Regresión múltiple y otras técnicas multivariadas

Tarea 07

Rivera Torres Francisco de Jesús
Rodríguez Maya Jorge Daniel
Samayoa Donado Víctor Augusto
Trujillo Barrios Georgina

Abril 03, 2019

Ejercicio 1

En el análisis del modelo de RLM, calcular ECM($\hat{\sigma}^2_{MCO}$) y ECM($\hat{\sigma}^2_{MV}$). A partir de estos resultados decidir qué estimador es mejor.

Demostración. Recordemos que si una variable aleatoria se distribuye ji-cuadrada con n grados de libertad $X \sim \chi^2_{(n)}$, entonces cumple que $\mathrm{E}(X) = n$ y $\mathrm{V}(X) = 2n$.

Sabemos que $\dfrac{(n-p-1)\hat{\sigma}^2_{MCO}}{\sigma^2} \sim \chi^2_{(n-p-1)}$, por lo tanto

$$(n-p-1) = \mathrm{E}\left[\frac{(n-p-1)\hat{\sigma}^2_{MCO}}{\sigma^2}\right] = \frac{(n-p-1)}{\sigma^2}\mathrm{E}\left(\hat{\sigma}^2_{MCO}\right) \quad \Rightarrow \quad \mathrm{E}\left(\hat{\sigma}^2_{MCO}\right) = \sigma^2$$

$$2(n-p-1) = \mathrm{V}\left[\frac{(n-p-1)\hat{\sigma}^2_{MCO}}{\sigma^2}\right] = \frac{(n-p-1)^2}{\sigma^4}\mathrm{V}\left(\hat{\sigma}^2_{MCO}\right) \quad \Rightarrow \quad \mathrm{V}\left(\hat{\sigma}^2_{MCO}\right) = \frac{2\sigma^4}{(n-p-1)}$$

Con lo anterior, se procede a calcular el error cuadrático medio (ECM) para $\hat{\sigma}^2_{MCO}$.

$$\mathrm{ECM}\left(\hat{\sigma}^2_{MCO}\right) = \mathrm{B}^2\left(\hat{\sigma}^2_{MCO}\right) + \mathrm{V}\left(\hat{\sigma}^2_{MCO}\right) = 0 + \frac{2\sigma^4}{(n-p-1)} = \frac{2}{(n-p-1)}\sigma^4$$

De forma análoga se procede con $\hat{\sigma}^2$. Sabemos que $\dfrac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{(n-p-1)}$, por lo tanto

$$(n-p-1) = \mathrm{E}\left[\frac{n\hat{\sigma}^2}{\sigma^2}\right] = \frac{n}{\sigma^2}\mathrm{E}\left(\hat{\sigma}^2\right) \quad \Rightarrow \quad \mathrm{E}\left(\hat{\sigma}^2\right) = \frac{(n-p-1)}{n}\sigma^2$$

$$2(n-p-1) = \mathrm{V}\left[\frac{n\hat{\sigma}^2}{\sigma^2}\right] = \frac{n^2}{\sigma^4}\mathrm{V}\left(\hat{\sigma}^2\right) \quad \Rightarrow \quad \mathrm{V}\left(\hat{\sigma}^2\right) = \frac{2(n-p-1)}{n^2}\sigma^4$$

Con lo anterior, se procede a calcular el error cuadrático medio (ECM) para $\hat{\sigma}^2$.

$$\begin{aligned}\text{ECM}\left(\hat{\sigma}^2\right) &= \text{B}^2\left(\hat{\sigma}^2\right) + \text{V}\left(\hat{\sigma}^2\right) = \left(\frac{(n-p-1)}{n}\sigma^2 - \sigma^2\right)^2 + \frac{2(n-p-1)}{n^2}\sigma^4 \\ &= \frac{(p+1)^2}{n^2}\sigma^4 + \frac{2(n-p-1)}{n^2}\sigma^4 \\ &= \frac{p^2+2n-1}{n^2}\sigma^4\end{aligned}$$

Tenemos así que $\text{ECM}\left(\hat{\sigma}^2_{MCO}\right) = \dfrac{2}{(n-p-1)}\sigma^4$ y $\text{ECM}\left(\hat{\sigma}^2\right) = \dfrac{p^2+2n-1}{n^2}\sigma^4$. □

Ejercicio 2

En el análisis del modelo RLM, la suma de cuadrados de regresión se define como

$$SCR = \sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2$$

donde $\hat{y}_i$ es la i-ésima componente del vector $\hat{\boldsymbol{y}}$ y $\bar{y} = \dfrac{1}{n}\sum_{i=1}^{n} y_i$. Mostrar que

$$SCR = \boldsymbol{y}^T\left(\boldsymbol{H} - \frac{1}{n}\boldsymbol{J}_n\right)\boldsymbol{y}$$

Demostración. Notemos que:

$$\hat{y} = X\hat{\beta} = Hy, \qquad \sum_{i=1}^{n}\hat{y}_i^2 = \hat{y}^T\hat{y} = (Hy)^T(Hy), \qquad \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2 = y^T\left(\frac{1}{n}J_n\right)y$$

con las igualdades previas se tiene:

$$\begin{aligned}SCR &= \sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2 = \sum_{i=1}^{n}\left(\hat{y}_i^2 - 2\hat{y}_i\bar{y} + \bar{y}^2\right) = \sum_{i=1}^{n}\hat{y}_i^2 - 2\bar{y}\sum_{i=1}^{n}\hat{y}_i + n\bar{y}^2 \\ &= \sum_{i=1}^{n}\hat{y}_i^2 - 2\bar{y}\sum_{i=1}^{n} y_i + n\bar{y}^2, \quad \text{por tarea 2} \\ &= \sum_{i=1}^{n}\hat{y}_i^2 - 2n\bar{y}^2 + n\bar{y}^2 = \sum_{i=1}^{n}\hat{y}_i^2 - n\bar{y}^2 = \sum_{i=1}^{n}\hat{y}_i^2 - n\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)^2 \\ &= \sum_{i=1}^{n}\hat{y}_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2 = (Hy)^T(Hy) - y^T\left(\frac{1}{n}J_n\right)y \\ &= y^T H^T H y - y^T\left(\frac{1}{n}J_n\right)y = y^T H y - y^T\left(\frac{1}{n}J_n\right)y, \quad \text{ya que } H \text{ es simétrica e idempotente} \\ &= y^T\left(H - \frac{1}{n}J_n\right)y\end{aligned}$$

□

Ejercicio 3

En el análisis del modelo RLM, mostrar que $(\boldsymbol{I}_n - \boldsymbol{H})\left(\boldsymbol{H} - \frac{1}{n}\boldsymbol{J}_n\right) = \mathbf{0}_{n\times n}$

Demostración.

$$\begin{aligned}
(I_n - H)\cdot\left(H - \frac{1}{n}J_n\right) &= I\left(H - \frac{1}{n}J_n\right) - H\left(H - \frac{1}{n}J_n\right) \\
&= IH - \frac{1}{n}IJ_n - HH + \frac{1}{n}HJ_n \\
&= H - \frac{1}{n}J_n - H + \frac{1}{n}HJ_n, \qquad \text{ya que } H \text{ es idempotente} \\
&= H - \frac{1}{n}J_n - H + \frac{1}{n}J_n, \qquad \text{ya que } \frac{1}{n}HJ_n = \frac{1}{n}J_n \text{ (tarea 6)} \\
&= \mathbf{0}_{n\times n}
\end{aligned}$$

□

Ejercicio 4

El conjunto de datos `house_selling_prices_OR.csv` contiene información sobre el precio de venta y características de una muestra de 200 observaciones. El objetivo es ajustar un modelo RLM paraexplicar la distribución del precio de venta en miles de dólares (`house_price`) como función del tamaño de la vivienda en metros cuadrados (`house_size`) y del número de habitaciones (`bedrooms`).

```
# Se proceden a cargar los datos
ej4 <- read_csv("datos/house_selling_prices_OR.csv",
                col_types = cols(house_price = col_double(),
                                 house_size = col_double(),
                                 lot_size = col_double(),
                                 bedrooms = col_integer(),
                                 bathrooms = col_double(),
                                 age = col_integer(),
                                 garage = col_integer(),
                                 contidion = col_integer(),
                                 age_cat = col_character()))
```

Con los resultados obtenidos, responder lo siguiente.

Inciso 4.a)

Reportar e interpretar las estimaciones de los coeficientes del modelo.

Considerando el modelo

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$$

donde y_i hace referencia al precio de la casa i-ésima (`house_price`), x_{1i} hace referencia al tamaño de la casa i-ésima (`house_size`), y x_{2i} hace referencia al número de cuartos en la casa i-ésima (`bedrooms`),

```
# Se genera el modelo
modelo_ej4 <- ej4 %>%
              lm(formula = house_price ~ house_size + bedrooms)

coeficientes <- coefficients(modelo_ej4)

# Se imprimen los resultados
summary(modelo_ej4)
```

```
##
## Call:
## lm(formula = house_price ~ house_size + bedrooms, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -306.90  -35.19   -0.77   30.47  376.81
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 60.09330   18.62312   3.227  0.00147 **
## house_size   0.67796    0.05116  13.251  < 2e-16 ***
## bedrooms    15.17173    5.32976   2.847  0.00489 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 80.27 on 197 degrees of freedom
## Multiple R-squared:  0.5244, Adjusted R-squared:  0.5196
## F-statistic: 108.6 on 2 and 197 DF,  p-value: < 2.2e-16
```

De lo anterior se concluye que los coeficientes son:

$$\beta_0 = 60.0933026 \qquad \beta_1 = 0.677956 \qquad \beta_2 = 15.1717253$$

En este contexto, el coeficiente β_0 no tene sentido ya que no existen viviendas con 0m^2 y 0 habitaciones.

El coeficiente β_1 se interpreta como el incremento del precio de la casa (en dolares) por cada incremento de 1m^2 en el tamaño de la casa. Siempre que el número de habitaciones permanezca constante. Por lo que, el incremento de un metro cuadrado en el tamaño de la vivienda implica un incremento de 0.68USD en el precio de la vivienda.

El coeficiente β_2 se interpreta como el incremento del precio de la casa (en dolares) por cada habitación extra en la casa. Siempre que el tamaño de la casa se mantenga constante (aunque esto puede sonar contraintuitivo, podría darse el caso de que se hagan modificaciones en la vivienda para reducir el tamaño de una habitación y generar otra nueva). Por lo que, al aumentar en uno

el número de habitaciones en la vivienda implica un incremento de 15.17USD en el precio de la vivienda.

Inciso 4.b)

Calcular intervalos de confianza simultáneos 95% para los coeficientes del modelo e interpretar los resultados.

```
X <- ej4 %>%
     mutate(intercepto = 1) %>%
     select(intercepto, house_size, bedrooms) %>%
     as.matrix()

# Estimación de los intervalos simultaneos
confint(modelo_ej4, level = (1 - 0.05/6))
```

```
##                0.417 %    99.583 %
## (Intercept) 10.4596319 109.7269733
## house_size   0.5415945   0.8143175
## bedrooms     0.9670314  29.3764192
```

Notamos que ningún intervalo de los coeficientes β_0, β_1 y β_2 contienen al cero. Por lo que conlcuimos que los coeficientes son significativos.

Inciso 4.c)

¿Tiene algún efecto el tamaño de la vivienda en el precio de venta?

```
# Extraemos los p-values de los coeficientes
tbl_coeff <- summary(modelo_ej4)$coefficients %>%
             as.data.frame() %>%
             rownames_to_column() %>%
             as_tibble() %>%
             rename(coeficientes = rowname)

t_values <- tbl_coeff %>%
            pull(`t value`)
```

De acuerdo con el valor de t de los coeficientes, se tiene que $|T_{\beta_1}| = 13.2505424 > 1.972079 = t_{197}^{1-0.05/2}$. Esto quiere decir que existe evidencia para rechazar la hipótesis nula ($H_0 : \beta_1 = 0$) y podemos concluir que hay una relación entre el tamaño de la vivienda y su precio.

Inciso 4.d)

¿Tiene algún efecto el número de habitaciones en el precio de venta de la vivienda?

De acuerdo con el valor de t de los coeficientes, se tiene que $|T_{\beta_2}| = 2.8466032 > 1.972079 = t_{197}^{1-0.05/2}$. Esto quiere decir que existe evidencia para rechazar la hipótesis nula ($H_0 : \beta_2 = 0$) y podemos

concluir que hay una relación entre el número de habitaciones de la vivienda y su precio.

Inciso 4.e)

Reportar la estimación de σ^2 y calcular un intervalo de confianza 95%.

```
# Calculo de sigma2
y <- ej4 %>%
     pull(house_price) %>%
     as.matrix()

H <- X %*% solve(t(X) %*% X) %*% t(X)

I <- diag(x = 1, nrow = nrow(X), ncol = nrow(X))

n <- nrow(X)
p <- 2
gl <- n - p - 1
SC_error <- t(y) %*% (I - H) %*% y
sigma2.hat <- as.double(SC_error / gl)
```

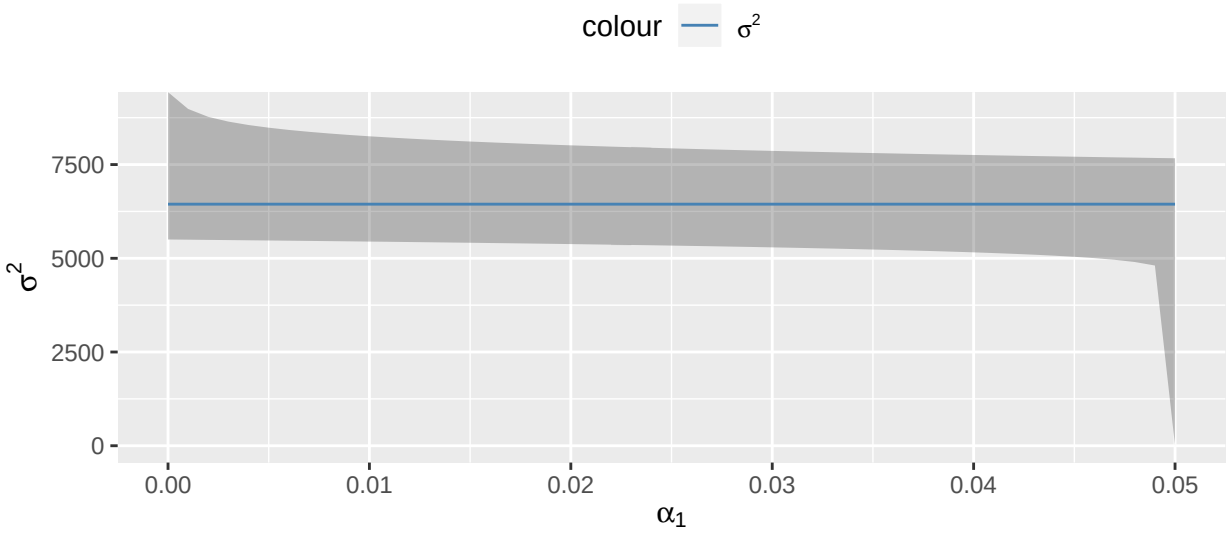
Sabemos que la estimación de sigma está dada por:

$$\hat{\sigma}^2_{MCO} = \frac{1}{n-p-1} SC_{error} = \boldsymbol{y}^T \left(\boldsymbol{H} - \frac{1}{n} \boldsymbol{J}_n \right) \boldsymbol{y} = 6443.34$$

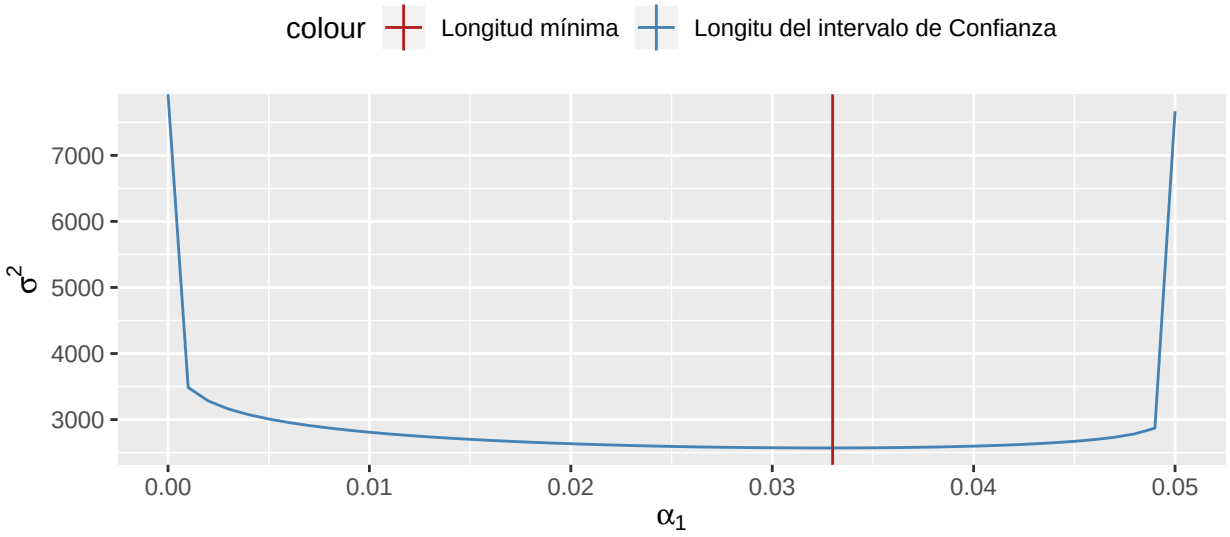
El intervalo de 95% de confianza está dado por:

$$\left(\frac{SC_{error}}{\chi^2_{n-p-1}(1-\alpha_2)}, \frac{SC_{error}}{\chi^2_{n-p-1}(\alpha_1)} \right), \qquad \text{con } \alpha_1 + \alpha_2 = \alpha, \alpha_1, \alpha_2 > 0$$

A continuación se muestran los intervalos de confianza para distrinas combinaciones de α_1 y α_2 tales que $\alpha_1 + \alpha_2 = \alpha = 0.05$.



Calculando la longitud de cada intervalo se obtiene el siguiente comportamiento:



Por lo que, el intervalo de confianza con longitud mínima para $\hat{\sigma}^2_{MCO}$ está dado por $\alpha_1 = 0.033$, y por ende $\alpha_2 = 0.017$. Obteniendo asi el intervalo de confianza:

$$(5258.6613575, 7829.0893554)$$

Inciso 4.f)

Estimar puntualmente y por intervalo la media del precio de venta de las viviendas de 250 metros cuadrados y tres habitaciones.

La estimación puntual está dada por:

```
x0 <- as.matrix(c(1, 250, 3), ncol = 1)
mu0 <- as.double(t(x0) %*% as.matrix(coeficientes))
```

$$\begin{aligned}\mu_0 &= \boldsymbol{x}_0^T \hat{\beta} \\ &= 60.0933026 + 0.677956 * (250) + 15.1717253 * (3) \\ &= 60.0933026 + 169.4890015 + 45.5151759 \\ &= 275.09748\end{aligned}$$

La estimación por intervalo está dada por:

$$x_0^T \hat{\beta} \pm t_{n-p-a}(\alpha/2)\hat{\sigma}_{MCO}\sqrt{1 + x_0^T \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1} x_0}$$

sustituyendo los valores anteriores se obtiene

```
# Calculamos intervalos de confianza
alfa <- 0.05
intervalo <- c(-1, 1) * as.double(qt(1 - alfa/2, gl) * sqrt(sigma2.hat) *
                                    sqrt(1 + t(x0) %*% solve(t(X) %*% X) %*% x0))
```

$$275.09748 \pm 158.7043497$$
$$(116.3931303, 433.8018297)$$

Ejercicio 5

El conjunto de datos `fl_crime.csv` contiene información sobre los 67 del estado de Florida, EUA. Para este ejercicio se debe ajustar un modelo RLM para explicar la distribución de la tasa de crímenes por cada 1000 habitantes (`crime_rate`) como función del porcentaje de adultos con educación superior (`edu`) y del grado de urbanización (`urban`). Con los resultados obtenidos, responder lo siguiente.

```
# Se proceden a cargar los datos
ej5 <- read_csv("datos/fl_crime.csv",
                col_types = cols(county = col_character(),
                                 crime_rate = col_integer(),
                                 edu = col_double(),
                                 urban = col_double(),
                                 income = col_double()))
```

Inciso 5.a)

Reportar e interpretar las estimaciones de los coeficientes del modelo

Considerando el modelo

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$$

donde y_i hace referencia a la tasa de crímenes del estado i-ésimo (`crime_rate`), x_{1i} hace referencia al porcentaje de adultos con educación superior en el estado i-ésimo (`edu`), y x_{2i} hace referencia al grado de urbanización del estado i-ésimo (`urban`),

```
# Se genera el modelo
modelo_ej5 <- ej5 %>%
              lm(formula = crime_rate ~ edu + urban)

coeficientes <- coefficients(modelo_ej5)

# Se imprimen los resultados
summary(modelo_ej5)
```

```
##
## Call:
## lm(formula = crime_rate ~ edu + urban, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -34.693 -15.742  -6.226  15.812  50.678
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  59.1181    28.3653   2.084   0.0411 *
## edu          -0.5834     0.4725  -1.235   0.2214
## urban         0.6825     0.1232   5.539 6.11e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.82 on 64 degrees of freedom
## Multiple R-squared:  0.4714, Adjusted R-squared:  0.4549
## F-statistic: 28.54 on 2 and 64 DF,  p-value: 1.379e-09
```

De lo anterior se concluye que los coeficientes son:

$$\beta_0 = 59.1180677 \qquad \beta_1 = -0.5833773 \qquad \beta_2 = 0.6825014$$

En este contexto, el coeficiente β_0 no tene sentido ya que no existen estados con un porcentaje de adultos con nivel de educación de 0%.

El coeficiente β_1 se interpreta como el decremento de la tasa de crímenes por cada incremento de 1% en el porcentaje de los adultos con un nivel de educación superior. Siempre que el grado de urbanización permanezca constante. Por lo que, el incremento de un punto porcentual en la

cantidad de adultos con un nivel de educación superior implica un decremento de -0.58 en la tasa de crímenes.

El coeficiente β_2 se interpreta como el incremento de la tasa de crímenes por cada incremento de 1 grado de urbanización. Siempre que porcentaje de adultos con nivel de educación superior permanezca constante. Por lo que, al aumentar en uno el grado de urbanización implica un incremento de 0.68 en la tasa de crímenes.

Inciso 5.b)

Calcular intervalos de confianza simultáneos 95% para los coeficientes del modelo e interpretar los resultados

```
# Estimación de los intervalos simultaneos
confint(modelo_ej5, level = (1 - 0.05/6))
```

```
##                 0.417 %    99.583 %
## (Intercept) -18.1133417 136.3494771
## edu          -1.8697614   0.7030067
## urban         0.3470254   1.0179775
```

Notamos que los intervalos de confianza tanto para el intercepto como para el coeficiente β_1, asociado a la variable `edu` contienen el 0. Por lo que concluimos que dichos coeficientes no son significativos.

El único coeficiente cuyo intervalo de confianza no contiene al 0 es β_2, el cual está asociado a `urban`

Inciso 5.c)

¿Tiene algún efecto la educación en la tasa de crímenes de los condados de Florida?

```
# Extraemos los p-values de los coeficientes
tbl_coeff <- summary(modelo_ej5)$coefficients %>%
             as.data.frame() %>%
             rownames_to_column() %>%
             as_tibble() %>%
             rename(coeficientes = rowname)

t_values <- tbl_coeff %>%
            pull(`t value`)
```

De acuerdo con el valor de t de los coeficientes, se tiene que $|T_{\beta_1}| = 1.2347679 < 1.9977297 = t_{64}^{1-0.05/2}$. Esto quiere decir que NO existe evidencia para rechazar la hipótesis nula ($H_0 : \beta_1 = 0$) y podemos concluir que no hay evidencia de que el nivel de educación pueda tener un efecto en la tasa de crímenes.

Inciso 5.d)

¿Tiene algún efecto la urbanización en la tasa de crímenes de los condados de Florida?

De acuerdo con el valor de t de los coeficientes, se tiene que $|T_{\beta_2}| = 5.5392182 > 1.9977297 = t_{64}^{1-0.05/2}$. Esto quiere decir que existe evidencia para rechazar la hipótesis nula ($H_0 : \beta_2 = 0$) y podemos concluir que hay evidencia de que el nivel de urbanización pueda tener un efecto en la tasa de crímenes.

Inciso 5.e)

Calcular la matriz de correlaciones de las tres variables involucradas en el modelo y reportar los resultados. Tratar de explicar los resultados de los incisos b) y c) a partir de estas correlaciones.

```
correlacion <- ej5 %>%
               select(crime_rate, edu, urban) %>%
               cor()

correlacion
```

```
##             crime_rate       edu     urban
## crime_rate  1.0000000 0.4669119 0.6773678
## edu         0.4669119 1.0000000 0.7907190
## urban       0.6773678 0.7907190 1.0000000
```

De la matriz de correlación se observa que las variables `edu` y `urban` tienen un alto nivel de correlación al obener un valor de 0.79, esto contradice el supuesto de no correlación.

También se observa que la variable objetivo `crime rate` tiene una mayor correlación con `urban` que con `edu`, esto puede explicar el por qué el nivel de educación no es significativo. Debido a que al tener las variables explicativas una alta correlación, entonces la variable explicativa que está mayormente correlacionada con la variable objetivo "aboservera" el nivel de significancia y permitirá explicar de mejor forma la variable objetivo, en este caso `crime rate`.

Inciso 5.f)

Reportar la estimación de σ^2 y calcular un intervalo de confianza 95%.

```
# Calculo de sigma2
X <- ej5 %>%
     mutate(intercepto = 1) %>%
     select(intercepto, edu, urban) %>%
     as.matrix()

y <- ej5 %>%
     pull(crime_rate) %>%
     as.matrix()

H <- X %*% solve(t(X) %*% X) %*% t(X)

I <- diag(x = 1, nrow = nrow(X), ncol = nrow(X))

n <- nrow(X)
```

```
p <- 2
gl <- n - p - 1
SC_error <- t(y) %*% (I - H) %*% y
sigma2.hat <- as.double(SC_error / gl)
```

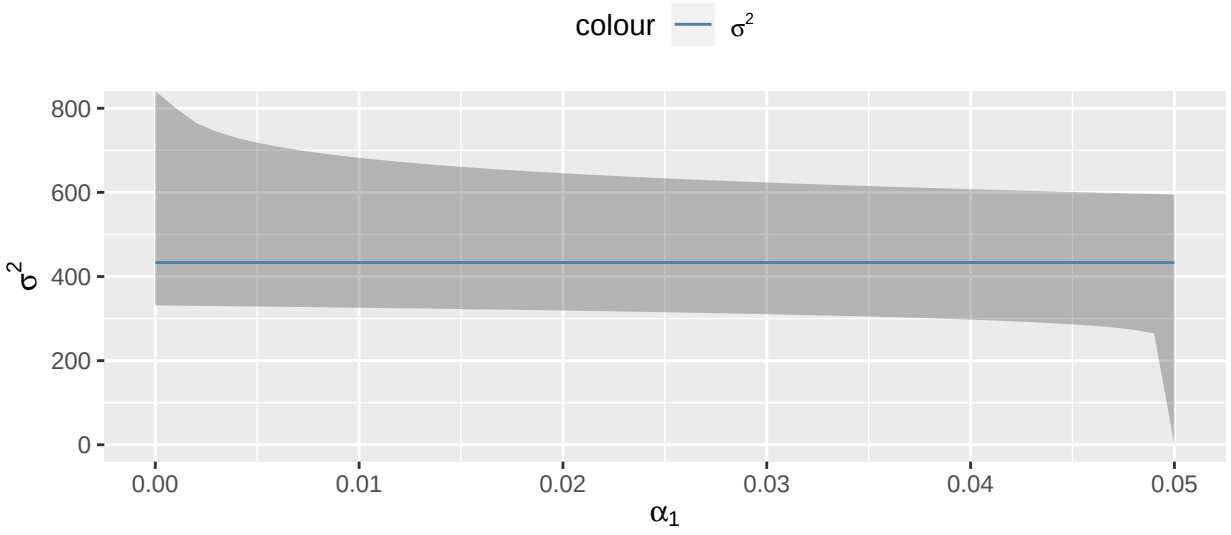
Sabemos que la estimación de sigma está dada por:

$$\hat{\sigma}^2_{MCO} = \frac{1}{n-p-1} SC_{error} = \boldsymbol{y}^T \left(\boldsymbol{H} - \frac{1}{n} \boldsymbol{J}_n \right) \boldsymbol{y} = 433.29$$

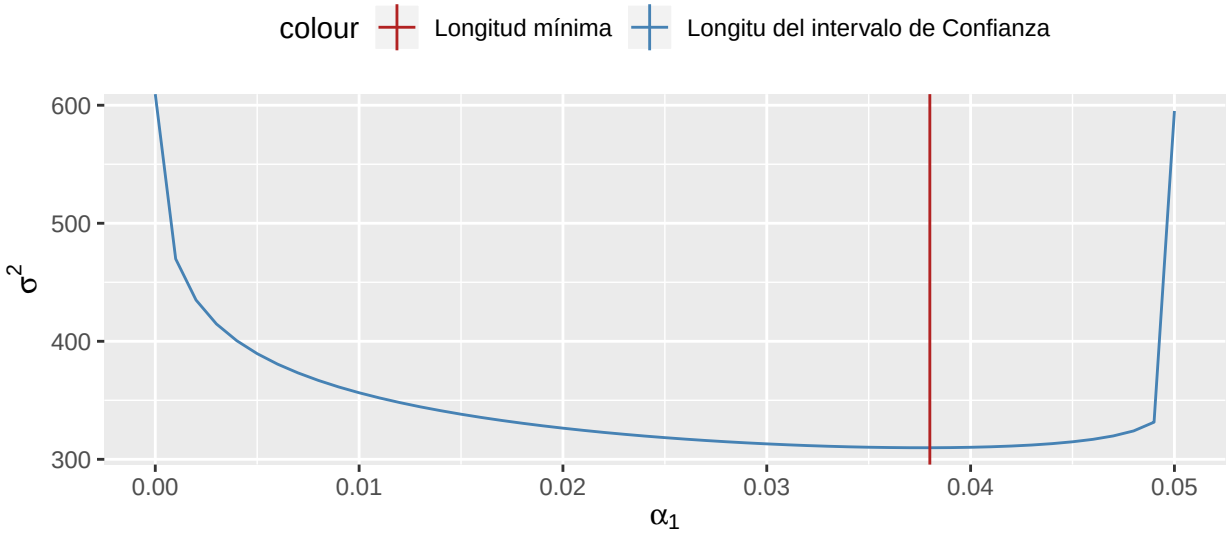
El intervalo de 95% de confianza está dado por:

$$\left(\frac{SC_{error}}{\chi^2_{n-p-1}(1-\alpha_2)}, \frac{SC_{error}}{\chi^2_{n-p-1}(\alpha_1)} \right), \qquad \text{con } \alpha_1 + \alpha_2 = \alpha, \alpha_1, \alpha_2 > 0$$

A continuación se muestran los intervalos de confianza para distrinas combinaciones de α_1 y α_2 tales que $\alpha_1 + \alpha_2 = \alpha = 0.05$.



Calculando la longitud de cada intervalo se obtiene el siguiente comportamiento:



Por lo que, el intervalo de confianza con longitud mínima para $\hat{\sigma}^2_{MCO}$ está dado por $\alpha_1 = 0.038$, y por ende $\alpha_2 = 0.012$. Obteniendo asi el intervalo de confianza:

$$(300.7083489, 610.5135529)$$

Inciso 5.g)

Estimar puntualmente y por intervalo la media de la tasa de crímenes para un 65% de adultos con educación superior y un grado de urbanización de 50%.

La estimación puntual está dada por:

```
x0 <- as.matrix(c(1, 65, 50), ncol = 1)
mu0 <- as.double(t(x0) %*% as.matrix(coeficientes))
```

$$\begin{aligned}\mu_0 &= \boldsymbol{x}_0^T \hat{\beta} \\ &= 59.1180677 + -0.5833773 * (65) + 0.6825014 * (50) \\ &= 59.1180677 + -37.9195277 + 34.1250711 \\ &= 55.3236112\end{aligned}$$

La estimación por intervalo está dada por:

$$x_0^T \hat{\beta} \pm t_{n-p-a}(\alpha/2)\hat{\sigma}_{MCO}\sqrt{1 + x_0^T \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1} x_0}$$

sustituyendo los valores anteriores se obtiene

```r
# Calculamos intervalos de confianza
alfa <- 0.05
intervalo <- c(-1, 1) * as.double(qt(1 - alfa/2, gl) * sqrt(sigma2.hat) *
                                    sqrt(1 + t(x0) %*% solve(t(X) %*% X) %*% x0))
```

$$55.3236112 \pm 42.1156386$$
$$(13.2079725, 97.4392498)$$