

MODELLI DI REGRESSIONE SEMPLICE e MULTIPLO ①  
DESCRIZIONE OUTPUT EXCEL (pag. 9 e 10)

DATASET = IMPIEGATI

Contiene dati su un campione di 424 impiegati; relativi ad alcune variabili, vengono prese in considerazione le seguenti:

- $SALARY/1000$  = retribuzione annua (in migliaia di dollari)
- $EDUC$  : anni di istruzione
- $PERVEXA$  : esperienze lavorative precedenti (in mesi)

Per ottenere l'output relativo al modello di regressione semplice (con variabile dipendente  $SALARY/1000$  e variabile esplicativa  $EDUC$ ) e l'output relativo al modello di regressione multipla (con variabile dipendente  $SALARY/1000$  e variabili esplicative  $EDUC$  e  $PERVEXA$ ) si procede come segue (dopo aver installato il componente aggiuntivo ANALISI DEI DATI):

- Dal menu DATI, si seleziona ANALISI DATI;
- dal menu che appare si seleziona REGRESSIONE;
- nella finestra che si apre si selezionano i dati relativi alla variabile dipendente (Y) e alla variabile indipendente (X) o alle variabili indipendenti tutte in blocco, nel caso della regressione multipla;
- se include le etichette, si selezionano la cella relativa; eventualmente, si sceglie un livello di confidenza aggiuntivo e quello fornito (0,95) in automatico.

1) REGRESSIONE SEMPLICE [i valori sono arrotondati a 1 quarto decimale] ②

$$Y = SALARY/1000$$

$$X = EDUC$$

- Nella I tabella vengono riportati l'indice

$R^2$  (pari a 0,4363), la radice quadrata della  $s^2$  (pari a 6, ovvero Se (indicated nell'output come  $ERRORE STANDARD$ , pari a 12,8335, per cui  $S^2 = 12,8335^2 = 164,6887$ ), il valore di  $n = 424$

la radice quadrata di  $R^2$  (indicated con  $R$  multiplo), che coincide con il valore assoluto del coefficiente di correlazione tra  $X$  e  $Y$ ; si tralasci qui  $R^2$  stesso.

- Nella II tabella vengono riportati i valori della DEVIANZA TOTALE (SST, indicato come TOTALE, uguale a 137946,4954), della DEVIANZA RESIDUA (indicated con  $RESIDUO$ , pari a 9738,2777), della DEVIANZA SPIEGATA SSR (indicated come REGRESSIONE, pari a 60178,2178).

Le rimanenti quantità pertinenti nella II tabella potranno essere tralasciate (nella regressione semplice).

- Nella III tabella vengono riportate le quantità necessarie per inferire su  $\beta_1$  (e  $\beta_0$ ); saranno l'attenzione su  $\beta_1$ , corrispondente alla  $III$  riga, indicata con  $EDUC$  (che è la variabile alloquale corrisponde il coefficiente  $\beta_1$ ).

3

Colonna coefficienti = stima di  $\beta_1$  (ovvero  $b_1 = 3.9089$ )  
(e stima di  $\beta_0$  (cioè  $b_0 = -18.3312$ )).

Da qui si ricava l'equazione  
stima  $\hat{y} = -18.3312 + 3.9089x$ .

Colonna errore standard, standard error di  $b_1$  ovvero  
 $SB_1 = 0.2045$ .

Colonna Stat t = valore della statistica test  
osservata ovvero  
 $T_{obs} = \frac{b_1}{SB_1} = 18.1150$

Colonna valore di p = p-value relativo al test  
significativo bilaterale corrispondente  
alle ipotesi:  $H_0: \beta_1 = 0$  contro  
 $H_1: \beta_1 \neq 0$ , per cui  
 $0.6337 \cdot 10^{-61}$  (quindi praticamente  
uguale a 0, per cui si rifiuta  
 $H_0$  a qualunque livello)

Colonna Inferiore 95% = estremi inferiore e superiore  
di superiore 95% dell'intervallo di confidenza  
al 95% per  $\beta_1$  (l'intervallo  
è quindi (3.5080, 4.3119))

Colonna Inferiore 99% = come il precedente, ma  
di superiore 99%

2) REGRESSIONE MULTIPLA [valori sempre arrotondati,  
al quarto decimale]

4

$Y = \text{SALE}/1000$   
 $X = \text{DVC}$   
 $X_2 = \text{PREVENP}$

• Nella prima tabella sono riportati gli indici:

$R^2 = 0.4415$ ,  $R = 0.6644$  (radice quadrata di  $R^2$ )  
uguale al valore assoluto del coefficiente di

correlazione lineare tra  $y$  e  $g$  indicato come

$R$  (multiplo),  $R^2$  correla = 0.4384, errore  
standard =  $SE = 12.7883$ ,  $n = 474$ .

• Nella seconda tabella sono riportati:

1° riga =  $gd1 = 2$ , numero di variabili esplicative ( $k$ );  
 $SSQ = 60884.1142$ , devianza spiegata =  $SSE$ ;

2° riga =  $gd1 = 474 = n - k - 1$

$MSQ = 77032.3872 = \text{DEVIANZA RESIDUA (SSE)}$

3° riga =  $gd1 = n - 1 = 473$   
 $MSQ = 137916.4554 = \text{DEVIANZA TOTALE (SST)}$

(si testano le altre quantità della seconda tabella).

• Nella terza tabella sono riportate le quantità relative  
all'inferenza (stima puntuale, standard error, intervalli,  
test) sui coefficienti  $\beta_1, \beta_2$  (della variabile DVC)  
e  $\beta_0$  (della variabile PREVENP); si vedrà la tabella  
relativa alla regressione semplice per la derivazione.

Sulla base dei due output, rispondiamo alle seguenti domande:

1) Nel modello di regressione che include la sola variabile EDUC,  $q_{werts}$  è significativo per la spiegazione della retribuzione? Se sì, quale è il suo effetto?

- 2) Nel modello che include EDUC e PEXPER, quest'ultima riporta un contributo significativo alla spiegazione della retribuzione (a livello 0.05)? Se sì, qual è il suo effetto? Qual è la percentuale di variabilità complessivamente spiegata dagli anni di istruzione e dall'esperienza precedente?
- 3) Si prevede, usando il modello con EDUC e PEXPER, la retribuzione media degli impiegati con 12 anni di istruzione e 200 mesi di esperienza?

### RISPOSTE

- 1) Si effettua la stessa previsione richiesta al punto precedente per il modello con la sola EDUC trascorrendo quindi il dato su PEXPER. Si tratta di un modello di regressione semplice. Occorre effettuare il test per  $H_0: \beta_1 = 0$  contro  $H_1: \beta_1 \neq 0$  dall'output si vede che il p-value per tale test è  $9.5387 \cdot 10^{-61} \approx 0$ , per cui si rifiuta  $H_0$  (ad ogni livello); c'è quindi evidenza che  $\beta_1 \neq 0$ , per cui EDUC è significativa, l'effetto si vede dallo stima di  $\beta_1$ , pari a  $b_1 = 3.9099$ , ad un incremento di un anno di istruzione è associato un incremento (medio) pari a  $3.9099$  dell'età di retribuzione.

5

2) Occorre effettuare, nell'ambito del modello di regressione multiple, il test per la significatività della singola variabile PEXPER, relativo cioè alle ipotesi:  $H_0: \beta_2 = 0$  contro  $H_1: \beta_2 \neq 0$ . Il p-value per questo test è  $0.0383 < 0.05$ , per cui si rifiuta  $H_0$  (a livello 0.05); dunque PEXPER è significativa in questo modello. Lo stima di  $\beta_2$  è  $0.0121$ , ad un incremento di un mese di esperienza, fatti gli anni di istruzione, è associato un incremento medio di retribuzione pari a  $12.1$  \$.

6

3) La percentuale di variabilità spiegata in questo modello di regressione multiple è pari al 44.15%. (dati dal valore di  $R^2$ ).

4) L'equazione stimata del modello è:  

$$\hat{y}_{n+1} = -20.9783 + 4.0203 \cdot X_1 + 0.0121 \cdot X_2$$
 La previsione è quindi:  

$$\hat{y}_{n+1} = -20.9783 + 4.0203 \cdot 12 + 0.0121 \cdot 200 = 28.6853$$

5) L'equazione stimata del modello (di regressione semplice) è:  

$$\hat{y}_n = -18.3312 + 3.9099 \cdot X$$
 La previsione è quindi:  

$$\hat{y}_{n+1} = -18.3312 + 3.9099 \cdot 12 = 28.5876$$

*Statistica della regressione*

R multiplo	0.6644218
R al quadrato	0.4414564
R al quadrato corretto	0.4390846
Errore standard	12.788694
Osservazioni	474

OUTPUT REGRESSIONE MULTIPLA

8

ANALISI VARIANZA

	<i>gdl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>Significatività F</i>
Regressione	2	60884.11424	30442.06	186.1322301	2.70364E-60
Residuo	471	77032.3812	163.5507		
Totale	473	137916.4954			

	<i>Coefficienti</i>	<i>Errore standard</i>	<i>Stat t</i>	<i>Valore di significatività</i>	<i>Inferiore 95%</i>	<i>Superiore 95%</i>	<i>Inferiore 99.0%</i>	<i>Superiore 99.0%</i>
Intercetta	-20.978304	3.087257655	-6.79513	3.28056E-11	-27.04480627	-14.9118009	-28.96290208	-12.99370513
educ	4.0203433	0.210649876	19.08543	1.42045E-60	3.606413509	4.434273168	3.475537926	4.565148751
prevexp	0.0120713	0.005810445	2.077516	0.038295234	0.000653688	0.023488891	-0.002956309	0.027098888

OUTPUT REGRESSIONE SEMPLICE

8

*Statistica della regressione*

R multiplo	0.66055891
R al quadrato	0.43633807
R al quadrato corretto	0.43514387
Errore standard	12.8335397
Osservazioni	474

ANALISI VARIANZA

	<i>gdl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>Significatività F</i>
Regressione	1	60178.21776	60178.2178	365.3813749	9.63974E-61
Residuo	472	77738.27768	164.699741		
Totale	473	137916.4954			

	<i>Coefficienti</i>	<i>Errore standard</i>	<i>Stat t</i>	<i>Valore di significatività</i>	<i>Inferiore 95%</i>	<i>Superiore 95%</i>	<i>Inferiore 99.0%</i>	<i>Superiore 99.0%</i>
Intercetta	-18.331178	2.821911555	-6.49601438	2.09629E-10	-23.87624179	-12.78611427	-25.62944694	-11.03290912
educ	3.90990671	0.204547037	19.1149516	9.63974E-61	3.507971232	4.311842181	3.380889674	4.438923739