Statistica 1: Pat-Z
Esercitazione 3

1. Le osservazioni $x_1, \ldots, x_n$ hanno media 52, mediana 52.1 e una deviazione standard di 7. In base a queste informazioni, quali delle seguenti affermazioni descrive meglio i dati?

   (a) La distribuzione è asimmetrica positiva.

   (b) La distribuzione è asimmetrica negativa.

   (c) La distribuzione è simmetrica.

   Risposta: (c)

   l'indice normalizzato di asimmetria è definito come:

   $$A = \frac{\mu - M_e}{\sigma}$$

   dove μ rappresenta la media, M_e la mediana e σ la deviazione standard. Poichè

   $$|\mu - M_e| \leqslant \sigma \Rightarrow -1 \leqslant A \leqslant 1$$

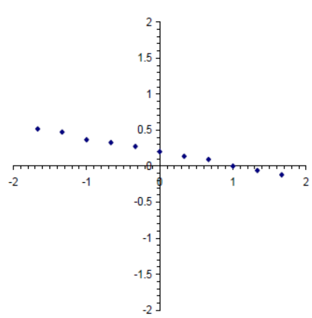
   e si avrà

   $$A < 0 \Rightarrow \text{asimmetria negativa}$$
   $$A = 0 \Rightarrow \text{simmetria}$$
   $$A > 0 \Rightarrow \text{asimmetria positiva}$$

   In questo caso abbiamo:

   $$A = \frac{52 - 52.1}{7} = 0.01$$

2. Si consideri il seguente diagramma di dispersione.



   La correlazione fra le variabili X e Y è vicina a:

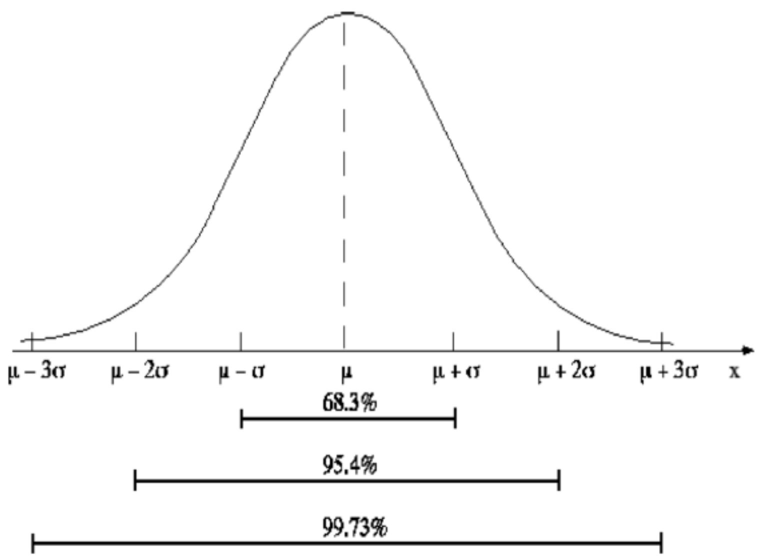
(a) 0.2
(b) -0.2
(c) 1
(d) -1
(e) 0

Risposta: (d)

infatti dall'analisi grafica si può notare come la relazione tra le variabili X e Y sia quasi perfettamente lineare, con pendenza negativa.

3. Si supponga che le osservazioni $x_1, \ldots, x_n$ abbiano distribuzione simmetrica con media 10. Si supponga che circa l' 84% delle osservazioni sia minore o uguale a 15. In accordo con la regola empirica, qual è un valore approssimato per la deviazione standard di queste osservazioni?

   Risposta: 5.
   Facendo riferimento al seguente grafico: Notiamo come l'84% circa dei dati sia minore della media più



   una deviazione standard. Per calcolare questa percentuale dobbiamo notare che il 68.3% dei dati si trova entro più o meno una deviazione standard dalla sua media, lasciando nelle code circa il 32%, ovvero il 16% per coda, in una distribuzione simmetrica. In questo caso dunque abbiamo che:

$$\mu + \sigma = 15$$

   ma dato che $\mu = 10$ questo ci da $\sigma = 5$.

4. Una proposta che ha ricevuto poca attenzione dalla Major League di baseball e quello di pagare i lanciatori secondo la seguente regola: ogni lanciatore riceve un salario di base di 4,25 milioni di dollari, meno 0,25 milioni di dollari per la sua media di lanci ottenuti (ERA) (A valori bassi di ERA sono associate delle prestazioni migliori). Se venisse adottata questa regola, quale sarebbe la correlazione tra i guadagni e l'ERA di un lanciatore? (Si supponga che ERA non possa superare il valore 17, in modo che non si abbiano dei guadagni negativi. Si puo anche assumere una deviazione standard di 1,2 per la variabile ERA.)

   Risposta: -1

   Definendo le varibili

$$Y = \text{Guadagni}$$

e

$$X = \text{ERA}$$

la loro relazione sarebbe definita dalla seguente relazione lineare:

$$Y = 4.25 - 0.25X$$

5. La scorsa settimana i 60 studenti della classe B hanno sostenuto gli esami di Finanza e Marketing. I risultati sono stati i seguenti:

   - Finanza: media = 25, deviazione standard = 2;
   - Marketing: media = 75, deviazione standard = 12;
   - Correlazione tra il punteggio in Finanza e il punteggio dello stesso studente in Marketing = 0.84.

   Mary, una studentessa della classe B, ha conseguito un 30 in Finanza e un 90 in Marketing. Siamo interessati a confrontare le sue prestazioni nei due esami in relazione al resto della classe. In modo particolare, vorremmo fare una dichiarazione su quale dei suoi punteggi si è posizionato più in alto rispetto agli altri punteggi dello stesso esame. Selezionare una delle seguenti opzioni e completare la frase.

   (a) Il punteggio di Mary in Finanza risulta probabilmente più alto del suo punteggio in Marketing in quanto ...

   (b) Il punteggio di Mary in Marketing risulta probabilmente più alto del suo punteggio in Finanza in quanto ...

   (c) Entrambi i punteggi ottenuti da Mary nei due esami risultano probabilmente alti perché...

   (d) Non possiamo fare nessun confronto sui punteggi ottenuti nei due esami perché...

   Risposta: (a) Il punteggio di Mary in Finanza risulta probabilmente più alto del suo punteggio in Marketing in quanto il suo punteggio in Finanza supera la media di più di due deviazioni standard, mentre il suo punteggio in Marketing è più alto della media poco più di una deviazione standard.

   In più possiamo immaginare che i punteggi si distribuiascono con una distribuzione normale e provare a standardizzare i due voti per confornti arli usando la classica formula:

$$Z = \frac{X - \mu}{\sigma}$$

   Nel primo caso avremo $Z = 2.5$, mentre nel secondo $Z = 1.25$.

6. Da un'indagine svolta presso una certa azienda è emerso che il 10% dei dipendenti sa usare Excel, il 20% Word, il 5% Power Point. Inoltre il 5% sa usare Word e Excel, il 3% Excel e Power Point, il 2% Power Point e Word e l'1% sa usare tutti e tre i programmi. Scegliendo a caso un dipendente, qual è la probabilità che sappia usare solo Word? E che sappia usare sia Excel che Power Point ma non sappia usare Word?

   Risposta:

   La probabilità che il dpiendente sappia usare solo Word (P_w) si ottiene considerando la percentuale di dipendenti che sanno usare Word e sottraendo la probilità che il dipendente sappia usare anche Excel e quella che sappia usare anche Excel e Power Point. Avremo dunque:

$$P_w = 0.2 - 0.05 - 0.01 = 0.14$$

   La probabilità che sappia usare Excel e Power Point, ma non Word (P_{ep}), si calcola considerando la probabilità che il dipendente sappia usare sia Excel che Power Point e sottraendo la probabilità che sappia usare tutti i programmi:

$$P_{ep} = 0.03 - 0 - 01 = 0 - 02$$

7. Da un'urna contenente 6 palline numerate da 1 a 6, se ne estraggono due con reinserimento. Descrivere uno spazio campionario per l'esperimento e calcolare la probabilita che la somma dei numeri sulle palline estratte:
   (a) sia 7 o 8?
   (b) sia 7 ottenuto con 2 seguito da 5?
   (c) sia 7 o 11? sia maggiore di 7? Ripetere l'esercizio nel caso in cui l'estrazione avvenga senza reinserimento.

   Risposta:

   Lo spazio campionario Ω è il seguente:

$$\begin{array}{c}
(1;1),(1;2),(1;3),(1;4),(1;5),(1;6)\\
(2;1),(2;2),(2;3),(2;4),(2;5),(2;6)\\
(3;1),(3;2),(3;3),(3;4),(3;5),(3;6)\\
(4;1),(4;2),(4;3),(4;4),(4;5),(4;6)\\
(5;1),(5;2),(5;3),(5;4),(5;5),(5;6)\\
(6;1),(6;2),(6;3),(6;4),(6;5),(6;6)
\end{array}$$

   Da qui è facile vedere che per la domanda (a) avremo 11 casi positivi:

$$\{(1;6),(2;5),(3;4),(4;3),(5;2),(6;1),(2;6),(3;5),(4;4),(5;3),(6;2)\}$$

   la probabilità sara' dunque data dagli esiti positivi sul totale degli esiti possibili, ovvero 11 su 36.

   (b): qui abbiamo ovviamente solo un esito positivo, ovvero $(2;5)$, per cui la probabilità sarà 1/36.

   (c) i casi in cui si abbia un 7 o un 11 sono:

$$\{(1;6),(2;5),(3;4),(4;3),(5;2),(6;1),(5;6),(6;5)\}$$

   la probabilità sarà dunque $8/36 = 2/9$
   La probabilità di avere un punteggio maggiore di 7 è data dai casi evidenziati in rosso, diviso il totale:

$$\begin{array}{c}
(1;1),(1;2),(1;3),(1;4),(1;5),(1;6)\\
(2;1),(2;2),(2;3),(2;4),(2;5),(2;6)\\
(3;1),(3;2),(3;3),(3;4),(3;5),(3;6)\\
(4;1),(4;2),(4;3),(4;4),(4;5),(4;6)\\
(5;1),(5;2),(5;3),(5;4),(5;5),(5;6)\\
(6;1),(6;2),(6;3),(6;4),(6;5),(6;6)
\end{array}$$

   Avremo dunque $15/36 = 5/12$

   Se l'esperimento viene condotto senza reinserimento, il nuovo spazio campionario sarà dato da:

$$\begin{array}{c}
(1;2),(1;3),(1;4),(1;5),(1;6)\\
(2;1),(2;3),(2;4),(2;5),(2;6)\\
(3;1),(3;2),(3;4),(3;5),(3;6)\\
(4;1),(4;2),(4;3),(4;5),(4;6)\\
(5;1),(5;2),(5;3),(5;4),(5;6)\\
(6;1),(6;2),(6;3),(6;4),(6;5)
\end{array}$$

e da qui è facile contare le occorrenze positive e dividerel per il totale dei casi, che non sarà più 36 ma 30, avremo quindi (a) = 10/30, (b)=1/30 e (c) = 8/30 e 12/30.

8. Un esperimento consiste nel chiedere a tre signore, scelte casualmente, se sono iscritte ad un social network.

   (a) Elencare gli elementi dello spazio campionario, usando le lettere Y per "si" e N per "no".

   (b) Elencare gli elementi di corrispondenti all'evento E = almeno due donne sono iscritte ad un social network".

   (c) Definire l'evento i cui elementi sono: $\{(Y;Y;Y);(N;Y;Y);(Y;Y;N);(N;Y;N)\}$

   Risposta:

   (a): $\Omega = \{(Y;Y;Y),(Y;Y;N),(Y;N;Y),(N;Y;Y),(Y;N;N),(N;Y;N),(N;N;Y),(N;N;N)\}$
   $(b) : E = \{(Y;Y;Y),(Y;Y;N),(Y;N;Y),(N;Y;Y)\}$
   (c): questi elementi si riferiscono all'evento E = almeno la seconda signora è iscritta a un social network"

9. Un oggetto su 3 prodotti in una catena di montaggio risulta difettoso. Se si prelevano 3 oggetti a caso, qual è la probabilità che esattamente uno di essi sia difettoso? E che almeno uno di essi sia difettoso?

   Risposta:

   La probabilità che esattamente uno sia difettoso si ottiene sommando la probabilità che solo il primo sia difettoso (P_1), alla probabilità che solo il secondo sia difettoso (P_2) e alla probabiltà che solo il terzo sia difettoso (P_3). Dato che

$$P_1 = 1/3 * 2/3 * 2/3 = 4/27$$

$$P_2 = 2/3 * 1/3 * 2/3 = 4/27$$

$$P_3 = 2/3 * 2/3 * 1/3 = 4/27$$

   la probabilità richiesta sarà data da $4/27 * 3 = 4/9$

   La probabilità che almeno uno sia difettose invece è ottenuta sommando la probabilità che solo uno di loro sia difettoso (appena calcolata), alla probabilità che due di loro siano difettosi (P_2), e quella che tutti e tre siano difettosi (P_3). Ora

$$P_2 = 1/3 * 1/3 * 2/3 + 1/3 * 2/3 * 1/3 + 2/3 * 1/3 * 1/3 = 6/27$$

   mentre

$$P_3 = 1/3 * 1/3 * 1/3 = 1/27$$

   dunque la probabilità richiesta sarà data da $12/27 + 6/27 + 1/27 = 19/27$

10. In un negozio ci sono 20 computer in vendita di cui 15 sono nuovi e 5 sono stati rinnovati. A prima vista i computer sono indistinguibili fra loro. Vengono acquistati 6 computer per il laboratorio di una scuola. Qual è la probabilità che fra i computer scelti ve ne siano 2 rinnovati?

    Risposta:

    Partiamo chiedendoci qual è la probabilità che dei 6 computer, esattamente i primi due siano rinnovati. Questa probabilità è data data da:

$$\frac{5}{20} \cdot \frac{4}{19} \cdot \frac{15}{18} \cdot \frac{14}{17} \cdot \frac{13}{16} \cdot \frac{12}{15} = 0,02348$$

Ora vogliamo sapere quante sono le possibili combinazioni di 6 computer in cui 2 qualsiasi siano difettose. Ricorrendo al calcolo combinatorio si sa che queste sono

$$\binom{6}{2} = \frac{6!}{(6-2)!2!} = 15$$

A questo punto la probabilità richiesta dall'esercizio si può ottenere considerando la probabilità che i 2 computer difettosi siano esattamente i primi 2 e moltiplicandola per il numero di possibili combinazioni di 2 computer difettosi su 6, ovvero 15. Avremo duqnue

$$0.02348 * 15 = 0.3522$$