CORSO DI STATISTICA (parte 2) - ESERCITAZIONE 3

Dott.ssa Antonella Costanzo

a.costanzo@unicas.it

Esercizio 1. La variabile casuale normale

Da un analisi di bilancio è emerso che, durante i giorni feriali (lunedì-sabato), l'incasso giornaliero (in migliaia di euro) di un certo supermercato può essere rappresentato attraverso una variabile casuale normale con media 20 e deviazione standard pari a 2. Per un qualunque giorno feriale,

a) calcolare la probabilità che l'incasso del supermercato sia inferiore a 19.000 euro;
b) calcolare la probabilità che l'incasso giornaliero sia compreso tra 20.000 e 21.500 euro.

Durante una settimana lavorativa da 6 giorni feriali, supponendo che gli incassi giornalieri siano tra loro indipendenti,
c) calcolare la probabilità che almeno 5 volte l'incasso giornaliero sia superiore a 19.000 euro;
d) calcolare la media e la varianza dell'incasso settimanale del supermercato.
e) calcolare la probabilità che l'incasso settimanale sia superiore a 110 mila euro.

Sol.

a) Sia X la variabile casuale che descrive l'incasso giornaliero in migliaia di euro del supermercato nel giorno feriale *i*, dove Xi ~ N(20, 4) con i = 1,2,…,6 (ossia lunedì, martedì,…,sabato). Allora:

$$P(X_i < 19) = P\left(Z < \frac{19-20}{2}\right) = P(Z < -0.5) = 1 - P(Z < 0.5) = 1 - \varphi(0.5) = 1 - 0.6915$$
$$= 0.3085$$

b) probabilità che l'incasso giornaliero sia compreso tra 20.000 e 21.500 euro

$$P(20 < X_i < 21.5) = P(0 < Z < 0.75) = \varphi(0.75) - \varphi(0) = 0.7734 - 0.5 = 0.2734$$

c) probabilità che <u>almeno 5 volte</u> l'incasso giornaliero sia superiore a 19.000 euro.

Per prima cosa calcoliamo la probabilità che l'incasso giornaliero sia superiore a 19.000 come complemento a 1 del risultato in a)

$P(Z > 19) = 1 - P(Z < 19) = 1 - 0.3085 = 0.6915$

Definiamo Y la variabile casuale "numero di volte che l'incasso giornaliero ha superato i 19.000 euro".

$Y \sim Bin(n, p)$ ossia $Y \sim Bin(n = 6, p = 0.6915)$

Ricorda la funzione di distribuzione di probabilità per una v.c. Binomiale $P(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}$

$P(Y \geq 5) = P(Y = 5) + P(Y = 6)$

Calcoliamo i singoli valori:

$$P(Y = 5) = \frac{6!}{5!\,1!}(0.6915)^5(0.3085)^{6-5} = 0.2927$$
$$P(Y = 6) = \frac{6!}{6!\,0!}(0.6915)^6(0.3085)^{6-6} = 0.1093$$

Quindi avremo:

$P(Y \geq 5) = P(Y = 5) + P(Y = 6) = 0.2927 + 0.1093 = 0.4020$

d) media e la varianza dell'incasso settimanale del supermercato.

Definizione = incasso settimanale del supermercato vuol dire la somma dei singoli incassi registrati nei giorni feriali (lunedi-sabato).

Sia S la variabile casuale "Incasso settimanale del supermercato" definita come:

$S = \sum_i X_i$ i=1,2,…,6

$$E(S) = E(\sum_i X_i) = \sum_i E(X_i) = n\mu = 6 * 20 = 120\ (migliaia\ di\ euro)$$
$$Var(S) = Var(\sum_i X_i) = \sum_i Var(X_i) = n * \sigma^2 = 6 * 4 = 24\ (migliaia\ di\ euro)$$

Nota: X_i i.i.d. (indipendenti e identicamente distribuite)

e) probabilità che l'incasso settimanale sia superiore a 110 mila euro. Siccome S è la somma di v.c. i.i.d. allora:

$S \sim N(120,24)$

$$P(S > 110) = P\left(Z > \frac{110 - 120}{4.90}\right) = P(Z > -2.04) = P(Z < 2.04) = \varphi(2.04) = 0.9793$$

Nota: sfrutto la simmetria della v.c. Normale standard per cui $P(Z > -2.04)$ equivale a calcolare $P(Z < 2.04)$

Esercizio 2. Teorema di Chebychev

Il punteggio al test di Statistica (da 0 a 100) riportato da un collettivo di 100 studenti ha una media pari a 70 e una deviazione standard di 5.

1. Qual è la percentuale di studenti che hanno riportato un punteggio al test compreso tra 60 e 80?
2. Qual è la percentuale di studenti che hanno riportato un punteggio superiore o inferiore a 3 volte σ?

Sol.

<u>Per il teorema di Chebychev:</u>

Se una distribuzione di probabilità ha media μ e scarto quadratico medio σ, allora la probabilità che il valore di una variabile casuale con tale distribuzione differisca da μ per più di kσ è inferiore a $\frac{1}{k^2}$. In particolare:

$$P(X < \mu - k\sigma \; e \; X > \mu + k\sigma) \leq \frac{1}{k^2}$$

Oppure, in modo equivalente:

$$P(|X - \mu| > k\sigma) \leq \frac{1}{k^2}$$

Oppure, ponendo $c = k\sigma$

$$P(|X - \mu| > c) \leq \frac{\sigma^2}{c^2}$$

Alternativamente è possibile determinare un limite inferiore alla frequenza con cui un carattere assume valori internamente all'intervallo:

$$P(\mu - k\sigma \leq X \leq \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$

Oppure, in modo equivalente:

$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$

Oppure, ponendo $c = k\sigma$

$$P(|X - \mu| < c) \geq 1 - \frac{\sigma^2}{c^2}$$

Il teorema è vale per tutte le variabili a prescindere dalla distribuzione di probabilità e consente di sfruttare le sole informazioni relative alla media e alla varianza.

Dalla disuguaglianza di Chebychev si deduce che, per qualunque carattere quantitativo, indipendentemente dalla forma della distribuzione:

A. Per k=2 almeno $1 - \frac{1}{2^2}$ ossia il 75% dei dati deve essere contenuto in un intervallo centrato intorno la media di ampiezza 4σ
B. Per k=3, almeno $1 - \frac{1}{3^2}$ ossia il 88.89% dei dati deve essere contenuto in un intervallo centrato intorno la media di ampiezza 6σ
C. Per k=4, almeno $1 - \frac{1}{4^2}$ ossia il 93.75% dei dati deve essere contenuto in un intervallo centrato intorno la media di ampiezza 8σ

Nel nostro caso avremo:

1) $\mu - k\sigma = 60$ da cui $70 - k(5) = 60$ k=2
   $\mu + k\sigma = 80$ da cui $70 + k(5) = 80$ k=2

Per il teorema di Chebichev, il 75% delle osservazioni è compreso nell'intervallo $\mu \pm 2\sigma$ per cui il 75% degli studenti ha ottenuto un punteggio compreso tra 60 e 80

2) La percentuale di studenti che hanno riportato un punteggio superiore o inferiore a 3 volte σ si ottiene

   $70 - 3(5) = 55$
   $70 + 3(5) = 85$

   Per il teorema di Chebychev

   $P(|X - \mu| > k\sigma) \leq \frac{1}{k^2}$ $P(X < 55 \; e \; X > 85) \leq \frac{1}{3^2}$ ossia 11% degli studenti

Esercizio 3. Valore atteso e varianza della somma di due v.c.

Si consideri il tempo di attesa in coda in un fast-food. Sia X il tempo di attesa prima di ordinare e Y il tempo di attesa tra il momento in cui è stato effettuato l'ordine e l'istante in cui si è serviti. Si supponga che le variabili casuali X e Y abbiano un coefficiente di correlazione pari a 0.22 e si considerino le seguenti ulteriori informazioni:

	Valore atteso	**Deviazione standard**
Variabile *X*	3.5 mins	1.5 mins
Variabile *Y*	1.6 mins	0.2 mins

a) Calcolare il valore atteso e la deviazione standard della variabile X+Y

b) Se le due variabili X e Y fossero state indipendenti, il risultato nel punto a) sarebbe stato diverso? Se si, in che modo? Motivare la risposta.

a)

$$E(X+Y) = E(X) + E(Y) = 3.5 + 1.6 = 5.1$$
$$V(X+Y) = V(X) + V(Y) + 2 \cdot \mathrm{cov}(X,Y) = V(X) + V(Y) + 2 \cdot \rho \cdot \sigma_X \cdot \sigma_Y =$$
$$= (1.5)^2 + (0.2)^2 + 2 \cdot 0.22 \cdot 1.5 \cdot 0.2 = 2.25 + 0.04 + 0.6 \cdot 0.22 = 2.29 + 0.132 = 2.422$$

b) Se X e Y sono indipendenti allora la varianza è: $V(X+Y) = V(X) + V(Y) = 2.29$

Quindi la deviazione standard è pari a: $\sqrt{2.29} = 1.5132$

Esercizio 4. Distribuzioni congiunte, marginali, condizionate di due v.c.

Le due variabili casuali X e Y hanno densità congiunta data dalla seguente tabella:

X|Y	**y=-2**	**y=1**	**y=3**
x=-2	0.1	0.1	0.05
x=4	0.2	0.5	0.05

Calcolare le densità marginali di X e Y, E(Y) E(X), E(XY), Var(X), Var(Y), Cov(X,Y), ϱ(X,Y). Trovare inoltre la media di Y condizionata a X=-2

Sol.

X|Y	y=-2	y=1	y=3	Totale
x=-2	0.1	0.1	0.05	**0.25**
x=4	0.2	0.5	0.05	**0.75**
Totale	**0.3**	**0.6**	**0.1**	**1**

Le distribuzioni marginali si presentano come delle variabili casuali ad una dimensione e dunque è possibile calcolare la media di X e di Y semplicemente con:

$$E[X] = \sum_{i=1}^{h} x_i\, p_i$$

$$E[Y] = \sum_{j=1}^{k} y_j\, p_j$$

Nel nostro caso avremo:

$$E[X] = 0.25(-2) * 0.75 * 4 = 2.5$$

$$E[Y] = 0.30(-2) * 0.6 * 1 + 0.1 * 3 = 0.3$$

Il valore atteso della distribuzione congiunta di X e Y è a dato da:

$$E[XY] = \sum_{i=1}^{h} \sum_{j=1}^{k} x_i y_j p_{ij}$$

$$E[XY] = (-2)(-2) * 0.1 + (-2)1 * 0.1 + (-2)3 * 0.05 + 4(-2) * 0.2 + 4(1) * 0.5 + 4(3) * 0.05 = 0.9$$

Le varianze di X e Y sono date rispettivamente da:

$$Var(X) = E(X^2) - [E(X)]^2$$

$$E(X^2) = \sum_{i=1}^{h} x_i^2\, p_i = (-2)^2 * 0.25 + 4^2 * 0.75 = 13$$

$$[E(X)]^2 = 2.5^2 = 6.25$$

Quindi $Var(X) = 13 - 6.25 = 6.75$ da cui $\sigma_X = 2.60$

$$Var(Y) = E(Y^2) - [E(Y)]^2$$

$$E(Y^2) = \sum_{j=1}^{k} y_j^2\, p_j = (-2)^2 * 0.3 + 1^2 * 0.6 + 3^2 * 0.1 = 2.7$$

$[E(Y)]^2 = 0.3^2 = 0.09$

$Var(Y) = 2.7 - 0.09 = 2.61$ da cui $\sigma_Y = 1.62$

La covarianza è espressa come:

$Cov(X, Y) = E[XY] - E[X]E[Y] = 0.9 - 2.5 * 0.3 = 0.15$

Infine, il coefficiente di correlazione:

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \frac{0.15}{2.60 * 1.62} = 0.035$$

Valore atteso di Y|X=-2

$$E(Y|X = -2) = -2\left(\frac{0.1}{0.25}\right) + 1\left(\frac{0.1}{0.25}\right) + 3\left(\frac{0.05}{0.25}\right) = 0.2$$

Esercizio 5. Verifica della condizione di indipendenza

Le tabelle seguenti riportano le distribuzioni marginali di due v.c. discrete X e Y

X	P(X=x)
0	2/8
1	3/8
2	2/8
3	1/8

Y	P(Y=y)
0	1/6
1	3/6
2	2/6

Costruire la distribuzione congiunta di X e Y nell'ipotesi che le due variabili siano indipendenti.

Sol.

Due variabili aleatorie X e Y distribuite congiuntamente sono dette indipendenti se e solo se la loro distribuzione di probabilità congiunta è uguale al prodotto delle loro distribuzioni di probabilità marginali:

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

per tutte le possibili coppie di valori di x e y.

Se due variabili casuali X e Y sono indipendenti e si conoscono le distribuzioni di probabilità marginali allora è nota anche la distribuzione di probabilità congiunta (X,Y). In particolare:

X|Y	**0**	**1**	**2**	**P(X=x)**
0	2/48	6/48	4/48	2/8
1	3/48	9/48	6/48	3/8
2	2/48	6/48	4/48	2/8
3	1/48	3/48	2/48	1/8
P(Y=y)	1/6	3/6	2/6	1

Nota: Due v.c. X e Y sono identicamente distribuite se le distribuzioni di probabilità marginali sono uguali tra loro. In tal caso X e Y sono indipendenti ma non identicamente distribuite.