

# ESERCIZI

## COMPITO 1

- 1) La tabella che segue mostra la distribuzione di un campione di soggetti in base al reddito percepito (in migliaia di euro) e al grado di soddisfazione nel rapporto con il coniuge:

A/B	Molto insoddisfatto	Poco insoddisfatto	Poco soddisfatto	Molto soddisfatto
<10	20	24	80	82
10-20	22	38	104	125
20-30	13	28	81	113
>30	7	18	54	92

- a) Si ricavino le distribuzioni condizionate della variabile reddito data la variabile “grado di soddisfazione”;
- b) Si determini un indice statistico che indichi l’eventuale dipendenza della variabile reddito dalla variabile “grado di soddisfazione”.
- 2) I seguenti dati costituiscono un campione di 20 tempi (espressi in minuti), impiegati per risolvere analoghi problemi, riferiti a due diversi uffici di una compagnia telefonica:

I ufficio

1.48	1.75	0.78	2.85	0.52	1.60	4.15	3.97	1.48	3.10
1.02	0.53	0.93	1.60	0.80	1.05	6.32	3.93	5.45	0.97

II ufficio

7.55	3.75	0.10	1.10	0.60	0.52	3.30	2.10	0.58	4.02
3.75	0.65	1.92	0.60	1.53	4.23	0.08	1.48	1.65	0.72

- a) Costruite, per ciascuno dei due uffici, il diagramma scatola e baffi;
- b) Basandosi sul precedente grafico, la distribuzione dei dati è da ritenere asimmetrica? Se sì, di che tipo di asimmetria si tratta?
- c) Calcolati opportuni indici di sintesi, si può ritenere che ci siano differenze tra i due uffici?
- 3) Cosa misura la covarianza e fra quali valori può variare?
- 4) Quali sono le variabili statistiche più informative e perché?

## COMPITO 2

- 1) Un gruppo di 76 studenti è stato sottoposto ad un test attitudinale per l'ammissione ad un corso di studi. Viene riportata la distribuzione di frequenze del *Numero di errori commessi* su un totale di 10 domande:

X	frequenze
0	1
1	2
2	5
3	8
4	14
5	20
6	13
7	6
8	4
9	2
10	1

- a) Stabilire, attraverso l'uso di un indice appropriato, se tale distribuzione può ritenersi simmetrica;
- b) Individuare il modello teorico più opportuno da adattare alla distribuzione osservata e verificarne l'adattamento.
- 2) In una città sono stati osservati giornalmente la *Condizione meteorologica* e il *Livello di traffico automobilistico* per un periodo di tempo di un anno. Si è potuta costruire così la seguente tabella doppia:

	LIVELLO DI TRAFFICO		
METEO	basso	medio	alto
sereno	84	26	11
variabile	29	98	29
pioggia	7	26	55

- a) Costruire le distribuzioni di frequenze marginali della suddetta tabella;
- b) Stabilire che tipo di media è possibile calcolare per ciascuna delle due distribuzioni e determinarle;
- c) Individuare se sussiste un'eventuale associazione fra le due variabili.
- 3) Quanti e quali sono i parametri che caratterizzano una distribuzione di probabilità normale e che valori assumono per la distribuzione normale standardizzata?
- 4) Qual è la differenza che sussiste fra gli indici di variazione e gli indici di dispersione?

### COMPITO 3

- 1) I seguenti valori rappresentano le somme (in dollari) ritirate da un bancomat da parte di 25 clienti di una banca:

40	100	140	50	100
150	50	100	160	70
100	110	200	80	110
80	120	90	120	100
130	100	140	160	70

- a) Costruire la distribuzione di frequenze considerando 6 classi, in modo tale che l'estremo inferiore della prima classe sia 30 e l'estremo superiore dell'ultima classe sia 210;
- b) Individuare il tipo di distribuzione da cui i dati provengono e verificarne l'adattamento.
- 2) Nella tabella che segue si riportano i valori di due tipi di valuta, il marco tedesco e lo yen giapponese, dal 1988 al 1997:

<b>anno</b>	<b>Marco tedesco</b>	<b>Yen giapponese</b>
1988	1.76	128.17
1989	1.88	138.07
1990	1.62	145.00
1991	1.66	134.59
1992	1.56	126.78
1993	1.65	111.20
1994	1.62	102.21
1995	1.50	103.35
1996	1.54	115.87
1997	1.80	130.38

Verificare, utilizzando un indice appropriato, se fra il marco tedesco e lo yen giapponese esisteva una qualche interdipendenza.

- 3) Quali vantaggi offre una distribuzione di frequenze rispetto a una serie di valori?
- 4) Quando la distribuzione binomiale tende alla distribuzione di Poisson e quando, invece, alla distribuzione normale?

## COMPITO 4

- 1) Due giudici di un concorso, a cui è stato domandato di ordinare 8 candidati A, B, C, D, E, F, G e H secondo la loro preferenza, hanno fatto le scelte riportate nella seguente tabella. Verificare fino a che punto i giudici si sono trovati d'accordo nelle loro scelte.

Primo giudice	5	2	8	1	4	6	3	7
Secondo giudice	4	5	7	3	2	8	1	6

- 2) La seguente tabella mostra la distribuzione dei pesi in tonnellate supportati al massimo da certi cavi prodotti da una società:

Peso massimo (tonnellate)	Numero di cavi
<9.7	2
9.7 - 10.2	5
10.2 - 10.7	18
10.7 - 11.2	27
11.2 - 11.7	19
11.7 - 12.2	6
>12.2	3

- a) Calcolare gli indici di asimmetria e di curtosi;  
b) adattare alla distribuzione di frequenze empiriche la distribuzione teorica che si ritiene più opportuna;  
c) verificare l'adattamento e commentare i risultati.
- 3) Quali sono i motivi che inducono all'uso degli indici di variabilità relativa?
- 4) Supposta una relazione di tipo lineare fra due variabili, come si comportano, da un punto di vista grafico, le due rette di regressione in relazione al valore assunto dal coefficiente di correlazione di Bravais-Pearson?

## COMPITO 5

- 1) La tabella che segue mostra il numero dei figli di un campione di donne in età superiore ai 40 anni, che si presume abbiano portato a termine il periodo fertile; le donne sono state classificate in base al loro livello di istruzione:

GRADO DI ISTRUZIONE	NUMERO DI FIGLI		
	0	1 o 2	3 o più
Livello A o equivalente	116	364	190
Inferiore al livello A	225	1143	721

- a) Determinare la percentuale di donne per ciascun livello di istruzione e numero di figli;
  - b) Verificare se c'è una relazione fra il livello di istruzione e il numero dei bambini nati.
- 2) Il Quoziente intellettivo (QI) è distribuito in modo normale con un valore medio di 100 e uno scarto quadratico medio di 15.
- a) Tra quali due valori si troverà il 68,2% del QI della popolazione?
  - b) Quale proporzione della popolazione avrà un QI al di sotto di 80?
  - c) Quale proporzione di popolazione avrà un QI tra 95 e 115?
- 3) Quando la distribuzione binomiale è simmetrica?
- 4) In base a quali criteri vengono scelte le medie più opportune?

## COMPITO 6

- 1) La tabella che segue mostra il numero di giorni, in un periodo di 50 giorni, durante i quali sono avvenuti, in una certa città,  $X$  incidenti automobilistici. Adattare alla distribuzione data la distribuzione teorica che si ritiene più opportuna e specificarne il motivo:

Numero di incidenti	Numero di giorni
0	21
1	18
2	7
3	3
4	1

- 2) Il direttore del personale di una grossa società suppone che ci sia una relazione tra l'assenteismo e l'età dei dipendenti. Si seleziona un campione di 10 lavoratori e si perviene ai seguenti risultati:

lavoratore	Età	Giorni di assenza
1	27	15
2	61	6
3	37	10
4	23	18
5	46	9
6	58	7
7	29	14
8	36	11
9	64	5
10	40	8

- a) supponendo l'esistenza di una relazione lineare, applicare il metodo dei minimi quadrati per la determinazione dei coefficienti  $a$  e  $b$  della retta di regressione;
- b) interpretare il significato dei due coefficienti;
- c) determinare quanti giorni di assenza può avere, in media, un lavoratore dell'età di 50 anni;
- d) calcolare la misura della bontà di adattamento della retta ai dati osservati.
- 3) Se ho una distribuzione di probabilità con asimmetria positiva, come si comportano media, mediana e moda?
- 4) Che caratteristiche ha una distribuzione di probabilità leptocurtica?

## COMPITO 7

- 1) In un palazzo di 50 appartamenti è stato rilevato il numero dei vani:

1	2	4	1	5
5	4	4	2	4
3	2	4	5	1
7	1	3	7	7
3	1	1	7	5
6	6	2	6	7
3	5	1	3	3
5	7	3	2	5
7	7	1	6	4
1	5	7	5	7

- a) Costruire la distribuzione di frequenze della suddetta serie di valori;
  - b) classificare la variabile rilevata e calcolare quelle costanti di sintesi che, per tale variabile, assumono pienezza di significato.
- 2) La seguente tabella mostra la distribuzione di frequenze del numero dei parti, rilevato in una clinica, nell'arco di un periodo di 100 giorni:

n. parti	frequenze
0	17
1	32
2	29
3	15
4	3
5	2
6	1
7	1

- a) adattare alla distribuzione osservata la distribuzione teorica che si ritiene più opportuna e spiegarne le motivazioni;
  - b) verificare, attraverso l'uso di un indice appropriato, la bontà dell'adattamento.
- 3) L'indipendenza in distribuzione fra due variabili statistiche implica l'indipendenza in media? Perché? Dimostrare che, per due variabili statistiche indipendenti in distribuzione, la covarianza è nulla.
- 4) Osservata una distribuzione di frequenza empirica asimmetrica, ha senso, in generale, calcolare un indice di curtosi?

## COMPITO 8

- 1) Su di un campione di 50 professori è stato rilevato il reddito annuo in migliaia di \$, come riportato nella seguente tabella:

26.209	23.413	26.120	29.187	24.450
23.253	19.313	23.449	31.728	24.904
26.399	21.455	25.110	31.728	34.134
19.876	25.072	29.598	34.161	24.740
21.619	22.981	33.675	30.010	27.540
23.602	21.669	27.129	33.675	26.120
23.602	24.740	28.775	35.133	24.772
22.447	23.602	30.831	30.657	32.701
21.864	24.772	32.701	22.897	32.701
23.602	25.784	31.728	25.818	27.129

- a) costruire la distribuzione di frequenza di tale variabile, considerando classi di ampiezza costante, e realizzarne una completa analisi descrittiva;
- b) verificare se i dati provengono da una distribuzione di probabilità normale; solo in tal caso provare ad adattarla e a calcolare l'indice di bontà di adattamento.

- 2) In un collettivo di 10 studenti è stato rilevato il voto riportato all'esame di Statistica (X) e quello riportato all'esame di Storia contemporanea (Y):

<i>Studente</i>	1	2	3	4	5	6	7	8	9	10
<i>Voto di Statistica</i>	28	22	18	18	20	30	20	23	23	27
<i>Voto di Storia contemporanea</i>	30	28	27	18	28	28	28	27	27	18

- a) costruire la distribuzione doppia di frequenze (X,Y), considerando, per ciascuna variabile, tre classi di ampiezza costante;
- b) calcolare il voto mediano dell'esame di Statistica;
- c) stabilire se vi è indipendenza in distribuzione fra le due variabili.

- 3) Che valori possono assumere, rispettivamente, la variabile casuale binomiale e la variabile casuale di Poisson?

- 4) Indicata con  $\sigma_{XY}$  la covarianza fra due variabili statistiche X e Y, dimostrare che  $\sigma_{XY} = M_{XY} - M_X M_Y$



## COMPITO 9

- 1) Nella tabella che segue troviamo alcune misurazioni riguardanti la temperatura a riposo, in gradi centigradi, e il numero dei battiti cardiaci di un gruppo di soggetti maschi:

temperatura	n. di battiti
35,7	70
36,2	82
36,3	78
36,6	58
36,7	78
36,7	73
36,8	86
36,9	68
37,0	70
37,1	78
37,3	83

- a) Rappresentare graficamente i punti osservati;  
 b) stabilire se esiste una relazione fra le due variabili ed eventualmente individuare la funzione che rappresenti al meglio tale relazione;  
 c) verificare se la suddetta funzione si adatta bene ai dati.
- 2) Un gruppo di 50 soggetti, di età compresa fra i 30 e i 60 anni, è stato suddiviso in 6 categorie in base al titolo di studio:

età	titolo di studio	età	titolo di studio	età	titolo di studio	età	titolo di studio	età	titolo di studio
34	2	38	4	45	1	43	2	56	4
35	3	36	5	48	1	41	3	59	6
38	3	33	5	41	1	45	2	51	5
31	1	32	4	37	6	41	1	48	5
37	3	31	4	42	1	45	2	54	4
34	4	34	6	43	1	42	4	57	6
33	2	39	4	48	1	47	2	55	3
31	3	35	5	49	1	44	3	58	5
32	2	36	4	44	1	42	1	51	2
39	3	32	5	47	1	49	2	53	5

- a) Indicata con X l'età e con Y il titolo di studio, costruire la distribuzione di frequenza doppia, considerando, per la variabile X, tre classi di ampiezza costante e pari a 10, in cui 30 è l'estremo inferiore per la prima classe e 60 l'estremo superiore dell'ultima classe;  
 b) considerare le distribuzioni condizionate del titolo di studio rispetto all'età, verificarne la simmetria, o stabilirne il tipo di asimmetria, attraverso l'uso di un indice appropriato.
- 3) Dimostrare che la varianza è invariante per traslazione.  
 4) Descrivere quali differenze esistono tra un grafico a colonne e un istogramma.

## COMPITO 10

- 1) La seguente tabella riporta la distribuzione di 70 giovani maschi a 20 anni secondo la statura (X) in metri ed il peso (Y) in kg:

X/Y	50 - 60	60 - 70	70 - 80	80 - 90
1.61 - 1.65	8	9	2	0
1.65 - 1.69	1	22	6	0
1.69 - 1.73	0	6	14	2

- a) misurare la dipendenza in media della variabile “peso” dalla variabile “statura”;  
 b) misurare l’interdipendenza fra le due variabili rilevate;  
 c) commentare i risultati.
- 2) Consideriamo, per le 20 regioni italiane, la raccolta differenziata dei rifiuti urbani, nell’anno 1998, attraverso le seguenti variabili:
- X= percentuale sul totale dei rifiuti differenziati;
  - Y= Kg per abitante di rifiuti differenziati.

I dati sono riportati nella seguente tabella:

REGIONI	X	Y	REGIONI	X	Y
Piemonte	11	49.1	Marche	7.5	37.7
Valle d’Aosta	10.3	51.6	Lazio	4.2	21.7
Lombardia	30.8	138.5	Abruzzo	2.6	11.3
Trentino Alto Adige	14.7	80.6	Molise	1.4	4.6
Veneto	19.5	88.2	Campania	1.6	6.6
Friuli Venezia Giulia	12.7	58	Puglia	2.7	9.7
Liguria	8.4	44.5	Basilicata	3.1	11.7
Emilia Romagna	14.8	84.8	Calabria	0.6	2.3
Toscana	13.1	71.1	Sicilia	1	4.9
Umbria	6.3	32.7	Sardegna	1	4.4

- a) stabilire attraverso un’analisi grafica, se è lecito supporre l’esistenza di una relazione lineare fra le due variabili;  
 b) determinare e interpretare, eventualmente, i coefficienti a e b della retta di regressione, attraverso il metodo dei minimi quadrati;  
 c) verificare, empiricamente, se la retta si adatta bene ai dati osservati.
- 3) Sapendo che i  $\frac{3}{5}$  dei semi contenuti in un sacco daranno fiori rossi, mentre i rimanenti daranno fiori gialli, calcolare la probabilità di ottenere esattamente 5 fiori rossi in un filare di 8 piante.
- 4) Quali proprietà deve possedere una funzione qualsiasi affinché possa essere definita “funzione di densità di probabilità”?

## COMPITO 11

- 1) Per ciascuno dei dipendenti di un'azienda sono stati rilevati i caratteri “numero di familiari a carico” e “anzianità di servizio” (in anni). Le informazioni raccolte sono riassunte nella seguente tabella:

	Familiari a carico				
Anzianità	0	1	2	3	4
[0, 1)	21	26	11	12	5
[1, 3)	9	19	14	7	17
[3, 5)	16	18	9	15	19
[5, 6)	7	4	8	7	8

- a) Qual è la percentuale dei dipendenti dell'intero collettivo che hanno 2 familiari a carico e anzianità di servizio maggiore di 5 anni?
- b) Qual è la percentuale dei dipendenti con anzianità di servizio tra 3 e 5 anni, tra coloro che hanno un familiare a carico?
- c) Qual è la percentuale dei dipendenti che hanno 4 familiari a carico tra quelli che hanno anzianità di servizio fino a 1 anno?
- d) La variabile “anzianità di servizio” si può ritenere dipendente in media dalla variabile “n. di familiari a carico”? Se sì, misurare tale dipendenza con un indice appropriato.

- 2) I seguenti dati riguardano il numero di mesi tra l'iscrizione e la laurea per 24 studenti di una determinata facoltà:

48, 84, 60, 51, 49, 48, 51, 53, 66, 76, 48, 50, 52, 54, 54, 58, 68, 60, 56, 72, 56, 54, 56, 52

- a) Calcolare il numero medio ed il numero mediano di mesi impiegati per laurearsi. Per quale ragione sono diversi?
  - b) Rappresentare graficamente la distribuzione con un diagramma a scatola e baffi (box-plot). Che forma presenta la distribuzione? Perché?
  - c) L'indice di asimmetria di Fisher conferma le conclusioni cui si è pervenuti al punto b)?
- 3) Se  $X$  è una variabile casuale normale di media 0.02 e varianza 4, qual è la probabilità di osservare un valore della variabile inferiore od uguale alla sua media?
- 4) Cosa misura il coefficiente di correlazione di Bravais-Pearson? In quale range varia e in quali casi assume esattamente il valore minimo e il valore massimo? E' un numero puro? Perché?

## COMPITO 12

- 1) La seguente tabella riporta l'età al matrimonio di 7 coppie di sposi:

Coppia	Età sposo	Età sposa
1	31	29
2	23	19
3	35	35
4	29	30
5	28	27
6	30	28
7	25	24

- a) Determinare se esiste concordanza tra l'età dello sposo e l'età della sposa, utilizzando sia il coefficiente di correlazione di Bravais-Pearson, sia l'indice  $\rho$  di Spearman. Commentare il risultato dei due indici;
- b) Rappresentare graficamente l'età dello sposo e l'età della sposa, determinare la migliore retta di interpolazione e misurare la bontà dell'adattamento.
- 2) Cinque monete sono state lanciate 1000 volte e in ciascun lancio è stato osservato il n. di teste. Nella seguente tabella è indicato il n. di lanci durante i quali sono stati ottenuti 0,1,2,3,4,5 teste:

Numero di teste	Numero di lanci
0	38
1	144
2	342
3	287
4	164
5	25

- a) adattare alla distribuzione osservata la distribuzione teorica che si ritiene più opportuna e spiegarne le motivazioni;
- b) verificare, attraverso l'uso di un indice appropriato, la bontà dell'adattamento;
- c) in base ai risultati ottenuti, è possibile ritenere che le monete siano non truccate?
- 3) Su che tipo di variabili è possibile calcolare le medie di posizione?
- 4) Fra gli indici di asimmetria studiati, qual è il più informativo e perché?

### COMPITO 13

- 1) Una compagnia di trasporti ha stabilito che in media i propri camion coprono una distanza pari a 50000 miglia in un anno, con uno scarto quadratico medio di 12000 miglia. Si suppone, inoltre, che la distanza coperta in un anno da parte dei camion della compagnia segua approssimativamente una distribuzione normale.
  - a) Scelto a caso un camion della compagnia, qual è la probabilità che nell'ultimo anno:
    - abbia coperto una distanza compresa tra 34000 e 50000 miglia;
    - abbia coperto una distanza superiore a 40000 miglia;
  - b) Quanti camion su 1000 ci aspettiamo che abbiano coperto una distanza compresa tra 30000 e 60000 miglia nell'ultimo anno?
  - c) Quante miglia ci aspettiamo che vengano percorse in un anno da una percentuale di camion pari al 50%?
  
- 2) In un collettivo di 420 volontari si è osservata la frequenza di attività di volontariato per classi di età, ottenendo la seguente distribuzione di frequenze relative percentuali:

Frequenza di attività di volontariato (Y)	classi di età (X)			
	[14, 20]	]20, 35]	]35, 55]	]55, 60]
Almeno una volta la settimana	10	15	10	5
Una o più volte al mese	10	20	20	10

- a) Quanti sono i volontari con età superiore a 20 anni e non superiore a 55?
  - b) Quanti sono i volontari che prestano la loro attività almeno una volta la settimana e che hanno un'età superiore a 55 anni e non superiore a 60?
  - c) Determinare il rapporto di correlazione dell'età dalla regolarità del servizio di volontariato.
  
- 3) Cosa vuol dire in Statistica “numero puro”? Elencare gli indici incontrati durante il corso, che possono essere definiti numeri puri, e descrivere in quali contesti vengono utilizzati e perché.
  
- 4) Determinare la media e la varianza di una combinazione lineare  $Y = \alpha + \beta X$ .

# COMPITO 14

- 1) La seguente distribuzione descrive il numero di particelle rilasciate durante il decadimento radioattivo del Polonio, in intervalli di 72 secondi:

<b>x</b>	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
<b>n</b>	57	203	383	525	532	408	273	139	45	27	10	4	0	1	1

- a) adattare alla distribuzione osservata la distribuzione teorica che si ritiene più opportuna e spiegarne le motivazioni;  
b) verificare, attraverso l'uso di un indice appropriato, la bontà dell'adattamento.
- 2) Una certa zona geografica è servita da sette compagnie telefoniche. Per ciascuna compagnia è stato rilevato il costo al minuto (in centesimi di euro) per telefonate interurbane nella fascia oraria a tariffa piena (X) ed il numero di minuti di conversazione (in milioni) delle telefonate interurbane effettuate dagli abbonati nell'ultimo anno (Y). I dati raccolti sono riassunti nella seguente tabella

Compagnia telefonica	costo al minuto X	n.min.di conversazione Y
A	7	0.50
B	10	1.07
C	11	1.10
D	9	0.81
E	8	0.72
F	12	1.23
G	14	1.55

- a) Ricavare i valori dei coefficienti della retta di regressione col metodo dei minimi quadrati.  
b) Qual è la percentuale di varianza di Y spiegata dalla regressione?  
c) Sappiamo che il costo di un minuto di telefonata interurbana con la compagnia H è pari a 13 centesimi di euro. Sulla base del modello, qual è stato il numero di minuti medio di conversazione delle telefonate interurbane effettuate nell'ultimo anno dagli abbonati ad H?
- 3) Data la distribuzione di frequenze relative del carattere Y

<b>y<sub>i</sub></b>	-3	-1	0	1	3
<b>f<sub>i</sub></b>	0.1	f <sub>2</sub>	f <sub>3</sub>	0.2	0.1

sapendo che la media aritmetica è pari a 0, quali sono i valori di f<sub>2</sub> ed f<sub>3</sub> ?

- 4) E' stata calcolata la retta di regressione che lega la variabile y alla x, ricavata col metodo dei minimi quadrati. Sono stati calcolati, inoltre, il coefficiente di correlazione lineare tra le due variabili, pari a 0.8, e la varianza dei residui, pari a 29.16. Quanto vale la varianza di Y?

## COMPITO 15

- 1) La seguente tabella riporta il peso, in kg, e la statura, in cm, di un gruppo di ragazze di undici anni di età:

Statura	Peso	Statura	Peso	Statura	Peso
135	26	141	28	149	46
146	33	136	28	147	36
153	55	154	36	152	47
154	50	151	48	140	33
139	32	155	36	143	42
131	25	133	31	148	32
149	44	149	34	149	32
137	31	141	32	141	29
143	36	164	47	137	34
146	35	146	37	135	30

- a) Costruire una distribuzione di frequenza doppia considerando, per ciascuna variabile, classi opportune di ampiezza costante;
- b) verificare, scegliendo un indice appropriato, se le due variabili sono interdipendenti;
- c) considerare la distribuzione di frequenza della variabile peso e provare ad adattare, se si ritiene possibile, la distribuzione teorica più idonea.
- 2) Con riferimento a molti processi industrializzati si utilizza il termine “work-in-process” (WIP). Negli impianti di fabbricazione di libri, il WIP rappresenta il tempo necessario per piegare, riunire, cucire e rilegare i fogli che provengono da una pressa. I dati che seguono sono relativi al tempo di lavorazione (tempo in giorni che intercorre tra quando i libri vengono stampati a quando sono impacchettati nei cartoni) per due campioni di 20 libri estratti da due impianti di fabbricazione:

<b>IMPIANTO A</b>	5,62	5,29	16,25	10,92	11,46	21,62	8,45	8,58	5,41	11,42
	11,62	7,29	7,5	7,96	4,42	10,5	7,58	9,29	7,54	8,92
<b>IMPIANTO B</b>	9,54	11,46	16,62	12,62	25,75	15,41	14,29	13,13	13,71	10,04
	5,75	12,46	9,17	13,21	6	2,33	14,25	5,37	6,25	9,71

Per ciascuno dei due impianti:

- a) calcolare le medie e gli indici di variabilità più appropriati;
- b) costruire il diagramma scatola e baffi;
- c) in base alle risposte date ai punti a) e b), si può ritenere che ci siano differenze tra i due impianti?
- 3) In corrispondenza di quali valori la distribuzione di probabilità normale presenta i punti di massimo e di flesso?
- 4) Come si definisce la funzione di ripartizione per una variabile casuale continua?

## COMPITO 16

- 1) Il manager di una catena di supermercati intende stabilire in quale maniera la vendita di cibo per animali è influenzata dallo spazio sugli scaffali destinato al prodotto. La seguente tabella riporta i valori dello spazio sugli scaffali (in piedi) e dell'ammontare delle vendite settimanali di cibo per animali (in migliaia di dollari), in 12 supermercati della medesima grandezza:

<b>Vendite settimanali</b>	1,6	2,2	1,4	1,9	2,4	2,6	2,3	2,7	2,8	2,6	2,9	3,1
<b>Spazio sugli scaffali</b>	5	5	5	10	10	10	15	15	15	20	20	20

- a) Disegnare il diagramma di dispersione per i dati della tabella;
  - b) nell'ipotesi che tra le due variabili sussista una relazione lineare, stimare con il metodo dei minimi quadrati i coefficienti di regressione  $b_0$  e  $b_1$ ;
  - c) fornire un'interpretazione di  $b_1$ ;
  - d) prevedere l'ammontare delle vendite settimanali di cibo per animali se lo spazio destinato ai prodotti è uguale a 8 piedi;
  - e) verificare, attraverso un indice appropriato, se la funzione lineare si adatta bene ai dati osservati.
- 2) La seguente tabella mostra il numero di tornado verificatisi, in una certa area geografica, tra gli anni 1959 e 1988:

<b>Anno</b>	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973
<b>Tornado</b>	3	4	5	1	3	1	5	1	2	2	7	4	5	6	6
<b>Anno</b>	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988
<b>Tornado</b>	6	3	7	5	8	3	4	3	3	8	6	7	9	6	5

- a) Costruire la distribuzione di frequenze della variabile “numero di tornado”;
  - b) individuare la distribuzione teorica che meglio rappresenta la distribuzione osservata e provare ad adattarla;
  - c) verificare, attraverso un indice di bontà di adattamento, se il modello teorico scelto descrive bene i dati osservati.
- 3) Se la distribuzione di un certo carattere ha media aritmetica pari a 5 e coefficiente di variazione pari a 2, quanto vale la varianza?
- 4) Se la media aritmetica di una distribuzione di frequenza è pari a zero vuol dire che:
- a) il numero delle unità su cui è stato rilevato un valore positivo della variabile è sempre pari al numero delle unità su cui è stato rilevato un valore negativo;
  - b) l'ammontare complessivo della variabile rilevato sulle unità che presentano valori positivi è pari all'ammontare complessivo rilevato sulle unità che presentano valori negativi;
  - c) la variabilità della distribuzione è nulla.



## COMPITO 17

- 1) In un collettivo di pazienti sono stati rilevati la quantità di colesterolo in milligrammi per 100 millilitri di sangue ed il genere. Dallo spoglio delle osservazioni si è ottenuta la seguente distribuzione doppia di frequenze:

<b>Colesterolo</b>	<b>Maschio</b>	<b>Femmina</b>
[120, 160]	40	20
]160, 180]	10	12
]180, 200]	20	10
]200, 240]	10	20
]240, 300]	45	10

- a) Rappresentare graficamente la distribuzione del colesterolo;  
 b) Calcolare la media del colesterolo per ciascun genere;  
 c) Calcolare la classe mediana del colesterolo per i maschi;  
 d) Stabilire quale delle due distribuzioni è più variabile.
- 2) Si consideri la distribuzione dei laureati del 1991 per gruppo di corso di laurea e condizione occupazionale nel 1993:

<b>X\Y</b>	<b>Occupato stabilmente</b>	<b>Occupato precariamente</b>	<b>Disoccupato</b>
Medicina	681	466	118
Economia	732	72	18
Lettere	770	585	147

- a) Determinare la distribuzione percentuale rispetto alla condizione occupazionale all'interno dei laureati in Medicina e confrontarla con quella dei laureati in Economia e con quella dei laureati in lettere;  
 b) Determinare la distribuzione percentuale rispetto al tipo di laurea degli occupati stabilmente e confrontarla con quella degli occupati precariamente e con quella dei disoccupati;  
 c) Determinare la distribuzione marginale del collettivo rispetto alla condizione occupazionale;  
 d) Determinare la distribuzione marginale del collettivo rispetto al corso di laurea;  
 e) Trovare le tabelle teoriche di perfetta indipendenza;  
 f) Calcolare l'indice  $X^2$  e commentare il risultato.
- 3) Da quanti e quali parametri sono caratterizzate, rispettivamente, le distribuzioni binomiale e di Poisson? In che relazione sono tali parametri con la media e la varianza delle due variabili casuali?
- 4) Quali sono i limiti dell'indice di associazione  $X^2$  e degli indici da esso derivati?

# COMPITO 18

- 1) E' stato rilevato il numero di schedine giocate al Superenalotto nell'ultimo mese da 100 famiglie, classificate per area geografica di residenza:

n. di schedine	Residenza		
	Nord	Centro	Sud
0-  5	10	10	0
5-  25	5	20	15
25-  60	0	10	30

- a) Tra tutti i residenti del Nord, qual è la frequenza relativa delle famiglie che hanno giocato un n. di schedine minore o uguale a 5?
- b) Tra i residenti al centro, qual è la classe modale?
- c) Qual è la percentuale delle famiglie residenti al sud che hanno giocato più di 25 schedine?
- d) Qual è il n. medio di schedine giocate per famiglia?
- 2) Con riferimento alla tabella precedente:
- a) Calcolare, se si ritiene possibile, l'indice  $X^2$  di Pearson ed eventualmente specificare quali indicazioni fornisce;
- b) individuare quale variabile può dipendere "in media" dall'altra e misurare tale dipendenza con un indice opportuno;
- c) spiegare i motivi per cui è possibile o meno definire il coefficiente di correlazione lineare di Bravais-Pearson tra le due variabili in esame.
- 3) Indichiamo con  $\sigma_X$  lo scarto quadratico medio della distribuzione della variabile X in una data popolazione. Se i valori della variabile Y si ottengono da quelli di X tramite la trasformazione  $Y = -\frac{1}{3}X$ , qual è il valore della varianza di Y?
- 4) Se, dopo aver rilevato n osservazioni  $x_1, x_2, \dots, x_n$ , ricaviamo che la media è  $M_X=1$  e la

varianza è  $\sigma_X^2=1$ , allora a quanto sarà pari  $\frac{1}{n} \sum_{i=1}^n x_i^2$  ?

# COMPITO 19

- 1) Consideriamo la distribuzione del numero di impiegati di un'azienda rispetto all'età ed al fatto che abbiano ottenuto o meno un avanzamento di carriera nell'ultimo anno:

età (in anni)	avanzamento di carriera	
	no	si
[18; 50)	36	84
[50; 65]	24	56

- Qual è la frequenza relativa degli ultra-cinquantenni che non hanno ottenuto un avanzamento di carriera nell'ultimo anno?
  - Qual è la frequenza relativa di coloro che hanno ottenuto un avanzamento di carriera nell'ultimo anno, tra tutti i dipendenti che hanno meno di 50 anni?
  - Qual è l'età media di coloro che hanno avuto un avanzamento di carriera nell'ultimo anno?
  - E' possibile stabilire se i due caratteri rilevati (età e avanzamento o meno di carriera) sono indipendenti linearmente? Perché?
  - E' possibile stabilire se uno dei due caratteri rilevati è indipendente in media dall'altro? Se sì, calcolare un indice che confermi l'eventuale indipendenza e commentarne il risultato.
- 2) Per 10 paesi gli incrementi percentuali verificatisi fra il 1985 e il 1990 relativamente a due grandezze economiche sono risultati i seguenti:

Paesi	Reddito	Spesa per il turismo
Australia	3,5	2,2
Canada	2,5	3,1
Francia	3,5	2,8
Germania	3,5	2,5
Italia	3,5	3,1
Giappone	5,0	6,3
Messico	4,5	0,6
Spagna	4,0	5,2
Gran Bretagna	3,5	1,2
U.S.A.	2,5	2,4

- Stimare l'equazione della retta che esprime la Spesa per il turismo in funzione del Reddito;
  - Misurare, mediante un opportuno indice, la bontà di adattamento della retta ai dati;
  - La variabile dipendente si può ritenere simmetrica?
- 3) Il tempo (in minuti) che un treno Regionale impiega per andare da Pianoverde a Monterosso ha distribuzione normale di media 30 e varianza 1. Sull'orario dell'azienda che gestisce la ferrovia è indicato un tempo di percorrenza di 32 minuti. Qual è la probabilità che un treno non arrivi in ritardo?
- 4) Se la covarianza tra due variabili  $X$  e  $Y$  è pari a 0.1, mentre  $\text{Var}(X) = 1$  e  $\text{Var}(Y) = 0.011$ , a cosa sarà uguale  $\text{Var}(X+Y)$ ?

## COMPITO 20

- 1) Dopo l'analisi di 80 nidi di una specie di volatili, viene costruita la distribuzione del numero di piccoli sopravvissuti:

<b>X</b>	<b>n</b>
0	8
1	15
2	20
3	22
4	15
totale	80

- a) Rappresentare graficamente la distribuzione osservata;  
 b) Provare ad adattare la distribuzione teorica che si ritiene più idonea;  
 c) Verificare, attraverso l'uso di un indice appropriato, se il modello teorico scelto si adatta bene ai dati empirici.
- 2) In un'indagine svolta su un campione di famiglie, è stata rilevata la distribuzione congiunta del reddito (in euro) del capofamiglia (Y) e del coniuge (X):

<b>X</b>	<b>Y</b>			
	<b>1000</b>	<b>1500</b>	<b>2000</b>	<b>2500</b>
<b>500</b>	10	14	13	5
<b>1000</b>	8	15	12	7
<b>1500</b>	3	5	7	4

- a) Calcolare la media e la varianza della distribuzione dei redditi del capofamiglia e del coniuge. Quale delle due presenta una più marcata variabilità?  
 b) Verificare se le due variabili sono interdipendenti linearmente.  
 c) Calcolare la media e la varianza di  $Z=X+Y$
- 3) Uno studente ha programmato di sostenere gli esami A e B in una determinata sessione. In base alla sua preparazione ritiene che la probabilità di superare l'esame A sia pari a 0.7, la probabilità di superare l'esame B sia 0.5, mentre la probabilità di superarli entrambi sia 0.4. Qual è la probabilità che lo studente superi almeno uno dei due esami?
- 4) Su ciascuno dei 63 studenti che seguono un certo corso di studi si rileva la variabile **X**=numero degli esami superati. I risultati ottenuti sono riassunti nella seguente tabella, dove con **F<sub>i</sub>** sono state indicate le frequenze relative cumulate:

<b>X</b>	0	1	2	3	4	5
<b>F<sub>i</sub></b>	0.1	0.3	0.3	0.6	0.8	1

Quanti studenti hanno superato esattamente due esami?

## COMPITO 21

- 1) Un produttore di pile per lampade tascabili vuol confrontare la durata di vita di due diverse tipologie di pile prodotte nei propri impianti. Le rilevazioni dei tempi di vita delle pile di tipo I e di quelle di tipo II sono riassunte nella seguente distribuzione di frequenze percentuali cumulate:

<b>tempo di vita (in ore)</b>	<b>% cumulata tipo I</b>	<b>% cumulata tipo II</b>
(200, 300]	10	30
(300, 400]	30	50
(400, 500]	50	90
(500, 600]	90	100
(600, 700]	100	100

- a) Qual è il valore della mediana della distribuzione della durata delle pile di tipo I e di quelle di tipo II?
  - b) Il 30% delle pile ha avuto una durata minore o uguale a quale valore?
  - c) In base alle informazioni ottenute da opportuni indici di sintesi, spiegare qual è nel complesso il tipo di pila che risulta più affidabile.
- 2) Durante un sondaggio sulle abitudini dei clienti abituali di un supermercato, su ciascuno dei clienti intervistati è stato rilevato il carattere “distanza”, misurata attraverso il tempo in minuti occorrente per raggiungere in auto il supermercato dall'abitazione, ed il numero di volte in cui il cliente si è recato nel supermercato per fare la spesa, nelle ultime due settimane. La distribuzione doppia è riassunta nella seguente tabella:

distanza	(0, 2]	(2, 5]	(5, 10]	(10, 15]
n. spese				
1	7	10	6	14
2	17	6	13	16
3	10	9	7	10
4	10	17	12	6
5	8	12	8	2

- a) Ricavare la distribuzione del carattere “distanza” nell'intero collettivo e rappresentarla graficamente in modo opportuno;
  - b) Misurare la dipendenza in media della variabile “n. di spese” dalla variabile “distanza”.
- 3) Si considerino 10 dipendenti, scelti a caso tra quelli che lavorano in un'azienda, e si indichi con  $X$  il numero delle donne. Il numero di donne presenti in azienda è pari al 70% del totale dei dipendenti.
- a) Qual è la distribuzione di probabilità della variabile casuale  $X$  ?
  - b) Qual è la probabilità che 9 dei dipendenti, tra i 10 considerati, siano donne?
  - c) Qual è il numero medio di dipendenti donna, tra i 10 considerati?
- 4) Nel palazzetto dello sport è in corso una partita di pallacanestro. La media delle altezze dei cinque giocatori della Virtus è di 186,6 cm, mentre la varianza è pari a  $15,44 \text{ cm}^2$ . Durante un cambio esce il giocatore alto 180 cm. Qual è la media e la varianza delle altezze dei 4 giocatori che restano in campo?