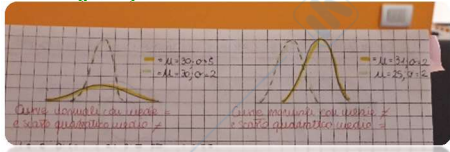
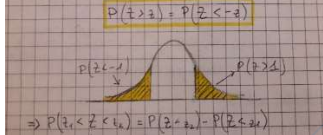


POPOLAZIONE	Insieme di individui/oggetti/valori che sono oggetto di analisi. Sono unità statistiche, sulla quale si osserva una variabile che è una caratteristica di ogni unità statistica appartenente alla popolazione
CAMPIONE	Sottoinsieme della popolazione. Non viene considerato come "unico", perché molti campioni differenti di popolazione possono essere presi in esame
ERRORE DI CAMPIONAMENTO	Differenza tra la caratteristica misurata su un'intera popolazione e la stessa riscontrata in un campione di quella stessa popolazione. Esso dipende dalla dimensione del campione e dalla variabilità che esiste nella popolazione
PARAMETRO	Valore numerico che descrive una caratteristica della popolazione. È una sola costante numerica
STIMATORE	Valore numerico che descrive una caratteristica di un campione. Può assumere valori diversi al variare del campione
N	Dimensione della popolazione (=numero unità statistica)
μ	Dimensione del campione
CAMPIONE CASUALE SEMPLICE	Scelta di un campione con estrazione tra le unità statistiche con la stessa probabilità di essere scelti
VARIABILI	Sono dati: qualitativi/ nominali/ordinali/quantitativi: continui-discreti
STATISTICA DESCRITTIVA	Permette di descrivere un campione o una popolazione sintetizzando i dati
STATISTICA INFERENZIALE	L'inferenza è fare previsioni, dedurre conclusioni sull'intera popolazione del campione. Si basa sulla teoria della probabilità, cioè verosimiglianza di osservare o selezionare un particolare campione della popolazione.
DISTRIBUZIONE DI FREQUENZA	Registra ogni categoria, valore, classe di valori, che una variabile potrebbe avere e il corrispondente numerico di volte che ognuna di esse ricorre nei dati. La frequenza della i-esima classe è indicata con f .
FREQUENZA RELATIVA (per dati qualitativi)	Consiste nel numero di valori in cui una osservazione si ritrova all'interno della classificazione stessa, rappresentata come una porzione del numero totale di osservazioni. La frequenza relativa può essere espressa come una frazione, decimale o percentuale. Si calcola: $rf_i = \text{frequenza } i\text{-esima classe } (f_i) / \text{numero totale delle osservazioni } (n)$
FREQUENZA RELATIVA COMULATA (per dati quantitativi)	È la somma della frequenza relativa di quella classe di tutte le classi precedenti. Rappresenta una porzione del numero totale delle osservazioni e può essere espressa come una frazione, un numero decimale o una percentuale
(Per dati continui)	Il numero di classi nella tabella di frequenza deve essere approssimativamente uguale alla radice quadrata del valore della dimensione del campione $\text{Numero della classi} = \sqrt{n}$
LE CLASSI	Le osservazioni devono ricadere in una sola delle classi definite. Le classi devono comprendere tutti i valori dell'insieme dei dati, e non devono essere sovrapposte
DIMENSIONE CLASSE	$\text{Scarto} = \text{valore massimo} - \text{valore minimo} (= \text{ampiezza della distribuzione})$ numero classi
FORMA DISTRIBUZIONE	Mostra come i dati sono distribuiti su ciascun lato rispetto al picco: simmetrica (se sono distribuiti uniformemente su entrambi i lati del picco) o asimmetrici (se non sono distribuiti uniformemente)
MEDIA ARITMETICA	Sintetizza la popolazione della distribuzione d'un insieme di dati. Si calcola: $\bar{x} = \frac{\text{somma di tutti i valori del campione}}{\text{Numero totale delle osservazioni}}$ Con il simbolo di sommatoria: $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ o $\sum_n x$ (media campionata) $\mu = \frac{\sum_{i=1}^n x_i}{n}$ (media della popolazione)
MEDIANA	Valore dell'osservazione centrale di una distribuzione ordinata ai dati Se n è dispari -> $n+1/2$ se n è pari -> $n/2$ e $n/2 + 1$

CONFRONTO MEDIA-MEDIANA	SE	DISTRIBUZIONE	
	MEDIA = MEDIANA	SIMMETRICA	
	MEDIA < MEDIANA	ASIMMETRICA A SX	
	MEDIA > MEDIANA	ASIMMETRICA A DX	
MODA	Valore con la frequenza più alta nel campione		
CLASSE MODALE	Intervallo di classe di una distribuzione di frequenza o di un istogramma che presenta la frequenza più alta. Si usa per i dati continui che possono assumere molti valori diversi		
CAMPO DI VARIAZIONE	È detto range, coinvolge i valori estremi della distribuzione: il più alto (MASSIMO) e il più basso (MINIMO). È la differenza fra l'osservazione maggiore e quella minore del campione. $R = \text{Max} - \text{Min}$.		
SCARTO QUADRATICO MEDIO	Detto devianza standard. È definito da una misura di dispersione detta varianza (s^2), che è la media dei quadrati degli scarti tra ciascun valore e la media campionaria. Lo scarto quadratico medio è la radice quadrata positiva della varianza (s) $\sigma^2 = \sum_{i=1}^N (X_i - \mu)^2$		
Regola empirica per interpretare lo scarto quadratico			
VALORI STANDARDIZZATI	Misura di quanti "scarti quadratici medi" un valore dista dalla media $Z = \frac{\text{scarto di un valore della media } (x - \mu)}{\text{Scarto quadratico medio } (\sigma)}$		
	Se il valore standardizzato è...	La regola empirica di che si verificherà...	Si può concludere che...
	Meno di -2 o più di 2	Circa il 5% delle volte	È insolito e forse è un valore anomalo
	Meno di -3 o più di 3	Meno dell'1% della volte	È molto insolito e probabilmente è un valore anomalo
COEFFICIENTE DI VARIAZIONE	Misura relativa della dispersione di un fenomeno $CV = \frac{\sigma}{\mu} \times 100$		
MISURE DI TENDENZA RELATIVA	3 TIPI: PRIMO QUARTILE (Q1) è un valore tale che il 25% dei dati è inferiore o uguale a esso. SECONDO QUARTILE (MEDIANA) 50-esimo percentile. TERZO QUARTILE (Q3) è un valore tale che il 75% dei dati è inferiore o uguale a esso		
SCARTO INTERQUARTILE	$Q_3 - Q_1$		
DATI BIVARIATI	Sono dati a due variabili, che si organizzano sulla base di una tabella di contingenza che è costituita da righe che rappresentano i possibili valori o categorie di una variabile e da colonne che rappresentano i possibili valori o categorie di una seconda variabile. I dati all'interno della tabella rappresentano il numero di volte in cui compare ogni coppia di valori o categorie. La variabile x è chiamata variabile indipendente , la variabile y è chiamata variabile dipendente .		

RETTE DEI MINIMI QUADRATI	Si usa se la relazione tra le variabili è lineare (una retta) e serve per trovare l'equazione della retta che meglio si adatta ai dati. Le distanze di ogni punto della retta sono chiamati scarti o errori. Il metodo dei minimi quadrati permette di trovare l'equazione della vita che minimizza la somma dei quadrati degli scarti tra i punti osservati e la retta. $\hat{y} = a + bx$. Lo scarto = $e = \hat{y} - y$
INTERPOPOLAZIONE	Si usa per predire i valori di y servendosi dei valori di x che si trovano all'interno dell'intervallo dei dati
ESTRAPOPOLAZIONE	Si usa per predire i valori di y attraverso i valori di x che si trovano al di fuori dell'intervallo dei dati
PROBABILITA'	Misura di quanto è verosimile il verificarsi di un evento, quindi di avere nel campione un certo risultato. Gli esiti possibili sono chiamati spazio campionario (=S) , che è l'insieme di tutti gli esiti possibili di un esperimento. L'esito può avere un numero finito (dati qualitativi o quantitativi) o infinito (dati qualitativi continui). La probabilità che un evento A si verifichi è indicata con P(A) $P(A) = \frac{\text{numero di casi in cui A può verificarsi } (n_A)}{\text{Numero totale di esiti possibili in S } (N)}$ $P(A) = 1 \rightarrow$ evento certo $P(A) = 0 \rightarrow$ evento nullo
STIMARE LA PROBABILITA'	Se i dati rappresentano fedelmente la popolazione si possono usare le frequenze relative come stime della vera probabilità per la popolazione. La probabilità calcolata in questo modo è chiamata probabilità empirica, che è calcolata dai dati campionari ed è una stima della vera probabilità.
LEGGI DELLA PROBABILITA'	Se vogliamo verificare due eventi contemporaneamente, dobbiamo considerare gli eventi $A \cup B = A \cup B$, che indica che A o B o entrambi possono verificarsi => UNIONE di due eventi / $A \cap B = A \cap B$, che indica che sia A che B si verificano => INTERSEZIONE di due eventi
REGOLA SEMPLICE DELLA SOMMA	$P(A \cup B) = P(A) + P(B)$ Due o più eventi mutuamente esclusivi
REGOLA GENERALE DELLA SOMMA	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ Gli eventi non sono mutuamente esclusivi
PROBABILITA' CONDIZIONATA	$P(A B) = \frac{P(A \cap B)}{P(B)}$ Probabilità che si verifichi A dato che si è verificato B
A e B sono indipendenti	$P(A \cap B) = P(A) \times P(B)$
DISTRIBUZIONE DI PROBABILITA'	$p(x) = P(X = x)$ per tutti i valori di x assegna una probabilità a ciascuno dei possibili valori che possono presentarsi
VARIABILE CASUALE BINOMIALE	Rappresenta il numero di successi in "n" prove o in un campione di dimensione "n"
DISTRIBUZIONE NORMALE	Curva normale = serve per identificare il valore della probabilità per un dato valore di x. Si basa su due parametri μ (= media della variabile casuale normale) e σ (= scarto quadratico medio) $X \sim N(\mu, \sigma)$ 
TABELLA NORMALE STANDARD	Contiene le probabilità per la variabile casuale Z. Per trasformare la variabile casuale x nella variabile nominale standard z: $Z = \frac{X - \mu}{\sigma}$

	
DISTRIBUZIONE CAMPIONARIE E GLI INTERVALLI DI CONFIDENZA	<p>STATISTICA INFERENZIALE ha come scopo quello di usare le informazioni campionarie per poter effettuare affermazioni probabilistiche sul comportamento della popolazione</p> <p>STIMATORE PUNTUALE = formula o regola usata per calcolare la stima puntuale per un particolare insieme di dati</p> <p>STIMA PUNTUALE = singolo numero calcolato dai dati campionari. È usato per stimare un parametro della popolazione</p>
TEOREMA CENTRALE DEL LIMITE (TCL)	<p>Nel campionamento casuale da una popolazione con media μ e scarto quadratico medio σ, quando n è abbastanza grande, la distribuzione di \bar{x} è</p> $\bar{x} = N(\mu; \sigma/\sqrt{n})$
FORMA DISTRIBUZIONE CAMPIONARIA	<p>Esaminiamo un istogramma delle osservazioni campionarie, che fornisce un quadro del campione e non della popolazione. Se il campione rappresenta graficamente appare a forma di campana, anche la popolazione sarà probabilmente a forma di campana.</p>
$\bar{x} = N(\mu; \dots)$	<p>Il centro dell'istogramma dei valori di \bar{x} deve essere μ. TCL afferma che se prendiamo tutti i possibili campioni di dimensione n, con n sufficientemente grande, calcoliamo per ogni campione la media campionaria e troviamo la media delle medie campionarie, otteniamo μ</p>
$\bar{x}(\mu; \frac{\sigma}{\sqrt{n}})$	<p>Lo scarto quadratico medio dei valori di \bar{x}, dipende da due fattori: il grado di variabilità della popolazione, σ, e la dimensione campionaria, n.</p> <p>All'aumentare dello scarto quadratico medio della popolazione aumenta l'errore standard di \bar{x}</p> $Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$
INTERVALLI DI CONFIDENZA PER LA MEDIA	<p>$P(I < \mu < S) = 1 - \alpha$</p> <p>l'intervallo di confidenza o stima intervallare è un intervallo di valore con associata una probabilità a livello di confidenza $1 - \alpha$. La probabilità quantifica la possibilità che l'intervallo contenga il vero parametro della popolazione</p>
INTERVALLO DI CONFIDENZA PER μ , POPOLAZIONE NORMALE, S.Q.M. NOTO CON $n > 1$	$I = \bar{x} - e = \bar{x} - \frac{z_{\alpha/2} \times \sigma}{\sqrt{n}}$ $S = \bar{x} + e = \bar{x} + \frac{z_{\alpha/2} \times \sigma}{\sqrt{n}}$
INTERVALLO DI CONFIDENZA PER μ , POPOLAZIONE NORMALE, S.Q.M. NON NOTO CON $n > 30$	$I = \bar{x} - e = \bar{x} - \frac{z_{\alpha/2} \times S}{\sqrt{n}}$ $S = \bar{x} + e = \bar{x} + \frac{z_{\alpha/2} \times S}{\sqrt{n}}$
INTERVALLO DI CONFIDENZA PER μ , POPOLAZIONE NON NORMALE, S.Q.M. NON NOTO CON $n < 30$	<p>Si usa la t di students $t_{(n-1)} = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$</p> $I = \bar{x} - e = \bar{x} - \frac{t_{\alpha/2, n-1} \times S}{\sqrt{n}}$ $S = \bar{x} + e = \bar{x} + \frac{t_{\alpha/2, n-1} \times S}{\sqrt{n}}$
INTERPRETARE L'INTERVALLO DI CONFIDENZA	<p>Quando aumentano il livello di confidenza (tipo dal 90% al 95%) aumentiamo la probabilità di avere un buon intervallo, ma allargandolo, prendiamo precisione</p>

STABIIRE LA MEDIA
CAMPIONARIA

Stimatore μ : $2e = \frac{2\sigma x Z_{\alpha/2}}{\sqrt{n}}$

$$n = \frac{(Z_{\alpha/2} x \sigma)^2}{(e)^2}$$

dove e è la massima distanza tra \bar{x} e μ -> errore di un valore assoluto

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari

www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari