Testo esercizi

Lezione 1 – Statistica descrittiva

Esercizio 1

La distribuzione del numero di titolari di conti correnti in una banca, rispetto all'età e all'eventuale possesso di carta di credito, è la seguente:

	Giovanissimi	Giovani	Anziani
Possiedono CC	40	52	36
Non possiedono CC	20	26	18

- Calcolare una distribuzione di frequenza percentuale che consenta di capire se il possesso di carta di credito dipende dall'età.
- Rappresentare i risultati ottenuti al punto precedente con degli opportuni diagrammi.

Esercizio 2

La distribuzione del numero di impiegati di una azienda, rispetto all'età e all'eventuale avanzamento di carriera nell'ultimo anno, è la seguente:

	Giovanissimi	Giovani	Anziani
Avanzamento	47	18	12
Non avanzamento	21	49	53

- Calcolare una distribuzione di frequenza percentuale che consenta di capire se l'avanzamento di carriera dipende dall'età.
- Rappresentare i risultati ottenuti al punto precedente con degli opportuni diagrammi.

Esercizio 3

La seguente tabella riporta i tempi di attesa (in minuti) di due campioni casuali di clienti di un operatore telefonico, distinti per call center. Confrontare i due poligoni di frequenze cumulate e non cumulate, per stabilire quale call center garantisce tempi minori e meno variabili.

C. C. 1	1.1	0.7	1.5	1.6	0.4	1.9	2.1	1.4	2.0	2.5
C. C. 2	0.5	1.0	1.3	2.1	0.6	1.0	1.7	1.2	1.4	0.3

Esercizio 4

I seguenti dati si riferiscono al numero di abbonamenti di una TV a pagamento venduti da parte di alcuni agenti di commercio nel 2014.

Agente	A	B	C	D	E	F	G	H	I	L
N Abb.	14	26	18	10	20	24	30	22	16	20

Tracciare il diagramma di Pareto per individuare i tre venditori migliori e il loro complessivo contributo percentuale ai volumi di vendita dell'azienda.

Lezione 2 – Statistica descrittiva

Esercizio 1

Un'azienda vuole valutare il tempo di guasto dell'impianto A2011 dall'ultimo collaudo (tempo in mesi). Si hanno a disposizione i seguenti dati, relativi a 13 stabilimenti in cui è installato:

0.25 0.50 2 4 6 12 24 30 30 36 42 53 100

- Calcola i 5 numeri di sintesi
- Calcola media e moda
- La distribuzione è simmetrica?
- Calcolare gli indici di variabilità.

Esercizio 2

Di seguito sono riportate le misurazioni di PM10 all'interno e all'esterno di ZTL.

Dentro ZTL	12.6	12.4	10.5	13.3	15.2	12.1	12.6	9.7	13.6	7.9	9.3	14.2
Fuori ZTL	12.0	17.2	10.0	14.5	14.5	19.4	16.3	12.2	17.1	14.6	13.9	/

Rappresentare l'istogramma delle frequenze dentro e fuori ZTL.

Lezione 3 – Probabilità

Esercizio 1

Data la seguente tabella di contingenza:

	B	B^c
A	35	25
A^c	30	10

- Disegna la corrispondente rappresentazione mediante diagramma di Venn.
- Calcola $P(A^c)$ e $P(B)$
- Calcola $P(A \cap B)$ e $P(A^c \cap B)$
- Calcola $P(A^c \cup B^c)$
- Calcola $P(A|B)$

Esercizio 2

La seguente tabella rappresenta lo stato civile dei residenti negli Stati Uniti, di almeno 18 anni, nel 2006. I dati sono espressi in milioni.

	Uomini	Donne
Mai sposati	30	25
Sposati	64	64
Vedovi	3	11
Divorziati	10	13

- a. Calcola la probabilità che un residente negli USA, di almeno 18 anni, sia vedovo.
- b. Calcola la probabilità che un residente negli USA, di almeno 18 anni, sia vedovo o divorziato.
- c. Calcola la probabilità che un residente negli USA, di almeno 18 anni, sia uomo o vedovo.

Esercizio 3

Un lotto di chip contiene il 3% di chip difettosi. Ogni chip viene testato prima di essere messo in vendita, ma il test che viene effettuato non è affidabile al 100%.

Un chip funzionante fallisce il test con probabilità 0.01, mentre un chip difettoso passa il test con probabilità 0.04.

- a. Qual è la probabilità che un chip fallisca il test, e quindi non venga messo in vendita?
- b. Qual è la probabilità che un cliente acquisti un chip effettivamente funzionante?

Esercizio 4

Una prova d'esame consiste in 18 quesiti a risposte multiple. Ogni quesito prevede 3 possibili risposte, di cui solo una esatta. Ogni risposta esatta vale +2, ogni risposta sbagliata vale -1 e ogni risposta non data vale 0.

Il voto finale è la somma dei punteggi ottenuti nei singoli quesiti.

- a. Qual è la probabilità di rispondere correttamente a tutte le domande tirando ad indovinare?
- b. Uno studente lascia 3 risposte in bianco e le altre le sceglie a caso. Qual è la probabilità che prenda un voto superiore a 26?

Esercizio 5

Sia X il numero di micro-incidenti che si verificano ogni giorno ad un incrocio stradale. È noto che la media di X è pari a 0.5.

- a. Indicare una ragionevole densità di probabilità per X e calcolare la probabilità che oggi non vi sia nessun incidente.
- b. Qual è la probabilità che oggi si verifichino almeno 2 incidenti?
- c. E quale quella che si verifichino esattamente 3 incidenti?

Esercizio 6

Due amici si sfidano giocando con un dado. Chi perde la scommessa deve pagare una cifra prestabilita all'amico.

Mario scommette 1500 euro che Roberto in 5 lanci consecutivi non otterrà mai 3.

Roberto scommette 700 euro che Mario in 5 lanci consecutivi otterrà al massimo una volta 3.

- a. Qual è la probabilità che Mario debba pagare 1500 euro a Roberto per la prima scommessa?
- b. Qual è la probabilità che Roberto debba pagare 700 euro a Mario per la seconda scommessa?

Lezione 4 – Intervalli di confidenza

Esercizio 1

- Assumendo che una popolazione sia normalmente distribuita, costruire un intervallo di confidenza al 95% per la stima della media della popolazione sulla base del campione di dimensione $n=7$:
 $1, 2, 3, 4, 5, 6, 21$
- Cambiare il numero 21 con 7 e ricalcolare l'intervallo di confidenza.
- Sulla base di questi risultati, descrivere l'effetto di un outlier sull'intervallo di confidenza.

Esercizio 2

Sia $X_1, \dots, X_n$ un campione casuale di una popolazione normale di media μ e varianza $\sigma^2 = 9$:

- Sapendo che $n=16$ e che $(3.53; 6.47)$ è un intervallo di confidenza simmetrico per μ , se ne determini il livello di confidenza.
- Sapendo che $\bar{x} = 5$, si deve costruire un intervallo di confidenza per μ al 90%.
- Si determini la minima ampiezza n del campione tale che un intervallo di confidenza per μ al livello 0.95 non risulti più ampio di 3.
- Ipotizziamo che σ^2 sia non nota e $S^2 = 9$. Calcolare un IC per μ al 90% con $\bar{x} = 5$.

Esercizio 3

Il direttore del personale di una grande azienda intende studiare l'assenteismo durante l'anno dei suoi dipendenti. Viene selezionato un campione casuale di 25 impiegati, dal quale emerge:

$\bar{x} = 9.7$ giorni di assenza

$S = 4.0$ giorni

12 Impiegati sono risultati assenti per più di 10 giorni.

- Costruire un intervallo di confidenza al 95% per μ .
- Costruire un intervallo di confidenza per la proporzione ξ di impiegati assenti per più di 10 giorni.
- Si supponga ora che il direttore del personale voglia condurre un'indagine anche nell'ufficio di una succursale. Si cerchi la numerosità campionaria necessaria ad avere un livello di confidenza del 95% e stimare μ con $e = \pm 1.5$ giorni e $\sigma = 4.5$ giorni.
- Quanti impiegati si devono selezionare per avere un livello di confidenza del 90% per stimare ξ con $e = \pm 0.075$.

Lezione 5 – Verifica d'ipotesi (Test a 1 campione)

Esercizio 1

In un impianto industriale si produce una sostanza chimica per cui è necessario tenere sotto controllo il grado di impurità. Sono accettabili impurità fino a 150 ppm.

Vengono prelevati piccoli campioni in un'ampolla, per verificare il grado di impurità. Da studi precedenti possiamo supporre che la quantità di impurità sia distribuita normalmente, con $\sigma = 20$ ppm.

- Viene estratto un campione di $n=10$ ampolle e si osserva che $\bar{x}_n = 154$ (quantità media di impurità presenti). Il processo produttivo può essere considerato "sotto controllo" al livello di significatività $\alpha = 0.05$?

Esercizio 2

Per progettare una cabina di pilotaggio di un aereo occorre considerare i valori antropometrici medi dei piloti. Supponiamo che la statura dei piloti sia distribuita secondo una legge $N(\mu, \sigma^2)$, con $\sigma = 6.1$ cm. La media stimata da un campione di $n=100$ piloti è $\bar{x}_n = 178.5$.

- Si calcoli un intervallo di confidenza al livello 0.95 per μ .
- Supponiamo $\mu_0 = 175$ cm e di voler testare $H_0: \mu = \mu_0$ contro $H_1: \mu \neq \mu_0$ al livello di significatività $\alpha = 0.05$.

Esercizio 3

Un produttore di vino vuole monitorare la percentuale di alcol nel vino che produce. Effettuando la misura dell'alcol in 5 bottiglie da un litro, si rilevano i valori (in cl):

10 12 9.5 10.1 10.5

- Supponendo che il contenuto X di alcol per ogni litro di vino sia normalmente distribuito, si esegua il test, con significatività 5%, con $H_0: \mu \geq 11$ contro $H_1: \mu < 11$
- Supponendo di sapere che $\sigma = 2$, si ripeta il test.
- Si calcoli il p-value in entrambe le situazioni precedenti.

Esercizio 4

Una moneta viene lanciata 1000 volte ottenendo 432 testa. Si esegua un test di livello 1% per stabilire se c'è evidenza sperimentale che la moneta sia truccata.

Lezione 6 – Verifica d'ipotesi (Test a 2 campioni)

Esercizio 1

Un'azienda vuole confrontare la durata di due tipi di batterie (che produce): modello A1 e modello B1. Siano le deviazioni standard note e rispettivamente $\sigma_{A1} = 1.8$ e $\sigma_{B1} = 2$.

La durata media di un campione di 100 A1 risulta essere di 4.1 ore, mentre quella di un campione di 120 B1 è di 4.5 ore.

- È possibile affermare che le batterie B1 sono più longeve delle A1 al livello di significatività 0.05?
- Calcola il p-value
- Determina un IC bilaterale per la differenza tra le durate medie al livello di confidenza del 99%
- Come cambierebbe la statistica test usata al punto a se conoscessimo solo le deviazioni standard campionarie $S_{A1}=1.8$ e $S_{B1}=2$ (sotto l'ipotesi di omoschedasticità)?

Esercizio 2

I consumi di due automobili A e B vengono sottoposti a confronto alla velocità costante di 110 Km/h.

Si assume che i consumi delle due automobili possano essere descritti da distribuzioni normali con la stessa varianza. L'auto A in 20 test mostra un consumo medio di 6,5 l/100 Km mentre l'auto B in 22 test consuma in media 6.61 l/100 Km

- Si fornisca una stima della varianza σ^2 del consumo di queste auto.

- b. Possiamo affermare che le due auto abbiano diverso consumo medio al livello di significatività $\alpha = 0.05$?
- c. Calcolare il p-value del test.
- d. Si calcoli un IC bilaterale al livello 0.95 per la differenza dei consumi medi.

Esercizio 3

Per decidere se i pneumatici di marca A sono migliori di quelli di marca B vengono entrambi testati sulle stesse 10 automobili da corsa sullo stesso circuito. Le due prove vengono effettuate a distanza di un'ora l'una dall'altra chiedendo ai piloti di guidare al meglio in entrambi i casi.

I tempi cronometrati sono i seguenti:

Auto	A	B
1	29.33	28.98
2	30.50	30.63
3	30.68	30.83
4	31.52	31.53
5	33.08	32.95
6	29.67	29.50
7	32.20	32.67
8	33.00	33.12
9	31.05	31.10
10	31.23	32.02

- a. Verifichiamo l'ipotesi che in media il tempo ottenuto usando pneumatici di marca A sia inferiore al tempo ottenuto usando quelli di marca B. Al livello di significatività $\alpha = 0.05$.
- b. Determinare un intervallo di confidenza bilaterale al livello 0.95 per la differenza dei tempi di gara.

Lezione 7 – Analisi di Regressione lineare semplice

Esercizio 1 – Esempio “teorico”

Si vuole stimare la relazione tra i pezzi di vendita Y (espressi in migliaia di euro) e la dimensione X (superficie espressa in mq) degli appartamenti di una certa area. I dati riferiti a un campione casuale di 10 case sono:

X	140	160	170	18705	110	155	235	245	142.5	170
Y	245	312	279	308	199	219	405	324	319	255

- a. Tracciare il diagramma di dispersione
- b. Specificare il modello
- c. Stimare i parametri del modello
- d. Stimare la capacità previsiva del modello
- e. Verificare la plausibilità delle assunzioni del modello
- f. Verificare le ipotesi sui coefficienti del modello
- g. Prevedere il valore di Y se X=200

Lezione 8 – Analisi di Regressione lineare semplice

Esercizio 1

Un data-set contiene dati per l'indice di fertilità totale (IFT), che corrisponde al numero totale di figli che ci si aspetta abbia una donna nell'arco della sua vita, e la percentuale di donne che assume contraccettivi in 8 paesi diversi.

Sia la variabile indipendente, X , la percentuale di donne che assume contraccettivi, e la variabile dipendente, Y , l'IFT:

i	X_i	Y_i
1	80	1.5
2	65	2.4
3	19	5.0
4	64	2.8
5	43	3.6
6	11	7.1
7	85	1.4
8	69	2.0

- Tracciare il diagramma di dispersione
- Calcolare il coefficiente di correlazione r
- Calcolare la retta di regressione che mette in relazione tra loro l'IFT e la variabile indipendente X
- Prevedere l'IFT (variabile Y) se la percentuale di donne che assume contraccettivo è pari al 20%

Esercizio 2

L'autorità garante delle comunicazioni negli Stati Uniti ha riscontrato una sostanziale differenza tra le vendite dichiarate dalle riviste e quelle riportate nei dati ufficiali (ossia le vendite effettive). Lo scopo delle riviste è quello di affermare di vendere più dell'effettivo in quanto più diffusione attrae più inserzioni, commettendo però operazioni illegittime.

La tabella riporta le vendite annuali (esprese in migliaia di euro) per 10 riviste americane. La variabile indipendente, X , si riferisce alle vendite dichiarate, mentre la variabile dipendente, Y , a quelle ufficiali:

Rivista	X_i	Y_i
YM	621	299.6
CosmoGirl	35907	207.7
Rosie	530	325
Playboy	492.1	336.3
Esquire	70.5	48.6
Teen People	567	400.3
More	125.5	91.2
Spin	50.6	39.1
Vogue	353.3	268.6
Elle	263.6	214.3

- Costruire il diagramma di dispersione
- Sapendo che $r=0.9494729$, calcolare R^2 e interpretarne il significato

- c. Eseguire un Test t per il coefficiente angolare, verificando se la relazione lineare tra X e Y è significativa ad un livello fissato $\alpha = 0.05$