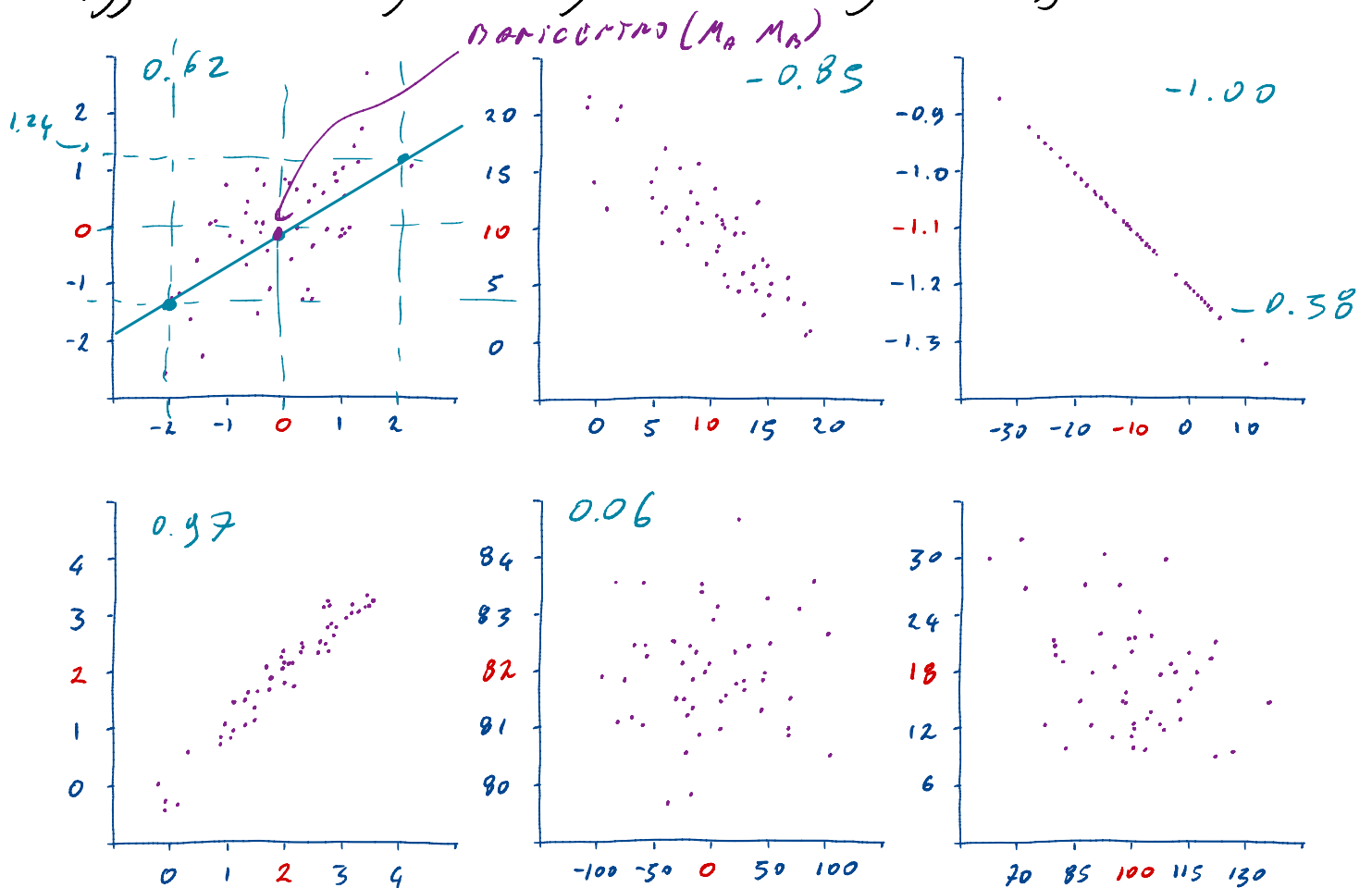


E_s D

I coefficienti di correlazione corrispondenti ai seguenti diagrammi di dispersione sono

~~-0.85~~ ~~-0.38~~ ~~-1.00~~ ~~0.06~~ ~~0.97~~ ~~0.62~~

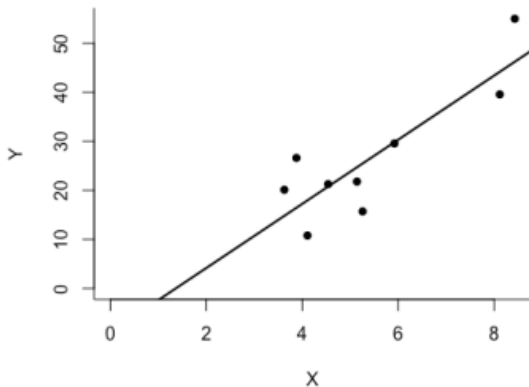
Associa ad ogni diagramma il suo coefficiente



I valori indicati sugli assi sono spazati di una deviazione standard delle rispettive variabili, con la media posta al centro. Per ciascuno dei diagrammi, calcola l'equazione della retta di correlazione e traccia questa retta sul diagramma stesso.

REGRESSIONE LINEARE

Esercizio 1



La relazione fra due variabili con N=9 soggetti è rappresentata nel grafico.

- a) Quale fra le tre seguenti relazioni lineari fra Y e X è la più plausibile:

$Y \approx -8.89 + 6.54 \cdot X$

$Y \approx +8.89 + 6.54 \cdot X$

$Y \approx -8.84 - 6.54 \cdot X$

- b) Usando i valori dei coefficienti della relazione più plausibile e sapendo che l'errore standard del coefficiente angolare è pari a 1.49, si può affermare con un livello di significatività dello 0.01 che c'è associazione lineare fra Y e X?
- c) Sapendo che $SDx=1.76$ e $cor(x,y)=0.86$ si calcoli SDy .

p-value

$[Y \sim -8.89 + 6.54 \cdot X; \text{si}; 13.4]$

Esercizio 2

La retta di regressione lineare della pressione sistolica in funzione dell'età, dell'indice di massa corporea (BMI) e del consumo di alcol (misurato in ml di alcol etilico per settimana), stimata su un campione di 346 uomini, ha la seguente equazione:

$\text{pressione sistolica} = 73.5 + 0.86 \cdot \text{età} + 0.80 \cdot \text{BMI} + 0.011 \cdot \text{alcol}$

- a) All'aumentare dell'età di 5 anni, la pressione sistolica aumenta o diminuisce? Di quanto?
- b) Qual è la pressione sistolica predetta dal modello di regressione per un uomo di 50 anni, un BMI di 25 e un consumo di alcol pari a 200 ml di alcol etilico per settimana?
- c) Sapendo che l'errore standard del coefficiente del BMI è 0.32, l'effetto del BMI è significativamente diverso da zero a un livello di significatività del 5%?
- d) Qual è il p-value del test sul coefficiente del BMI?

$[4.3; 138.7; \text{si}; 0.012]$

Esercizio 3

In uno studio che ha lo scopo di determinare se le infezioni orali in età pediatrica sono associate a fattori di rischio cardiovascolare in età adulta, si misura lo spessore intima-media (IMT) dell'arteria carotide in un campione di 755 soggetti e si effettua una regressione lineare con una variabile indice standardizzato di presenza di infezioni orali in età pediatrica (X) ottenendo un coefficiente angolare di 0.158 il cui errore standard è 0.055.

- Di quanto differisce lo spessore IMT fra due soggetti aventi X pari a 1 e -1, rispettivamente?
- Si può affermare ad un livello di significatività dell'1% che esiste associazione lineare fra infezioni orali in età pediatrica e rischio di eventi cardiovascolari in età adulta?
- Quale è il p-value associato alla stima del coefficiente lineare di X?

[0.316; sì; 0.004]

Esercizio 4

Si valuta la dipendenza dall'età della distanza alla quale un segnale stradale è visibile in un gruppo di 10 automobilisti. Applicando un modello di regressione lineare ai dati, si ottengono le seguenti stime ed errori dei coefficienti della regressione:

intercetta: $152.7 \text{ m} \pm 11.1 \text{ m}$

coefficiente angolare: $-0.73 \text{ m/anno} \pm 0.22 \text{ m/anno}$

- Di quanto varia la distanza di visibilità in un individuo dopo 10 anni dal conseguimento della patente di guida?
- Sapendo che $SD_{età} = 20.3$ e $SD_{distanza} = 19.7$, quale è la correlazione fra distanza ed età?
- Si può affermare con un livello di significatività di 5% che la distanza di visibilità dipende linearmente dall'età? e al livello di significatività di 1%?

[-7.3; -0.75; sì, no]

$$B^* = r A^*$$

$$\frac{B - \mu_B}{\sigma_B} = r \frac{A - \mu_A}{\sigma_A}$$

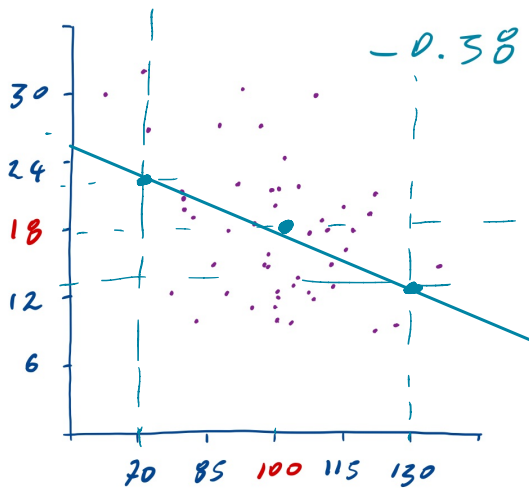
$$r = 0.62$$

$$\mu_A = \mu_B = 0$$

$$\sigma_A = \sigma_B = 1$$

$$\frac{B - 0}{1} = 0.62 \frac{A - 0}{1}$$

$$B = 0.62 A$$



$$r = -0.38$$

$$\mu_A = 100 \quad \sigma_A = 15$$

$$\mu_B = 18 \quad \sigma_B = 6$$

$$B^* = r A^* \quad \frac{B - \mu_B}{\sigma_B} = r \frac{A - \mu_A}{\sigma_A}$$

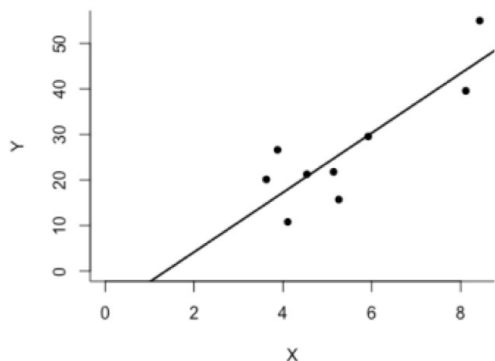
$$= -0.152 \cdot A + 33.2$$

$$\frac{B - 18}{6} = -0.38 \frac{A - 100}{15}$$

$$B = -0.38 \frac{6}{15} A + 18 + 0.38 \frac{100 \cdot 6}{15}$$

$$= -0.152 \cdot A + 33.2$$

Esercizio 1



La relazione fra due variabili con $N=9$ soggetti è rappresentata nel grafico.

- a) Quale fra le tre seguenti relazioni lineari fra Y e X è la più plausibile:

$Y \approx -8.89 + 6.54 \cdot X$

$Y \approx +8.89 + 6.54 \cdot X$

$Y \approx -8.84 - 6.54 \cdot X$

- b) Usando i valori dei coefficienti della relazione più plausibile e sapendo che l'errore standard del coefficiente angolare è pari a 1.49, si può affermare con un livello di significatività dello 0.01 che c'è associazione lineare fra Y e X ?

- c) Sapendo che $SDx=1.76$ e $cor(x,y)=0.86$ si calcoli SDy .

p-value

$[Y \sim -8.89 + 6.54 \cdot X; \text{si: } 13.4]$

$$Y = -8.89 + 6.54 X$$

$$\sigma_x = 1.76 \quad r = 0.86$$

$$\sigma_y ?$$

$$\frac{Y - \mu_Y}{\sigma_Y} = r \frac{X - \mu_X}{\sigma_X}$$

$$Y = \underbrace{\mu_Y - r \frac{\sigma_Y}{\sigma_X} \mu_X}_{-8.89} + \underbrace{r \frac{\sigma_Y}{\sigma_X}}_{6.54} X$$

$$0.86 \frac{\sigma_Y}{1.76} = 6.54$$

$$\sigma_Y = \frac{6.54 \cdot 1.76}{0.86}$$

$$= 13.38$$

$$t = r \sqrt{\frac{n-2}{1-r^2}} = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{r}{SE(r)}$$

$$t = 0.86 \sqrt{\frac{7}{1-0.86^2}} \approx 4.46 > 3.50$$

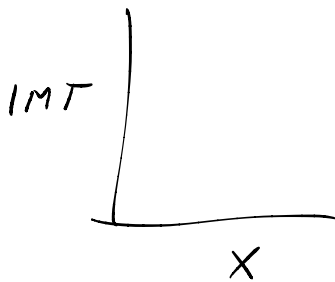
⌞

$$p\text{-val.} < 0.01$$

Esercizio 3

In uno studio che ha lo scopo di determinare se le infezioni orali in età pediatrica sono associate a fattori di rischio cardiovascolare in età adulta, si misura lo spessore intima-media (IMT) dell'arteria carotide in un campione di 755 soggetti e si effettua una regressione lineare con una variabile indice standardizzato di presenza di infezioni orali in età pediatrica (X) ottenendo un coefficiente angolare di 0.158 il cui errore standard è 0.055.

- Di quanto differisce lo spessore IMT fra due soggetti aventi X pari a 1 e -1, rispettivamente?
- Si può affermare ad un livello di significatività dell'1% che esiste associazione lineare fra infezioni orali in età pediatrica e rischio di eventi cardiovascolari in età adulta?
- Quale è il p-value associato alla stima del coefficiente lineare di X?



$$IMT = X \cdot 0.158 + C$$

$$IMT_1 = 1 \cdot 0.158 + C$$

$$IMT_{-1} = (-1) \cdot 0.158 + C$$

$$IMT_1 - IMT_{-1} = 0.158 + C - (-0.158) - C$$

$$= 0.316$$

$$t = \frac{0.158}{0.055} = 2.87$$

Statistica descrittiva e tabelle di contingenza

ESERCIZIO 1 (PS1-08/05/2018)

Su un campione di 11 persone si rileva il numero di porzioni medio settimanale di frutta, ottenendo i seguenti valori:

14, 12, 11, 7, 14, 2, 7, 14, 6, 7, 7

- Si calcolino media, moda e mediana della distribuzione
- Si calcolino il primo ed il secondo terzile della distribuzione campionaria (T1 e T2)
- Si costruisca la tabella delle frequenze assolute e relative delle categorie di consumo di frutta secondo le seguenti categorie: $\leq T1$; $(T1, T2]$; $>T2$ dove T1 è il primo terzile e T2 il secondo terzile

[9.2, 7, 7; 7, 12;

Numero di porzioni	Frequenza assoluta	Frequenza relativa
0-7	6	0.55
8-12	2	0.18
12+	3	0.27

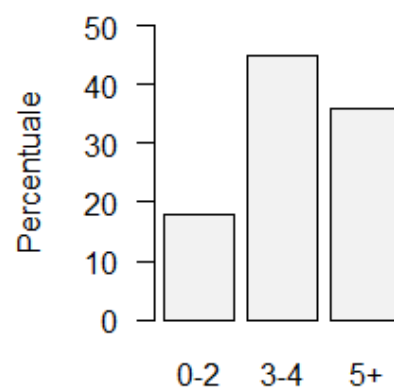
ESERCIZIO 2 (PS1-26/02/2018)

Ad un gruppo di donne di 65 anni si chiede quante mammografie hanno eseguito fino a quel momento e si ottengono le seguenti risposte: 4, 5, 6, 2, 3, 1, 4, 5, 3, 3, 8.

- si calcoli la SD della distribuzione
- si calcoli il range interquartile della distribuzione
- si categorizzi il numero di mammografie eseguite nelle seguenti categorie: ≤ 2 , 3 o 4, 5+ e si costruisca la tabella delle frequenze assolute e delle percentuali dei valori
- Si costruisca un diagramma a barre delle percentuali della variabile peso categorizzata

[1.9; 3-5;

Numero di mammografie	N (%)
0-2	2 (18)
3-4	5 (45)
5+	4 (36)



ESERCIZIO 3 (PS1-6/07/2018)

Per curare una certa patologia si utilizzano prevalentemente 3 farmaci che indichiamo con A, B e C. Da uno studio condotto su 400 pazienti di cui 40% sotto i 50 anni, risulta che il farmaco A è assunto da 60 sotto i 50 anni e da 40 con 50 anni o più; che il farmaco B è utilizzato da 50 pazienti sotto i 50 anni ed il farmaco C da 100 pazienti con 50 anni o più.

- Si costruisca la tabella di contingenza dei dati
- Quale è la probabilità che un paziente affetto dalla patologia in questione utilizzi il farmaco C?
- Quale è la probabilità che un paziente sotto i 50 anni affetto dalla patologia in questione utilizzi il farmaco C?
- Se una farmacia che serve una popolazione di 10000 abitanti deve acquistare una scorta dei tre farmaci per un mese, sapendo che la prevalenza della patologia in questione è del 2%, quante scatole contenenti la dose mensile dovrà acquistare per ciascuna tipologia?

	A	B	C	
<50 anni	60	50	50	160
>=50 anni	40	100	100	240
	100	150	150	400

; 0.38; 0.31; A:50, B:75, C:75]

Da: Li, Q et al (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*.

Table 1. Characteristics of Patients with Novel Coronavirus–Infected Pneumonia in Wuhan as of January 22, 2020.*

Characteristic	Before January 1 (N=47)	January 1 –January 11 (N=248)	January 12 –January 22 (N=130)
Median age (range) — yr	56 (26–82)	60 (21–89)	61 (15–89)
Age group — no./total no. (%)			
<15 yr	0/47	0/248	0/130
15–44 yr	12/47 (26)	39/248 (16)	33/130 (25)
45–64 yr	24/47 (51)	106/248 (43)	49/130 (38)
≥65 yr	11/47 (23)	103/248 (42)	48/130 (37)
Male sex — no./total no. (%)	31/47 (66)	147/248 (59)	62/130 (48)
Exposure history — no./total no. (%)			
Wet market exposure	30/47 (64)	32/196 (16)	5/81 (6)
Huanan Seafood Wholesale Market	26/47 (55)	19/196 (10)	5/81 (6)
Other wet market but not Huanan Seafood Wholesale Market	4/47 (9)	13/196 (7)	0/81
Contact with another person with respiratory symptoms	14/47 (30)	30/196 (15)	21/83 (25)
No exposure to either market or person with respiratory symptoms	12/47 (26)	141/196 (72)	59/81 (73)
Health care worker — no./total no. (%)	0/47	7/248 (3)	8/122 (7)

* Reduced denominators indicate missing data. Percentages may not total 100 because of rounding.

- In quale dei tre periodi di comparsa della polmonite dovuta a 2019-NCoV è maggiore l'età mediana dei pazienti?
- Qual è la classe modale di età prima del 1° gennaio?
- Che tipo di variabile è il sesso? Qual è la percentuale delle femmine prima del 1° gennaio?
- La proporzione di pazienti che non erano stati esposti né a mercati né a persone con sintomi respiratori è maggiore prima del 1° gennaio o nel periodo 1-11 gennaio?

[12-22 gennaio; 45-64 anni; binaria, 34%; 1-11 gennaio]

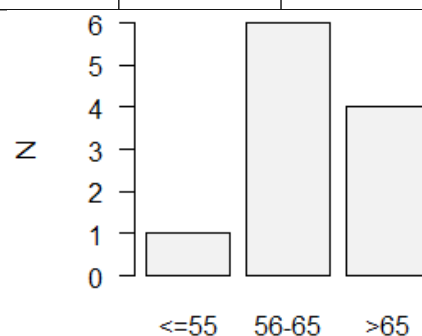
ESERCIZIO 4 (Inf-15122017)

Nella clinica ginecologica di un consultorio durante una giornata di visite si presentano 11 donne per le quali si misurano i seguenti pesi (kg): 59, 66, 70, 53, 67, 68, 62, 62, 62, 61, 63

- Quale è il peso medio del campione? Quale è il peso mediano del campione?
- Si categorizzi il peso nelle categorie ≤ 55 , (55,65], >65 e si costruisca la tabella delle frequenze assolute e delle frequenze relative della variabile risultante.
- Si costruisca un diagramma a barre delle frequenze assolute della variabile peso categorizzata

[63, 62;

Peso	Frequenza assoluta	Frequenza relativa
0-55	1	0.09
56-65	6	0.55
>65	4	0.36



ESERCIZIO 5 (PS1-6/07/2018)

In una scuola viene somministrato un test per valutare il livello di socializzazione degli studenti, ottenendo i seguenti punteggi:

8, 10, 7, 11, 10, 19, 2, 5, 20, 13, 11, 8, 2, 3, 4, 14, 10

- Si calcolino media, moda e mediana della distribuzione di frequenza
- Si calcoli la percentuale di soggetti che hanno realizzato un punteggio compreso fra 10 e 15 inclusi
- Si calcoli il primo terzile della distribuzione di frequenza

[9.2, 10, 10; 41%; 7]

ESERCIZIO 6 (PS1-26/02/2018)

In uno studio, 30 pazienti affetti da nevralgia del trigemino vengono suddivisi in due bracci di trattamento di uguale dimensione: i pazienti del primo gruppo ricevono un trattamento farmacologico a base di carbamazepina mentre i pazienti del secondo gruppo ricevono sia il trattamento a base di carbamazepina che un trattamento laser.

Sapendo che il 33% dei pazienti del primo gruppo sono uomini e che in totale ci sono 19 donne:

- Si costruisca la tabella di contingenza sesso x trattamento.
- Quale è la proporzione di donne nel primo gruppo di trattamento? E la percentuale di uomini del secondo?
- Estraendo a caso un paziente dal campione dello studio, quale è la probabilità che sia un uomo del secondo gruppo?

	Carbamazepina	Carbamazepina+lase	
--	---------------	--------------------	--

		r	
Uomini	5	6	11
Donne	10	9	19
	15	15	30

; 0.67, 0.40; 0.20]

ESERCIZIO 1 (PS1-08/05/2018)

Su un campione di 11 persone si rileva il numero di porzioni medio settimanale di frutta, ottenendo i seguenti valori:

14, 12, 11, 7, 14, 2, 7, 14, 6, 7, 7

- Si calcolino media, moda e mediana della distribuzione
- Si calcolino il primo ed il secondo terzile della distribuzione campionaria (T1 e T2)
- Si costruisca la tabella delle frequenze assolute e relative delle categorie di consumo di frutta secondo le seguenti categorie: $\leq T1$; $(T1, T2]$; $> T2$ dove T1 è il primo terzile e T2 il secondo terzile

2 6 7 7 7 7 11 12 14 14 14

↑

$M = 9.18$ mediana moda = 7

$$i\text{-esimo centile} = \frac{i \cdot (n+1)}{100} - \text{esimo individuo}$$

$$\text{terzile} = \frac{i (n+1)}{3}$$

$$1^{\circ} \text{ terzile} = 4^{\circ} \text{ valore} = 7 = T1$$

$$2^{\circ} \text{ terzile} = 8^{\circ} \text{ valore} = 12 = T2$$

	f. ass.	f. rel.
$- \leq 7$	6	55 %
$7 < - \leq 12$	2	18 %
$12 < -$	3	27 %

Distribuzione normale

ESERCIZIO 1 (Inf-02022018)

Si supponga che il punteggio di un questionario per quantificare la capacità di gestire lo stress nella popolazione sia distribuito normalmente con media 120 e deviazione standard 20.

- Quale è la probabilità che un individuo estratto a caso dalla popolazione riporti un punteggio inferiore a 100?
- Quale è la probabilità che un individuo estratto a caso dalla popolazione riporti un punteggio superiore a 140?
- Se si misura la capacità di gestire lo stress in 50 persone, quante ci si attende di trovarne con valori del questionario compresi fra 100 e 140?

[0.16; 0.16; 34]

ESERCIZIO 2 (Med AK - 25/9/2019)

Il livello di beta-hCG in una popolazione di donne con problemi di fertilità è distribuito normalmente con media 196 mIU/ml e deviazione standard 80 mIU/ml.

- Calcolare la probabilità che una donna abbia valori inferiori a 170 mIU/ml.

Se si estrae un campione di numerosità 100,

- quale è il numero atteso di soggetti con valori del marcatore superiore a 250?
- quale è la probabilità di trovare almeno 30 soggetti con valore del marcatore superiore a 250?

[0.37; 25; 0.13]

ESERCIZIO 3 (Med AK - 03/07/2019)

Un marcatore sierico X è distribuito normalmente nella popolazione sana con media $\mu=40$ e varianza $\sigma^2=100$.

- Calcolare il quartili della distribuzione, ossia i valori q_1 , q_2 , q_3 del marcatore per i quali $P(X \leq q_1)=0.25$, $P(X \leq q_2)=0.50$ e $P(X \leq q_3)=0.75$.
- Qual è la probabilità che un soggetto sano abbia un valore del marcatore inferiore a 60?
- Se si estrae un campione di numerosità 25, quale è il numero atteso di soggetti con valori del marcatore compresi fra 38 e 42?

[33.3, 40, 46.7; 0.977; 4]

Da: Saber H et al.(2017). Serum insulin-like growth factor 1 and the risk of ischemic stroke: the Framingham study. *Stroke*, 48(7), 1760-1765.

Table 1. Baseline Characteristics of Entire Study Sample

N	757
Age, mean±SD	79±5
Female, n (%)	470 (62)
Systolic blood pressure, mean±SD	143±21
Body mass index, mean±SD	26.78±4.68
Waist-hip ratio, mean±SD	0.93±0.09
Current smoking, n (%)	62 (8)
Diabetes mellitus, n (%)	86 (12)
Antihypertensive, n (%)	363 (49)
Prevalent CVD, n (%)	241 (32)
Prevalent AF, n (%)	81 (11)
IGF-1, mean±SD	145±60
Incident stroke, n (%)	119 (16)
Ischemic stroke, n (%)	99 (13)

AF indicates atrial fibrillation; CVD, cardiovascular disease; IGF-1, insulin like growth factor 1; N, number; and SD, standard deviation.

Data l'elevata numerosità del campione, consideriamo le medie $\bar{x}$ e le deviazioni standard s campionarie riportate in tabella come una buona approssimazione delle corrispondenti medie μ e deviazioni standard σ di popolazione. Assumendo che la pressione sistolica (SBP) sia distribuita normalmente,

- Qual è la probabilità per un soggetto di questa popolazione di avere un valore di SBP superiore a 130?
- Qual è il range interquartile di SBP?
- Qual è il valore soglia x di SBP tale che il 10% della popolazione ha un valore maggiore di x ?
- Se estraessimo a caso un campione di 40 individui, quanti fumatori ci aspetteremmo di trovare?

[0.73; 129, 157; 170; 3]

ESERCIZIO 4 (Inf-15122017)

Si vogliono selezionare un certo numero di infermieri per un corso di alta specializzazione. Per l'ammissione al corso si deve svolgere un test il cui punteggio è distribuito normalmente con media 45 e deviazione standard 2.

- Se al corso sono ammessi solo infermieri che riportano un punteggio superiore a 48, quale è la probabilità di essere ammessi?
- Se sono posti in lista di attesa per l'ammissione coloro che conseguono un punteggio compreso fra 47.5 e 48, quale è la probabilità di essere posti in lista di attesa?
- Supponendo che alla selezione partecipino 3600 infermieri, quanti ci si attende che riportino valori superiori a 48 e quindi vengano selezionati?

[0.067; 0.039; 241]

ESERCIZIO 5 (MedAK – 28/02/2018)

La distribuzione del livello totale di estradiolo nel siero in una popolazione di donne è normale con media 55 pmol/litro e SD 20 pmol/litro.

- a) Quale è la probabilità che una donna estratta a caso dalla popolazione abbia valori compresi fra 45 e 75 pmol/litro?
- b) Misurando i livelli di estradiolo nel siero in 300 donne, quante si attende di trovarne con valori superiori a 75 pmol/litro?
- c) Su un campione di 300 donne, qual è la probabilità di trovarne al massimo 35 con valori superiori a 75 pmol/litro?

[0.53; 48; 0.020]

ESERCIZIO 6 (Med AK - 8/5/2019)

I valori del colesterolo LDL in una popolazione di anziani ospedalizzati sono normalmente distribuito con media e deviazione standard pari a 141 e 34 mg/dL. Secondo le linee guida, valori di LDL < 100 mg/dL sono da considerarsi ottimali, valori compresi fra 100 e 130 sono borderline e valori superiori a 130 sono considerati alti.

- a) Se si estrae a caso un soggetto dalla popolazione, quale è la probabilità che abbia valori borderline?
- b) In campione di 300 persone, quante se ne attendono avere valori alti del colesterolo LDL?

[0.26; 188]

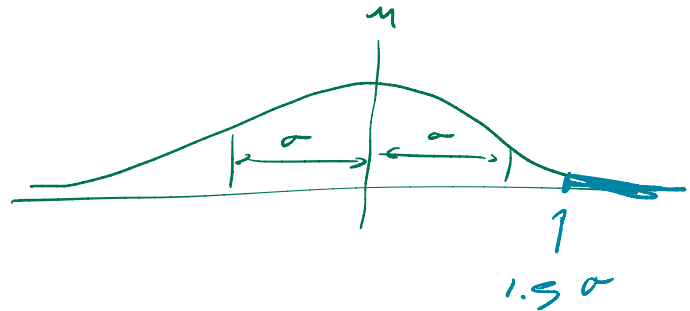
in modo gaussiano

ESERCIZIO 4 (Inf-15122017)

Si vogliono selezionare un certo numero di infermieri per un corso di alta specializzazione. Per l'ammissione al corso si deve svolgere un test il cui punteggio è distribuito normalmente con media 45 e deviazione standard 2.

- a) Se al corso sono ammessi solo infermieri che riportano un punteggio superiore a 48, quale è la probabilità di essere ammessi? 6.68 %
- b) Se sono posti in lista di attesa per l'ammissione coloro che conseguono un punteggio compreso fra 47.5 e 48, quale è la probabilità di essere posti in lista di attesa? 3.9 %
- c) Supponendo che alla selezione partecipino 3600 infermieri, quanti ci si attende che riportino valori superiori a 48 e quindi vengano selezionati? 2240

$$z_1 = \frac{x - \mu}{\sigma} = \frac{48 - 45}{2} = 1.5$$



$$z_2 = \frac{47.5 - 45}{2} = 1.25$$

$$P(\text{ind.} > 47.5) = \frac{100 - 28.82}{2} = 10.58 \%$$