PROBABILITA' E STATISTICA 2013

Tendenza Centrale:

- ***Media:*** è calcolata sommando tutte le osservazioni in una serie di dati e dividendo per il numero totale delle misurazioni.
- ***Mediana:*** per dati ordinali, discreti e continui. La mediana è definita come il cinquantesimo percentile di una serie di misurazioni.
- ***Moda:*** è utilizzata come misura di sintesi per tutti i tipi di dati. La moda di una serie di valori è l'osservazione che si verifica con maggiore frequenza.

Misure di dispersione:

- ***Campo di variazione: (range)*** Un numero che può essere utilizzato per descrivere la variabilità in una serie di dati è il campo di variazione. Il campo di variazione o range di un gruppo di misurazioni è definito come la differenza tra l'osservazione più grande e quella più piccola.
- ***Campo di variazione interquartile:*** è calcolato sottraendo il venticinquesimo percentile dei dati dal settantacinquesimo percentile e comprende, pertanto, il 50% delle osservazioni centrali.
- ***Varianza:*** misura l'entità della variabilità o dispersione dalla media delle misurazioni.
  $$S^2=\frac{1}{n}\sum_{i=1}^{n}(xi-\bar{x})$$
- ***Deviazione standard:*** di una serie di dati è la radice quadrata della varianza.
  $$S=\sqrt{S^2}$$
- ***Coefficiente di variazione:*** mette in relazione la deviazione standard di una serie di valori con la sua media.
  $$CV=\frac{S}{\bar{x}}\times 100$$
- ***Media raggruppata:*** $$\bar{x}=\frac{\sum_{i=1}^{k} mi*fi}{\sum_{i=1}^{k} fi}$$
- ***Varianza raggruppata:*** $$S^2=\frac{\sum_{i=1}^{k}(mi-\bar{x})^2 fi}{(\sum_{i=1}^{k} fi)-1}$$

Evento: Un evento è l'elemento di base al quale può essere applicata la probabilità; esso è il risultato di un'osservazione o di un esperimento. Operazioni sugli eventi:

- ***L'intersezione*** di A e B, indicata come $A\cap B$, è definita come l'evento 'sia A che B'.
- ***L'unione*** di A e B, indicata come $A\cup B$, è l'evento 'A o B, o entrambi'.
- ***Il complemento*** di un evento A, indicato con A^c o $\bar{A}$, è l'evento 'non A'.

La probabilità di un evento A è la frequenza relativa con cui l'evento si verifica, o la proporzione di volte con cui l'evento si verifica, in una lunga serie di esperimenti ripetuti in condizioni virtualmente identiche.

Definizione Frequentista: Se un esperimento è ripetuto n volte in condizioni sostanzialmente identiche, e se l'evento A si verifica m volte, all'aumentare di n il rapporto m/n si avvicina ad un limite fisso che è la probabilità di A.

$$P(A)=\frac{m}{n}$$

Eventi:

- ***Mutuamente esclusivi:*** se due eventi non possono verificarsi contemporaneamente.
- ***Indipendenti:*** il verificarsi di uno non ha alcuna influenza sul verificarsi o non dell'altro.
  $P(A|B)=P(A)$
- ***Somma della Probabilità:*** la probabilità del verificarsi dell'uno o dell'altro evento è uguale alla somma delle probabilità di ciascuno dei due eventi.
  $P(A\cup B)=P(A)+P(B)$
- ***Prodotto della Probabilità:*** afferma che la probabilità che si verifichino entrambi gli eventi A e B è uguale alla probabilità di A moltiplicato la probabilità di B, dato che A si è già verificato.
  $$P(A|B)=\frac{P(A\cap B)}{P(B)}$$

Teorema di Bayes: consente di ricalcolare una probabilità in base a nuove informazioni

$$P(Ai|B)=\frac{P(Ai)P(B|Ai)}{P(A1)P(B|A1)+...+P(An)P(B|An)}$$

Variabile Casuale: di solito rappresentate con lettere maiuscole:
- ***Continue:*** può assumere un qualsiasi valore nell'ambito di uno specifico intervallo.
- ***Discrete:*** può assumere solo un numero finito o numerabile di valori.

Distribuzioni di Probabilità: applica la teoria della probabilità per descrivere il comportamento di una variabile casuale.
- Nel caso di variabili discrete, essa specifica tutti i possibili risultati della variabile casuale insieme alla probabilità che ciascuno di essi si verifichi.
- Nel caso di variabili continue, essa ci consente di determinare le probabilità associate a determinati range di valori.

Media della Popolazione: valore medio di una variabile casuale.

Varianza della Popolazione: è la dispersione dei valori relativi alla media della popolazione.

Deviazione standard della Popolazione: radice della varianza della popolazione.

Distribuzione Binomiale: si considera un variabile dicotomica Y, tale variabile può assumere solo valore 0/1. (variabile di Bernoulli)

$P(Y=1)=p$

$P(Y=0)=1-p$

$$P(X=x)=\frac{n!}{x!(n-x)!}p^x(1-p)^{(n-x)} \qquad P(X=x)=\binom{n}{x}p^x(1-p)^{(n-x)}$$

Distribuzione di Poisson: è utilizzata per modellare eventi discreti che si verificano raramente nel tempo o nello spazio; per questo motivo è talvolta chiamata la distribuzione di eventi rari.

Poisson implica:
- La probabilità che un singolo evento si verifichi in un intervallo è proporzionale alla lunghezza dell'intervallo.
- Teoricamente, in un singolo intervallo è possibile che l'evento si verifichi un numero infinito di volte. Non esiste un limite al numero degli esperimenti.
- Gli eventi si verificano indipendentemente nello stesso intervallo e tra intervalli consecutivi.

$$P(X=x)=\frac{e^{-\lambda}\lambda^x}{x!} \qquad \lambda=2{,}71828$$

Distribuzione Normale: distribuzione Gaussiana o curva a campana.
- Tra -1/1 copriamo il 68,2%

- Tra -2/2 copriamo il 95,4%

$$Z=\frac{X-\bar{x}}{s}$$

Inferenza Statistica: processo attraverso il quale si traggono conclusioni su un'intera popolazione in base ad un campione.

Teorema del limite centrale: consente di quantizzare l'incertezza insita nell'inferenza statistica senza fare molte assunzioni che non possono essere verificate.

- La media della distribuzione campionaria è uguale alla media **μ** della popolazione.
- La deviazione standard della distribuzione delle medie campionarie è uguale a $\sigma/\sqrt{n}$. Questa quantità è nota come *errore standard* della media.
- La forma della distribuzione campionaria è approssimativamente normale, posto che n sia sufficientemente grande.

$$Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$$

Stima:

- ***Puntuale:*** implica il calcolo di un singolo numero per stimare il parametro in esame.
- ***Intervallare:*** fornisce un range di possibili valori entro i quali si ritiene sia compreso il parametro in esame con un certo grado di confidenza. Questo range di valori è denominato *intervallo di confidenza*.

Intervalli di Confidenza Bilaterali: Per una variabile casuale normale standardizzata, il 95% delle osservazioni è compreso tra -1,96 e 1,96.

$$P=(-1{,}96\leq Z\leq 1{,}96)=0{,}95$$

Intervalli di Confidenza Unilaterali: Per una variabile casuale normale standardizzata, il 95% giace al di sopra di z=-1,645.

$$P=(Z\leq -1{,}645)=0{,}95$$

Distribuzione t di Student: si usa quando la media **μ** non è nota, quindi neanche **σ** non è nota.

$$t_{(n-1)}=\frac{\bar{X}-\mu}{s/\sqrt{n-1}}$$

Test Chi-quadrato: confronta le frequenze osservate in ciascuna categoria della tabella di contingenza, rappresentate da *O*, con le frequenze attese, posto che l'ipotesi nulla sia vera, indicate con *E*. Esso è utilizzato per stabilire se le differenze tra le frequenze osservate e quelle attese, *O-E*, siano troppo grandi per essere attribuite al caso.
La distribuzione chi-quadrato non è simmetrica. Una variabile casuale che segue una distribuzione chi-quadrato non può essere negativa; essa può assumere valori da zero ad infinito ed è asimmetrica a destra. Ancora una volta, tuttavia, l'area totale sotto la curva è uguale a 1.

$$X^2=\sum_{i=1}^{rc}\frac{(Oi-Ei)^2}{Ei}$$

Correlazione di Pearson: è lo stimatore della correlazione della popolazione o semplicemente *coefficiente di correlazione*, indicato con ***r*** ed è calcolato utilizzando la formula:

$$r=\frac{\sum_{i=1}^{rc}(xi-\bar{x})(yi-\bar{y})}{\sqrt{[\sum_{i=1}^{rc}(xi-\bar{x})^2][\sum_{i=1}^{rc}(yi-\bar{y})^2]}} \qquad t=r\sqrt{\frac{n-2}{1-r^2}}$$

t: errore standard stimato di *r*.