Ejercicio 4 - Práctico 2

Valentina Caldiroli

Modelos lineales

```
datos <- read.table("Tabla 7.4.txt")
datos<- datos %>% filter(V1 != 'y1')
y1 <- as.numeric(datos$V1)
y2 <- as.numeric(datos$V2)
x1 <- as.numeric(datos$V3)
x2 <- as.numeric(datos$V4)
x3 <- as.numeric(datos$V5)
datos <-data.frame(y1 =y1, y2 = y2, x1 = x1, x2 = x2, x3= x3)
```

Se cargaron los datos para comenzar a trabajar en este ejercicoi, observando 3 variables explicativas, y 2 variables de respuestas, tenemos entonces:

$x_1 = \ Temperatura\ (Cř)$

$x_2 = \ Concentración\ de\ un\ reactivo\ (\%)$

$x_3 = \ Tiempo\ de\ reacción\ (horas)$

$y_1 = \ Porcentaje\ del\ material\ inicial\ que\ no\ presenta\ cambios$

$y_2 = \ Porcentaje\ que\ se\ convierte\ al\ producto\ deseado$

Parte a:

Comenzaremos trabajando con y_1, en donde podemos estimar una recta que ajuste al modelo como:

$$y_1 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

Calcularemos entonces las estimaciones de los coeficientes β_i mediante mínimos cuadrados ordinarios.

```
modelo <- lm(y1~x1+x2+x3)
b <- coef(modelo)
b
```

```
## (Intercept)          x1          x2          x3
##  332.110983   -1.545961   -1.424559   -2.237366
```

Obtuvimos la matriz b de β_i, lo que nos da un modelo de la forma:

$$y_1 = 332.11 - 1.545x_1 - 1.424x_2 - 2.237x_3$$

Para la estimación de σ^2, utilizaremos la fórmula:

$$\sigma^2 = \frac{SCR}{n-k-1}$$

Para comenzar, calcularemos la suma de los cuadrados de los residuos

```
residuos <- residuals(modelo)
SCR <- sum(residuos^2)
SCR
```

```
## [1] 80.17354
```

Como sabemos $n = 19$ y $k = 3$, entonces podemos sustituir en la fórmula y obtener

```
sigma2 <- SCR /df.residual(modelo)
sigma2
```

```
## [1] 5.344903
```

$$\sigma^2 = \frac{80.1735}{19-3-1} = 5.3449$$

Parte b:

Para obtener la varianza de b, elevamos al cuadrado el error estandar del modelo, dato que conseguimos con la función $summary()$ que se aplica al modelo

```
summary(modelo)
```

```
##
## Call:
## lm(formula = y1 ~ x1 + x2 + x3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6420 -1.1551 -0.0821  1.1563  4.9055
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 332.11098   18.69293  17.767 1.74e-11 ***
## x1           -1.54596    0.09903 -15.610 1.10e-10 ***
## x2           -1.42456    0.14765  -9.648 7.99e-08 ***
## x3           -2.23737    0.33982  -6.584 8.67e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.312 on 15 degrees of freedom
## Multiple R-squared:  0.9551, Adjusted R-squared:  0.9462
## F-statistic: 106.5 on 3 and 15 DF,  p-value: 2.459e-10
```

```r
2.312^2
```

```
## [1] 5.345344
```

Nos devuelve que $Var(b) = 2.312^2 = 5.3453$

Parte c:

Los datods pedidos se encuentran expresado en el resumen del modelo que se halló en el punto anterior, y obtuvimos como resultados entonces $R^2 = 0.9551$ y $R^2_{aj} = 0.9462$

Parte d:

$$y_1 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1^2 + \beta_5 x_2^2 + \beta_6 x_3^2 + \beta_7 x_1 x_2 + \beta_8 x_1 x_3 + \beta_9 x_2 x_3 + \epsilon$$

En este caso plantearemos la función $lm()$, para modelizar lo escrito anteriormente. Luego se le aplicará la función $coef()$ al modelo, para que nos brinde todas las betas.

```r
modelo1 <- lm(y1~x1+x2+x3+I(x1^2)+I(x2^2)+I(x3^2)+x1*x2+x1*x3+x2*x3, datos)
coef(modelo1)
```

```
##  (Intercept)           x1           x2           x3      I(x1^2)      I(x2^2)
## 964.92906310  -7.44212765 -11.50770430  -2.14012737   0.01245711   0.03321878
##      I(x3^2)        x1:x2        x1:x3        x2:x3
##  -0.29401396   0.05350700   0.03803998  -0.10163307
```

Obtuvimos entonces el siguiente resultado:

$$y_1 = 964.929 - 7.4421 x_1 - 11.5077 x_2 - 2.1401 x_3 + 0.0124 x_1^2 + 0.0332 x_2^2 - 0.2940 x_3^2 + 0.0535 x_1 x_2 + 0.0380 x_1 x_3 - 0.1016 x_2 x_3$$

Y para calcular σ^2 utilizaremos la función $residuals()$ del modelo para poder hallar los residuos y poder sumarlos y elevarlos al cuadrado para obtener SCR

```r
residuos <- residuals(modelo1)
SCR <- sum(residuos^2)
SCR
```

```
## [1] 46.20826
```

```r
sigma2 <- SCR /df.residual(modelo1)
sigma2
```

```
## [1] 5.134251
```

Y llegamos entonces a la conclusión de que

$$\sigma^2 = \frac{SCR}{19 - 9 - 1} = 5.134$$

Parte e:

Para hallar tanto el R^2 como R^2_{aj}, lo que haremos es realizarle un *summary*() al modelo 1, y nos brinda el siguiente resumen

```
summary(modelo1)
```

```
##
## Call:
## lm(formula = y1 ~ x1 + x2 + x3 + I(x1^2) + I(x2^2) + I(x3^2) +
##     x1 * x2 + x1 * x3 + x2 * x3, data = datos)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.5502 -1.1637  0.0333  0.8984  3.5041
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 964.92906  645.42185   1.495    0.169
## x1           -7.44213    7.08316  -1.051    0.321
## x2          -11.50770    8.08301  -1.424    0.188
## x3           -2.14013   15.34570  -0.139    0.892
## I(x1^2)       0.01246    0.01995   0.624    0.548
## I(x2^2)       0.03322    0.05066   0.656    0.528
## I(x3^2)      -0.29401    0.22478  -1.308    0.223
## x1:x2         0.05351    0.03593   1.489    0.171
## x1:x3         0.03804    0.09923   0.383    0.710
## x2:x3        -0.10163    0.15117  -0.672    0.518
##
## Residual standard error: 2.266 on 9 degrees of freedom
## Multiple R-squared:  0.9741, Adjusted R-squared:  0.9483
## F-statistic: 37.68 on 9 and 9 DF,  p-value: 4.414e-06
```

Y obtenemos como resultado que $R^2 = 0.9741$ y $R^2_{aj} = 0.9483$. Decimos entonces que al tener un R^2 cercano a 1, el modelo está explicandoo en gran parte a la y_1, dado que si $R^2 = 1$, hablamos de que la y es explicada completamente por el modelo.

Parte f:

Comenzaremos a utilizar ahora como variable de respuesta a y_2. Entonces planteamos un modelo de la siguiente forma:

$$y_2 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$

```
modelo2 <- lm(y2~x1+x2+x3, datos)
beta <- coef(modelo2)
beta
```

```
## (Intercept)          x1          x2          x3
## -26.0352643   0.4045506   0.2929900   1.0338005
```

Obtuvimos entonces un modelo con la siguiente forma

$$y_2 = -26.035 + 0.4045x_1 + 0.2929x_2 + 1.0338x_3$$

Y conociendo al vector x_0, calculamos y_0 como indica la letra utilizando la estimación de las β_i que obtuvimos realizando el modelo con la función $lm()$ y luego la función $coef()$.

```
x0 <- c(1,165,32,5)
y0 <- t(x0)%*%beta
y0
```

```
##          [,1]
## [1,] 55.26027
```

Como $y_0 = 55.26027$, si le palicamos la esperanza a este valor podemos concluir que $E_{y0} = 55.26027$, dado que la esperanza de una constante es la propia constante. Debemos entonces encontrar un intervalo de confianza para este valor.

Un intervalo de confianza al $100(1-\alpha)\%$, puede construirse como:

$$x_0'\hat{\beta} \pm t_{\alpha/2,n-k-1}S\sqrt{x_0'(X'X)^{-1}x_0}$$

```
X <- matrix(c(y2/y2, x1,x2,x3), nrow = 19, ncol = 4)
I <- diag(1, nrow = 4)
aux <- t(X)%*%X
aux1 <- solve(aux, I)
t(x0)%*%aux1%*%x0
```

```
##           [,1]
## [1,] 0.1997508
```

```
S = sqrt(sum(residuals(modelo2)^2)/df.residual(modelo2))
S
```

```
## [1] 4.078088
```

$$55.26027 \pm t_{(\alpha/2),16}4.078\sqrt{0.1997}$$

Parte g:

Un intervalo de predicción para y_0 puede definirse de la siguiente manera:

$$x_0'\hat{\beta} \pm t_{\alpha/2,n-k-1}S\sqrt{1 + x_0'(X'X)^{-1}x_0}$$

Como tenemos todos los datos podemos sustituir y llegar al siguiente resultado:

$$55.26027 \pm t_{(\alpha/2),16}4.078\sqrt{1 + 0.1197}$$

Parte h:

Debemos entonces testear la siguiente hipótesis nula

$$H_0) 2\beta_1 = 2\beta_2 = \beta 3$$

```
A    <- matrix(c(0,2,0,0,0,0,2,0,0,0,0,1), ncol =4)
linearHypothesis(modelo2, A, test = 'F')
```

```
## Linear hypothesis test
##
## Hypothesis:
## 2 x2 = 0
## 2 ((Intercept) = 0
## x3 = 0
##
## Model 1: restricted model
## Model 2: y2 ~ x1 + x2 + x3
##
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     18 335.76
## 2     15 249.46  3    86.295 1.7296 0.2038
```

Obtuvimos un $p - valor = 0.203$, demostrando entonces que este contraste no es significativo ni al 5% ni al 10%.