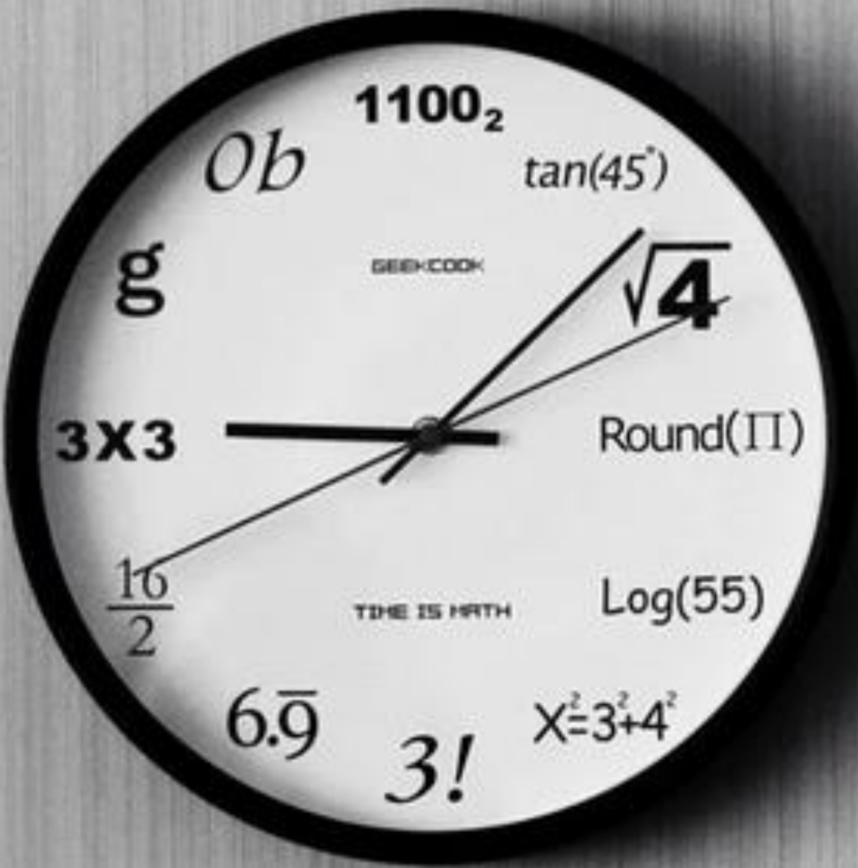
Statistics

Last Minute Review



DR VASANT KOTHARI

Statistics: Last Minute Review

Dr Vasant Kothari
B Tech, M Tech (Textile), MBA (Marketing)
M Phil (Management), MA (Psychology)*
PhD (Management)

SPRING SEASON PUBLICATIONS
Navi Mumbai, INDIA

*

Statistics: Last Minute Review

First Published – 2018

Spring Season Publications
Kharghar Sec 10, Navi Mumbai, MS, INDIA, 410210
www.springseason.in

Content

Chapter 1

Parametric Vs Non Parametric

Parametric Test: Parametric tests normally involve data expressed in absolute numbers or values rather than ranks.

Parametric tests are restricted to data that:

1. shows a normal distribution
2. are independent of one another
3. are on the same continuous scale of measurement

Parametric tests are useful as these tests are most powerful for testing the significance or trustworthiness of the computed sample statistics.

Non-Parametric Test: There may be a situations where we cannot meet the assumptions and conditions and thus cannot use parametric statistical procedures. In such situation we are bound to apply non-parametric statistics.

The first meaning of non-parametric covers techniques that do not rely on data belonging to any particular distribution. In this the statistics is based on the ranks of observations and do not depend on any distribution of the population.

Basically, Non-parametric statistics

1. deals with small sample sizes
2. are not bound by any assumptions
3. are user friendly compared with parametric statistics and economical in time

Non-parametric tests are used on data that:

1. show an other-than normal distribution
2. are dependent or conditional on one another
3. in general, do not have a continuous scale of measurement

Parametric Test	Non-Parametric Test
For making inferences about various population values (parameters), we generally make use of parametric and non-parametric tests	
Sample size more than 30	Normally sample size less than 30
Focuses on the mean difference	Focuses on the difference between medians
Original data score is used	Data need to change from scores to ranks or signs
Makes assumptions	No assumptions are made
Information about the population is completely known	No information about the population
Data should be normally distributed	Distribution free tests
Applicable only for variable	Applied to both variable and attributes
Not applicable for nominal scale data	Exist for nominal and ordinal scale data
Powerful	Less powerful
Efficient	Less efficient
Less likely to make a Type-I Error	More likely to make a Type-I Error
Data should be normally distributed	Distribution free tests
If the mean accurately represents the center of distribution and sample size is large enough, consider a parametric test because they are more powerful	If the median better represents the center of distribution, consider the nonparametric test even in case of a large sample

	Parametric Test	Non-Parametric Test
Assumed distribution	Normal	Any
Assumed variance	Homogeneous	Any
Typical data	Ratio or Interval	Ordinal or Normal
Data set relationships	Independent	Any
Usual central measure	Mean	Median
Benefits	Can draw more conclusions	Simplicity; Less affected by outliers
Correlation	Pearson	Spearman
Independent measures, 2 groups (Comparison of 2 group)	t-test	Mann-Whitney U test
Independent measures, >2 groups (Comparison of several group)	One-way ANOVA	Kruskal-Wallis test

Assumptions of Parametric Test

- The populations are normally distributed
- The selected population is representative of general population
- The data is in interval or ratio scale
- The observation must be independent
- These populations must have the same variance

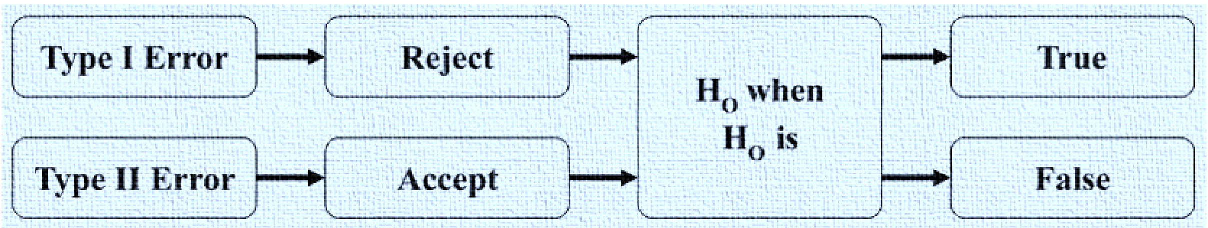
Assumptions of Non-Parametric Test

- Data don't follow any specific distribution
- Data measured on any scale
- No assumptions about the population are made
- The variable is continuous
- Sample size is quite small
- Assumption like normality of the distribution of scores in the population are doubtful
- The data can be expressed in the form of ranks
- The nature of the population from which samples are drawn is not known to be normal
- The variables are expressed in nominal form

Type II Error: Failure to reject H0 when H1 is true is called a Type II error. Example, Acquitted the defendant when he is guilty!

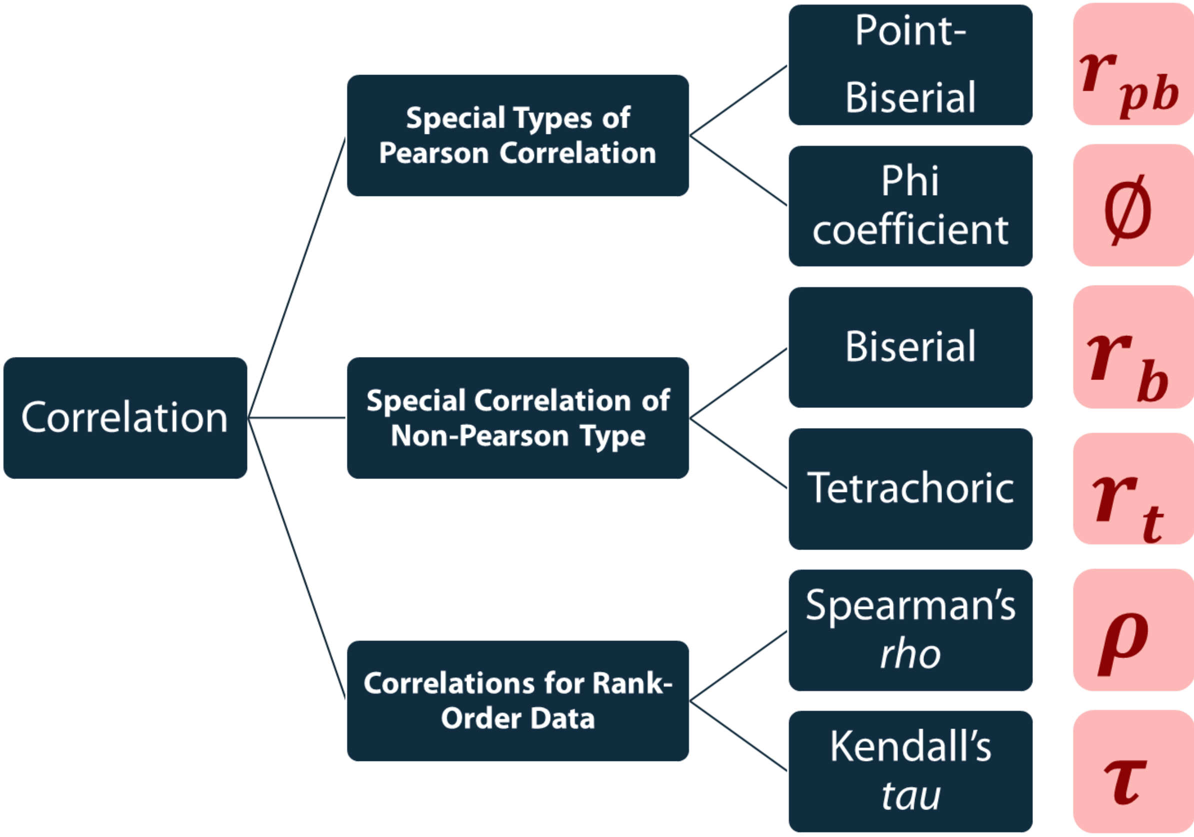
		Conclusion Drawn	
		Accept H_O	Reject H_O
True State of Nature	H_O True	Correct 1-α	Type I Error A
	H_O False	Type II Error β	Correct 1-β

Type I Error	Type II Error
The probability of committing a Type I error is called the significance level	The probability of not committing a Type II error is called the Power of the test
Often denoted by α (Alpha)	often denoted by β (Beta)
The value of alpha is always set before the experiment or study is undertaken	Beta is not usually stated at the beginning of the hypothesis testing procedure
It is the probability of overreacting	It is probability of under reacting



Chapter 7

Types of Correlation



Point-Biserial Correlation: The point biserial correlation coefficient is the correlation calculated between a continuous random variable and a dichotomous variable. Point Biserial correlation is calculated in a similar way to Pearson's Product moment correlation.

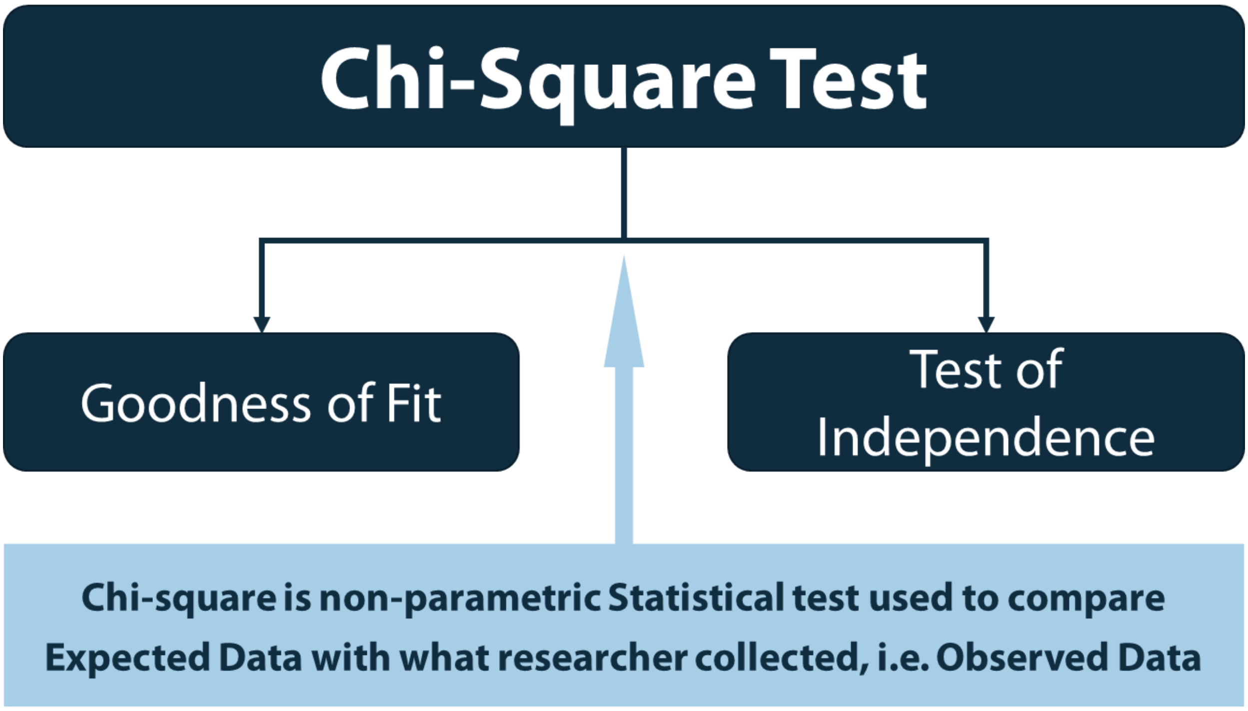
Subject	1	2	3	4	5	6	7	8	9
Gender	1	1	1	1	1	0	0	0	0
Marks	46	74	58	67	62	71	65	69	59

Dichotomous Variables Male = 1 & Female = 0

Continuous Random Variable

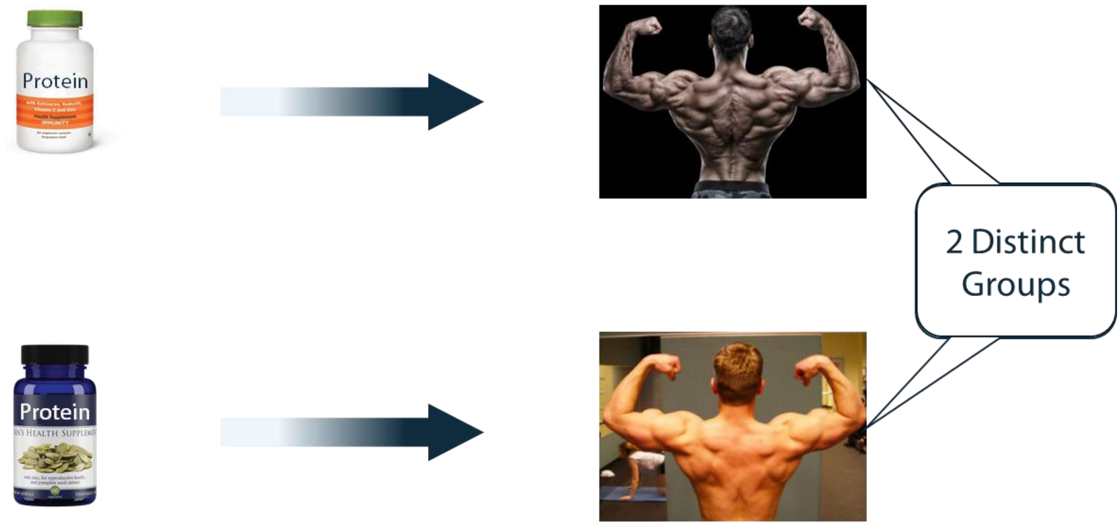
Chapter 13

Chi-Square Test



Goodness of Fit	Test of Independence
It is used to determine the extent to which observed data matches the values expected by theory	It is used to determine whether there is a significant association between the two variables
Variable A H_0: Observed = Expected H_a: Observed $\neq$ Expected	**Variable A & Variable B** H_0: Not Dependent H_a: Dependent

Independent Measures



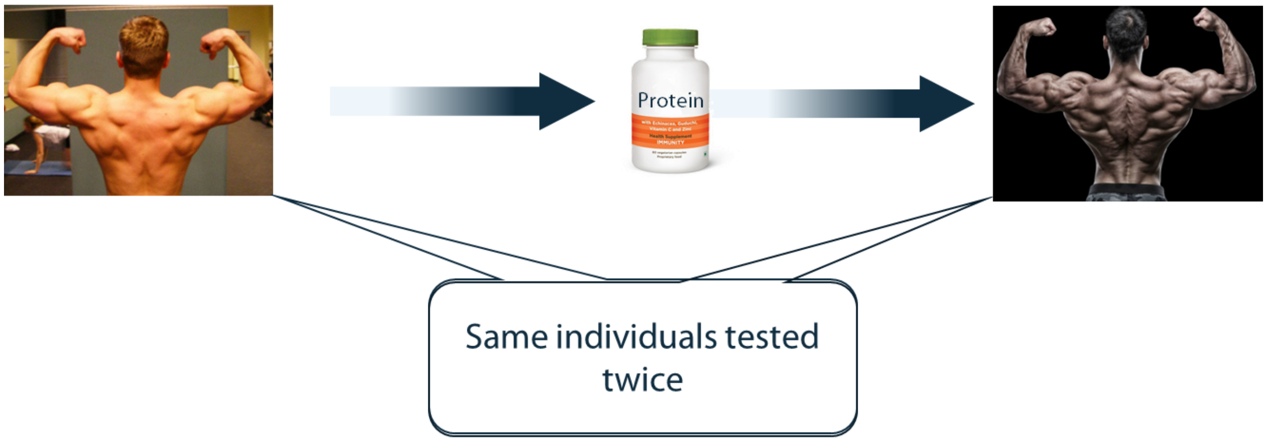
Independent-means t-Test

Parametric Test

Mann-Whitney 'U' Test

Nonparametric Test

Dependent Measures



Dependent-means t-Test

Parametric Test

Wilcoxon Test

Nonparametric Test