

Si avvisano gli studenti che durante l'Attività di TUTORATO prevista per il 23/10/17 saranno svolti gli esercizi di seguito riportati e alcuni esercizi di Statistica descrittiva appartenenti ai Temi d'Esame dell'A.A. 2015/16 disponibili nella pagina web del corso.

ESERCIZIO 2 (punti 3)

1/07/2016

A

Il centro studi della banca sta indagando sui comportamenti d'investimento dei propri clienti. Per un campione di 130 clienti si sono rilevate quindi X, l'età (in anni), e Y, la percentuale (sul totale del patrimonio affidato in gestione alla banca) investita in attività classificate come a basso rischio. Alcuni dati di sintesi sono riportati qui di seguito:

$$\bar{x} = 33.5214 \quad \bar{y} = 48.2784$$

$$\sum_{i=1}^{130} (x_i - \bar{x})^2 = 16196.12 \quad \sum_{i=1}^{130} (y_i - \bar{y})^2 = 21098.3 \quad \sum_{i=1}^{130} (x_i - \bar{x})(y_i - \bar{y}) = -15616.21$$

Si stimino i coefficienti della retta di regressione lineare di Y su X.

ES. 10

Si consideri un campione di 6 circoli estratti dalla popolazione descritta nell'esercizio precedente. I dati relativi alle variabili X e Y per i 6 circoli estratti sono riportati nel seguito:

X	5	6	6	10	5	7
Y	145	182	140	280	178	257

$$\sigma_X^2 = 2,92$$

$$\sigma_Y^2 = 2'838$$

- a) Si calcoli la covarianza campionaria di X e Y.
- b) Utilizzando un modello lineare, si vuole prevedere il numero di soci mediante il numero di campi di cui il circolo dispone. Si costruisca il modello lineare di Y su X e si preveda quindi il numero di soci di un circolo con 7 campi da tennis. Per comodità, vengono fornite le varianze (campionarie) di X e Y: $s_X^2 = 3.5$, $s_Y^2 = 3405.6$.
- c) È lecito prevedere, mediante un modello lineare, il numero di soci di un circolo con 18 campi da tennis? Si giustifichi la risposta.

ESERCIZIO 4 (punti 5)

15/01/16

Per un campione di 10 negozi di una catena di retail italiana si sono rilevati i seguenti dati relativi al fatturato di ogni negozio (in migliaia di euro) realizzato nel 2014 e il numero di scontrini emessi (in migliaia) nel medesimo anno:

ID negozio	Fatturato	Numero scontrini	Fatturato ²	(Numero scontrini) ²	(Fatturato) × (Numero scontrini)
1	1700	41	2890000	1681	69700
2	3300	87	10890000	7569	287100
3	480	20	230400	400	9600
4	1130	15	1276900	225	16950
5	1500	48	2250000	2304	72000
6	455	12	207025	144	5460
7	2430	74	5904900	5476	179820
8	220	7	48400	49	1540
9	1455	41	2117025	1681	59655
10	1080	35	1166400	1225	37800
Totale	13750	380	26981050	20754	739625

- a) Attraverso un opportuno indice, si valuti il grado di associazione lineare tra il fatturato e il numero di scontrini e fornire un commento sulla base del risultato ottenuto.
- b) Sulla base di questi dati, il management della catena intende formulare delle previsioni per il fatturato di alcuni negozi che apriranno a inizio del prossimo anno. A tal fine si procede alla stima dei coefficienti del modello di regressione lineare tra Y = "Fatturato" e X = "Numero scontrini". Si riportino le stime dei coefficienti di regressione.

ES. 4

Su 9 modelli di automobile vengono rilevate alcune caratteristiche, di seguito specificate:

- X: "tipo carburante" (0 = benzina, 1 = gasolio),
- Y: "potenza in Kw",
- Z: "consumo in Km/litro",
- W: "numero posti per cui è omologata l'auto",
- T: "numero porte".

I dati rilevati sono riportati nella seguente tabella.

Modello	X	Y	Z	W	T
A	0	80	11.4	5	5
B	0	98	11	5	4
C	1	76	16.5	5	4
D	1	125	14	7	5
E	1	178	10	4	3
F	0	65	15	5	5
G	1	78	17.2	7	5
H	1	54	18.3	5	3
I	0	74	12.5	4	3

Si riportano, per agevolare i calcoli, le seguenti quantità di sintesi:

$$\sum y_i = 828; \sum y_i^2 = 87790; \sum w_i = 47; \sum w_i^2 = 255; \sum z_i = 125.9; \sum z_i^2 = 1831.19;$$

$$\sum z_i w_i = 669.4; \sum y_i z_i = 11003.8.$$

- a) Si dica se la distribuzione di frequenze della potenza è da considerarsi simmetrica o obliqua (a destra o a sinistra).
- b) Si dica se, mediamente, hanno consumi più elevati le auto a benzina o quelle a gasolio (si ricordi che il consumo è espresso in Km/litro).
- c) Si dica, giustificando la risposta, se per spiegare la variabilità del consumo attraverso un modello lineare, sia più opportuno utilizzare il numero di posti o la potenza.
- d) Si dica, giustificando la risposta, se i caratteri W e T sono indipendenti.
- e) Si disegni un grafico opportuno per rappresentare il numero di posti (carattere W).

ES. 42

Durante la stagione invernale gli autisti di un'azienda che effettua servizio di trasporto urbano hanno ricevuto un timer per registrare i tempi effettivi di percorrenza su certe tratte ed in certe fasce orarie dei giorni feriali. Dall'elaborazione del campione è stata ricavata la seguente tabella: la parte sinistra riguarda la distribuzione di frequenze, la parte destra alcune statistiche. Ciascun tempo di percorrenza è stato espresso come "scostamento percentuale rispetto al tempo di percorrenza fissato dall'orario ufficiale".

Cond. meteo	X = Scostamento % rispetto all'orario ufficiale						Statistiche	
	[-10, 0)	[0, 10)	[10, 25)	[25, 50)	[50, 100)	[100, 200)	Media	Std dev
Pioggia	16	25	43	26	11	20	59.3	31.4
No pioggia	60	186	122	111	72	38	131	47.6
Totale	76	211	165	137	83	58		

- a) Fornire una rappresentazione grafica della distribuzione (marginale) di X e commentare la forma della distribuzione.
- b) Stabilire se vi sia maggiore variabilità dei tempi di percorrenza in caso di pioggia.
- c) Ricavare quanti viaggi, in percentuale, hanno una durata che si discosta di oltre il 25% dall'orario ufficiale.