

Una compagnia di assicurazione (con riferimento a polizze RCA) vuole studiare la dipendenza degli indennizzi versati per incidenti da alcuni possibili fattori. A questo scopo considera un campione di 200 pratiche, rilevando su ciascuna le seguenti variabili:

- AC (Accident Compensation): entità dell'indennizzo;
- AGE : età del conducente;
- SP (Speed): velocità del veicolo al momento dell'incidente.

I dati relativi al campione, opportunamente elaborati, danno luogo ai 2 output riportati al termine delle domande.

1. Si vuole spiegare la variabilità degli indennizzi attraverso la sola velocità del veicolo. Questo modello è significativo a livello 0.01? Ovvero la velocità ha un effetto significativo sull'indennizzo?
2. Se la risposta alla domanda precedente è positiva, si descriva l'effetto della velocità sull'indennizzo. Si precisi tale effetto con un intervallo al 90%.
3. A quanto ammonta la parte di variabilità degli indennizzi spiegata dalla velocità?
4. Si preveda, tramite questo modello, l'indennizzo medio associato ad incidenti in cui la velocità del veicolo è pari a 100 km/h.
5. Si aggiunga ora alla velocità, per la spiegazione degli indennizzi, l'età del conducente. Si scriva l'equazione del modello stimato.
6. La variabile età è significativa, a livello 0.05, nel modello considerato al punto 5.? Se sì, qual è l'effetto dell'età sull'indennizzo (in questo modello)?
7. Usando ancora il modello con tutte le variabili, si preveda l'indennizzo per un incidente in cui il conducente ha 50 anni di età e procede a 100 km/h.

#### OUTPUT 1

| <i>Statistica della regressione</i> |            |
|-------------------------------------|------------|
| R multiplo                          | 0.700877   |
| R al quadrato                       | 0.49122857 |
| R al quadrato corretto              | 0.48865902 |
| Errore standard                     | 137.714411 |
| Osservazioni                        | 200        |

#### ANALISI VARIANZA

|             | <i>gdl</i> | <i>SQ</i>   | <i>MQ</i> | <i>F</i>    | <i>Significatività F</i> |
|-------------|------------|-------------|-----------|-------------|--------------------------|
| Regressione | 1          | 3625641.622 | 3625642   | 191.1727989 | 7.09488E-31              |
| Residuo     | 198        | 3755121.258 | 18965.26  |             |                          |
| Totale      | 199        | 7380762.88  |           |             |                          |

|            | <i>Coefficienti</i> | <i>Errore standard</i> | <i>Stat t</i> | <i>Valore di significatività</i> | <i>Inferiore 95%</i> | <i>Superiore 95%</i> |
|------------|---------------------|------------------------|---------------|----------------------------------|----------------------|----------------------|
| Intercetta | -153.929901         | 51.68104295            | -2.97846      | 0.003259102                      | -255.845821          | -52.01398109         |
| SP         | 6.11297823          | 0.442119632            | 13.82653      | 7.09488E-31                      | 5.24111059           | 6.984845874          |

## OUTPUT 2

| <i>Statistica della regressione</i> |             |
|-------------------------------------|-------------|
| R multiplo                          | 0.708334429 |
| R al quadrato                       | 0.501737664 |
| R al quadrato corretto              | 0.496679163 |
| Errore standard                     | 136.6301495 |
| Osservazioni                        | 200         |

## ANALISI VARIANZA

|             | <i>gdl</i> | <i>SQ</i>   | <i>MQ</i> | <i>F</i>    | <i>Significatività F</i> |
|-------------|------------|-------------|-----------|-------------|--------------------------|
| Regressione | 2          | 3703206.725 | 1851603   | 99.18702719 | 1.5835E-30               |
| Residuo     | 197        | 3677556.155 | 18667.8   |             |                          |
| Totale      | 199        | 7380762.88  |           |             |                          |

|            | <i>Coefficienti</i> | <i>Errore standard</i> | <i>Stat t</i> | <i>Valore di significatività</i> | <i>Inferiore 95%</i> | <i>Superiore 95%</i> |
|------------|---------------------|------------------------|---------------|----------------------------------|----------------------|----------------------|
| Intercetta | -30.2224276         | 79.44923539            | -0.3804       | 0.704058929                      | -186.9025989         | 126.4577438          |
| AGE        | -1.65125851         | 0.810080966            | -2.03839      | 0.042847575                      | -3.248802202         | -0.053714825         |
| SP         | 5.740192463         | 0.475236793            | 12.07859      | 1.69312E-25                      | 4.802987948          | 6.677396978          |

1. Occorre effettuare il test per le ipotesi  $H_0 : \beta_{SP} = 0$  contro  $H_1 : \beta_{SP} \neq 0$ . Dall'output si rileva che il p-value del test è  $7.09488 \cdot 10^{-31}$ , per cui si rifiuta l'ipotesi nulla a livello 0.01 (ed a qualunque altro livello), e dunque l'effetto della velocità è significativo.

2. La stima del coefficiente di SP è 6.1130; quindi ad un incremento di 1 km/h della velocità al momento dell'incidente è associato un incremento (medio) dell'indennizzo pari a 6.1130. L'intervallo al 90% per tale effetto (cioè per  $\beta_{SP}$ ) è:  $(6.1130 - 1.645 \cdot 0.4421, 6.1130 + 1.645 \cdot 0.4421) = (5.3857, 6.8403)$ ; è stato usato nell'intervallo il quantile della distribuzione normale, essendo i gradi di libertà della T di Student pari a 198.

3. La parte di variabilità spiegata dalla velocità è il 49.12%, come indicato dal coefficiente  $R^2 = 0.4912$ .

4. La previsione richiesta è:  $\hat{y}_{n+1} = -153.9299 + 6.1130 \cdot 100 = 457.3701$ .

5. L'equazione del modello stimato è:

$$\hat{y} = -30.2224 - 1.6513 \text{ AGE} + 5.7402 \text{ SP}$$

6. Occorre effettuare un test (il test t) per l'ipotesi  $H_0 : \beta_{AGE} = 0$  contro  $H_1 : \beta_{AGE} \neq 0$ . Dall'output relativo al modello con tutte le variabili si vede che il p-value per questo test è  $0.0428 < 0.05$ , quindi si rifiuta l'ipotesi nulla a livello 0.05 e dunque l'età è significativa. La stima del suo coefficiente è  $-1.6513$ , per cui ad un incremento di 1 anno di età del conducente corrisponde una riduzione dell'indennizzo medio pari a 1.6513, fissata la velocità del veicolo.

7. La previsione è data da  $\hat{y}_{n+1} = -30.2224 - 1.6513 \cdot 50 + 5.7402 \cdot 100 = 461.2326$