ECONOMETRIA COD. 6089
PRIMO PROBLEM SET

Esercizio 1

Dimostrate che quando c'è una costante tra i regressori la media campionaria della y è uguale a quella della $\widehat{y}$ predetta.

Esercizio 2

Considerate il modello

$$\mathbf{y} = \alpha + \boldsymbol{\beta} X + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim i.i.d\left(0, \sigma^2 I\right)$$

a) Definite il coefficiente di determinazione R^2

b) Discutete intuitivamente il fatto che l'eliminazione di una variabile da un modello lineare peggiora sempre il valore di R^2.

c) Dimostrate che $R^2 = corr^2\left(\mathbf{y}, \widehat{y}\right).$

Esercizio 3

Dato il vettore y e la matrice X

$$\mathbf{y} = \begin{bmatrix} 3 \\ 1 \\ 8 \\ 3 \\ 5 \end{bmatrix}; X = \begin{bmatrix} 1 & 3 & 5 \\ 1 & 1 & 4 \\ 1 & 5 & 6 \\ 1 & 2 & 4 \\ 1 & 4 & 6 \end{bmatrix}$$

Calcolate:

a) il prodotto $X'X$

b) il prodotto $X'\mathbf{y}$

c) lo stimatore dei minimi quadrati $\widehat{\boldsymbol{\beta}}$ per il modello: $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ (suggerimento: non invertite la matrice $X'X$, ma svolgete i calcoli usando il metodo di eliminazione)

d) Scrivete la retta di regressione $\widehat{\mathbf{y}} = X\widehat{\boldsymbol{\beta}}$

Esercizio 4

Considerate il seguente modello lineare:

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{1}$$

dove X è la matrice che contiene gli altri k regressori, compresa la costante.

Considera anche il seguente modello lineare dove tutti i regressori sono divisi per una costante a $\left(u \text{ è } \frac{\epsilon}{\alpha}\right)$:

$$\mathbf{y} = \frac{1}{a}X\tilde{\boldsymbol{\beta}} + \mathbf{u} \tag{2}$$

a) Mostrate che la relazione tra gli stimatori OLS dei modelli (1) e (2) è la seguente:

$$\widehat{\tilde{\boldsymbol{\beta}}}_{OLS} = a\widehat{\boldsymbol{\beta}}_{OLS}$$

b) Mostrate che l'R^2 nei due modelli è lo stesso. Potrebbe essere negativo?

c) Calcolate la varianza di $\widehat{\tilde{\boldsymbol{\beta}}}_{OLS}$ se k=3, $a = 2$ e $\mathrm{Var}\left(\widehat{\boldsymbol{\beta}}_{OLS}\right) = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 10 \end{pmatrix}$.

Esercizio 5

Sotto tutte le ipotesi del modello di regressione lineare, costruire la regione di rifiuto ($\alpha = 0.05$), per l'ipotesi $H_0 : \boldsymbol{\beta}_1 = 4\boldsymbol{\beta}_2$ contro $H_1 : \boldsymbol{\beta}_1 \geq 4\boldsymbol{\beta}_2$ nel modello $\mathbf{y} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 X_1 + \boldsymbol{\beta}_2 X_2 + \boldsymbol{\epsilon}$.

Esercizio 6

a) Regredendo la variabile Y su regressori X_1 e X_2 otteniamo i seguenti risultati:

	coefficiente	std.error
X_1	0.50	0.50
X_2	0.80	0.40

Sappiamo inoltre che il numero di osservazioni N è 42, la somma dei quadrati dei residui (RSS_U) è 80, mentre la somma dei quadrati dei residui nell'ipotesi che i coefficienti di X_1 e X_2 siano pari a zero (RSS_R) è 120.

Calcolate il test-F.

b) Dato il seguente output di STATA, commentate i valori dei test-t e del test-F. Cosa potete concludere riguardando alla regressione proposta?

```
reg   y x1 x2

      Source |       SS       df       MS              Number of obs =      50
-------------+------------------------------           F(  2,    47) =  148.36
       Model |  8.6155e+10     2  4.3077e+10           Prob > F      =  0.0000
    Residual |  1.3646e+10    47   290348771           R-squared     =  0.8633
-------------+------------------------------           Adj R-squared =  0.8574
       Total |  9.9801e+10    49  2.0368e+09           Root MSE      =   17040

------------------------------------------------------------------------------
    y        |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        X1   |   3614.933     2304.47     1.57   0.123    -1021.063   8250.929
        X2   |  -3378.924     7218.92    -0.47   0.642    -17901.52   11143.67
      _cons  |   .1122415    .1419259    17.23   0.000     .1270837   .0073778
------------------------------------------------------------------------------
```

Esercizio 7

Regredendo la variabile Y sui regressori X_1 e X_2 si ottiene:

Regressore	Coefficiente	Std. Error
X_1	1.5	0.50
X_2	2	0.25

a) Definite il test-F per testare l'ipotesi $H_0 = \beta_1 = c_1; \beta_2 = c_2; ...; \beta_k = c_k$ contro l'ipotesi alternativa che almeno una delle componenti di H_0 sia falsa.

b) Costruite il test-F per l'ipotesi nulla che il coefficiente di X_1 sia nullo e che il coefficiente di X_2 sia pari a uno, contro l'ipotesi alternativa che almeno una restrizione non sia vera e fornite il valore della realizzazione della sua statistica test (supponiamo $X_1^{'} X_2 = X_2^{'} X_1 = 0$).

c) Dato il valore del test-F calcolato al punto b) e supponendo che la somma dei quadrati nel modello non ristretto RSS_U sia pari a 10 e un campione di ampiezza 52, calcolate RSS_R.

Esercizio 8

Considerate il seguente modello:

$$\mathbf{y} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 x + \boldsymbol{\beta}_2 z + \boldsymbol{\epsilon}$$

Ecco la stima in STATA

```
 reg y x z

      Source |       SS       df       MS              Number of obs =     953
-------------+------------------------------           F(  2,   950) =  966.71
       Model |  128389.471     2  64194.7355           Prob > F      =  0.0000
    Residual |  63085.3254   950  66.4056057           R-squared     =  0.6705
-------------+------------------------------           Adj R-squared =  0.6698
       Total |  191474.796   952  201.128988           Root MSE      =   8.149

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |    7.08117   .2573069    27.52   0.000     6.576214    7.586125
           z |   .0078033   .0002896    26.94   0.000     .0072349    .0083717
       _cons |   7.476546   .7369036    10.15   0.000     6.030399    8.922693
------------------------------------------------------------------------------
```

a) Commentate i coefficienti di x e z.. Che effetto hanno sulla variabile y? Sono significativi?

b) Costruite l'intervallo di confidenza al 95% di significatività per x. Ottenete lo stesso risultato della tabella?
(*Il valore critico di una normale nel caso di livello di confidenza del 95% è pari a 1.96*).

c) Commentate la statistica F.

d) Definite R^2 ed R^2_{adj}. Calcolateli, usando i valori della tabella di STATA e confrontate il risultato ottenuto con quello riportato in tabella.

Esercizio 9

Consideriamo il seguente modello:

$$\mathrm{col}\,gpa = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 hesperc + \boldsymbol{\beta}_2 sat + \boldsymbol{\epsilon}$$

dove col gpa misura su una scala da 0 a 4 i voti degli studenti di un college americano, *hesperc* esprime in termini percentili la promozione alla scuola superiore (*hesperc*=5 significa che lo studente era nel 5% dei bravi della sua classe) e *sat* rappresenta il punteggio di un test attitudinale per l'ingresso in università, in cui vengono testate le abilità matematiche e verbali.

a) Spiegate perché ci aspettiamo una stima del coefficiente $\boldsymbol{\beta}_1 < 0$ e del coefficiente $\boldsymbol{\beta}_2 > 0$

b) Ecco il modello stimato in STATA:

```
reg  colgpa  hsperc  sat

      Source |       SS       df       MS              Number of obs =    4137
-------------+------------------------------           F(  2,  4134) =  777.92
       Model |  490.606706     2  245.303353           Prob > F      =  0.0000
    Residual |  1303.58897  4134  .315333567           R-squared     =  0.2734
-------------+------------------------------           Adj R-squared =  0.2731
       Total |  1794.19567  4136  .433799728           Root MSE      =  .56155

------------------------------------------------------------------------------
      colgpa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      hsperc |  -.0135192   .0005495   -24.60   0.000    -.0145965    -.012442
         sat |   .0014762   .0000653    22.60   0.000     .0013482    .0016043
       _cons |   1.391757   .0715424    19.45   0.000     1.251495    1.532018
------------------------------------------------------------------------------
```

Ed ecco una descrizione delle tre variabili:

```
sum   colgpa hsperc sat

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      colgpa |      4137    2.652686    .6586347          0          4
      hsperc |      4137    19.23707    16.56873   .1666667         92
         sat |      4137    1030.331    139.4014        470       1540
```

Spiegate la significatività economica e statistica dei coefficienti di *hesperc* e *sat*.

c) Calcolate il valore di $\widehat{\mathrm{col}\,gpa}$, considerando che *hsperc*=20 e *sat*=1 050.

d) Supponiamo che due studentesse, Alice e Melissa, si siano diplomate piazzandosi con lo stesso percentile, ma Alice ha ottenuto un punteggio di 140 punti maggiore di Melissa al test SAT. Che differenze ci sono nel valore stimato di $\widehat{\mathrm{col}\,gpa}$?

e) Tenendo fissa la variabile $hsperc$, che differenza c'è nel punteggio del test SAT al fine di ottenere una variazione su $colgpa$ pari a 0.5? Commentate.