

Statistica

Ede Boanini

31 dicembre 2025

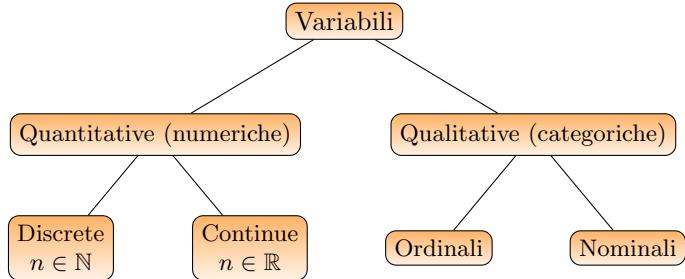
Indice

1	Introduzione	3
1.1	Classificazione delle Variabili	3
1.2	Distribuzioni di Frequenza	3
1.2.1	Tipi di Frequenza	4
2	Statistica Descrittiva	5
2.1	Diagrammi a barre vs Istogrammi	5
2.2	Media	6
2.3	Moda	6
2.4	Mediana	6
2.5	Quartili	7
2.6	Campo di Variazione / Range	8
2.7	Differenza Interquartile	8
2.8	Varianza	8
2.9	Deviazione standard	8
2.10	Coefficiente di variazione	8
3	Calcolo Combinatorio	9
3.1	Permutazioni Semplici	9
3.2	Permutazioni con Ripetizione	9
3.3	Disposizioni Semplici	10
3.4	Disposizioni con Ripetizione	11
3.5	Combinazioni	11
4	Probabilità	13
4.1	Operazione insiemi	13
4.2	Proprietà operazione tra eventi	13
4.3	Tipi di eventi	13
4.3.1	Eventi compatibili	13
4.3.2	Eventi incompatibili	13
4.3.3	Eventi complementari	13
4.4	Definizione	14

4.5	Assiomi	14
4.5.1	Conseguenze degli assiomi	14
4.5.2	Probabilità Totale	14
5	Probabilità Condizionata e Indipendenza	15
5.1	Probabilità Condizionata	15
5.2	Probabilità Congiunta	15
5.3	Eventi Indipendente ed Eventi Dipendenti	15
6	Variabili Casuali	15
6.1	Famiglie Parametriche	15
7	Inferenza Statistica	15
7.1	Stima Puntuale	15
7.2	Stima Intervallare	15
7.3	Verifica delle Ipotesi	15

1 Introduzione

1.1 Classificazione delle Variabili



Differenza tra ordinali e nominali:

- **Ordinali:** categorie che hanno un ordine, puoi solo dire se un valore è minore o maggiore rispetto ad un altro.
 - *Livello di istruzione: elementare < media < ...*
 - *Grado di soddisfazione: nullo < basso < medio < ...*
 - *Classifica di una gara: quinto < quarto < ...*
 - *Matricola: 17345 < 17346 < ...*
- **Nominali:** categorie che non hanno un ordine.
 - *Colore occhi: blu, verdi, marroni, ...*
 - *Genere: M, F*
 - *Marche auto: Toyota, Ford, ...*
 - *Nazionalità: Giapponese, Italiano, ...*

1.2 Distribuzioni di Frequenza

È una tabella che contiene modalità e frequenze.

Modalità di X (x_i)	Frequenze assolute f_i
x_1	f_1
x_2	f_2
...	...
x_n	f_n
N	

1.2.1 Tipi di Frequenza

1. **Frequenza assoluta:** numero di ripetizioni di una certa modalità (es: quanti studenti hanno preso 28 all'esame)

$$freq_{assoluta} = f_i$$

2. **Frequenza relativa:**

$$freq_{relativa} = \frac{f_i}{N}$$

3. **Frequenza percentuale:**

$$freq\% = \frac{f_i}{N} \cdot 100$$

oppure

$$freq\% = freq_{relativa} \cdot 100$$

4. **Frequenza cumulata:** somma progressiva delle frequenze assolute o relative.

$$freq_{cumulataAssoluta} = \sum_{i=1}^n f_i$$

$$freq_{cumulataRelativa} = \sum_{i=1}^n \frac{f_i}{N} = \sum_{i=1}^n freq_{relativa_i}$$

5. **Frequenza cumulata percentuale:**

$$freq_{cumulataAssoluta\%} = \sum_{i=1}^n f_i \cdot 100$$

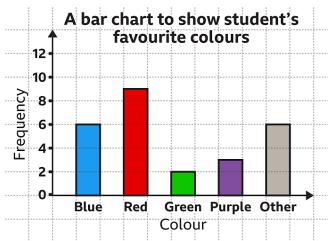
$$freq_{cumulataRelativa\%} = \sum_{i=1}^n \frac{f_i}{N} \cdot 100 = \sum_{i=1}^n freq_{relativa_i} \cdot 100$$

2 Statistica Descrittiva

2.1 Diagrammi a barre vs Istogrammi

Definizione 2.1 (Diagrammi a barre). Descrivono la distribuzione di frequenza di una o più variabili qualitative (categoriche). Le barre devono avere tutte la stessa base ed essere equi-spaziate (lasciare un pò di spazio tra una barra e l'altra).

- altezza barre: frequenza

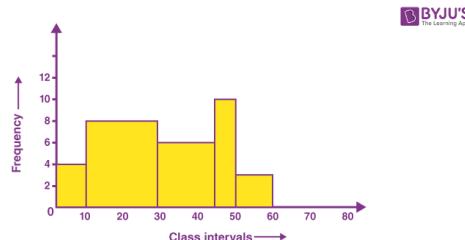


Definizione 2.2 (Istogrammi). Descrivono la distribuzione di frequenza di una o più variabili quantitative. Ogni barra rappresenta una classe e la sua frequenza.

- altezza barre: densità di frequenza

$$densità_{freq} = \frac{\text{Frequenza}}{\text{Ampiezza classe}}$$

- base barre: ampiezza delle classi



Osservazione: Definire k classi di uguale ampiezza

$$\text{Ampiezza classe} = \frac{\max - \min}{k}$$

I dati sulla statura di 48 adulti vanno da un minimo di 160 a 180 cm. Come fare k classi di ugual ampiezza?

1. Scelgo k (es: $k = 5$)
2. Uso formula $\text{Ampiezza classe} = \frac{\max - \min}{k}$ (es: $\text{Ampiezza classe} = \frac{180 - 160}{5} = 4$)

cm); quindi ogni classe avrà ampiezza 4.

3. Gli estremi inferiori delle classi sono (contando ampiezza 4):

- 160
- 164
- 168
- 172
- 176

Conclusione: le $k = 5$ classi di ugual ampiezza sono:

$$[160, 164), [164, 168), [168, 172), [172, 176), [176, 180]$$

2.2 Media

Qual è il centro dei dati? **valore tipico** attorno a cui si concentrano i dati.
 f_i indica la frequenza assoluta.

- **Formula della media per distribuzione di frequenze:** (variabili discrete)

$$\bar{x} = \frac{\sum_{i=1}^n (x_i \cdot f_i)}{N} = \sum_{i=1}^n (x_i \cdot freq_{relativa_i})$$

- **Formula della media per distribuzione di frequenze:** (variabili continue)

$$\bar{x} = \frac{\sum_{i=1}^n (m_i \cdot f_i)}{N} = \sum_{i=1}^n (m_i \cdot freq_{relativa_i})$$

dove a, b estremi dell'intervallo e $m_i = \frac{a+b}{2}$ il valore centrale della classe.

2.3 Moda

Qual è il centro dei dati? **valore tipico** attorno a cui si concentrano i dati.
La Moda è il valore che si ripete più spesso nei dati.

- **Formula della moda per distribuzione di frequenze:** (variabili discrete)

$$Moda = x_i \text{ con maggior frequenza}$$

- **Formula della media per distribuzione di frequenze:** (variabili continue)

$$Moda = \frac{a + b}{2}$$

Esempio Moda

Per esempio, per l'esame di analisi 2 ci sono stati tanti studenti che hanno preso tra il 20 e il 25 (classe), allora [20-25] è la classe modale. Pertanto, nel nostro esempio $Moda = \frac{20+25}{2} = 22.5$

2.4 Mediana

Qual è il centro dei dati? **valore tipico** attorno a cui si concentrano i dati.
La Mediana è il valore che è più grande (o uguale) della prima metà dei dati e allo stesso tempo, più piccolo (o uguale) della seconda metà dei dati.
È il **valore che sta in mezzo a dati ordinati**; quindi per poter stimare la Me è necessario ordinare i dati:

- **Formula della mediana per distribuzione di frequenze:** (variabili discrete)

1. Ordina i dati
2. Trova indice i :
 - se N pari: $i_1 = \frac{N}{2}$, $i_2 = \frac{N}{2} + 1$
 - se N dispari: $i = \frac{N}{2}$
3. La mediana è il valore associato all'indice trovato ($i = x_i$):
 - Se ho due indici i_1, i_2 , allora $Me = \frac{x_1+x_2}{2}$
 - Se ho un solo indice i , allora $Me = x_i$

- **Formula della mediana per distribuzione di frequenze:** (variabili continue)

1. Calcola frequenza cumulata di ogni classe

$$freq_{cumulataAssoluta} = \sum_{i=1}^n f_i$$

2. Trova indice i :
 - se N pari: $i_1 = \frac{N}{2}$, $i_2 = \frac{N}{2} + 1$
 - se N dispari: $i = \frac{N}{2}$
3. Osserva i in che classe cade (vedi frequenza cumulata), allora $Me = classe$.
Oppure, se abbiamo due indici i_1, i_2 con valori x_1, x_2 , allora $Me = \frac{x_1+x_2}{2}$

2.5 Quartili

Il p -esimo percentile è il valore che ha $\%p$ dei dati sotto/dietro di sé.

- Q_1 = 25-esimo percentile
(25% dei dati sotto questo valore)
- Q_2 = 50-esimo percentile = Mediana
(50% dei dati sotto questo valore)
- Q_3 = 75-esimo percentile
(75% dei dati sotto questo valore)

Divido la distribuzione in 4 parti uguali, per questo si chiamano "quartili".

- **Come trovare il Q_k per distribuzione di frequenze:** (variabili discrete)

1. Ordina i dati
2. Trova indice: $i = \frac{N+1}{4} \cdot k$
3. Q_k è il valore associato all'indice:
 - Se $i \in \mathbb{N}$, allora $Q_k = x_i$
 - Se $i \in \mathbb{Q}$, allora $Q_k = \frac{\text{somma dei valori associati}}{2}$

Esempio: se $i = 6.75$ allora $i_1 = 6, i_2 = 7$, e i valori associati a $i_1 = 20, i_2 = 25$, allora $Q_k = \frac{x_1+x_2}{2} = \frac{20+25}{2} = 22.5$

- **Come trovare il Q_k per distribuzione di frequenze:** (variabili continue)

1. Calcola frequenza cumulata di ogni classe

$$freq_{cumulataAssoluta} = \sum_{i=1}^n f_i$$

2. Trova indice: $i = \frac{N+1}{4} \cdot k$

3. Osserva i in che classe cade (vedi frequenza cumulata)

4. Allora avremo:

$$Q_k = L + \frac{i - f_{cumulata}}{f_i} \cdot h$$

dove:

- L : estremo inferiore della classe attuale (dell'indice)
- i : indice
- $f_{cumulata}$: frequenza cumulata classe precedente
- f_i : frequenza assoluta classe attuale
- h : ampiezza classe attuale

2.6 Campo di Variazione / Range

Distanza tra min e max.

$$Range = max - min$$

2.7 Differenza Interquartile

Si usano i quartili per capire la variabilità centrale.

$$IQR = Q_3 - Q_1$$

2.8 Varianza

Quanto sono variabili i dati? i dati sono vicini o molto sparsi.

Quanto i dati si allontanano dalla loro media (μ oppure \bar{x}).

- **Formula varianza per distribuzione di frequenze:** (variabili discrete)

- Varianza della popolazione (P):

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{N}$$

- Varianza campionaria ($C \subseteq P$):

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{N - 1}$$

2.9 Deviazione standard

Quanto sono variabili i dati? i dati sono vicini o molto sparsi.

$$\sigma = \sqrt{\sigma^2}$$

dove σ^2 è la varianza.

2.10 Coefficiente di variazione

Si calcola quando μ, \bar{x}, σ sono positivi e si esprime in percentuale.

$$CV = \frac{\sigma}{\mu} \quad \text{oppure} \quad CV = \frac{\sigma}{\bar{x}}$$

3 Calcolo Combinatorio

Ordine conta: $abc \neq cba$ vuol dire "2 modi diversi di ordinare gli elementi".

Ordine non conta: $abc = cba$ vuol dire "c'è solo 1 modo per ordinare gli elementi".

Trucco:

- **Permutazioni:** uso tutti gli n oggetti. Ordine conta
- **Disposizioni:** non uso tutti gli n oggetti ma solo r oggetti scelti dall'insieme dove $r < n$. Ordine conta
- **Combinazioni:** si parla di gruppi di k elementi. Ordine non conta

	Ordine conta?	Oggetti usati	Formula
Permutazioni	✓	Uso tutti gli n oggetti	$P_n = n!$
Disposizioni	✓	Uso r oggetti su n	$D(n, r) = \frac{n!}{(n - r)!}$
Combinazioni	✗	Uso gruppi di k oggetti su n	$C_{n,k} = \frac{n!}{r!(n - r)!}$

3.1 Permutazioni Semplici

Una permutazione semplice è un modo di ordinare in successione oggetti distinti (qui non esistono oggetti uguali tra loro, sono tutti distinti).

Teorema

Il **numero di permutazioni** di n oggetti distinti è il numero di modi diversi per ordinare tali oggetti.

$$P_n = n!$$

Esempio

Se ho 10 libri, allora avrò:

$$P_{10} = 10! = 3628800 \text{ modi diversi di ordinare 10 libri}$$

3.2 Permutazioni con Ripetizione

Una permutazione con ripetizione è un modo di ordinare oggetti tra cui alcuni uguali tra loro.

Teorema

Il **numero di permutazioni** di n oggetti alcuni uguali tra loro è il numero di modi diversi per ordinare tali oggetti.

$$P_{\substack{\text{numero di scatole} \\ \text{numero tot di oggetti}}} = P_n^r = \frac{n!}{r_1!r_2!\cdots r_k!}$$

Devo pensarla così:

1. Dividi gli oggetti distinti come se fossero scatole distinte, senza ripetizione (k = conta quante scatole sono)

- Inserisci ogni oggetto nella corrispettiva scatola
(r = conta quanti oggetti ha ogni scatola)

Esempio

Se io ho la parola STATISTICA, ho $n = 10$ allora:

- Dividi gli oggetti distinti come se fossero scatole distinte
(k = conta quante scatole sono)
S, T, A, I, C quindi $k = 5$ scatole
- Inserisci ogni ripetizione nella corrispettiva scatola
(r = conta quanti oggetti ha ogni scatola)
 - scatola S: la parola ha 2 "S" ripetute, quindi $r_1 = 2$
STATISTICA
 - scatola T: la parola ha 3 "T" ripetute, quindi $r_2 = 3$
STATISTICA
 - scatola A: la parola ha 2 "A" ripetute, quindi $r_3 = 2$
STATISTICA
 - scatola I: la parola ha 2 "I" ripetute, quindi $r_4 = 2$
STATISTICA
 - scatola C: la parola ha 1 "C" ripetuta, quindi $r_5 = 1$
STATISTICA

Quindi:

$$P_{10}^5 = \frac{10!}{r_1!r_2!r_3!r_4!r_5!} = \frac{10!}{2!3!2!2!1!} = 75600 \text{ modi diversi di ordinare la parola}$$

3.3 Disposizioni Semplici

Nel caso delle disposizioni, non uso tutti gli n oggetti (come nelle permutazioni) ma solo un sottoinsieme scelto k di n dove $k \leq n$.

Teorema

Il **numero di disposizioni** di oggetti scelti k tra n oggetti totali distinti è il numero di modi diversi per ordinare k oggetti.

$$D_{n,k} = \frac{n!}{(n-k)!}$$

- Scelgo k oggetti tra n oggetti totali
- $D_{n,k}$ è il numero di modi diversi per ordinare k oggetti

Esempio

In quanti modi diversi posso sistemare su una libreria 7 libri scelti da un insieme di 20 libri?

$$D_{20,7} = \frac{20!}{(20-7)!} = 390700800$$

3.4 Disposizioni con Ripetizione

Nel caso delle disposizioni con ripetizione, non uso tutti gli n oggetti (come nelle permutazioni) ma solo un sottoinsieme scelto k di n dove $k \leq n$.

Teorema

Il **numero di disposizioni con ripetizione** di oggetti scelti k tra n oggetti totali distinti è il numero di modi diversi per ordinare k oggetti in cui alcuni possono ripetersi nella stessa sequenza.

$$D_{n,k}^R = n^k$$

Esempio

Quante password di 5 caratteri si possono creare con un alfabeto di 26 lettere?

1. Scelgo $k = 5$ sottoinsieme di $n = 26$
2. Alcune lettere possono ripetersi nella stessa sequenza

$$D_{26,5}^R = 26 \cdot 26 \cdot 26 \cdot 26 \cdot 26 = 26^5 = 11881376$$

3.5 Combinazioni

Nel caso delle combinazioni, si parla di gruppi il cui numero corrisponde esattamente a quanti oggetti k ho scelto da n . La scelta k corrisponde solo al numero da cui è formato ogni gruppo.

Ovvero, se io scelgo 10 oggetti su 20, allora ogni gruppo dovrà avere esattamente 10 oggetti. L'ordine non conta perché l'importante è la presenza dell'oggetto all'interno del gruppo, che sia primo o ultimo non cambia niente.

La domanda è: quanti modi diversi ho di formare questi gruppi?

Teorema

Il **numero di combinazioni** è il numero di modi diversi per formare gruppi di k elementi ($k \leq n$).

$$C_{n,k} = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

1. Scelgo un numero k dove $k \leq n$
2. Ogni gruppo dovrà avere esattamente k oggetti
3. Quanti modi diversi ho di formare questi gruppi che contengono esattamente k oggetti?

Esempio

Ho 3 frutti diversi ma la io voglio fare merenda solo con 2. Quante tipe di merende posso creare con 2 frutti?

1. Scelgo $k = 2$ dove $2 < 3$
2. Ogni gruppo avrà esattamente 2 frutti

$$C_{3,2} = \binom{3}{2} = \frac{3!}{2!(3-2)!} = \frac{1 \cdot 2 \cdot 3}{2!1!} = \frac{6}{2} = 3$$

Ho 3 modi diversi di formare gruppi di 2 frutti, quindi ho 3 tipi di merende diverse.

Esempio

Voglio giocare a basket e devo formare 2 gruppi di persone per la partita. So che ogni squadra deve avere esattamente 5 giocatori. Le persone che si son presentate come candidate sono 40. Quanti modi diversi ho per formare una squadra? e per formarne due? sapendo che la persona non puo essere contemporaneamente scelta in entrambe?

1. Scelgo $k = 5$ dove $5 < 40$
2. Ogni gruppo avrà esattamente 5 persone

$$C_{40,5} = \binom{40}{5} = \frac{40!}{5!(40-5)!} = 657708 \text{ modi diversi per formare una sola squadra}$$

Ho 657708 modi diversi di formare gruppi di 5 persone, quindi posso scegliere 1 tra 657708 da mandare in campo.

3. Se invece voglio formare due squadre (sapendo che la stessa persona non può essere in entrambe) ? quante scelte avrei?

$$\text{Prima squadra: } C_{40,5} = \binom{40}{5} = \frac{40!}{5!(40-5)!} = 657708$$

Dato che le persone non possono ripetersi, se io ho formato la prima squadra, avrò per la seconda 35-5 persone ancora disponibili (perchè 5 le ho già scelte nella 1^a squadra):

$$\text{Seconda squadra: } C_{35,5} = \binom{35}{5} = \frac{35!}{5!(35-5)!} = 324632$$

Quindi i diversi modi per formare le due squadre saranno in totale $657708 \cdot 324632 = 213846580000$

4 Probabilità

4.1 Operazione insiemi

$$\begin{aligned} A_1 \cup A_2 &= \{\omega \mid \omega \in A_1 \vee \omega \in A_2\} \\ A_1 \cap A_2 &= \{\omega \mid \omega \in A_1 \wedge \omega \in A_2\} \\ A_1 - A_2 &= \{\omega \mid \omega \in A_1 \wedge \omega \notin A_2\} \\ \bar{A} &= \{\omega \mid \omega \notin A\} \end{aligned}$$

4.2 Proprietà operazione tra eventi

Siano A_1, A_2, A_3 tre eventi.

- **Proprietà commutativa:** l'ordine non cambia il risultato

$$E_1 \cup E_2 = E_2 \cup E_1$$

$$E_1 \cap E_2 = E_2 \cap E_1$$

- **Proprietà associativa:**

$$(E_1 \cup E_2) \cup E_3 = E_1 \cup (E_2 \cup E_3)$$

$$(E_1 \cap E_2) \cap E_3 = E_1 \cap (E_2 \cap E_3)$$

- **Proprietà distributiva:**

$$(E_1 \cup E_2) \cap E_3 = (E_1 \cap E_3) \cup (E_2 \cap E_3)$$

$$(E_1 \cap E_2) \cup E_3 = (E_1 \cup E_3) \cap (E_2 \cup E_3)$$

- **Leggi di De Morgan:**

$$\overline{E_1 \cap E_2} = \overline{E_1} \cup \overline{E_2}$$

$$\overline{E_1 \cup E_2} = \overline{E_1} \cap \overline{E_2}$$

4.3 Tipi di eventi

4.3.1 Eventi compatibili

Due eventi che possono verificarsi congiuntamente.

$$A_1, A_2 \text{ compatibili} \iff A_1 \cap A_2 \neq \emptyset$$

4.3.2 Eventi incompatibili

Due eventi che non possono verificarsi congiuntamente.

$$A_1, A_2 \text{ incompatibili} \iff A_1 \cap A_2 = \emptyset$$

4.3.3 Eventi complementari

Due eventi che non possono verificarsi congiuntamente e tale che uno dei due si verifica di sicuro.

$$A_1, A_2 \text{ complementari} \iff \begin{cases} A_1 \cap A_2 = \emptyset \\ A_1 \cup A_2 = \Omega \end{cases}$$

4.4 Definizione

Sia A un evento e Ω lo spazio campionario. Definisco $P(A)$ la probabilità che si verifichi A dove $0 \leq P(A) \leq 1$:

$$P(A) = \frac{\text{numero di casi favorevoli}}{\text{numero di casi possibili}}$$

4.5 Assiomi

1. **Primo assioma:** la probabilità di un evento A è un numero reale non negativo
2. **Secondo assioma:** la probabilità dell'intero spazio campionario è uguale a 1

$$P(\Omega) = 1$$

3. **Terzo assioma:** Se A_1, A_2 sono eventi **incompatibili**, allora la probabilità dell'unione dei due eventi è la somma delle loro probabilità

$$P(A_1 \cup A_2) = P(A_1) + P(A_2)$$

4.5.1 Conseguenze degli assiomi

1. **Probabilità del complementare di un evento:**

$$P(\bar{A}) = 1 - P(A)$$

2. **Probabilità dell'evento impossibile:**

$$P(\emptyset) = 0$$

3. **Proprietà di monoticità:** Se B è un evento incluso in un evento A , allora la probabilità di B è minore o uguale alla probabilità di A

$$B \subseteq A \implies P(B) \leq P(A)$$

4. **Probabilità dell'unione di eventi incompatibili:** la probabilità dell'unione di eventi incompatibili è la somma delle loro probabilità

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$$

4.5.2 Probabilità Totale

Teorema

Siano A_1, A_2 due eventi, la probabilità dell'unione dei due eventi è uguale alla somma delle due probabilità meno la loro intersezione:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

- Se i due eventi sono incompatibili $A_1 \cap A_2 = \emptyset$ allora,

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = P(A_1) + P(A_2) - \emptyset = P(A_1) + P(A_2)$$

(assioma 3)

- Se i due eventi sono compatibili $A_1 \cap A_2 \neq \emptyset$ allora,

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

(resta uguale)

5 Probabilità Condizionata e Indipendenza

5.1 Probabilità Condizionata

Teorema

Si definisce probabilità condizionata la probabilità che si verifichi un evento E sapendo che si è già verificato l'evento B :

$$P(E | B) = \frac{P(E \cap B)}{P(B)}$$

5.2 Probabilità Congiunta

5.3 Eventi Indipendente ed Eventi Dipendenti

6 Variabili Casuali

6.1 Famiglie Parametriche

7 Inferenza Statistica

7.1 Stima Puntuale

7.2 Stima Intervallare

7.3 Verifica delle Ipotesi