

# Taller de práctica Examen

## 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 + children + sex, data = data)`. Los valores de SCT y SCReg son  $1.9607422 \times 10^{11}$  y  $1.8898072 \times 10^{10}$ , respectivamente.
  - a) Calcule SCE.
  - b) Calcule  $R^2$ .
  - c) Calcule  $R^2$  ajustado.
2. Se ejecutó en R el comando: `lm(bmi ~ age + smoker, data = data)`. Aquí  $SCT = 4.972 \times 10^4$ ,  $SCReg = 596$ ,  $S_{\hat{\beta}_0} = 0.5842$ ,  $S_{\hat{\beta}_1} = 0.0118$  y  $S_{\hat{\beta}_2} = 0.4111$ . Además, la ecuación de regresión ajustada es

$$\hat{Y}_{bmi} = 28.8790 + 0.0475X_{age} - 0.0980I_{smoker = no}$$

- a) Calcule MCE y MCReg.
- b) Calcule el estadístico  $F_0$ . Estudie la prueba de hipótesis asociada al 95 % de confianza, para ello, utilice 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
$F_{1-\alpha,k,n-k-1}$	4.6211	3.9235	3.5158	3.2266	3.0025	2.8193	2.6646	2.5305	2.4123	2.3066
$F_{1-\alpha/2,k,n-k-1}$	0.0101	0.0202	0.0305	0.0408	0.0513	0.0619	0.0726	0.0834	0.0943	0.1054

- c) 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 2: Valores críticos según significancia

	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
$t_{1-\alpha/2,n-k-1}$	2.5795	2.3291	2.1724	2.0558	1.9617	1.8824	1.8134	1.7520	1.6966	1.6460
$t_{\alpha,n-k-1}$	-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 ~ age + sex + region + smoker, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15726.5  -2066.5  -1296.1   -212.3   28606.9
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   21509.50     704.17  30.546  <2e-16 ***
## age           275.15      12.46  22.081  <2e-16 ***
## sexfemale     -75.15     350.96  -0.214    0.830
## regionnorthwest -294.59    502.45  -0.586    0.558
## regionsoutheast  390.35    489.07   0.798    0.425
## regionsouthwest -436.80    502.45  -0.869    0.385
## smokerno      -23790.26    435.98 -54.567  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6398 on 1331 degrees of freedom
## Multiple R-squared:  0.7221, Adjusted R-squared:  0.7209
## F-statistic: 576.5 on 6 and 1331 DF,  p-value: < 2.2e-16
```

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 y no nulidad conjunta. 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 ~ charges + sex + region + smoker	true autocorrelation is not 0
5	Durbin-Watson	0.0457	charges ~ age + region + smoker	true autocorrelation is not 0
6	Durbin-Watson	0.0701	charges ~ age + sex + region + smoker	true autocorrelation is not 0
7	Breusch-Pagan	0.0416	age ~ charges + sex + region + smoker	-
8	Breusch-Pagan	0.0303	charges ~ age + sex + region + smoker	-
9	Breusch-Pagan	0.0877	charges ~ age + region + smoker	-