

Termini e Concetti della Statistica

1. Per **statistica inferenziale**, s'intende:

- I metodi per organizzare, rappresentare e descrivere i dati.
- I metodi utilizzati per descrivere la distribuzione di probabilità.
- I metodi e le procedure utilizzate in un processo decisionale.**
- I metodi per estrarre un campione da una popolazione.

Statistica

- Statistica descrittiva
- Statistica inferenziale

2. In statistica, una popolazione è composta da:

- Tutte le persone che vivono in una nazione.
- Tutte le persone che vivono nella regione che vogliamo studiare.
- Tutti i soggetti o oggetti la cui caratteristica è motivo di studio.**
- Un numero limitato di elementi selezionati a scopo di studio.

3. Quale dei seguenti è un esempio di dati qualitativi?

- Il fatturato estero di un'azienda nel secondo trimestre del 2011.
- Il punteggio di laurea ottenuto dagli studenti della Laurea Magistrale.
- Il numero di visitatori entrati al parco di Mirabilandia nel 2010.
- Il trend osservato per il numero di immatricolazioni al CdL in Medicina negli ultimi anni.**

- Decrescente, Crescente, Stazionario,
- Stazionario o Invariato?

4. Quale delle seguenti affermazioni indica un campionamento **con e senza** reinserimento?

- Cinque amici noleggiavano una bicicletta per fare un giro in città (**con/senza**).
- La scelta casuale di uno studente durante il I giro di interrogazioni (**con/senza**).
- I numeri del lotto estratti sulla ruota di Venezia (**con/senza**).
- I numeri del lotto estratti sulle ruote di Milano, Venezia e Palermo (**con/senza**).

Mal interpretato, perché?
La formulazione corretta?

5. Una variabile è:

- Il risultato di un calcolo che varia da campione a campione.
- Una caratteristica oggetto studio, che assume valori diversi per elementi diversi.**
- Il campione prelevato da una popolazione in modo tale che tutti i suoi elementi siano diversi.
- Una caratteristica oggetto studio, che assume valori diversi per elementi identici.**

6. Un campione casuale, è un campione prelevato in modo tale che:

- Ogni membro della popolazione ha probabilità pari al 10% di essere incluso nel campione.
- Tutti gli elementi di una popolazione sono inclusi nel campione.
- Alcuni membri della popolazione non hanno alcuna possibilità di essere inclusi nel campione.
- Ogni membro della popolazione ha qualche possibilità di essere inclusi nel campione.**

7. Come parte di uno studio sulla produttività, al termine della settimana, chiediamo a tre dipendenti (Aldo, Giovanni e Giacomo) di indicare il tempo impiegato per la pausa caffè del mattino. Quale dei seguenti elenchi descrive gli elementi, del campione?

- Lunedì, Martedì, Mercoledì, Giovedì, Venerdì**
- 13,6 minuti
- 15 minuti, 12 minuti, 11 minuti, 14 minuti, 16 minuti
- Aldo, Giovanni, Giacomo

Esercizio 7.

- Variabile: pausa caffè mattutina.
- Misura: durata in minuti della pausa.
- Data set: valori dichiarati da Aldo, ... per ciascun gg della settimana.
- Elementi: lunedì, martedì, mercoledì,

8. La tabella indica il numero di passeggeri che hanno usufruito dei voli offerti dall'America Air Line, dal 1995 al 2004:

Anno	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Passeggeri (M)	4.4	4.7	5.1	5.4	5.9	6.9	6.9	7.6	8.2	9.0

Con riferimento alla Tabella indica cosa si intende per:

- Variabile **Passeggeri**
- Misura **No passeggeri, imbarchi, biglietti acquistati, M (milioni è?)**
- Data Set **Passeggeri per Anno**
- Elemento **Anno**

9. I valori in Tabella rappresentano una Sezione di Dati (SD) o una Serie Temporale (ST)? **Serie Temporale**

10. I passeggeri che fruiscono di uno sconto negli anni 2001 e 2002 sono una Sezione di Dati o una Serie Temporale? **SD**

Organizzazione e Rappresentazione dei Dati

11. I dati grezzi sono i dati che:

- a) Sono presentati sotto forma di una tabella delle frequenze.
- b) Sono disposti in un ordine casuale.
- c) Forniscono informazioni separate per ogni elemento oggetto di indagine.
- d) Sono disposti in modo ordinato crescente o decrescente.

} Ambiguità !!!!

12. In un istogramma delle frequenze, la frequenza di una classe è:

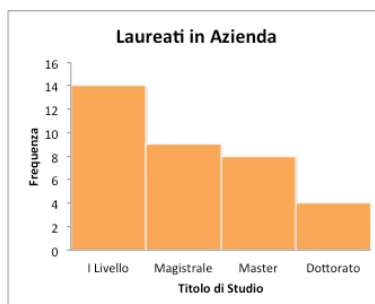
- a) L'altezza della barra corrispondente.
- b) La larghezza della barra corrispondente.
- c) L'altezza moltiplicata per la larghezza della barra corrispondente.
- d) L'altezza divisa per la larghezza della barra corrispondente.

13. Il grafico illustra la distribuzione delle frequenze del titolo di studio – laurea di I Livello, Magistrale, Master e Dottorato – conseguito da 35 ingegneri che lavorano in una grande azienda.

I - Il grafico è noto come:

- a) Grafico a barre
- b) Istogramma
- c) Box-and-Wisker plot
- d) Diagramma a Torta

} Ambiguità !!!!



II - La distribuzione delle frequenze ha forma:

- a) Uniforme
- b) Simmetrica
- c) Asimmetrica verso destra
- d) Asimmetrica verso sinistra

15. La tabella riporta l'elenco dei passeggeri del Titanic divisi per genere, età, classe di imbarco, ruolo e esito del naufragio. Indica quale/quali sono le variabili "esplicative" e quale/quali sono le variabili "risposta".

No	Cognome Nome	Genere	Età	Classe	Ruolo	Esito Naufragio
1	...	Maschio	Bambino	I	Passeggero	Sopravvissuto
2	...	Femmina	Adulto	III	Passeggero	Morto
3	...	Femmina	Adulto	III	Passeggero	Morto
4	...	Maschio	Adulto	Alloggi	Personale	Morto
5	...	Femmina	Bambino	II	Passeggero	Sopravvissuto
6	...	Maschio	Adulto	I	Passeggero	Sopravvissuto
...

- Variabili esplicative: Genere, Età, Classe, Ruolo
- Variabili risposta: Esito del Naufragio

16. Ad un gruppo di persone adulte è stato chiesto di indicare il loro stato di salute scegliendo la definizione opportuna fra quelle indicate in tabella. I valori numerici indicano la percentuale delle risposte

Eccellente	Molto buono	Buono	Discreto	Non buono
17	36.2	32.5	12	2.3

La variabile "stato di salute" è:

- a) Quantitativa
- b) Ordinale
- c) Categoriale
- d) Continua

17. Volendo rappresentare le risposte ottenute tramite un grafico, quale sceglieresti?

- 37. Istogramma
- 38. Dot-plot
- 39. Grafico a torta
- 40. Grafico a barre

18. La tabella riporta il fatturato trimestrale di tre grandi magazzini della città in due modi diversi. L'operazione che trasforma le variabili nelle colonne 1-3 nella rappresentazione di colonna 4-5 è nota come:

Iper Coop	Le Mura	Nuovo Doro	Fatturato	Magazzino
32	23	11	32	Iper Coop
35	25	13	35	Iper Coop
26	18	9	26	Iper Coop
29	20	12	29	Iper Coop
			23	Le Mura
			25	Le Mura
			18	Le Mura
			20	Le Mura
			11	Nuovo Doro
			13	Nuovo Doro
			9	Nuovo Doro
			12	Nuovo Doro

- a) Column Sorting
b) **Column Stack**
c) Column Unstack
d) Column Transpose

19. In tabella sono riportati i risultati di un'indagine eseguita su un campione 80 persone adulte ai quali è stato chiesto di indicare quante carte di credito possiedono.

Carte di Credito	Persone
0 - 3	18
4 - 7	26
8 - 11	22
15 - 16	11
16 - 21	3

La percentuale delle persone che possiedono un numero di Carte di credito **pari o superiore** a 8 è uguale a:

- a) 17.5 %
b) 82.5%
c) 27.5%
d) 22.0%

$$P(x \geq 8) = 45\%$$

$$P(x \geq 15) = 17.5\%$$

Risposta Non Data

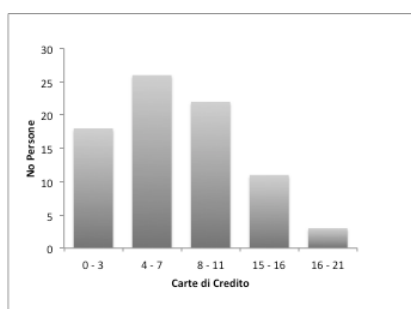
20. Nella tabella dell'esercizio precedente la variabile No di Carte di Credito è divisa in:

- a) Categorie
b) Intervalli
c) **Classi**
d) Valori

Il No di Carte di Credito è una **variabile quantitativa discreta**, e nella distribuzione delle frequenze è divisa in classi

Misure Numeriche Descrittive

21. Nell'istogramma che segue la barra corrispondente a un valore di frequenza pari a 26 rappresenta:



- a) La mediana dell'istogramma
b) La media dell'istogramma
c) **La moda dell'istogramma**
d) **Il valore massimo**

Errore!!

22. Per **outliers** o valore anomalo si intende

- a) Un valore che deve essere eliminato dal data-set
b) **Un valore molto diverso rispetto agli altri valori**
c) Un valore molto più grande dei restanti valori
d) Un valore affetto da un errore di misura

23. In un **data-set ordinato** i quartili separano il gruppo di dati

- a) **Quattro parti di uguale dimensione**
b) Novantanove parti di uguale dimensione.
c) n-parti di uguale dimensione
d) Quattro parti di dimensione diversa

- Un data set ordinato è rappresentato graficamente con un Box-Plot
- I quartili sono indicatori di posizione che dividono il data set in 4 parti, ciascuna di queste contiene un numero di valori pari ad 1/4 delle dimensioni del campione
- Le osservazioni con valori molto più piccolo o molto più grande dei restanti elementi del data set sono dette outliers.

24. La media dei valori di un data set di dati non raggruppati è:
- Il valore in posizione centrale in un insieme di dati ordinati.
 - La somma dei valori presenti nel data-set, divisa per il loro numero totale.**
 - La differenza tra il valore massimo e minimo presenti nel data-set.
 - La media delle deviazioni di ciascun valore dalla media.
25. La varianza di una data set è:
- La media al quadrato della distanza di ciascun valore, x_i , rispetto al valore medio**
 - La differenza fra i valori massimo e minimo
 - La somma delle differenze fra ciascun valore, x_i , e la media del data-set
 - La radice quadrata della devianza
26. Se tutti i valori in un data-set sono uguali, il valore della deviazione standard del data set è:
- positivo
 - maggiore di 1
 - zero**
 - tra zero e 1
27. Il tempo di pendolarismo (in minuti) da casa al luogo di lavoro per un campione di cinque persone è di 47, 53, 28, 31 e 16. La media calcolata per questi dati è una:
- regola di calcolo
 - distribuzione di frequenza
 - parametro
 - statistica**
28. Un **diagramma scatola baffi** è la rappresentazione di un data set ordinate I cui elementi sono:
- I valori massimo e minimo, il primo e terzo quartile e la mediana**
 - La variabile esaminata divisa in classi non sovrapposte e la rispettiva frequenza
 - I singoli valori della variabile esaminata e la rispettiva frequenza
 - Le sezioni di una circonferenza associate alle percentuali presenti nel data set

Diagramma scatola
baffi o Box-Plot

29. I valori in tabella riportano il numero di contravvenzioni ricevute da 13 cittadini nella loro carriera di autisti:

1	3	4	4	7	7	8	8	8	11	13	14	17
---	---	---	---	---	---	---	---	---	----	----	----	----

- La moda del data set è pari a 8
- La mediana del data set è pari a 8
- Il range del data set è pari a 16

Descriptive Statistics: Contravvenzioni

Variable	Total Count	Mean	Median	Range	Mode	N for Mode
Contravvenzioni	13	8.08	8.00	16.00	8	3

30. La media dei valori in tabella è:

1	3	4	4	7	7	8	8	8	11	13	14	17
---	---	---	---	---	---	---	---	---	----	----	----	----

- Circa 8**
- Circa 7
- Circa 11
- Uguale a 7

Probabilità, Variabili casuali e Distribuzione di Probabilità

31. Un campione è:
- L'insieme di molti spazi campionari
 - Un punto che rappresenta una popolazione in un campione
 - Un elemento dello spazio campionario
 - Una raccolta di osservazioni
32. Scelte a caso due famiglie, si osserva se possiedono una segreteria telefonica. Il numero totale dei risultati di questo esperimento è:
- sei
 - otto
 - due
 - quattro

33. Una agenzia di noleggio automobili ha 44 auto disponibili 18 delle quali è equipaggiata con un sistema di navigazione che utilizza il GPS. Se selezioniamo un'auto casualmente
- La probabilità di estrarre un'auto equipaggiata con il GPS è _____
 - La probabilità di estrarre un'auto non equipaggiata con il GPS è _____
 - La somma delle probabilità calcolate a punto a) e b) è uguale a _____
 - Gli eventi a) e b) sono detti eventi _____
34. Una società ha intervistato 25 persone di entrambi i generi per conoscere la loro opinione sulle recenti riforme previdenziali introdotte dal governo. La somma delle probabilità che gli intervistati siano favorevoli, o contrari può raggiungere un valore massimo pari a:
- 0,00
 - 0,85
 - 1,00
 - più di 1,0
35. Quale dei seguenti è un esempio di una variabile casuale discreta?
- Il peso di una scatola di biscotti
 - La lunghezza del telaio di una finestra
 - Il numero di cavalli di proprietà di un contadino
 - La distanza da casa al lavoro per un lavoratore

36. La tabella seguente riporta la distribuzione di probabilità di una variabile casuale discreta x .

x	0	1	2	3	4	5	6	7
$P(x)$	0.04	0.11	0.18	0.24	0.14	0.17	0.09	0.03

I – La Probabilità di $x = 3$ è:	a	b	c	d
	0.57	0.24	0.43	0.18

II – La Probabilità di $x \leq 5$ è:	a	b	c	d
	0.71	0.88	0.14	0.17

38. L'area totale sotto la curva della distribuzione di probabilità normale o di Gauss, $f(x)$, è:
- minore di 1
 - maggiore di 1
 - esattamente pari a 1
 - tra 0 e 1
39. Data la distribuzione normale standard, la variabile $z = (x-\mu)/\sigma$ indica la distanza tra la media e un punto dell'asse orizzontale in multipli o sottomultipli della:
- media
 - deviazione standard
 - varianza
 - centro della curva
40. Una variabile casuale che può assumere solo i valori "successo" e "fallimento" è detta
- Variabile standard
 - Variabile discreta
 - Variabile binomiale
 - Variabile continua

Stima di un parametro e Verifica di ipotesi

41. Dividi il numero n di elementi estratti da un campione per il numero totale, N , di elementi nel campione; il risultato della divisione è:
- La media del campione
 - Una proporzione del campione
 - La distribuzione del campione
 - La probabilità del campione

42. La distribuzione di probabilità della media campionaria \bar{X} è la:
- distribuzione di frequenza di \bar{X}
 - distribuzione di probabilità di \bar{X}
 - distribuzione campionaria di \bar{X}
43. distribuzione normale di X . L'età media di tutti gli studenti dell'Università di FE è di 24 anni. L'età media di un campione casuale di 100 studenti selezionati tra gli iscritti a Unife è pari a 23,6 anni. La differenza di $23,6 - 24 = 0,4$ è detta errore
- sistemático;
 - di campionamento;
 - di trascrizione;
 - di probabilità
44. La stima è una procedura con cui assegnare un valore numerico a:
- Un parametro di popolazione sulla base delle informazioni raccolte da un campione
 - Un campione statistico in base alle informazioni raccolte da un campione
 - Un parametro di popolazione sulla base delle informazioni raccolte da una popolazione
 - Un campione statistico in base alle informazioni raccolte da una popolazione
45. Si vuole ricavare una stima del tempo medio di attesa in coda al fast-food. Trenta clienti sono stati intervistati in momenti diversi del giorno, e hanno dato un'indicazione del tempo di attesa prima di effettuare l'ordine: la media di questi trenta valori è utilizzata per stimare il tempo medio di attesa per tutti i clienti, perché:
- è una stima priva di imprecisioni
 - è una stima puntuale
 - è una stima di intervallo
 - non abbiamo commesso errori
46. Quale step non fa parte della procedura utilizzata per la stima del valore di un parametro della popolazione: quale?
- Selezionare un campione rappresentativo
 - Raccogliere le informazioni da ciascun elemento del campione.
 - Calcolare del valore della statistica campionaria.
 - Calcolare il valore esatto del parametro della corrispondente popolazione.
47. Nella verifica di ipotesi sulla media o sulla proporzione di una popolazione, l'ipotesi nulla H_0 afferma che:
- La media è considerata falsa fino a che non viene dichiarata vero
 - La media è considerata vera fino a che non viene dichiarata a falsa
 - Il campione dal quale è calcolata la media si presume falso fino a che non è dichiarato vero
 - Il campione dal quale è calcolata la media si presume essere vero fino a che non è dichiarata falsa
48. Le cinque fasi elencate di seguito sono permettono di eseguire un test d'ipotesi utilizzando confrontando il valore della statistica del test con un valore critico. Indica l'ordine corretto di esecuzione
- Calcolare il valore della statistica del test.
 - Selezionare la distribuzione di probabilità per il test.
 - Prendere la decisione sulla base del confronto.
 - Determinare la regione di rigetto e non rigetto.
 - Definire l'ipotesi nulla H_0 e l'ipotesi alternativa H_1 .
- a) 4, 1, 2, 3, 5 b) 5, 2, 4, 1, 3 c) 5, 4, 2, 1, 3 d) 5, 1, 2, 3, 4
49. Quale fra le seguenti affermazioni è riferita al grafico
- Le due variabili x e y hanno una forte correlazione positiva.
 - Le due variabili x e y hanno una forte correlazione negativa.
 - Le due variabili x e y hanno una debole correlazione positiva.
 - Le due variabili x e y hanno una debole correlazione negativa.
50. L'analisi di regressione lineare:
- Individua una relazione lineare tra una variabile dipendente x e una variabile indipendente y .
 - Individua una relazione tra due variabili x, y che non necessariamente è una linea retta.
 - Individua una relazione lineare fra tre o più variabili.
 - Individua il grado di correlazione fra tre o più variabili.

