Taller de práctica Prueba 2

Parte 1

La siguiente base contiene 1338 observaciones respecto a la contratación de seguros médicos. Las columnas de la base de datos son las siguientes:

- `age`: edad en años del beneficiario principal.
- `sex`: sexo ("female", "male") del beneficiario principal.
- `bmi`: índice de masa corporal del beneficiario principal.
- `children`: número de niños cubiertos por el seguro.
- `smoker`: indica si es ("yes") o no ("no") fumador el beneficiario principal.
- `region`: el área residencial del beneficiario en los EEUU, "northeast", "southeast", "southwest", "northwest".
- `charges`: costos médicos (en dólares) individuales facturados por el seguro de salud.

Considere que la base de datos está almacenada en el objeto `data`.

1. Se ejecutó en R el comando: `lm(charges ~ age, data = data)`. Los valores de SCT y SCReg son 1.9607422×10^{11} y 1.8898072×10^{10}, respectivamente.

   a) Calcule SCE.

   b) Calcule R^2.

2. Se ejecutó en R el comando: `lm(bmi ~ age, data = data)`. Aquí SCT $= 4.972 \times 10^4$, SCReg $= 596$, $S_{\widehat{\beta}_0} = 0.5842$ y $S_{\widehat{\beta}_1} = 0.0118$. Además, la ecuación de regresión ajustada es

$$\widehat{Y}_{\text{bmi}} = 28.8790 + 0.0475 X_{\text{age}}$$

   a) Calcule MCE y MCReg.

   b) Estudie la prueba de no nulidad para cada uno de los parámetros de la regresión al 98 % de confianza. Interprete utilizando la siguiente tabla.

Tabla 1: Valores críticos según significancia

	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
$t_{1-\alpha/2,n-2}$	2.5795	2.3291	2.1724	2.0558	1.9617	1.8824	1.8134	1.7520	1.6966	1.6460
$t_{\alpha,n-2}$	-2.3291	-2.0558	-1.8824	-1.7520	-1.6460	-1.5558	-1.4767	-1.4059	-1.3415	-1.2822

Parte 2

A continuación, haciendo uso de la base de datos explicada en la Parte 1, se despliega una salida de R asociada a un modelo de regresión lineal simple. Considere que el modelo se ha guardado en el objeto `modelo`.

```
##
## Call:
## lm(formula = charges ~ bmi, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -20956  -8118  -3757   4722  49442
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1192.94    1664.80   0.717    0.474
## bmi           393.87      53.25   7.397 2.46e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11870 on 1336 degrees of freedom
## Multiple R-squared:  0.03934, Adjusted R-squared:  0.03862
## F-statistic: 54.71 on 1 and 1336 DF,  p-value: 2.459e-13
```

Luego,

1. Escriba la ecuación de regresión poblacional.
2. Escriba la ecuación de regresión ajustada.
3. Interprete los parámetros estimados.
4. Estudie las pruebas de hipótesis de no nulidad de cada parámetro. Utilice una confianza del 95 %. Escriba las hipótesis involucradas.
5. Interprete las métricas del modelo.
6. Utilice la siguiente tabla (que contiene un resumen de salidas en R) para estudiar los supuestos de Homocedasticidad, Independencia y Normalidad del modelo de la pregunta anterior. Indique la fila del código seleccionado para cada supuesto, justificando su elección, además, escriba las pruebas de hipótesis involucradas e interprete utilizando una confianza del 96 %.

	Prueba de hipótesis	Valor-p	Datos utilizados	Hipótesis alternativa
1	Shapiro-Wilk	0.0837	modelo$fitted.values	-
2	Shapiro-Wilk	0.0321	rstandard(modelo)	-
3	Anderson-Darling	0.068	modelo$fitted.values	-
4	Durbin-Watson	0.0698	age $\sim$ charges	true autocorrelation is not 0
5	Durbin-Watson	0.0457	charges $\sim$ bmi	true autocorrelation is not 0
6	Durbin-Watson	0.0701	charges $\sim$ age	true autocorrelation is not 0
7	Breusch-Pagan	0.0416	age $\sim$ charges	-
8	Breusch-Pagan	0.0303	charges $\sim$ bmi	-
9	Breusch-Pagan	0.0877	charges $\sim$ region	-

Parte 3

Considere el siguiente modelo de regresión lineal simple:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

1. Obtenga el estimador de mínimos cuadrados de β_0 y β_1.
2. Muestre que SCT = SCReg + SCE.