# PROVA SCRITTA DI STATISTICA (COD 4038-5047-371)
## 8 aprile 2005
# SOLUZIONI MODALITÀ (A)

**Esercizio 1**. (**4 punti)** Vengono osservati 100 clienti dell'Agenzia n. 13 della banca B1 Spa, rilevando il numero $X$ di utilizzi della carta di credito in un certo periodo. I valori osservati $x_1,\ldots,x_{100}$ danno luogo alle seguenti sintesi:

$$\sum_{i=1}^{100} x_i = 4000 \qquad \sum_{i=1}^{100} x_i^2 = 200000 .$$

a) Si determini lo scarto quadratico medio del numero di utilizzi. (1 punto)
b) Utilizzando la disuguaglianza di Cebishev, si forniscano opportune informazioni sulla frequenza dei clienti che hanno effettuato un numero di utilizzi compreso tra 15 e 65. (2 punti)
c) Per 120 clienti dell'Agenzia n. 17 della stessa banca, si è rilevato, nello stesso periodo, un numero medio di utilizzi della carta di credito pari a 150 con uno scarto quadratico medio pari a 100. Si dica, motivando la risposta, in quale delle due Agenzie il numero di utilizzi della carta di credito presenta la maggiore variabilità. (1 punto)

**Soluzione**

**a)** $\boxed{\sigma_X = \sqrt{\frac{200000}{100} - \left(\frac{4000}{100}\right)^2} = \sqrt{2000-1600} = \sqrt{400} = 20}$

**b)** $\boxed{Fr\{15 \le X \le 65\} = Fr\{40-25 \le X \le 40+25\} \ge 1 - \frac{20^2}{25^2} = 1 - \left(\frac{4}{5}\right)^2 = \frac{9}{25} = 0.36}$

**c)** $c.v.(Y) = \frac{100}{150} = \frac{2}{3} = 0.6667 > c.v.(X) = \frac{20}{40} = \frac{1}{2} = 0.5$

$\Rightarrow$ $Y =$ "n. utilizzi Ag. 17" presenta maggiore variabilità di $X =$ "n. utilizzi Ag. 13"

**Esercizio 2. (4 punti)** Data la seguente distribuzione congiunta di due caratteri *X* e *Y*,

| ***X* \ *Y*** | **-1** | **0** | **1** |
|---|---|---|---|
| **0** | 0.22 | 0 | 0.3 |
| **1** | 0.13 | 0.22 | 0 |
| **10** | 0 | 0 | 0.13 |

a) Si determini la funzione di regressione di *Y* su *X.* (1 punto)
b) Si stabilisca se *Y* è regressivamente indipendente (o indipendente in media) da *X*, giustificando la risposta. (1 punto)
c) Si determini la retta di regressione di *Y* su *X* e se ne valuti la bontà di adattamento sulla base di un opportuno indice. (2 punti)

**Soluzione**

**a)** $$M(Y/X=x) = \begin{cases} -1\times\frac{0.22}{0.52} + 1\times\frac{0.3}{0.52} = 0.1538 & x=0 \\ -1\times\frac{0.13}{0.35} + 0\times\frac{0.22}{0.35} = -0.3714 & x=1 \\ 1\times\frac{0.13}{0.13} = 1 & x=10 \end{cases}$$

**b)** $M(Y/X=x)$ non è costante $\Rightarrow$ $Y$ non è regressivamente indipendente da $X$

**c)** $\hat{y} = \hat{b}x + \hat{a} = 0.0977x - 0.0812,\ \rho^2(X,Y) = 0.1310 \Rightarrow$ bontà di adattamento bassa

calcoli intermedi:

$M(X)=1.65, M(Y)=0.08, M(XY)=1.17, Cov(X,Y)=1.038, V(X)=10.628, V(Y)=0.7736$

**Esercizio 3. (5 punti)** Ciascuno di due contratti finanziari "A" e "B", con la stessa scadenza, può andare a buon fine o può non andare a buon fine. La probabilità che il contratto A vada a buon fine è pari a 0.7; la probabilità che il contratto B vada a buon fine è pari a 0.6; la probabilità che entrambi i contratti vadano a buon fine è infine pari a 0.4.

a) Si calcoli la probabilità che almeno uno dei due contratti finanziari vada a buon fine. (1 punto)

b) Si determini la distribuzione della variabile aleatoria $X$ = "numero di contratti finanziari a buon fine". (2 punti)

c) Considerate due variabili casuali qualsiasi $X$ e $Y$, si dimostri la relazione:

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\,\text{Cov}(X,Y). \qquad \text{(2 punti)}$$

**Soluzione**

**a)** $P(\text{almeno un contratto a buon fine}) = P(A \cap \bar{B}) + P(\bar{A} \cap B) + P(A \cap B) = 0.9$
dove: $P(A \cap \bar{B}) = P(A) - P(A \cap B) = 0.7 - 0.4 = 0.3$,
$P(\bar{A} \cap B) = P(B) - P(A \cap B) = 0.6 - 0.4 = 0.2$

**b)** $p_X(x) = \begin{cases} 0.1 & x = 0 \\ 0.5 & x = 1 \\ 0.4 & x = 2 \\ 0 & \text{altrove} \end{cases}$

**c)** $V(X+Y) = E\{[(X+Y) - E(X+Y)]^2\} = E\{[(X - E(X)) + (Y - E(Y))]^2\} =$
$= E\{(X - E(X))^2 + (Y - E(Y))^2 + 2(X - E(X))(Y - E(Y))\} =$
$= V(X) + V(Y) + 2Cov(X,Y)$

**Esercizio 4. (5 punti)** Sia data una popolazione $X$ gaussiana (o Normale) con valore atteso $\mu$ e varianza $\sigma^2$=9. Sia inoltre $(X_1, \ldots, X_4)$ un campione bernoulliano di ampiezza $n$=4 estratto dalla suddetta popolazione.

a) Si calcoli la seguente probabilità: $\Pr\{\bar{X}_4 > \mu + 0.5\}$, ossia la probabilità che la media campionaria $\bar{X}_4$ sovrastimi il parametro $\mu$ per più di 0.5. (2 punti)

b) Considerato lo stimatore $T_4 = \frac{2}{3}\bar{X}_4 - \frac{2}{3}$, si stabilisca se è non distorto per $\mu$; si calcoli inoltre la varianza dello stimatore. (2 punti)

c) Si fornisca la definizione di stimatore consistente (in senso forte o in media quadratica) di un generico stimatore per un parametro $\theta$. (1 punto)

**Soluzione**

**a)** $P(\bar{X}_4 > \mu + 0.5) = P\left(\frac{\bar{X}_4 - \mu}{\sqrt{9/4}} > \frac{\mu + 0.5 - \mu}{\sqrt{9/4}}\right) = P\left(N(0,1) > \frac{1}{3}\right) = 1 - \Phi(0.33) = 1 - 0.6293 = 0.3707$

**b)** $E(T_4) = \frac{2}{3}E(\bar{X}_4) - \frac{2}{3} = \frac{2}{3}\mu - \frac{2}{3} = \frac{2}{3}(\mu - 1) \neq \mu;\quad V(T_4) = \frac{4}{9}V(\bar{X}_4) = \frac{4}{9} \times \frac{9}{4} = 1$

**c)** uno stimatore $T_n$ è consistente per $\theta$ se: $\lim_{n\to\infty} E\left[(T_n - \theta)^2\right] = 0 \ \forall\theta$
ovvero, uno stimatore $T_n$ è consistente per $\theta$ se: $(a)$ $\lim_{n\to\infty} E(T_n) = \theta \ \forall\theta$, $(b)$ $\lim_{n\to\infty} V(T_n) = 0 \ \forall\theta$

**Esercizio 5.** **(7 punti)**

E' stato selezionato un campione di $n = 100$ studenti di una certa università che hanno fatto scambi internazionali e si è chiesto loro se giudicano positivamente la loro esperienza: 80 studenti intervistati hanno dato risposta affermativa. Si indichi ora con $p$ la proporzione (incognita) della popolazione di studenti scambisti che giudicano positivamente la loro esperienza.

a) Si indichi quale distribuzione può essere specificata per la popolazione. (1 punto)
b) Utilizzando uno stimatore non distorto, si determini una stima puntuale di $p$. (1 punto)
c) Si determini l'intervallo di confidenza al 99% per $p$. (2 punti)
d) Si scriva la regione di rifiuto per le ipotesi $H_0 : p \le 0.7$ contro $H_1 : p > 0.7$ al livello di significatività $\alpha = 0.005$. Si decida inoltre a favore o contro $H_0$. (2 punti)
e) Sapendo che il p-value del test per verificare $H_0 : p = 0.75$ contro $H_1 : p \ne 0.75$ risulta pari a 0.3174, si decida a favore o contro $H_0$, fissando $\alpha = 0.005$. (1 punto)

**Soluzione**

**a)** la popolazione è bernoulliana di parametro $p$

**b)** stimatore non distorto per $p$: $\bar{X}_{100} \Rightarrow$ stima puntuale di $p$: $\hat{p} = \bar{x}_{100} = \frac{80}{100} = 0.8$

**c)** $IC_{0.99}(p) = (0.697, 0.903)$, calcoli intermedi:

$$\hat{p} \mp z_{.995}\sqrt{\frac{\hat{p}(1-\hat{p})}{100}} = 0.8 \mp 2.576\sqrt{\frac{0.8\times 0.2}{100}} = 0.8 \mp 2.576\frac{0.4}{10} = 0.8 \mp 2.576\frac{0.4}{10} = 0.8 \mp 0.1030$$

**d)** Regione di rifiuto: $\bar{X}_{100} \ge 0.7 + 2.576\sqrt{\frac{0.7\times 0.3}{100}} = 0.7 + 2.576\frac{0.4583}{10} = 0.7 + 0.1180 = 0.818$

decisione: non si rifiuta $H_0$ perché $\bar{x}_{100} = 0.8 < 0.818$

**e)** decisione: non si rifiuta $H_0$ perché $p - value = 0.3174 > \alpha = 0.005$

**Esercizio 6.** **(2 punti)**

La tabella seguente riporta parte dell'output di EXCEL relativo al modello lineare $Y = \beta_0 + \beta_1 x + \varepsilon$. Utilizzando i dati forniti, si determini l'intervallo di confidenza per $\beta_1$ al 90%.

| *Statistica della regressione* | |
|---|---|
| R multiplo | 0,715995625 |
| R al quadrato | 0,512649735 |
| R al quadrato corretto | 0,463914708 |
| Errore standard | 17,52229057 |
| Osservazioni | 12 |

ANALISI VARIANZA

| | *gdl* | *SQ* | *MQ* |
|---|---|---|---|
| Regressione | 1 | 3229,69333 | 3229,69333 |
| Residuo | 10 | 3070,30667 | 307,030667 |
| Totale | 11 | 6300 | |

| | *Coefficienti* | *Errore standard* | *Stat t* |
|---|---|---|---|
| Intercetta | -72,87347201 | 53,54087015 | -1,361081204 |
| Variabile X | 25,78597469 | 7,950488007 | 3,24331974 |

$IC_{0.90}(\beta_1) = (11.38, 40.192)$, calcoli intermedi:

$25.786 \mp t_{0.95}^{10}\, 7.9505 = 25.786 \mp 1.812 \times 7.9505 = 25.786 \mp 14.406$

**PROVA SCRITTA DI STATISTICA (COD 4038-5047-371)**
**8 aprile 2005**

# SOLUZIONI MODALITÀ (B)

__Esercizio 1__. (**4 punti**) Vengono osservati 150 clienti dell'Agenzia n. 17 della banca B2 Spa, rilevando il numero $X$ di utilizzi della carta di credito in un certo periodo. I valori osservati $x_1,\dots,x_{150}$ danno luogo alle seguenti sintesi:

$$\sum_{i=1}^{150} x_i = 6000 \qquad \sum_{i=1}^{150} x_i^2 = 300000 .$$

a) Si determini lo scarto quadratico medio del numero di utilizzi. (1 punto)
b) Utilizzando la disuguaglianza di Cebishev, si forniscano opportune informazioni sulla frequenza dei clienti che hanno effettuato un numero di utilizzi compreso tra 10 e 70. (2 punti)
c) Per 100 clienti dell'Agenzia n. 13 della stessa banca, si è rilevato, nello stesso periodo, un numero medio di utilizzi della carta di credito pari a 150 con uno scarto quadratico medio pari a 50. Si dica, motivando la risposta, in quale delle due Agenzie il numero di utilizzi della carta di credito presenta la maggiore variabilità. (1 punto)

__Soluzione__

**a)** $\boxed{\sigma_X = \sqrt{\frac{300000}{150} - \left(\frac{6000}{150}\right)^2} = \sqrt{2000-1600} = \sqrt{400} = 20}$

**b)** $\boxed{Fr\{10 \le X \le 70\} = Fr\{40-30 \le X \le 40+30\} \ge 1 - \frac{20^2}{30^2} = 1 - \left(\frac{2}{3}\right)^2 = \frac{5}{9} = 0.5556}$

**c)** $c.v.(Y) = \frac{50}{150} = \frac{1}{3} = 0.3333 < c.v.(X) = \frac{20}{40} = \frac{1}{2} = 0.5$
$\Rightarrow$ $Y =$"n. utilizzi Ag. 13" presenta minore variabilità di $X =$"n. utilizzi Ag. 17"

__Esercizio 2. (4 punti)__ Data la seguente distribuzione congiunta di due caratteri $X$ e $Y$,

| **X \ Y** | **-1** | **0** | **1** |
|---|---|---|---|
| **0** | 0.22 | 0 | 0.3 |
| **1** | 0.13 | 0.22 | 0 |
| **10** | 0 | 0 | 0.13 |

a) Si determini la funzione di regressione di $X$ su $Y$. (1 punto)
b) Si stabilisca se $X$ è regressivamente indipendente (o indipendente in media) da $Y$, giustificando la risposta. (1 punto)
c) Si misuri la connessione tra i due caratteri, mediante un opportuno indice. (2 punti)

__Soluzione__

**a)** $$M(X/Y=y) = \begin{cases} 0\times\frac{0.22}{0.35} + 1\times\frac{0.13}{0.35} = 0.3714 & y=-1 \\ 1\times\frac{0.22}{0.22} = 1 & y=0 \\ 0\times\frac{0.3}{0.43} + 10\times\frac{0.13}{0.43} = 3.0233 & y=1 \end{cases}$$

**b)** $M(X/Y=y)$ non è costante $\Rightarrow$
$X$ non è regressivamente indipendente da $Y$

**c)** $\tilde{\varphi}^2 = \frac{\varphi^2}{2} = 03687,$ calcoli intermedi:

$$\varphi^2 = \sum_{\forall(i,j)} \frac{\left(p_{ij} - p_i p_j\right)^2}{p_i p_j} = \begin{bmatrix} 0.0079+ & 0.1144+ & 0.0261+ \\ 0.0005+ & 0.2656+ & 0.1505+ \\ 0.0455+ & 0.0286+ & 0.0982+ \end{bmatrix} = 0.7373$$

**Esercizio 3. (5 punti)** Ciascuno di due contratti finanziari "A" e "B", con la stessa scadenza, può andare a buon fine o può non andare a buon fine. La probabilità che il contratto A vada a buon fine è pari a 0.7; la probabilità che il contratto B vada a buon fine è pari a 0.6; la probabilità che entrambi i contratti vadano a buon fine è infine pari a 0.4.

a) Si calcoli la probabilità che almeno uno dei due contratti finanziari *non* vada a buon fine. (1 punto)

b) Si determini la distribuzione della variabile aleatoria $X$ = "numero di contratti finanziari *non* a buon fine". (2 punti)

c) Considerate due variabili casuali qualsiasi $X$ e $Y$, si dimostri la relazione:

$$\mathrm{Var}(X-Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)-2\,\mathrm{Cov}(X,Y)\,. \qquad \text{(2 punti)}$$

**Soluzione**

**a)** $P\left(\text{almeno un contratto non a buon fine}\right)=P\left(A\cap\overline{B}\right)+P\left(\overline{A}\cap B\right)+P\left(\overline{A}\cap\overline{B}\right)=0.6$

dove: $P\left(A\cap\overline{B}\right)=P(A)-P(A\cap B)=0.7-0.4=0.3$,

$P\left(\overline{A}\cap B\right)=P(B)-P(A\cap B)=0.6-0.4=0.2$, $P\left(\overline{A}\cap\overline{B}\right)=1-(0.3+0.2+0.4)=0.1$

**b)** $$p_X(x)=\begin{cases}0.4 & x=0\\ 0.5 & x=1\\ 0.1 & x=2\\ 0 & \text{altrove}\end{cases}$$

**c)** $V(X-Y)=E\{[(X-Y)-E(X-Y)]^2\}=E\{[(X-E(X))-(Y-E(Y))]^2\}=$

$=E\{(X-E(X))^2+(Y-E(Y))^2-2(X-E(X))(Y-E(Y))\}=$

$=V(X)+V(Y)-2Cov(X,Y)$

**Esercizio 4. (5 punti)** Sia data una popolazione $X$ gaussiana (o Normale) con valore atteso $\mu$ e varianza $\sigma^2$=25. Sia inoltre $(X_1,\dots,X_5)$ un campione bernoulliano di ampiezza $n$=5 estratto dalla suddetta popolazione.

a) Si calcoli la seguente probabilità: $\Pr\left\{\overline{X}_5>\mu+1\right\}$, ossia la probabilità che la media campionaria $\overline{X}_5$ sovrastimi il parametro $\mu$ per più di 1. (2 punti)

b) Considerato lo stimatore $T_5=\frac{3}{2}\overline{X}_5+\frac{3}{2}$, si stabilisca se è non distorto per $\mu$; si calcoli inoltre la varianza dello stimatore. (2 punti)

c) Si fornisca la definizione di stimatore consistente (in senso forte o in media quadratica) di un generico stimatore per un parametro $\theta$. (1 punto)

**Soluzione**

**a)** $P\left(\overline{X}_5>\mu+1\right)=P\left(\frac{\overline{X}_5-\mu}{\sqrt{25/5}}>\frac{\mu+1-\mu}{\sqrt{25/5}}\right)=P\left(N(0,1)>\frac{1}{\sqrt{5}}\right)=1-\Phi(0.45)=1-0.6736=0.3264$

**b)** $E(T_5)=\frac{3}{2}E\left(\overline{X}_5\right)+\frac{3}{2}=\frac{3}{2}\mu+\frac{3}{2}=\frac{3}{2}(\mu+1)\neq\mu;\ \ V(T_5)=\frac{9}{4}V\left(\overline{X}_5\right)=\frac{9}{4}\times\frac{25}{5}=\frac{45}{4}=11.25$

**c)** uno stimatore $T_n$ è consistente per $\theta$ se: $\lim_{n\to\infty}E\left[\left(T_n-\theta\right)^2\right]=0\ \forall\theta$

ovvero, uno stimatore $T_n$ è consistente per $\theta$ se: $(a)$ $\lim_{n\to\infty}E(T_n)=\theta\ \forall\theta$, $(b)$ $\lim_{n\to\infty}V(T_n)=0\ \forall\theta$

**<u>Esercizio 5.</u> (7 punti)**

E' stato selezionato un campione di $n = 100$ studenti di una certa università che hanno fatto scambi internazionali e si è chiesto loro se giudicano positivamente la loro esperienza: 60 studenti intervistati hanno dato risposta affermativa. Si indichi ora con $p$ la proporzione (incognita) della popolazione di studenti scambisti che giudicano positivamente la loro esperienza.

*a)* Si indichi quale distribuzione può essere specificata per la popolazione. (1 punto)
*b)* Utilizzando uno stimatore non distorto, si determini una stima puntuale di *p.* (1 punto)
c) Si determini l'intervallo di confidenza al 99% per *p*. (2 punti)
d) Si scriva la regione di rifiuto per le ipotesi $H_0 : p \geq 0.7$ contro $H_1 : p < 0.7$ al livello di significatività $\alpha = 0.005$. Si decida inoltre a favore o contro $H_0$. (2 punti)
e) Sapendo che il p-value del test per verificare $H_0 : p = 0.75$ contro $H_1 : p \neq 0.75$ risulta pari a 0.0006, si decida a favore o contro $H_0$, fissando $\alpha = 0.005$. (1 punto)

**<u>Soluzione</u>**

**a)** $\boxed{\text{la popolazione è bernoulliana di parametro } p}$

**b)** $\boxed{\text{stimatore non distorto per } p\text{: } \bar{X}_{100} \Rightarrow \text{stima puntuale di } p\text{: } \hat{p} = \bar{x}_{100} = \frac{60}{100} = 0.6}$

**c)** $IC_{0.99}(p) = (0.4738, 0.7262)$, calcoli intermedi:

$$\hat{p} \mp z_{.995}\sqrt{\frac{\hat{p}(1-\hat{p})}{100}} = 0.6 \mp 2.576\sqrt{\frac{0.6 \times 0.4}{100}} = 0.6 \mp 2.576\frac{\sqrt{0.24}}{10} = 0.6 \mp 2.576\frac{0.4899}{10} = 0.6 \mp 0.1262$$

**d)** Regione di rifiuto: $\bar{X}_{100} \leq 0.7 - 2.576\sqrt{\frac{0.7 \times 0.3}{100}} = 0.7 - 2.576\frac{0.4583}{10} = 0.7 - 0.1180 = 0.582$

decisione: non si rifiuta $H_0$ perché $\bar{x}_{100} = 0.6 > 0.582$

**e)** $\boxed{\text{decisione: si rifiuta } H_0 \text{ perché } p - value = 0.0006 < \alpha = 0.005}$

**<u>Esercizio 6.</u> (2 punti)**

La tabella seguente riporta parte dell'output di EXCEL relativo al modello lineare $Y = \beta_0 + \beta_1\, x + \varepsilon$. Utilizzando i dati forniti, si determini l'intervallo di confidenza per $\beta_1$ al 90%.

| *Statistica della regressione* | |
|---|---|
| R multiplo | 0,414566 |
| R al quadrato | 0,171865 |
| R al quadrato corretto | 0,05356 |
| Errore standard | 3,334768 |
| Osservazioni | 9 |

ANALISI VARIANZA

| | *gdl* | *SQ* | *MQ* |
|---|---|---|---|
| Regressione | 1 | 16,15528 | 16,15528 |
| Residuo | 7 | 77,84472 | 11,12067 |
| Totale | 8 | 94 | |

| | *Coefficienti* | *Errore standard* | *Stat t* |
|---|---|---|---|
| Intercetta | 29,36025 | 11,96609 | 2,453621 |
| Variabile X | -0,95031 | 0,788449 | -1,20529 |

$IC_{0.90}(\beta_1) = (-2.4443, 0.5437)$, calcoli intermedi:

$-0.9503 \mp t_{0.95}^{7}\, 0.7884 = -0.9503 \mp 1.895 \times 0.7884 = -0.9503 \mp 1.494$