

PROCEDIMENTI EXCEL

EXCEL

➤ Estendere la selezione al limite dei dati

CTRL + MAIUSC + tasto freccia

➤ Formule – operatori aritmetici

+	Addizione	=A1+B1 Somma i valori contenuti nelle celle A1 e B1
-	Sottrazione	=B2-B3 Sottrae dal valore contenuto nella cella B2 il valore contenuto nella cella B3
*	Moltiplicazione	=C1*4 Moltiplica per 4 il valore contenuto nella cella C1
/	Divisione	=B2/A1 Divide il valore contenuto nella cella B2 per quello contenuto nella cella A1
^	Elevamento a potenza	=H3^2 Eleva al quadrato il valore contenuto nella cella H3

➤ Riferimenti assoluti

\$ → blocca

➤ La funzione SUM (SOMMA)

=SUM (intervallo) → per selezionare intervallo uso :

➤ La funzione IF (SE)

=IF (test;[se_vero];[se_falso])

➤ La funzione ROUND (ARROTONDA)

=ROUND(num; num_cifre)

➤ La funzione CONCAT (CONCAT)

=CONCAT(testo1; [testo2]; ...)

➤ Ordinamento dei dati

1. Selezionare una cella all'interno dei dati.
2. Selezionare Home > Sort and Filter.
Oppure selezionare Data > Sort.
3. Selezionare un'opzione:

- Ordina dalla A alla Z dispone la colonna selezionata
- Ordina dalla Z alla A dispone la colonna selezionata
- Ordinamento personalizzato ordina i dati in più colonne applicando diversi criteri di ordinamento.

➤ **Filtraggio dati**

Data e cliccare su Filter (oppure Home > Sort and Filter > Filter).

Il comando trasforma ogni cella della prima riga in un menu a tendina. Cliccando sulla freccia, possiamo togliere la spunta dai valori che desideriamo escludere.

LEZIONE 2 – DISTRIBUZIONI DI FREQUENZA E RAPPRESENTAZIONI GRAFICHE

➤ **Distribuzioni di frequenza**

Il metodo più semplice per costruire una distribuzione di frequenza in Excel è una tabella Pivot.

- Per creare una Tabella Pivot in Excel aprire il foglio di lavoro che contiene i dati grezzi (es. dati1.xlsx) e selezionare Insert → PivotTable.

- Nella finestra di dialogo PivotTable:

1. Lasciate selezionata l'opzione Select a table or range e inserite l'intervallo delle celle (es. B1:B21).
2. Selezionate l'opzione New Worksheet e fate click su OK.

- Nella finestra PivotTable Fields:

1. Trascinate l'etichetta della variabile che deve essere analizzata e inseritela nel riquadro ROWS.
2. Trascinate una seconda volta questa stessa etichetta nel riquadro VALUES. Questa seconda etichetta viene trasformata trasformata in Count of (nome della variabile), a indicare che nell'area DATI sarà visualizzato un conteggio delle frequenze di ciascuna modalità.

- Per aggiungere una colonna con le **frequenze relative**:

1. Trascinate un'altra volta l'etichetta della variabile che deve essere analizzata nel riquadro VALUES.
2. Cliccare sulla tendina a fianco della seconda etichetta nel campo VALUES e selezionare nel successivo menu l'opzione Value Field Settings.
3. Nella successiva finestra di dialogo nella scheda Summarize Values By impostare Count e nella scheda Show Values As al posto di nessun calcolo impostare % of Grand Total.

➤ **Distribuzione di frequenza in classi**

1. Creare la tabella pivot (frequenze assolute) dei dati non raggruppati in classi.
2. Nella tabella pivot fare clic con il pulsante destro del mouse su un valore e scegliere Group.
3. Nella finestra Group modificare se necessario il valore minimo, il valore massimo e l'ampiezza nelle classi.
4. Selezionare OK.

➤ **Diagramma a torta**

- Selezionare la tabella delle frequenze (assolute o relative).
- Selezionare Insert → Insert Pie or Doughnut Chart e selezionare un grafico
- Fare clic sul grafico e quindi sulle icone accanto al grafico per modificarne l'aspetto.

➤ **Diagramma a barre**

- Selezionare la tabella delle frequenze (assolute o relative).
- Selezionare Insert → Insert Column or Bar Chart e selezionare un grafico
- Fare clic sul grafico e quindi sulle icone accanto al grafico per modificarne l'aspetto.

➤ **PivotChart e PivotTable**

Il comando PivotChart & PivotTable ci consente di creare simultaneamente una distribuzione di frequenza e la sua rappresentazione grafica partendo dallo stesso dataset.

Nella finestra PivotTable Fields:

1. Trascinate l'etichetta della variabile che deve essere analizzata e inseritela nel riquadro Axis (Categories).
2. Trascinate una seconda volta questa stessa etichetta nel riquadro VALUES per definire il tipo di frequenza desiderata.

➤ **Istogramma**

1. A partire dai dati grezzi

- Selezionare i dati.
- Selezionare Insert → Histogram e scegliere Histogram
- Per modificare il numero di classi cliccare con il tasto destro del mouse sull'asse x e selezionare Format Axis.

2. A partire dalle distribuzioni di frequenza

- Selezionare la tabella delle frequenze (assolute o relative).
- Selezionare Insert → Insert Column or Bar Chart e selezionare Clustered Column
- Cliccare con il tasto destro su una barra e selezionare Format Data Series dal menu di scelta rapida.
- Nel riquadro Format Data Series, modificare i dati in Gap Width per eliminare gli spazi tra le barre.

LEZIONE 3 – INDICI DI POSIZIONE E DI VARIABILITA'

➤ Indici di posizione

=AVERAGE(intervallo dei dati da analizzare)

Calcola la media

=MEDIAN(intervallo dei dati da analizzare)

Calcola la mediana

=MODE(intervallo dei dati da analizzare)

Calcola la moda

=QUARTILE.EXC(intervallo dei dati da analizzare; numero quartile)

Calcola i quartili

Utilizzare il numero 1 come *numero quartile* per calcolare il primo quartile, 2 per il secondo (la mediana) o 3 per il terzo.

➤ Indici di variabilità

=VAR.P(intervallo dei dati da analizzare)

Calcola la varianza di una popolazione

=STDEV.P(intervallo dei dati da analizzare)

Calcola la dev.st di una popolazione

=VAR.S(intervallo dei dati da analizzare)

Calcola la varianza di un campione

=STDEV.S(intervallo dei dati da analizzare)

Calcola la dev.st di un campione

➤ Funzioni statistiche di excel

INGLESE	ITALIANO
=AVERAGE()	=MEDIA()
=MEDIAN()	=MEDIANA()
=MODE()	=MODA()
=QUARTILE.EXC()	=ESC.QUARTILE()
=VAR.P()	=VAR.P()
=STDEV.P()	=DEV.ST.P()
=VAR.S()	=VAR.C()
=STDEV.S()	=DEV.ST.C()

➤ **Data analysis tool**

- Selezionare Data → Data Analysis
- Nella finestra di dialogo Data Analysis selezionare Descriptive Statistics e fare clic su OK.
- Nella finestra di dialogo Descriptive Statistics, inserire l'intervallo delle celle che contengono i dati nel campo Input Range.
- Selezionare NewWorksheet e Summary Statistics e cliccare OK

➤ **Descriptive Summary (PHStat)**

- Selezionare Add-Ins → PHStat → Descriptive Statistics → Descriptive Summary
- Nella finestra di dialogo Descriptive Summary, inserire l'intervallo delle celle che contengono i dati nel campo Raw Data Cell Range.
- Il comando Descriptive Summary accetta anche dati relativi a più gruppi.
- I dati relativi a diversi gruppi possono essere organizzati in due modi:
 - I. Multiple Groups – Unstacked – Più gruppi contenuti in più colonne, una per ogni gruppo.
 - II. Multiple Groups – Stacked – Più gruppi impilati in una sola colonna. In questo caso è necessario specificare dove è contenuta la variabile che definisce l'appartenenza ai gruppi (nel campo Grouping Variable Cell Range).

➤ **Creare un boxplot con outlier**

- Selezionare i dati
- Selezionare Insert → Recommended Charts
- Nella finestra di dialogo Insert Chart, selezionare All Charts → Box and Whisker e clic OK.

➤ **Creare un boxplot (PHStat)**

Selezionare Add-Ins → PHStat → Descriptive Statistics → Boxplot

- Nella finestra di dialogo Boxplot, inserire l'intervallo delle celle che contengono i dati nel campo Raw Data Cell Range.

Nota. Nello stesso grafico possono essere rappresentati più boxplot che sintetizzano la distribuzione del carattere considerato in gruppi diversi.

I dati relativi a diversi gruppi possono essere organizzati in due modi: Unstacked e Stacked.

➤ **Standardizzazione**

= STANDARDIZE(x; media; dev_standard)

LEZIONE 4 – DISTRIBUZIONE NORMALE

➤ **Verificare la normalità**

1. Utilizzo di grafici per analizzare la forma della distribuzione (boxplot, istogramma o poligono delle frequenze).
2. Valutazione delle statistiche descrittive.
3. Valutazione di un normal probability plot ovvero di una rappresentazione grafica utile valutare la normalità dei dati. Un metodo comunemente utilizzato è quello chiamato q-q plot (quantile-quantile plot).

➤ **Normal Probability Plot (PHStat)**

Selezionare Add-Ins → PHStat → Probability & Prob. Distributions → Normal Probability Plot.

- Nella finestra di dialogo, inserire l'intervallo delle celle che contengono i dati nel campo Variable Cell Range.
- Per aggiungere la linea retta che rappresenta la perfetta normalità

LEZIONE 5 – INTERVALLI DI CONFIDENZA

➤ **Intervallo di confidenza per la media**

1. FUNZIONI STATISTICHE DI EXCEL

- Funzione per calcolare il margine d'errore IC con varianza della popolazione nota:
=CONFIDENCE.NORM(alpha; standard_dev; size)
- Funzione per calcolare il margine d'errore IC con varianza della popolazione non nota:
=CONFIDENCE.T(alpha; standard_dev; size)

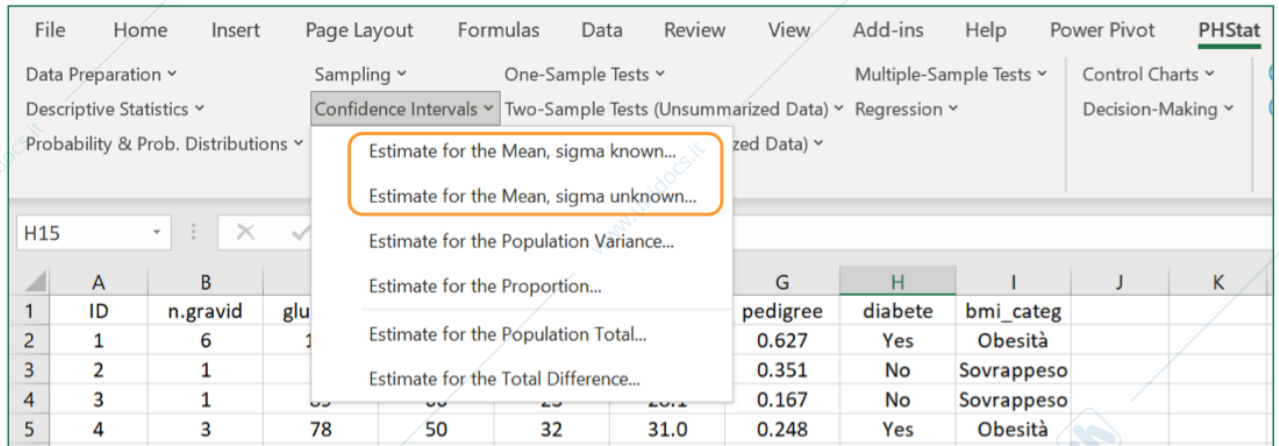
2. DATA ANALYSIS TOOL

Calcolo del ME dell'intervallo di confidenza per la stima della media con varianza non nota:

- Selezionare Data → Data Analysis

- Nella finestra di dialogo Data Analysis selezionare Descriptive Statistics e fare clic su OK.
- Nella finestra di dialogo Descriptive Statistics, inserire l'intervallo delle celle che contengono i dati nel campo Input Range.
- Selezionare Summary Statistics e Confidence Level for Mean, specificare il livello di confidenza e cliccare OK.

3. PHStat



- **Calcolo dell'intervallo di confidenza per la stima della media con varianza nota:**

Esempio 1a

Confidence Interval Estimate for the Mean	
Data	
Population Standard Deviation	0.6
Sample Mean	3.12
Sample Size	5
Confidence Level	95%
Intermediate Calculations	
Standard Error of the Mean	0.2683
Z Value	-1.9600
Interval Half Width	0.5259
Confidence Interval	
Interval Lower Limit	2.59
Interval Upper Limit	3.65

Input

- Calcolo dell'intervallo di confidenza per la stima della media con varianza non nota:

Esempio 1b

Confidence Interval Estimate for the Mean	
Data	
Sample Standard Deviation	0.6
Sample Mean	3.12
Sample Size	5
Confidence Level	95%
Intermediate Calculations	
Standard Error of the Mean	0.268328157
Degrees of Freedom	4
t Value	2.7764
Interval Half Width	0.7450
Confidence Interval	
Interval Lower Limit	2.38
Interval Upper Limit	3.86

Input

- Calcolo dell'intervallo di confidenza per la stima della media con varianza non nota:

diabete.xlsx

Intervallo di confidenza al 95% della pressione media

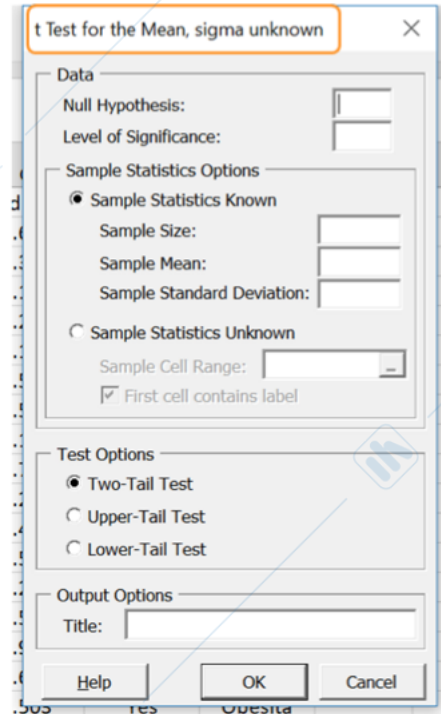
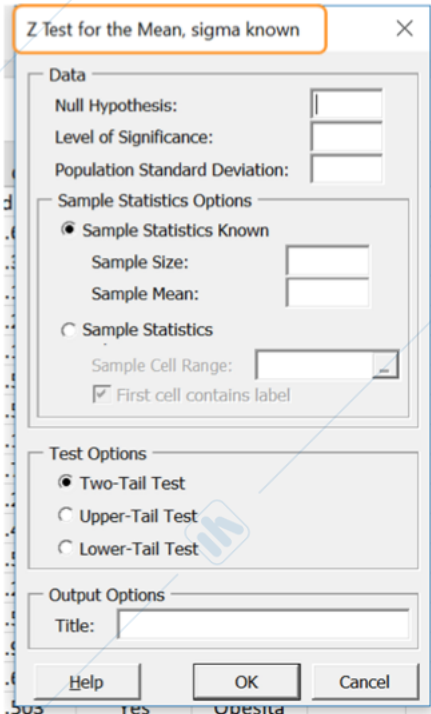
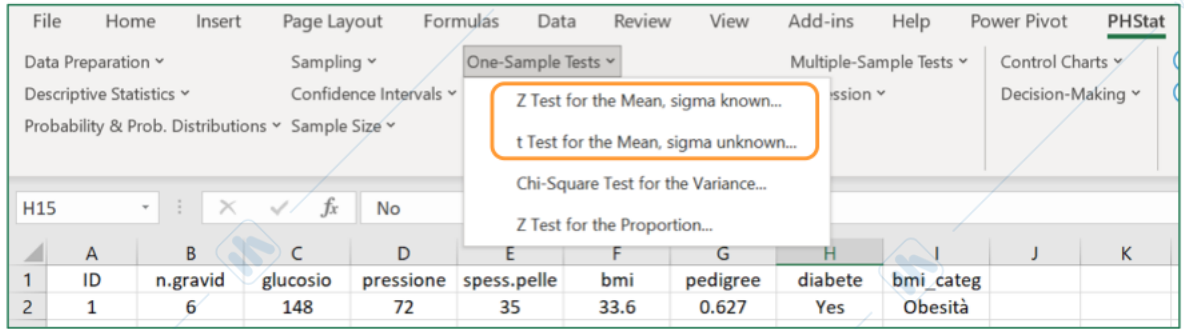
Confidence Interval Estimate for the Mean	
Data	
Sample Standard Deviation	12.79930673
Sample Mean	72
Sample Size	332
Confidence Level	95%
Intermediate Calculations	
Standard Error of the Mean	0.702453216
Degrees of Freedom	331
t Value	1.9672
Interval Half Width	1.3818
Confidence Interval	
Interval Lower Limit	70.27
Interval Upper Limit	73.04

Input

L'intervallo di confidenza al 95% è [70.27 ; 73.04]

LEZIONE 6 – VERIFICA DI IPOTESI

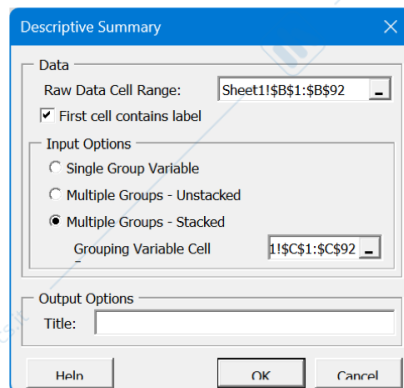
- **Verifica di ipotesi sulla media**
 - Verifico normalità



LEZIONE 7 – CONFRONTO TRA DUE GRUPPI

➤ Confronto misure di sintesi

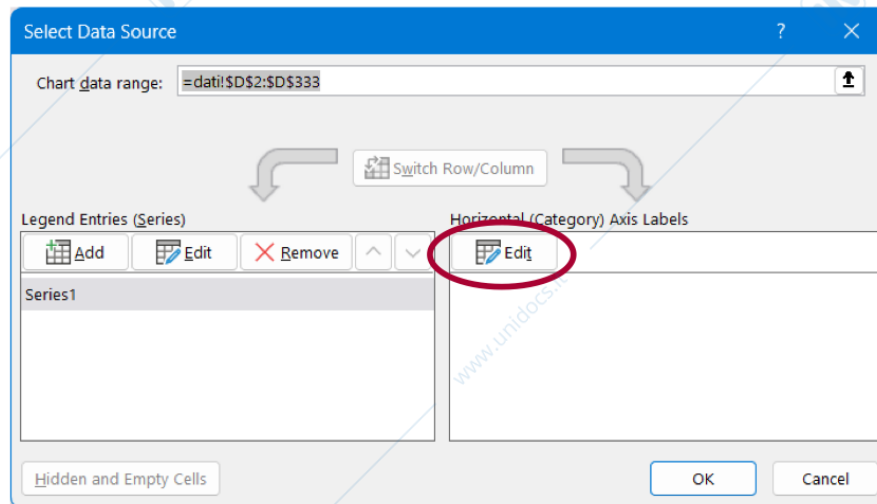
PHStat → Descriptive Statistics → Descriptive Summary → Multiple Groups – Stacked



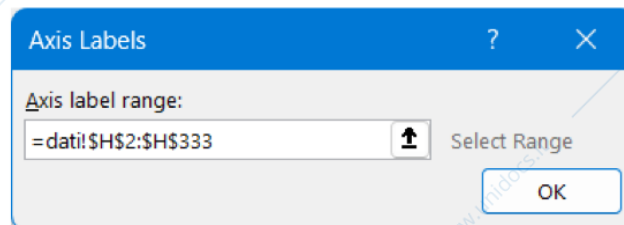
	Yes	No
Mean	74.7706422	70.13004484
Median	74	70
Mode	70	70
Minimum	30	24
Maximum	110	108
Range	80	84
Variance	172.3451	153.3118
Standard Deviation	13.1280	12.3819
Coeff. of Variation	17.56%	17.66%
Skewness	-0.1433	-0.1091
Kurtosis	0.9846	0.9290
Count	109	223
Standard Error	1.2574	0.8292

➤ Confronto boxplot (PHStat)

1. Costruire il box-plot dei valori della variabile pulsazioni cardiache.
2. Fare clic con il pulsante destro del mouse sul grafico, quindi scegliere Select Data
3. Nella finestra di dialogo Select data Source, cliccare sul pulsante Edit nella parte destra della finestra (Horizontal (Category) Axis Labels).



4. Nella finestra Axis Labels, inserire l'intervallo di celle contenente i dati sulla base della quale si vuole stratificare l'analisi.



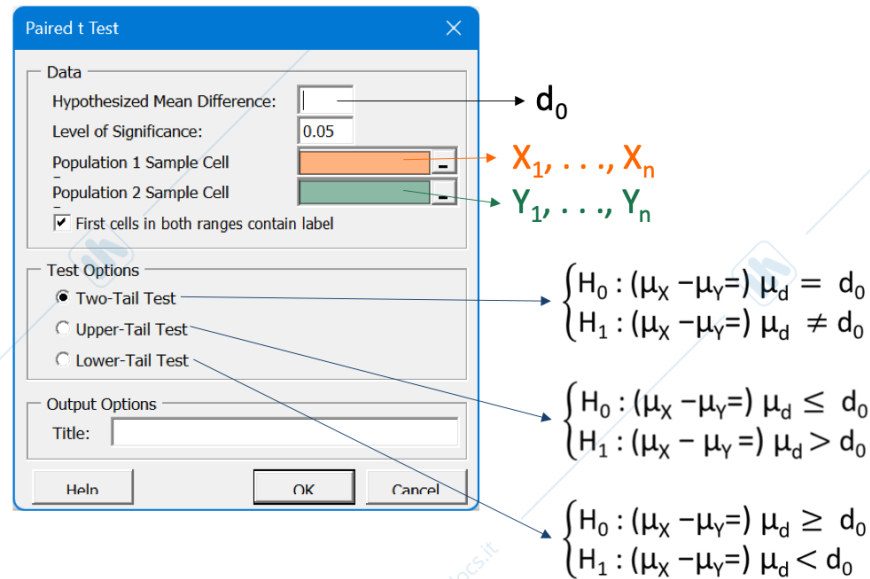
➤ Test di confronto tra due medie con campioni dipendenti (PHStat)

MODO 1

- Verifico normalità
- PHStat → Two-Sample Tests (Unsummarized Data) → Paired t Test

MODO 2

- Verifico normalità
- PhStat con one sample t test della var d



➤ Confronto tra due medie con campioni indipendenti

- Verifico normalità
- PHStat → Two-Sample Tests (Unsummarized Data) → F Test for Differences in Two Variances...
- Se varianze uguali → Two-Sample Tests (Unsummarized Data) → Pooled variance T-test
Se varianze diverse → Two-Sample Tests (Unsummarized Data) → Separate-variance T-test

LEZIONE 8 – RELAZIONE TRA DUE VARIABILI CATEGORICHE

➤ Tabelle di contingenza

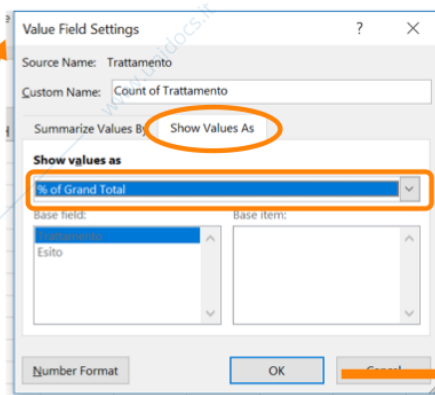
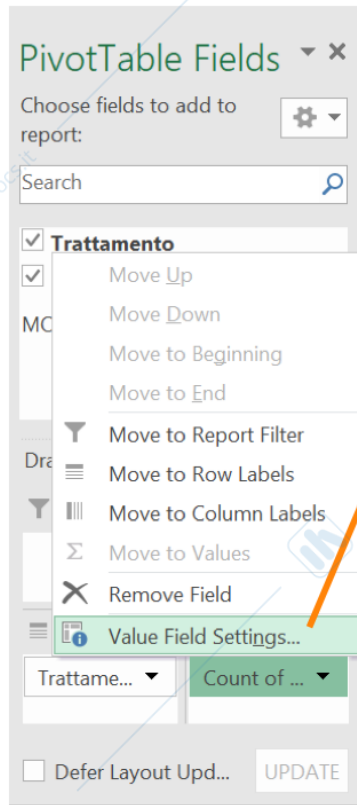
- Aprire il foglio di lavoro che contiene i dati grezzi (es. dati_trattamento.xlsx) e selezionare Insert → PivotTable.
- Nella finestra di dialogo PivotTable:
 1. Lasciate selezionata l'opzione Select a table or range e inserite l'intervallo delle celle (es. B1:C501).
 2. Selezionate l'opzione New Worksheet e cliccare su OK.
- Nella finestra PivotTable Fields:
 1. Trascinate l'etichetta della variabile di riga e inseritela nel riquadro ROWS.
 2. Trascinate una seconda volta questa stessa etichetta nel riquadro VALUES.
 3. Trascinate l'etichetta della variabile di colonna e inseritela nel riquadro COLOUMNS.

PHStat

- Selezionare Add-Ins → PHStat → Descriptive Statistics → Two-Way Tables & Charts (questa procedura sfrutta la funzione di Excel PivotTable).

- Nella finestra di dialogo Two-Way Tables & Charts
 1. Inserire l'intervallo di dati relativo alla variabile di riga nel riquadro Row Variable Cell Range.
 2. Inserire l'intervallo di dati relativo alla variabile di Colonna nel riquadro Column Variable Cell Range.
 3. Click OK.

➤ Frequenze relative



	A	B	C	D
1				
2				
3	Count of Trattamento	Column Labels		
4	Row Labels	Guarito	Non guarito	Grand Total
5	A	9.00%	11.00%	20.00%
6	B	12.00%	18.00%	30.00%
7	C	32.00%	18.00%	50.00%
8	Grand Total	53.00%	47.00%	100.00%

➤ Frequenze condizionate

The image illustrates the process of configuring a PivotTable to show conditional percentages. It shows the 'PivotTable Fields' task pane on the left, the 'Value Field Settings' dialog box in the center, and two resulting pivot tables on the right. The first pivot table shows '% of Row Total' for the 'Guarito' field, and the second shows '% of Column Total'.

Row Labels	Guarito	Non guarito	Grand Total
A	45.00%	55.00%	100.00%
B	40.00%	60.00%	100.00%
C	64.00%	36.00%	100.00%
Grand Total	53.00%	47.00%	100.00%

Row Labels	Guarito	Non guarito	Grand Total
A	16.98%	23.40%	20.00%
B	22.64%	38.30%	30.00%
C	60.38%	38.30%	50.00%
Grand Total	100.00%	100.00%	100.00%

➤ Grafici a barre

- Selezionare la tabella delle frequenze (assolute, relative o condizionate).
- Selezionare Insert → Insert Column or Bar Chart e selezionare un grafico.
- Fare clic sul grafico e quindi sulle icone accanto al grafico per modificarne l'aspetto. (uso diagramma a barre sovrapposte con frequenze condizionate per confronto)

➤ Indice di associazione PHStat

- Selezionare Add-Ins → PHStat → Multiple-Sample Tests → Chi-Square Test

LEZIONE 9 – RELAZIONE TRA DUE VARIABILI NUMERICHE

➤ Diagramma a dispersione

- Selezionare tutte le celle contenenti i valori delle due variabili.
- Selezionare la scheda Insert.
- Nella sezione Charts, cliccare su Scatter e selezionare l'opzione Scatter.
- Aggiungere titoli agli assi.

➤ Covarianza

- Covarianza della popolazione:
=COVARIANCE.P(Primo intervallo di celle; Secondo intervallo di celle)
- Covarianza campionaria:
=COVARIANCE.S(Primo intervallo di celle; Secondo intervallo di celle)

➤ **Coefficiente di correlazione di Pearson**

- Coefficiente di correlazione lineare
=CORREL(Primo intervallo di celle; Secondo intervallo di celle)

➤ **Matrice di correlazione**

- Selezionare Data → Data Analysis
- Nella finestra di dialogo Data Analysis selezionare Correlation e fare clic su OK.
- Nella finestra di dialogo Correlation, inserire l'intervallo delle celle che contengono i dati nel campo Input Range.
- Selezionare NewWorksheet e cliccare OK.

➤ **Costruire una variabile stratificatrice di indicazione**

- Costruisco due limiti → liminf; limsup
- Funzione IF (-3 e 3 dev standard per identificare outlier) → se ≤ -3 o > 3 1 se no 0
- Applico a tutte le celle