



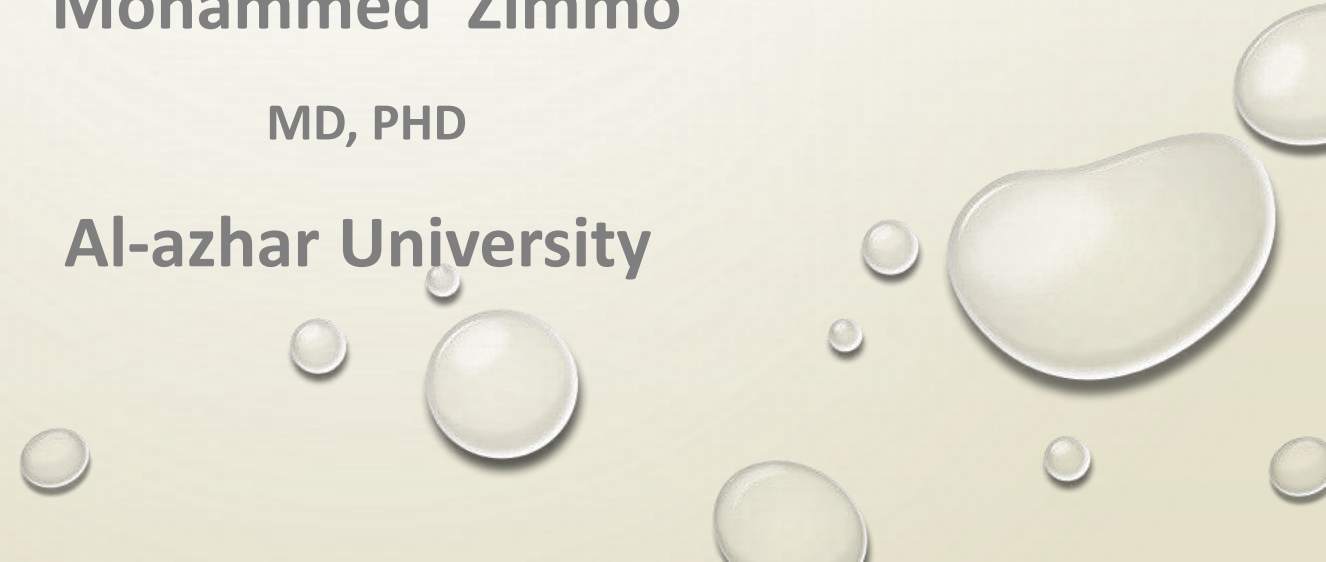
Statistics Tests

Part 2

Mohammed Zimmo

MD, PHD

Al-azhar University



Correlation

- Correlation is defined as the strength and direction of the **relation between two variables**
- The strength and direction are measured by correlation coefficient (r) with value ranges between (+1 and -1)
- The value measures the strength and the sign indicates the direction

Coefficient (r)

- **Positive sign**
 - Indicates positive correlation
- **Negative sign**
 - Indicates negative correlation
- **Zero:**
 - If there is absolutely no relationship between the two variables
 - The correlated variables could be presented by a scatter-grams (Bivariate distribution)

Coefficient (r)

- **P value**
 - Will indicate whether correlation is statistically significant
- **Coefficient of determination**
 - Expresses the proportion of variance in one variable that can be explained by the variance in the second variable
 - It is measured by squaring (r) and equal to (r^2).

SPSS for Correlation

survey1 - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

1 : gender

	gender	age	marital
1	1	35	2
2	0	37	2
3	1	43	2
4	0	35	2
5	1	35	2
6	1	35	1
7	1	43	2

Analyze menu options:

- Reports
- Descriptive Statistics
- Compare Means
- General Linear Model
- Correlate**
 - Bivariate...**
 - Partial...
 - Distances...
- Regression
- Classify
- Data Reduction
- Scale
- Nonparametric Tests
- Multiple Response

Bivariate Correlations



- # gender
- # marital
- # smoke
- # work
- # polaff
- # depress
- # exer
- # eat



Variables:

- # age
- # educ

OK

Paste

Reset

Cancel

Help

Correlation Coefficients

☒ Pearson ☐ Kendall's tau-b ☐ Spearman

Test of Significance

☒ Two-tailed ☐ One-tailed

☒ Flag significant correlations

Options...

survey1 - SPSS Data Editor

File Edit View Data Transform Analyze

Graphs Utilities Windows



1 : gender

1

	gender	age	marital	educ
1	1	35	2	21
2	0	37	2	18
3	1	43	2	22
4	0	35	2	20
5	1	35	2	20
6	1	35	1	19
7	1	43	2	18
8	.	45	2	18
9	0	52	2	18
10	0	69	2	15
11	1	21	1	12
12	0	19	1	14
13	0	37	1	17
14	0	35	2	19
15	0	40	4	12

Gallery
Interactive

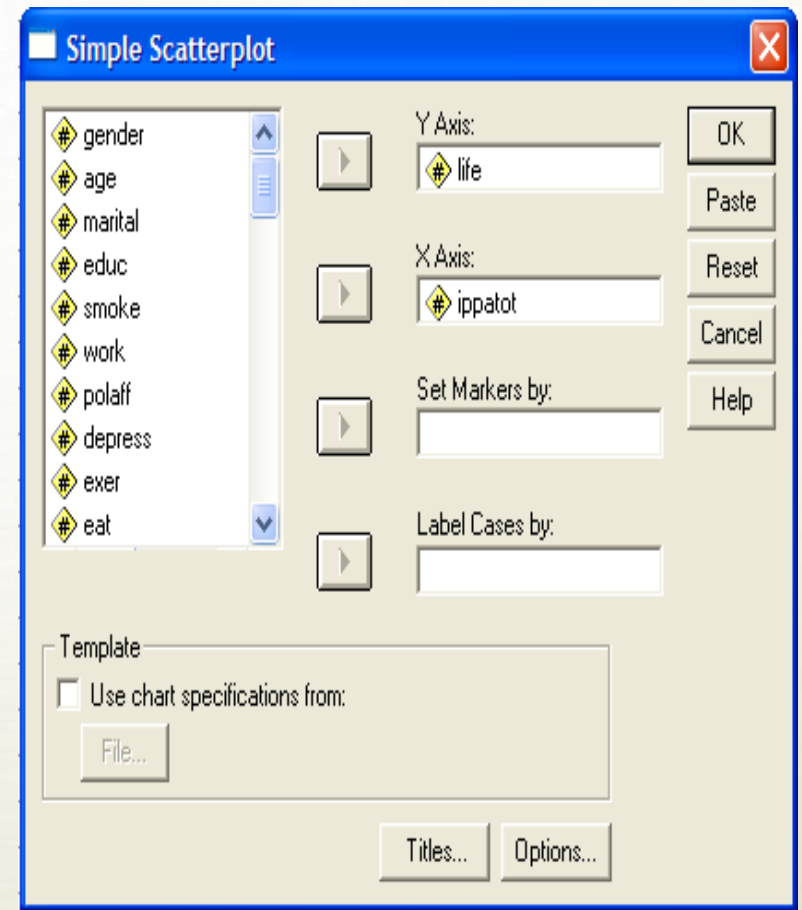
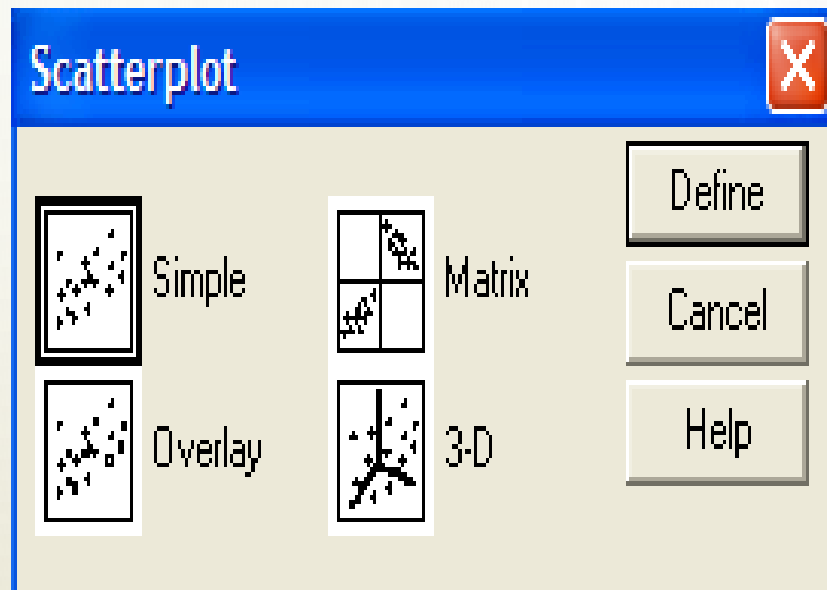
Bar...
Line...
Area...
Pie...
High-Low...

Pareto...
Control...

Boxplot...
Error Bar...

Scatter...

Histogram...
P-P...
Q-Q...
Sequence...
ROC Curve...
Time Series



Pearson's Correlation Coefficient

Is Diastolic BP related to Weight?

Diastolic BP (mm)

**90
140
68
110
100
95**

Weight (kg)

**82
114
56
62
83
110**

$r = 0.805$ $p < 0.01$

Pearson's Correlation Coefficient

- $R = 0.805$ does not mean weight gain causes increase in BP or vice versa
- Correlation does not prove cause and effect

Logistic Regression

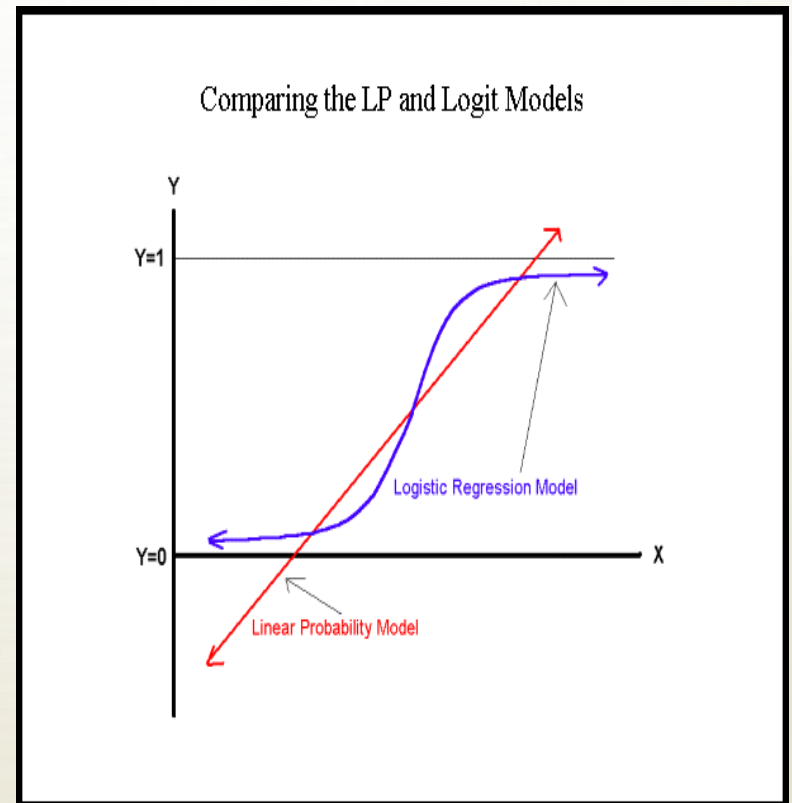
- Correlation is defined as the strength and direction of the **relation between two variables**
- The strength and direction are measured by correlation coefficient (r) with value ranges between (+1 and -1)
- The value measures the strength and the sign indicates the direction

Logistic Regression

- Binomial (or binary) logistic regression is a form of regression which is used when the **dependent is a dichotomy** and the independents are of any type
- Logistic Regression is used to determine which independent variables (Xs) **affect the probability of an outcome of the dependent variable**

Logistic Regression

- Logistic regression does not assume linearity of relation between the independent variables and the dependent
- Does not require normally distributed variables

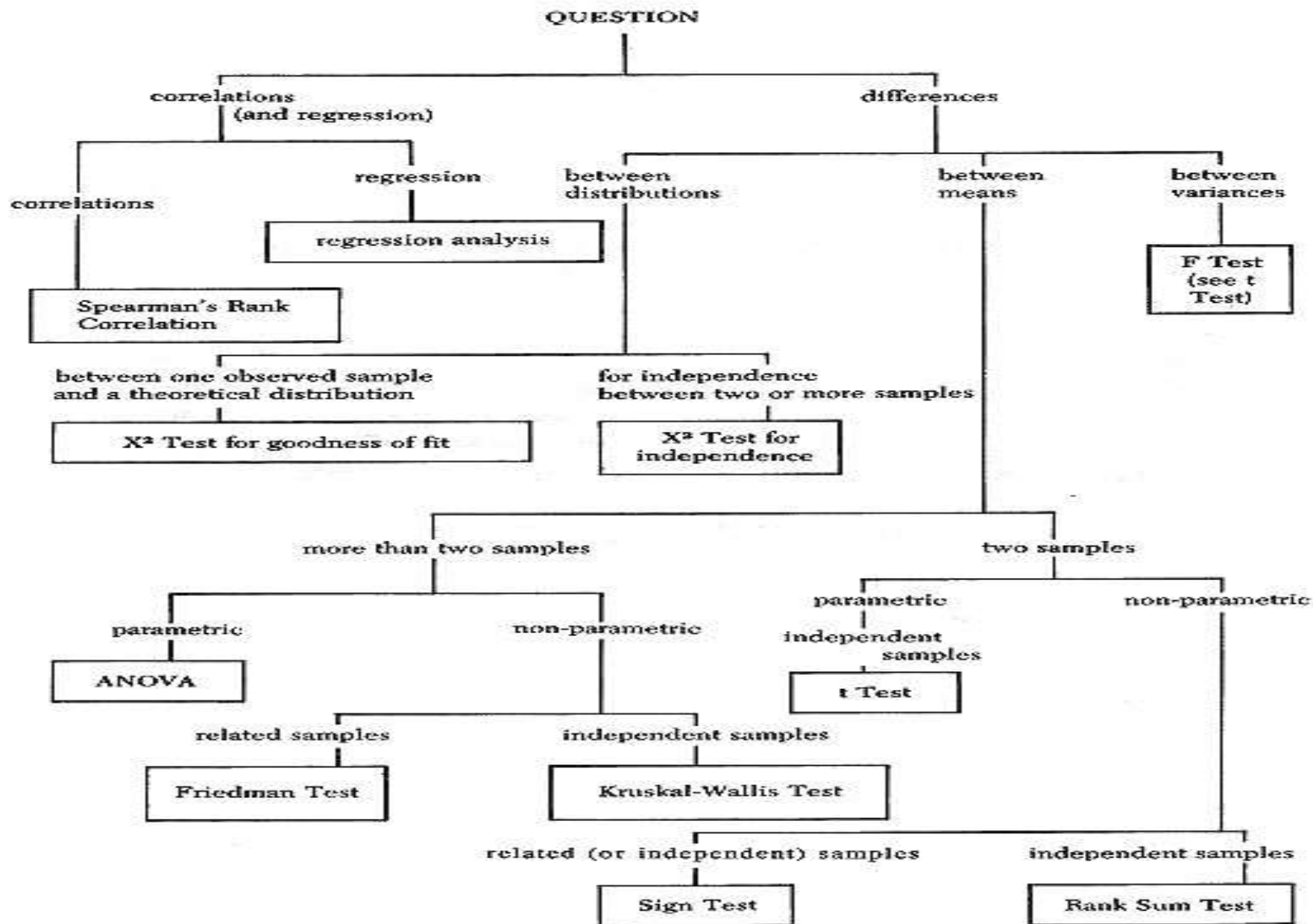


Why we use logistic regression?

1. Prediction: logistic regression can be used to predict a **dependent variable on the basis** of independents.
2. Determine the percent of variance in the dependent variable explained by the independents; to rank the relative importance of independents
3. To control for **confounder**
4. To assess **interaction** effects

Type of Variable	Descriptive Statistics		Compare Groups		Relation of Groups
Nominal (Categorical) (Non-parametric)	Central Tendency	Variability	1 or 2 groups	≥ 3 groups	χ^2
	Mode		χ^2 Fisher's Exact Test		
Ordinal (Ranks)	Mode Median	Range	Mann-Whitney Wilcoxon	Kruskal-Wallis	Spearman's Rank Order
Interval or Ratio (Continuous) (Parametric)	Mode Median Mean Confidence Interval	Standard deviation Standard error Variance Range	t-test	ANOVA	Pearson's r Regression

Data		Independent	Dependent
Continuous	Normal	t test ANOVA / ANCOVA	Paired t test
	Non-normal (non-parametric)	Mann-Whitney U test Kruskal-Wallis test	Wilcoxon test
Categorical		Fisher's exact test χ^2 test/CMH test Logistic regression	McNemar's test (binary)/
Time to event		Kaplan-Meier analysis Log rank test Cox proportional HR model	



Thank You

