i) Generar una muestra i.i.d.

$$\left\{ \begin{pmatrix} X_{1i} \\ X_{2i} \\ X_{3i} \end{pmatrix} \right\}_{i=1}^{500} \text{ donde } \begin{pmatrix} X_{1i} \\ X_{2i} \\ X_{3i} \end{pmatrix} \sim N_3 \left(\begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 0.8 & 0.4 & -0.2 \\ 0.4 & 1.0 & -0.8 \\ -0.2 & -0.8 & 2.0 \end{pmatrix} \right).$$

```
library(mvtnorm)
set.seed(6)
rm(list = ls())

n <- 500
mu <- c(1, 0, 2)
Sigma <- matrix(c(0.8, 0.4, -0.2,
                  0.4, 1, -0.8,
                  -0.2, -0.8, 2),
                nrow = 3,
                byrow = T)
Z <- data.frame(
  rmvnorm(
    n,
    mean = mu,
    sigma = Sigma)
  )
names(Z) <- c("X_1", "X_2", "X_3")

Z <- as.matrix(Z)
X <- cbind(1, Z[,-1])

n <- nrow(Z)
k <- ncol(Z)
```

ii) Generar una muestra $\{U_i\}_{i=1}^{500}$ de v.a.s i.i.d. con $E(U_i) = 0$, tal que U_i y el vector aleatorio $\begin{pmatrix} X_{1j} & X_{2j} & X_{3j} \end{pmatrix}$ sean independientes $\forall i \neq j$, $i, j \in \{1, 2, \ldots, 500\}$. Usar cualquier distribución para U_i que se quiera siempre y cuando no sea Normal.

```
U <- rlogis(n, scale = 3)
```

iii) Usando los datos obtenidos en i) y ii), generar la muestra $\{Y_i\}_{i=1}^{5}00$ donde

$$Y_i = -1 + 2.5X_{1i} - 2X_{2i} + 3X_{3i} + U_i, \quad i = 1, 2, \ldots, 500$$

```
Y <- -1 + 2.5*Z[,1] - 2*Z[,2] + 3*Z[,3] + U
```

iv) Con los datos obtenidos en i) y iii), estimar por mínimos cuadrados ordinarios el modelo de regresión lineal

$$Y_i = b_1 + b_2X_{1i} + b_3X_{2i} + b_4X_{3i} + U_i, \quad i = 1, 2, \ldots, 500$$

```r
regresion <- lm(Y~X_1 + X_2 + X_3, data = data.frame(Y, Z))
print(xtable::xtable(summary(regresion), digits = 6), comment = FALSE)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.892335	0.525984	-1.696505	0.090418
X_1	2.156798	0.309130	6.977003	0.000000
X_2	-1.966675	0.316132	-6.221053	0.000000
X_3	3.089357	0.203891	15.152037	0.000000

v) Encontrar un intervalo de confianza de b_2 con nivel de confianza de 90% (aproximadamente)

a) Sin suponer homocedasticidad

Podemos utilizar el pivote

$$\frac{c'\hat{b_n} - c'b}{\sqrt{c'\frac{1}{n}\hat{V}_n c}} \xrightarrow{D} N(0,1)$$

donde $c = \begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix}$ y $\hat{V}_n = (\frac{1}{n}X'X)^{-1}(\frac{1}{n}\sum_{j=1}^{n} U_j^2 x_j x_j')(\frac{1}{n}X'X)^{-1}$

Por tanto, con un $\alpha = 10\%$, existe un cuantil q de la normal estándar tal que

$$\begin{aligned}
\mathbb{P}\left[-q < \frac{c'\hat{b_n} - c'b}{\sqrt{c'\frac{1}{n}\hat{V}_n c}} < q \right] = 0.9 \iff & \mathbb{P}\left[-q\sqrt{c'\frac{1}{n}\hat{V}_n c} < c'\hat{b_n} - c'b < q\sqrt{c'\frac{1}{n}\hat{V}_n c} \right] = 0.9 \\
\iff & \mathbb{P}\left[-\left(c'\hat{b_n} + q\sqrt{c'\frac{1}{n}\hat{V}_n c} \right) < -c'b < -\left(c'\hat{b_n} - q\sqrt{c'\frac{1}{n}\hat{V}_n c} \right) \right] = 0.9 \\
\iff & \mathbb{P}\left[c'\hat{b_n} + q\sqrt{c'\frac{1}{n}\hat{V}_n c} > c'b > c'\hat{b_n} - q\sqrt{c'\frac{1}{n}\hat{V}_n c} \right] = 0.9 \\
\iff & \mathbb{P}\left[c'\hat{b_n} - q\sqrt{c'\frac{1}{n}\hat{V}_n c} < c'b < c'\hat{b_n} + q\sqrt{c'\frac{1}{n}\hat{V}_n c} \right] = 0.9 \\
\iff & \mathbb{P}\left[\hat{b_{n_2}} - q\sqrt{\frac{1}{n}\hat{V_{n_{22}}}} < b_2 < \hat{b_{n_2}} + q\sqrt{\frac{1}{n}\hat{V_{n_{22}}}} \right] = 0.9
\end{aligned}$$

El intervalo de confianza entonces es

$$\left(\hat{b_{n_2}} - q\sqrt{\frac{1}{n}\hat{V_{n_{22}}}} \quad , \quad \hat{b_{n_2}} + q\sqrt{\frac{1}{n}\hat{V_{n_{22}}}} \right)$$

Donde $\hat{V_{n_{22}}} = (\frac{1}{n}X'X)^{-1}(\frac{1}{n}\sum_{j=1}^{n} U_j^2 x_j x_j')(\frac{1}{n}X'X)^{-1}_{22}$

Podemos calcularlo:

```r
alpha <- 0.1
b_hat_2 <- regresion$coefficients[2]
q <- qnorm(p = alpha/2, lower.tail = F)
```

```r
X <- as.matrix(cbind(1, Z))
XX <- t(X)%*%X
Uj.sq <- regresion$residuals^2
suma_Uj.sq_xj_xj <- matrix(rep(0, 4), nrow = 4, ncol = 4)
for (j in 1:n) {
  suma_Uj.sq_xj_xj <- suma_Uj.sq_xj_xj + Uj.sq[j]*X[j,]%*%t(X[j,])
}
mean_Uj.sq_xj_xj <- (1/n)*suma_Uj.sq_xj_xj
V_hat_hetero <- solve(1/n*XX)%*%mean_Uj.sq_xj_xj%*%solve(1/n*XX)
V_hat_22_hetero <- V_hat_hetero[2,2]
V_hat_hetero
```

```
##                      X_1       X_2        X_3
##     145.42992 -35.015880 -24.92654 -43.304769
## X_1 -35.01588  49.912547 -24.80637  -5.198505
## X_2 -24.92654 -24.806367  59.80729  23.078705
## X_3 -43.30477  -5.198505  23.07870  24.356462
```

```r
CI_het <- c(b_hat_2 - q*sqrt(V_hat_22_hetero/n),
        b_hat_2 + q*sqrt(V_hat_22_hetero/n))
CI_het
```

```
##      X_1      X_1
## 1.637105 2.676491
```

Es decir, el intervalo de confianza para b_2 sin suponer homocedasticidad es:

$$\left(1.6371047, \quad 2.6764914\right)$$

b) Suponiendo homocedasticidad

El intervalo de confianza sigue siendo el mismo, con la diferencia de que ahora $\hat{V_{n_{22}}} = \hat{\sigma}_u^2(\frac{1}{n}X'X)^{-1}_{22}$. Calculando:

```r
var.residual <- summary(regresion)$sigma^2
V_hat_homo <- var.residual*solve(1/n*XX)
V_hat_22_homo <- V_hat_homo[2,2]
V_hat_homo
```

```
##                      X_1       X_2        X_3
##     138.32981 -37.069638 -15.15984 -38.127353
## X_1 -37.06964  47.780544 -22.04175  -3.449607
## X_2 -15.15984 -22.041749  49.96977  17.091204
## X_3 -38.12735  -3.449607  17.09120  20.785682
```

```r
CI_homo <- c(b_hat_2 - q*sqrt(V_hat_22_homo/n),
             b_hat_2 + q*sqrt(V_hat_22_homo/n))
CI_homo
```

```
##      X_1      X_1
## 1.648325 2.665271
```

Es decir, el intervalo de confianza para b_2 suponiendo homocedasticidad es:

$$\Big(1.6483252, \quad 2.6652709\Big)$$

v) Encontrar un intervalo de confianza de $b_1 - 3b_2$ con nivel de confianza de 95% (aproximadamente)

  a) Sin suponer homocedasticidad

En este caso requerimos un pivote que nos permita considerar formas no lineales de b; podemos utilizar el siguiente:

$$\frac{g(\hat{b_n}) - g(b)}{\sqrt{\frac{\partial g(\hat{b_n})}{\partial \hat{b_n}} \frac{1}{n} \hat{V_n} \frac{\partial g(\hat{b_n})}{\partial \hat{b_n}}'}} \xrightarrow{D} N(0,1)$$

Donde $g(\hat{b_n}) = b_1 - 3b_2$, $\frac{\partial g(\hat{b_n})}{\partial \hat{b_n}} = \begin{pmatrix}1 & -3 & 0 & 0\end{pmatrix}$ y $\hat{V_n} = (\frac{1}{n}X'X)^{-1}(\frac{1}{n}\sum_{j=1}^{n} U_j^2 x_j x_j')(\frac{1}{n}X'X)^{-1}$. Dado esto, el denominador del pivote se convierte en:

$$\begin{aligned}
(1,-3,0,0)\frac{1}{n}\hat{V_n}\begin{pmatrix}1\\-3\\0\\0\end{pmatrix} &\iff \frac{1}{n}\left(\hat{V}_{n_{11}} - 3\hat{V}_{n_{12}}, \quad \hat{V}_{n_{12}} - 3\hat{V}_{n_{22}}, \quad \hat{V}_{n_{13}} - 3\hat{V}_{n_{23}}, \quad \hat{V}_{n_{14}} - 3\hat{V}_{n_{24}}\right)\begin{pmatrix}1\\-3\\0\\0\end{pmatrix} \\
&\iff \frac{1}{n}\Big(\hat{V}_{n_{11}} - 3\hat{V}_{n_{12}} - 3(\hat{V}_{n_{12}} - 3\hat{V}_{n_{22}})\Big) \\
&\iff \frac{1}{n}\Big(\hat{V}_{n_{11}} - 6\hat{V}_{n_{12}} + 9\hat{V}_{n_{22}}\Big)
\end{aligned}$$

El pivote entonces es:

$$\frac{g(\hat{b_n}) - g(b)}{\sqrt{\frac{1}{n}\Big(\hat{V}_{n_{11}} - 6\hat{V}_{n_{12}} + 9\hat{V}_{n_{22}}\Big)}} \xrightarrow{D} N(0,1)$$

Por tanto, con un $\alpha = 5\%$, existe un cuantil q de la normal estándar tal que

$$\mathbb{P}\left[-q < \frac{g(\hat{b_n}) - g(b)}{\sqrt{\frac{1}{n}\left(\hat{V}_{n_{11}} - 6\hat{V}_{n_{12}} + 9\hat{V}_{n_{22}}\right)}} < q\right] = 0.95$$

$$\iff \mathbb{P}\left[-q\sqrt{\frac{1}{n}\left(\hat{V}_{n_{11}} - 6\hat{V}_{n_{12}} + 9\hat{V}_{n_{22}}\right)} < g(\hat{b_n}) - g(b) < q\sqrt{\frac{1}{n}\left(\hat{V}_{n_{11}} - 6\hat{V}_{n_{12}} + 9\hat{V}_{n_{22}}\right)}\right] = 0.95$$

$$\iff \mathbb{P}\left[-\left(g(\hat{b_n}) + q\sqrt{\frac{1}{n}\left(\hat{V}_{n_{11}} - 6\hat{V}_{n_{12}} + 9\hat{V}_{n_{22}}\right)}\right) < -g(b) < -\left(g(\hat{b_n}) - q\sqrt{\frac{1}{n}\left(\hat{V}_{n_{11}} - 6\hat{V}_{n_{12}} + 9\hat{V}_{n_{22}}\right)}\right)\right] = 0.95$$

$$\iff \mathbb{P}\left[g(\hat{b_n}) + q\sqrt{\frac{1}{n}\left(\hat{V}_{n_{11}} - 6\hat{V}_{n_{12}} + 9\hat{V}_{n_{22}}\right)} > g(b) > g(\hat{b_n}) - q\sqrt{\frac{1}{n}\left(\hat{V}_{n_{11}} - 6\hat{V}_{n_{12}} + 9\hat{V}_{n_{22}}\right)}\right] = 0.95$$

$$\iff \mathbb{P}\left[g(\hat{b_n}) - q\sqrt{\frac{1}{n}\left(\hat{V}_{n_{11}} - 6\hat{V}_{n_{12}} + 9\hat{V}_{n_{22}}\right)} < g(b) < g(\hat{b_n}) + q\sqrt{\frac{1}{n}\left(\hat{V}_{n_{11}} - 6\hat{V}_{n_{12}} + 9\hat{V}_{n_{22}}\right)}\right] = 0.95$$

El intervalo de confianza entonces es

$$\left(g(\hat{b_n}) - q\sqrt{\frac{1}{n}\left(\hat{V}_{n_{11}} - 6\hat{V}_{n_{12}} + 9\hat{V}_{n_{22}}\right)} \quad , \quad g(\hat{b_n}) + q\sqrt{\frac{1}{n}\left(\hat{V}_{n_{11}} - 6\hat{V}_{n_{12}} + 9\hat{V}_{n_{22}}\right)}\right)$$

Calculando

```
g_b_hat <- regresion$coefficients[1] - 3*regresion$coefficients[2]
alpha <- 0.05
q <- qnorm(p = alpha/2, lower.tail = F)

CI_het2 <- c(g_b_hat - q*sqrt(1/n * (V_hat_hetero[1,1] - 6*V_hat_hetero[1,2] + 9*V_hat_hetero[2,2])),
        g_b_hat + q*sqrt(1/n * (V_hat_hetero[1,1] - 6*V_hat_hetero[1,2] + 9*V_hat_hetero[2,2])))
CI_het2
```

```
## (Intercept) (Intercept)
##   -9.849240   -4.876218
```

Es decir, el intervalo de confianza para $b_1 - 3b_2$ sin suponer homocedasticidad es:

$$\left(-9.8492401, \quad -4.8762182\right)$$

b) Suponiendo homocedasticidad

El intervalo de confianza sigue siendo el mismo, con la diferencia de que ahora $\hat{V}_n = \hat{\sigma}_u^2(\frac{1}{n}X'X)^{-1}$. Calculando:

```
CI_homo2 <- c(g_b_hat - q*sqrt(1/n * (V_hat_homo[1,1] - 6*V_hat_homo[1,2] + 9*V_hat_homo[2,2])),
        g_b_hat + q*sqrt(1/n * (V_hat_homo[1,1] - 6*V_hat_homo[1,2] + 9*V_hat_homo[2,2])))
CI_homo2
```

```
## (Intercept) (Intercept)
##   -9.827570   -4.897888
```

Es decir, el intervalo de confianza para $b_1 - 3b_2$ suponiendo homocedasticidad es:

$$\Big(-9.82757, \quad -4.8978883 \Big)$$