Modelli Statistici per l'Economia, a.a. 2017-17, Prof. Andrea Tancredi

Compito del 0-0-0000

- Esercizio R. Nella pagina web `http://www.memotef.uniroma1.it/node/6787` è presente il data set icecream.dat. I dati provengono da un'analisi di mercato riguardante il consumo di gelati condotta negli Stati Uniti. I dati si riferiscono al periodo che va dal 1951 al 1953 e rappresentano le seguenti variabili raccolte con frequenza approssimativamente mensile: `IC`, pari al consumo di gelato misurato in pinte per persona, `price`, che rappresenta il prezzo in dollari di una pinta di gelato, `temp` pari alla temperatura media nel periodo di riferimento, `Lag.temp` che fornisce la temperatura media nel periodo precedente, `Year` che indica l'anno e `Income` pari al reddito settimanale familiare in dollari

  1. Scaricare il data set nella propria directory di lavoro e riportare sul foglio la prima riga del data set

     ```
     > dati=read.table("icecream.dat",header=TRUE)
     > dati[1,]
     ```

     ```
       date    IC price income temp Lag.temp Year
     1    1 0.386  0.27     78   41       56    0
     ```

  2. Stimare un modello di regressione semplice $Y_i = \beta_0 + \beta_1 x_i + u_i$, $i = 1, \ldots, n$ avente il consumo di gelato come variabile risposta e la temperatura come covariata (Riportare sul foglio le stime dei minimi quadrati dei parametri β_0 e β_1)

     ```
     > m=lm(IC~temp,data=dati)
     > m
     ```

     ```
     Call:
     lm(formula = IC ~ temp, data = dati)

     Coefficients:
     (Intercept)         temp
         0.20686      0.00311
     ```

  3. Riportare i valori dell'indice R^2 e dello standard error di $\hat{\beta}_1$ ovvero della stima del coefficiente di regressione della variabile temperatura. Che informazioni forniscono queste due quantità?

     ```
     > summary(m)
     ```

     ```
     Call:
     lm(formula = IC ~ temp, data = dati)

     Residuals:
          Min       1Q   Median       3Q      Max
     -0.06941 -0.02448 -0.00737  0.02913  0.12052

     Coefficients:
     ```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.206862   0.024700    8.37  4.1e-09
temp        0.003107   0.000478    6.50  4.8e-07

Residual standard error: 0.0423 on 28 degrees of freedom
Multiple R-squared:  0.602,        Adjusted R-squared:  0.587
F-statistic: 42.3 on 1 and 28 DF,  p-value: 4.79e-07
```

Il valore di R^2 è 0.602. Lo standard error di $\hat{\beta}_1$ è 0.0005.

4. Riportare l'intervallo di confidenza al 95% per β_1.

   L'intervallo di confidenza è $[0.003107 \pm 1.96 \cdot 0.000478] = [0.002, 0.004]$

5. Riportare il p-value associato alla verifica dell'ipotesi $H_0 : \beta_1 = 0$ rispetto all'ipotesi alternativa $H_A : \beta_1 \neq 0$. Sulla base del modello stimato, ritieni che la variabile temperatura influenzi il consumo di gelato?

   Il P-value è 4.8×10^{-7}*, ovvero un valore molto piccolo. Possiamo quindi rifiutare l'ipotesi nulla e stabilire che la temperatura influenza il consumo di gelato.*

6. Stimare un modello di regressione multipla avente il consumo di gelato come variabile risposta e le variabili temperatura e prezzo come covariate (Riportare sul foglio le stime dei minimi quadrati delle variabili temperatura e prezzo)

```
> m2=lm(IC~temp+price,data=dati)
> m2

Call:
lm(formula = IC ~ temp + price, data = dati)

Coefficients:
(Intercept)         temp        price
    0.59655      0.00303     -1.40176
```

   Le stime dei minimi quadrati sono quindi $\hat{\beta}_1 = 0.00303$ *and* $\hat{\beta}_2 = -1.40176$

7. Ritieni che la variabile `price` debba essere effettivamente inclusa nel modello che spiega il consumo di gelato

```
> summary(m2)

Call:
lm(formula = IC ~ temp + price, data = dati)

Residuals:
     Min       1Q   Median       3Q      Max
-0.08226 -0.02051  0.00184  0.02272  0.10076

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.59655    0.25831    2.31    0.029
temp         0.00303    0.00047    6.45  6.6e-07
price       -1.40176    0.92509   -1.52    0.141
```

```
Residual standard error: 0.0413 on 27 degrees of freedom
Multiple R-squared:  0.633,        Adjusted R-squared:  0.606
F-statistic: 23.3 on 2 and 27 DF,  p-value: 1.34e-06
```

Poichè il p-value associato al test $H_0 : \beta_2 = 0$ *rispetto all'ipotesi alternativa* $H_A : \beta_2 \neq 0$ *(dove* β_2 *è il coefficiente di regressione associato alla variabile* ***price****) è 0.141, possiamo dire che non c'è evidenza sufficiente per rifiutare l'ipotesi che* $\beta_2 = 0$ *e quindi, allo stato attuale, possiamo escludere la variabile* ***price*** *dal modello.*