el foco en el valor de  $\beta_1$ , éste es el **Estimate** de `genderMale`, 1.179. Tal y como vimos en el análisis exploratorio, el diferencial entre hombres y mujeres es aproximadamente un año. Igualmente, observamos que este resultado es estadísticamente significativo ( $p\text{-value} = 0.016$ ).

Otro dato interesante para contextualizar la importancia del factor extra-legal de género es el  $R^2$ , que se interpreta como la proporción de la variabilidad en la duración de las sentencias explicada por las variables independientes incluidas en el modelo. En nuestro caso, el género del condenado solo explica el 0.9% de la variabilidad en la duración de las sentencias. Visto así es más difícil argumentar que este factor extra-legal se pueda considerar relevante.

Más importante aún, no hemos descartado que la asociación entre el género y la duración de las sentencias pueda ser espúrea, i.e. debido a un tercer factor de confusión. Para explorar esta hipótesis procedemos a introducir variables independientes - a modo de controles - en nuestro modelo. Empezamos con `agg_child`, un factor agravante que parecía estar más presente en los robos von violencia cometidos por hombres. Para añadir controles solo tenemos que replicar el comando anterior introduciendo el nombre de la variable de control precedido de `+`. En este caso para reproducir el modelo utilizamos `tab_model` en lugar de `summary`, simplemente para demostrar que se existen diferentes opciones - a mi me gusta más `tab_model` aunque ofrece menos información.

```
#Modelos lineal múltiple, ya que introducimos un control.
control1 = lm(length ~ gender + agg_child, data = data)
tab_model(control1)
```

Cuando el delito tuvo lugar en una residencia y un niño o alguna otra persona vulnerable estaba presente la pena aumenta en 6.7 años. El  $R^2$  ahora es más de 20 veces mayor que en el anterior modelo, lo que sirve para poner en contexto la importancia de factores legales que definen el caso. Respecto a nuestra pregunta de investigación, observamos que el coeficiente para género del condenado se ha reducido, de 1.18 a 1.03, indicando que al menos una parte de la asociación que observamos anteriormente era espúrea.

Vuestro turno: Estimad un nuevo modelo, esta vez incluyendo también `harm_category` (un factor utilizado en las ‘guidelines’ para distinguir tres niveles de daño, categoría 1 es la más grave). Observáis algún cambio relevante respecto a nuestra pregunta de investigación?

## Estrategias de modelización

A la hora de decidir que variables se deberían utilizar como controles el consejo más común es: incluir variables que afecten a la variable dependiente. Sin embargo, esta estrategia puede ser contraproducente. Tenemos que considerar también el principio de parsimoneidad y el rol causal de los controles que introducimos (Pina-Sánchez et al., 2025).

Al introducir variables en nuestro modelo reducimos sus grados de libertad. Esto significa que la precisión con que podemos establecer la significatividad estadística de cualquiera de esas variables se reduce. El problema aumenta cuando las muestras son pequeñas, o, de manera más exacta, cuando el número de casos en la muestra no es suficientemente grande en relación con el número de variables introducidas.

En tales situaciones podemos tener un problema de multicolinealidad. Esto viene a decir que existe demasiado solapamiento entre variables independientes para estimar con precisión la correlación parcial entre una de esas variables independientes afectadas por multicolinealidad y la variable dependiente. La mejor manera de conceptualizar este problema es de manera visual. Si siguieramos introduciendo variables independientes ( $X$ ) en el modelo representado en la Figura 1, llegaría un momento en el que el área de asociación entre  $X_1$  e  $Y$  sería minúsculo o no existente, y por tanto imposible de estimar la correlación parcial entre esas dos variables. De hecho, podemos ver como eso ya es el caso para  $X_2$ , que es completamente redundante.

```
include_graphics("regression.png")
```

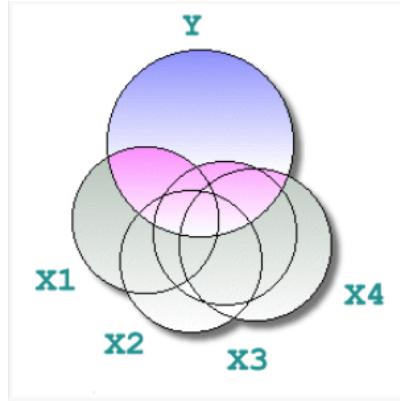


Figure 1: Representación visual del problema de multicolinealidad

Para valorar la presencia de este problema en nuestros modelos podemos utilizar el comando `vif` (variance inflation factor). El VIF indica el factor por el que las medidas de incertidumbre de nuestro modelo (e.g., ‘standard errors’, ‘p-values’) están infladas debido al solapamiento entre variables independientes.

Para investigar la presencia de multicolinealidad podemos estimar un modelo más grande, donde en lugar de incluir las categorías de daño y culpabilidad utilizamos los indicadores que los jueces deben considerar para establecer esas categorías, además procedemos añadir todos los otros factores agravantes y mitigantes.

```
control3 = lm(length ~ gender + culp_useweapon + culp_prodblade + culp_prodfirearm +
  culp_prodother + culp_threaten + culp_sigforce + culp_minforce +
  culp_leadrole + culp_sigrole + culp_sophisticated +
  culp_littleplanning + harm_physical_minimal + harm_physical_neither +
  harm_psych_serious + harm_psych_minimal + harm_psych_neither +
  harm_detrim_serious + harm_detrim_minimal + harm_detrim_neither +
  harm_value_high + harm_value_low + harm_value_neither + harm_soiling +
  harm_limited + agg_precons + agg_bail + agg_victimvulnerability +
  agg_preventing + agg_prolonged + agg_restraint + agg_location +
  agg_timing + agg_concealidentity + agg_underinfluence +
  agg_concealevidence + agg_communityimpact + agg_failureto comply +
  agg_licence + agg_highvalue + agg_significantplanning + agg_leading +
  agg_child + agg_victimcompelled + agg_otherfactors + mit_convictions +
  mit_remorse + mit_character + mit_disorder + mit_age + mit_steps +
  mit_littleplanning + mit_otherfactors,
  data = data)

summary(control3)
#Valorando la presencia de multicolinealidad en nuestro modelo.
vif(control3)
```

El VIF para `gender` es muy bajo, 1.1, lo que indica que la incertidumbre añadida por utilizar un modelo tan saturado de controles es únicamente un 10%. Dado este diagnóstico, y teniendo en cuenta que todas las variables incluidas son relevantes a nivel teórico (i.e. deben ser consideradas por los jueces al decidir la sentencia), aconsejaría no eliminar ninguna de esas variables del modelo. Ahora bien...

Pregunta: ¿Qué haríais si vuestra pregunta de investigación fuera la de establecer si el factor ‘restraint, detention or additional degradation of the victim’ (`agg_restraint`) actúa realmente como un factor de agravación de la pena?

Otro criterio que se tiende a ignorar es el de no introducir variables que se encuentren en la trayectoria causal entre la variable independiente (i.e. la causa, *gender*) y la variable dependiente (i.e. el efecto, *length*). Hasta hace poco, los investigadores en el campo de sentencing asumían - probablemente sin darse cuenta - que todos los factores legales eran exógenos al juez. En realidad, sin embargo, muchas de las características del caso están definidas por el mismo juez a cargo de imponer la sentencia, por lo cual es posible que el hipotético prejuicio judicial bajo investigación esté presente tanto en la sentencia final como en las características del caso<sup>4</sup>. De ser así, al controlar por factores definidos por el juez en estudios que busca estimar el valor de algún tipo de prejuicio (por cuestiones de raza, clase, sexo, etc.), es posible que estemos controlando por parte del efecto que buscamos estimar (Figura 2) y por lo tanto infraestimandolo.

```
include_graphics("Figure 4.png")
```

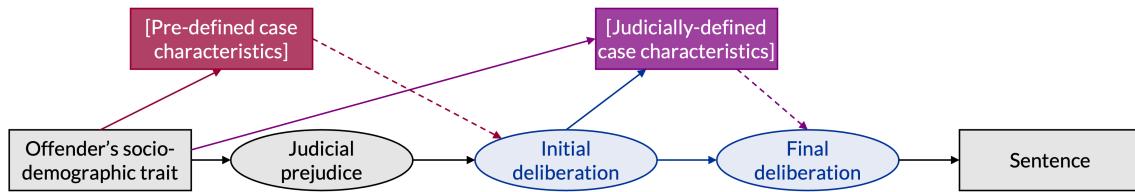


Figure 2: Representación del proceso causal en la imposición de sentencias

Si se da esta situación es imposible estimar exactamente el nivel de discriminación en las sentencias. Lo que si podemos hacer es reportar un intervalo donde es probable que ese hipotético efecto se encuentre. Una manera simple de hacer eso es reportar el coeficiente de interés estimado a través de dos modelos, uno incluyendo todas las características del caso, y otro incluyendo únicamente aquellas que pueden considerarse inamovibles, no afectadas por la discrecionalidad del juez. Para realizar este tipo de distinción es clave poseer conocimiento legal del tipo de delito bajo investigación, aún así, se trata de un proceso subjetivo, afectado por un grado de incertidumbre importante. En Pina-Sánchez et al. (2025) indicamos como reflejar parte de esa incertidumbre utilizando una curva de especificación, pero ese proceso lo dejamos para otra curso más avanzado.

Como modelo donde se encuentran todas las características del caso utilizamos *control3*. Este modelo se podría mejorar algo más ya que no incluye factores legales importantes como si el acusado se declaró culpable, o si se trata de un caso en tentativa, pero nos sirve como ejemplo de modelo “saturado”.

Para el nuevo modelo incluyendo solo factores del caso que se puedan entender como objetivos, o predefinidos, usamos el uso de armas (*culp\_useweapon*, *culp\_prodblade*, *culp\_prodfirearm*, *culp\_prodother*), la presencia de historial delictivo (*agg\_precons*), si el delito tuvo lugar en la residencia de la víctima (*agg\_location*), o tuvo lugar mientras que el condenado estaba bajo libertad provisional o quebrantando alguna orden judicial (*agg\_bail*, *agg\_licence*, *agg\_failuretocomply*), si el condenado se encontraba bajo la influencia de alcohol u otras drogas (*agg\_underinfluence*), o si buscó esconder evidencias del delito (*agg\_concealevidence*), y descartamos factores más discretionales como el uso de fuerza (*culp\_threaten*, *culp\_sigforce*, *culp\_minforce*), el rol adoptado durante la comisión del delito (*culp\_leadrole*, *culp\_sigrole*), grado de predeterminación (*culp\_sophisticated*, *culp\_littleplanning*, *mit\_littleplanning*, *agg\_significantplanning*), el grado de daño físico, mental, a la víctima o su propiedad (*harm\_physical\_minimal*, *harm\_physical\_neither*, *harm\_psych\_serious*, *harm\_psych\_minimal*, *harm\_psych\_neither*, *harm\_detriment\_serious*, *harm\_detriment\_minimal*, *harm\_detriment\_neither*, *harm\_value\_high*, *harm\_value\_low*, *harm\_value\_neither*, *harm\_soiling*, *harm\_limited*) y cualquier factor agravante o mitigante por cuestiones personales (e.g., *mit\_convictions*, *mit\_remorse*, *mit\_character*, *mit\_disorder*, *mit\_age*, *mit\_steps*, *mit\_otherfactors*).

<sup>4</sup>Guilfoyle & Pina-Sánchez (2025) establecen la sospecha de que éste es el caso en la Crown Court con relación a las disparidades étnicas, ya que la mayor parte de características del caso están relativamente uniformemente distribuidas entre grupos étnicos, la única excepción son los factores mitigantes de tipo personal, donde la discrecionalidad judicial es mayor.

```
#Un modelo más parsimonioso donde solo las características del caso que se puedan
#definir de manera objetiva son incluidas.
control4 = lm(length ~ gender + culp_useweapon+ culp_prodblade + culp_prodfirearm +
              culp_prodoother + agg_precons + agg_location + agg_bail +
              agg_licence + agg_failuretocomply + agg_underinfluence +
              agg_concealevidence, data = data)
```

En lugar de reproducir el resultado del modelo entero utilizando `summary` o `tab_model`, y ya que solo queremos reportar el coeficiente para la variable `gender`, podemos simplemente pedir el valor de estos coeficientes a R para los dos modelos que estamos comparando, el modelo con todas las características del caso (`control3`, modelo saturado) y el modelo donde solo incluimos características que se pueden considerar objetivas (`control4`, modelo parsimonioso).

```
#El valor del coeficiente de género en el modelo saturado.
control3$coefficients[2]
```

```
## genderMale
##      0.64
```

```
#El valor del coeficiente de género en el modelo más parsimonioso.
control4$coefficients[2]
```

```
## genderMale
##      0.64
```

Vemos que para ambos modelos el diferencial de género es idéntico (hasta el segundo decimal). Esto son buenas noticias en cuanto a que la incertidumbre que emana de la elección de diferentes modelos es nula, lo cuál facilita el proceso inferencial. En este caso podemos simplemente descartar el modelo saturado y enfocarnos en el modelo más parsimonioso. Si utilizamos `summary` o `tab_model` también vemos que el coeficiente de género no es estadísticamente significativo. Es decir, a pesar de que en nuestra muestra observamos un diferencial de género que apunta a que condenados por robo con violencia masculinos son discriminados - o lo que es lo mismo, que las mujeres son favorecidas - no podemos concluir que este resultado sea real - es decir, se observe en la población de donde se derivo la muestra. Por lo tanto no podemos concluir que la ‘Crown Court’ discrimine contra hombres al sentenciar robos con violencia.

Para estar más seguros de la validez de esta conclusión podemos llevar a cabo tests de robustez / análisis de sensitividad.

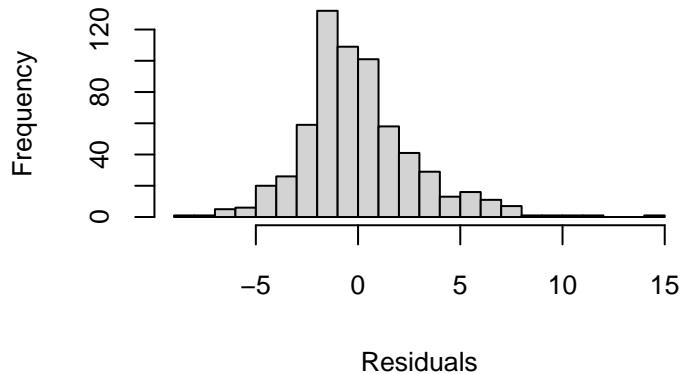
## Tests de robustez

Vamos a considerar dos supuestos que invocamos al estimar modelos de regresión: i) normalidad de los residuos y ii) la ausencia de variables de confusión que no hayan sido controladas en el modelo.

Al comienzo del ejercicio observamos como la distribución de la duración de las sentencias no parecía del todo la de una normal, sobre todo cuando incluimos las sentencias diferentes de prisión como 0s. Esto puede dar lugar a problemas de especificación afectando las medidas de incertidumbre del modelo y en algunos casos (e.g. cuando existen sentencias que ejercen demasiada influencia en el modelo) pueden dar lugar a sesgos en el coeficiente de interés.

Para explorar en qué medida la violación del supuesto de normalidad supone un riesgo, podemos explorar visualmente los residuos (i.e. la diferencia entre las predicciones de nuestro modelo y la duración observada). Esto es importante ya que el supuesto de normalidad se refiere a normalidad en la distribución de los residuos, no de la variable dependiente, algo que solemos olvidar.

```
#Visualizando la distribución de los residuos de nuestro modelo.
res = residuals(control4)
hist(res, breaks = 30, main = "", xlab = "Residuals", cex.lab = 0.85, cex.axis = 0.8)
```



Como podemos ver, la distribución de los residuos se aproxima bastante a una normal. Existen algunos casos extremos en la cola derecha de la distribución que pueden estar ejerciendo demasiada influencia, y el valor mediano está por debajo de 0, indicando la posibilidad de que un modelo capaz de reflejar tal distribución sesgada a la derecha podría hacer un mejor trabajo.

La manera más simple de considerar tal especificación sería mediante la transformación logarítmica de la duración de las sentencias. Dado que el logaritmo de 0 es indefinido, antes de hacer esa transformación sumamos 0.0001 a la duración de todas las sentencias en nuestra muestra.

En último lugar, para determinar cuál es la mejor especificación (cuál de los dos modelos se ajusta a los datos mejor) podemos utilizar estadísticos como el AIC o el BIC. El modelo que tiene un valor más bajo en este tipo de estadísticos se considera la mejor opción.

```
#Esta transformación es para poder utilizar un modelo log-lineal, en lugar de un modelo
#mas complicado (e.g. zero-inflated log-linear).
data$loglength = log(data$length + 0.0001)
#Una aproximación a una especificación log-lineal.
loglineal = lm(loglength ~ gender + culp_useweapon+ culp_prodblade + culp_prodfirearm +
                 culp_prodother + agg_precons + agg_location + agg_bail +
                 agg_licence + agg_failureto comply + agg_underinfluence +
                 agg_concealevidence, data = data)

#Tests
AIC(control4, loglineal)

##          df    AIC
## control4 16 3146
## loglineal 16 3462
```

```
AIC(control4, loglineal)
```

```
##          df   AIC
## control4 16 3146
## loglineal 16 3462
```

Vemos que el modelo más simple (el modelo lineal que veníamos utilizando) es la mejor opción. Así que podemos concluir que el modelo no está mal especificado con respecto al supuesto de la normalidad de los residuos.

Un supuesto mucho más importante, es el de que no hemos dejado ninguna variable de confusión (aquellas que afectan tanto a la variable dependiente como a la independiente) sin controlar.

Hemos visto como al eliminar 41 factores legales - aquellos que consideramos que podrían haber estado definidos por el juez - no se introdujo ningún sesgo en nuestro coeficiente de interés. Sin embargo, no hemos explorado todas los factores legales disponibles en la base de datos. Convendría por lo tanto seguir el proceso de modelización introduciendo esos factores como variables de control. Esto no lo vamos a hacer aquí por razones de espacio. En su lugar, vamos a considerar que podemos hacer en el caso en el que alguna/s de las variables de confusión no estén disponibles en nuestra base de datos.

Para el caso de la determinación de las penas por robos con violencia, podemos anticipar un buen número de variables de confusión que no hemos controlado. Existen factores legales que son muy difíciles de cuantificar, como el grado de peligrosidad o el potencial de rehabilitación del condenado, además de muchos otros factores extra-legales que simplemente no se incluyeron en el cuestionario, como la nacionalidad, clase social, estado laboral, etc.

Cuando nos enfrentamos a esta realidad - siempre presente en el análisis de sentencias utilizando datos observacionales - tendemos simplemente a reflejarla como una limitación al dinal del artículo. En Pina-Sánchez et al. (2025) proponemos explorar esta limitación utilizando el ‘robustness value’ desarrollado por Cinelli & Hazlett (2020). El ‘robustness value’ se puede utilizar para valorar cuan importante deben de ser las variables de confusión que no logramos controlar para que el coeficiente de interés ( $\beta_1$ , el diferencial de género) sea igual a 0. Para calcular este valor utilizamos el comando `sensemakr`.

```
#Análisis de sensibilidad por factores de confusión.
sensemakr(model = control4, treatment = "genderMale")
```

```
## Sensitivity Analysis to Unobserved Confounding
##
## Model Formula: length ~ gender + culp_useweapon + culp_problade + culp_prodfirearm +
##                  culp_proother + agg_precons + agg_location + agg_bail +
##                  agg_licence + agg_failuretocomply + agg_underinfluence +
##                  agg_concealevidence
##
## Null hypothesis: q = 1 and reduce = TRUE
##
## Unadjusted Estimates of 'genderMale':
##   Coef. estimate: 0.64
##   Standard Error: 0.4
##   t-value: 1.6
##
## Sensitivity Statistics:
##   Partial R2 of treatment with outcome: 0.004
##   Robustness Value, q = 1 : 0.062
##   Robustness Value, q = 1 alpha = 0.05 : 0
```

```
##  
## For more information, check summary.
```

Los estadísticos de sensitividad que obtenemos nos indican: i) el género del condenado explica únicamente un 0.4% de la variabilidad en la duración de las sentencias - aproximadamente la mitad de lo que observamos cuando nos basamos en un modelo simple sin controles - y ii) para que el diferencial de género que estimamos sea totalmente espúreo ( $\beta_1 = 0$ ), los factores de confusión que no controlamos deberían explicar al menos el 6.2% de la variabilidad en la duración de las sentencias y en la variable género del condenado.

Al recibir esta información podemos re-evaluar nuestra conclusión. Esto sigue siendo un proceso subjetivo/cualitativo, pero por lo menos ahora podemos llevar a cabo esta valoración mejor informados.

Por ejemplo, deberíamos considerar cómo los hombres están más presentes en la población extranjera que en la nacional, a la vez que los condenados extranjeros parecen ser tratados con mayor severidad (Kemp & Varona, 2023; Riba et al., 2023). Y lo mismo podríamos decir con respecto a cuestiones de grado de peligrosidad, o potencial de rehabilitación.

En conclusión, dado que muchas de las variables que no hemos controlado cumplen ambas condiciones de estar moderadamente correlacionadas tanto con el género del condenado como con la duración de las sentencias, podemos cuestionar nuestros resultados no sólo por la falta de significatividad estadística, sino porque es altamente probable que no reflejen más que una correlación espúrea, causada por terceros factores que no logramos controlar.

## Referencias

- Bushway, S., Johnson, B. D., & Slocum, L. A. (2007). Is the Magic Still There? The Use of the Heckman Two-Step Correction for Selection Bias in Criminology. *Journal of Quantitative Criminology*, 23(2), 151–178. <https://doi.org/10.1007/s10940-007-9024-4>
- Cinelli, C., & Hazlett, C. (2020). Making Sense of Sensitivity: Extending Omitted Variable Bias. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(1), 39–67. <https://doi.org/10.1111/rssb.12348>
- Guilfoyle, E., & Pina-Sánchez, J. (2025). Racially Determined Case Characteristics: Exploring Disparities in the Use of Sentencing Factors in England and Wales. *The British Journal of Criminology*, 65(2), 241–260. <https://doi.org/10.1093/bjc/azae039>
- Kemp, S., & Varona, D. (2023). Foreign and Dangerous? Unpacking the Role of Judges and Prosecutors in Sentencing Disparities in Spain. *The British Journal of Criminology*, 63(4), 984–1002. <https://doi.org/10.1093/bjc/azac068>
- Pina-Sánchez, J., & Gosling, J. P. (2020). Tackling selection bias in sentencing data analysis: A new approach based on a scale of severity. *Quality & Quantity*, 54(3), 1047–1073. <https://doi.org/10.1007/s11135-020-00973-z>
- Pina-Sánchez, J., Hamilton, M., & Tennant, P. W. G. (2025). Estimating Discrimination in Sentencing: Distinguishing between Good and Bad Controls. *European Journal of Empirical Legal Studies*, 2(3). <https://doi.org/10.62355/ejels.53977>
- Riba, J. M. L., Cabello, Ú. R., Gómez, D. V., & Vasilescu, C. (2023). Equality Before the Courts? Studying Citizenship Disparities in Sentencing in Catalonia. *European Journal on Criminal Policy and Research*, 29(4), 601–624. <https://doi.org/10.1007/s10610-022-09530-w>