# Esercitazione 9 del corso di Statistica (parte seconda)

Dott.ssa Paola Costantini

17 Marzo 2009

**Esercizio 1**

Al fine di verificare le spese trimestrali per la tenuta di c/c bancari, la Banca d'Italia esamina le spese di tenuta conto praticate da 8 banche italiane riscontrando una spesa trimestrale media di 32 Euro con una varianza campionaria pari a 1.8.
Allo stesso tempo, procede ad effettuare lo stesso tipo di rilevazione presso 10 banche straniere operanti in Italia, rilevando una spesa trimestrale media di 32.9 Euro con una varianza campionaria pari a 1.7.
Si vuole verificare se le spese trimestrali medie di tenuta conto sono le stesse presso i due tipi di banche.

SVOLGIMENTO

Si richiede di effettuare un test sulla differenza tra medie non conoscendo le varianze del carattere osservato nelle due popolazioni. In tal caso, per poter effettuare il test, è necessario ipotizzare che le due varianze siano uguali $\left(\sigma_1^2 = \sigma_2^2 = \sigma^2\right)$.
Le ipotesi da sottoporre a verifica sono le seguenti:

$$H_0 : \mu_1 - \mu_2 = 0 \qquad\qquad H_1 : \mu_1 - \mu_2 \neq 0$$

Per poter effettuare il test bisogna anche ipotizzare che il carattere osservato abbia una distribuzione normale presso le due popolazioni.
La statistica test da considerare è la seguente:

$$T = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \approx t_{(n_1+n_2-2)}$$

che ha una distribuzione $t$ di Student con $n_1+n_2-2$.
Fissato un livello di significatività $\alpha = 0.05$ è possibile determinare le regioni di accettazione e di rifiuto:

Regione di accettazione: $\left\{-t_{\alpha/2} < t_i < t_{1-\alpha/2}\right\}$; Regione di rifiuto: $\left\{t_i < -t_{\alpha/2}\right\} \cup \left\{t_i > t_{1-\alpha/2}\right\}$.

Dai dati a disposizione risulta:

$$n_1 = 8; \qquad n_2 = 10; \qquad \overline{x}_1 = 32; \qquad \overline{x}_2 = 32.9; \qquad s_1^2 = 1.8 \qquad s_2^2 = 1.7.$$

Il test è a due code; i valori critici della distribuzione *t* di Studente per questo livello di significatività sono $t_{0.025} = -2.12$ e $t_{0.975} = 2.12$.
La regola di decisione consiste nel rifiutare l'ipotesi nulla se il valore campionario della statistica test è inferiore a -2.12 o superiore a 2.12.

Il valore campionario della statistica test è: $t = \dfrac{32 - 32.9}{\sqrt{\dfrac{8(1.8)+10(1.7)}{8+10-2}\left(\dfrac{1}{8}+\dfrac{1}{10}\right)}} = -1.35$

Al livello di significatività del 5% l'ipotesi nulla viene accettata, concludendo che le spese trimestrali medie di tenuta conto sono le stesse presso i due tipi di banche.

Esercizio 2

Un economista del Ministero degli Esteri desidera verificare se gli accordi di negoziazione tra Italia e Giappone siano rispettati. In particolare egli sospetta che i produttori giapponesi fissino un prezzo più basso per i prodotti venduti sul mercato italiano rispetto a quello usato sul mercato interno, ostacolando al contempo le importazioni di prodotti italiani con forti ostacoli di tipo burocratico. Si interessa in particolare al mercato dell'auto e vuole testare l'ipotesi che prezzi più alti siano applicati in Giappone rispetto all'Italia per le autovetture di produzione giapponese. Esamina a tal fine due campioni relativi a pratiche di acquisto di tali autovetture nello stesso periodo di tempo (50 per il mercato italiano e 30 per il mercato giapponese). Convertendo i prezzi di vendita in Giappone usando il cambio corrente Yen/Euro, ottiene i risultati elencati nella seguente tabella:

| | ITALIA | GIAPPONE |
|---|---|---|
| Ampiezza campione | 50 | 30 |
| Media campionaria | € 16545 | € 17243 |

Siano inoltre noti i seguenti valori per le rispettive popolazioni di riferimento:

| | ITALIA | GIAPPONE |
|---|---|---|
| Deviazione standard | € 1989 | € 1843 |

a) Costruire un test di ipotesi usando un livello $\alpha$=0.5
b) Si calcoli inoltre il livello di significatività osservato del test (p-value)

**SVOLGIMENTO**

**a)**
1) Ipotesi nulla
Ho: $(\mu_1-\mu_2)=0$
2) Ipotesi alternativa
Ha: $(\mu_1-\mu_2)<0$
3) Statistica test

Statistica test → $$\frac{\left(\overline{x_1}-\overline{x_2}\right)-\left(\mu_1-\mu_2\right)}{\sigma_{\left(\overline{x_1}-\overline{x_2}\right)}}=\frac{\left(\overline{x_1}-\overline{x_2}\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim Z$$

4) Regola di decisione
Utilizzando le tavole della Z si ha infatti:
$\alpha$=0.05 → $Z_\alpha=-1.645$

5) Assunzioni (ipotesi mantenute)
I due campioni sono selezionati in maniera indipendente dalle due popolazioni. Le rispettive ampiezze campionarie $n_1$ e $n_2$ sono sufficientemente grandi affinché sia $\bar{x}_1$ che $\bar{x}_2$ siano distribuite approssimativamente come normali.

6) Esperimento sul campione

$\bar{x}_1 = 16545$ $\sigma_1 = 1989$ $n_1 = 50$

$\bar{x}_2 = 17243$ $\sigma_2 = 1843$ $n_2 = 30$

Statistica test → $$\frac{\left(\bar{x}_1 - \bar{x}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(16545 - 17243)}{\sqrt{\frac{1989^2}{50} + \frac{1843^2}{30}}} = -1.59$$

7) Conclusione
La statistica test, per il campione considerato, produce un valore nella regione di accettazione: i dati empirici non permettono di accettare l'ipotesi che il prezzo praticato in Giappone per le autovetture in questione sia più alto di quello praticato in Italia.

## Esercizio 3

| Tipo di contratto / Utilizzo della laurea | Stabile | Atipico | Inserimento/ formazione | Senza contratto | Tot |
|---|---|---|---|---|---|
| In misura elevata | 5 | 3 | 1 | 0 | 9 |
| In misura ridotta | 2 | 5 | 1 | 0 | 8 |
| Per niente | 2 | 4 | 0 | 2 | 8 |
| totale | 9 | 12 | 2 | 2 | 25 |

Verificare ad un livello di significatività del 95% se i due caratteri sono indipendenti

Ipotesi

$$H_0 = \sum_i \sum_j \frac{(nij - \hat{n}ij)^2}{\hat{n}ij} = 0$$

$$H_1 = \sum_i \sum_j \frac{(nij - \hat{n}ij)^2}{\hat{n}ij} > 0$$

Statistica Test = $$\sum_i \sum_j \frac{(nij - \hat{n}ij)^2}{\hat{n}ij}$$

Valore critico = $\chi^2_{0,05,6}$ 6=(r-1)*(c-1) = 12,592

Valore test = $\chi^2 = = \mathbf{7,45}$

Accetto l'ipotesi nulla, quindi i due caratteri sono indipendenti. Il valore test è diverso da zero, ma non significativamente diverso da zero.

## Esercizio 4

Si ha motivo di ritenere che un nuovo farmaco A abbia la proprietà di abbassare il livello di glicemia nel sangue. In ciascuno dei 6 pazienti diabetici osservati, con lo stesso livello di glicemia iniziale, in trattamento con A da diversi periodi di tempo, sono stai rilevati la diminuzione di glicemia conseguita DG e il periodo di tempo T, in giorni, di durata del trattamento.
Si vuole spiegare, mediante un modello di regressione lineare, l'abbassamento del livello di glicemia (Y) in funzione del tempo (X).

### Soluzione

Tabella di calcolo:

| DG (Y) | T (X) | $(y_i - \bar{y})$ | $(x_i - \bar{x})$ | $(y_i - \bar{y})(x_i - \bar{x})$ | $(x_i - \bar{x})^2$ | $(y_i - \bar{y})^2$ |
|---|---|---|---|---|---|---|
| 15 | 12 | -6 | -7 | 42 | 49 | 36 |
| 20 | 15 | -1 | -4 | 4 | 16 | 1 |
| 18 | 15 | -3 | -4 | 12 | 16 | 9 |
| 25 | 21 | 4 | 2 | 8 | 4 | 16 |
| 22 | 23 | 1 | 4 | 4 | 16 | 1 |
| 26 | 28 | 5 | 9 | 45 | 81 | 25 |
| **126** | **114** | | | **115** | **182** | **88** |

$$\mu_x = \frac{114}{6} = 19 \qquad \mu_y = \frac{126}{6} = 21 \quad \sum_{1=1}^{n} x_i^{\,2} = 2348 \qquad \sum_{1=1}^{n} y_i^{\,2} = 2734$$

$$\sigma^2{}_x = VAR(X) = E[X^2] - \mu^2 = \frac{2348}{6} - 19^2 = 30{,}34 \rightarrow \sigma = \sqrt{30{,}34} = 5{,}5$$

$$\sigma^2{}_y = VAR(YX) = E[Y^2] - \mu^2 = \frac{2734}{6} - 21^2 = 14{,}66 \rightarrow \sigma = \sqrt{14{,}66} = 3{,}83$$

$$\mu(x \cdot y) = (15 \cdot 12 + 20 \cdot 15 + \ldots\ldots + 26 \cdot 28/6) = 2509/5 = 418{,}167$$

$Cov_{x,y} = \mu(x \cdot y) - (\mu_x \cdot \mu_y) = 418{,}167 - (19 \cdot 21) = 19{,}167$

$Corr_{x,y} = \rho_{x,y} = \frac{Cov_{x,y}}{\sigma_x \cdot \sigma_y} = \frac{19{,}167}{5{,}5 \cdot 3{,}83} = 0{,}91 \quad \rho^2 = 0{,}91^2 = 0{,}826$

STIMA DELLA RETTA DI REGRESSIONE

$\hat{y} = a + b_1 x_i$

$b = \frac{Cov_{x,y}}{\sigma^2{}_x} = \frac{19{,}167}{30{,}34} = 0{,}632$

$a = \overline{y} - b\overline{x} = 21 - (0{,}632 \cdot 19) = 9$

$\hat{y} = a + bx_i = 9 + 0{,}632 \cdot x_i$

Quando si costruisce un modello di regressione l'obiettivo è quello di spiegare le variazioni della variabile dipendente (Y) mediante le variazioni della variabile esplicativa (X). Maggiore è la percentuale della varianza della Y che si riesce a spiegare con la variabile X, più soddisfacente sarà il modello. L'informazione della percentuale della varianza di Y spiegata dal modello di regressione è fornita dall'indice di determinazione $R^2$, che varia tra 0 e 1. Esso è dato dal rapporto tra devianza spiegata e devianza totale del modello.

$R^2 = \frac{DevSpiegata}{DevTotale} = \frac{D(R)}{D(Y)}$

$DevTotale = D(Y) = \sum (y_i - \overline{y})^2 = 88$

$Dev\,\mathrm{Re}\,sidua = D(E) = \sum (y_i - \overline{y})^2 - (1 - \rho^2) = 88(1 - 0{,}826) = 15{,}312$

$DevSpiegata = D(R) = D(Y) - D(E) = 88 - 15{,}312 = 72{,}688$

$R^2 = \frac{D(R)}{D(Y)} = \frac{72{,}688}{88} = 0{,}826$

Dal risultato di può notare come $R^2 = \rho^2$ .

Inoltre, è possibile ottenere $R^2 = 1 - \frac{D(E)}{D(Y)} = 1 - \frac{15{,}132}{88} = 0{,}826$

Costruire un test sulla pendenza (β), coefficiente che misura l'effetto che una variazione unitaria della variabile esplicativa X produce sulla variabile dipendente Y.

Di regola, dopo aver stimato il modello di regressione si sottopone ad ipotesi nulla:

$$H_0 = \beta = 0$$

In tal caso (β=0), il valore atteso della Y è costante e pari a $\hat{\beta}_0$ per qualsiasi valore di X. Ciò implica l'assenza di un legame in media tra Y ed X e di conseguenza il modello di regressione è inutile.

Supponiamo che la mia ipotesi alternativa sia:

$$H_1 = \beta \neq 0$$

La statistica test è data da:

$$t = \frac{b}{s / \sqrt{\left(x_i - \bar{x}\right)^2}} = \frac{0{,}632}{1{,}96 / \sqrt{182}} = \frac{0{,}632}{0{,}145} = 4{,}35$$

Essendo $t_{n-2,\alpha/2} = t_{4,0,025} = 2{,}776$, respingiamo $H_0$ e concludiamo che la somministrazione del nuovo farmaco effettivamente incide sul livello di glicemia.

Ora costruiamo un test sul coefficiente di correlazione $\rho$.

Sottoponiamo a test le ipotesi:

$H_0 : R^2 = 0$ vs $H_1 : R^2 > 0$

Sotto l'ipotesi nulla la statistica test

$$Xtest = \frac{R^2 (n-2)}{(1-R^2)}$$

Ha una distribuzione *t* di Student con n-2 gradi di libertà e il valore osservato della statistica è dato da:

$$Vtest == \frac{0,826(4)}{1-0,826} = \frac{3,304}{0,174} = 18,98$$

Per n=6 la statistica test si distribuisce come una v.c $F_{1,\, n-2}$ Al livello di significatività $\alpha$=0,05, il valore critico risulta $F_{0,05,1,4} = 7,71$

Anche in questo caso rifiutiamo $H_0$ e concludiamo che esiste una relazione lineare tra il tempo di somministrazione del nuovo farmaco e il livello di glicemia nel sangue.