

Advanced Statistics

The Advanced Statistics workshop covers Inferential Statistics. Inferential Statistics uses probability to assess the likelihood of whether the observed difference or relationship is real or because of random chance. A probability value (p-value) lesser than a permissible maximum chance of error (e.g., 5%) set by the researcher indicates a high likelihood of the observed difference or relationship being real and not because of random chance.

Why do we use inferential statistics?

An observed difference in the average (e.g., mean, median) value of a dependent variable (e.g., weight) across different groups (e.g., teenagers, adults) obtained from **descriptive statistics** of a sample data is not necessarily statistically significant. This is because the descriptive statistics summarizes the sample data, and not the population. Also, there is some level of randomness or error associated with sampling.

The question of whether the observed difference from the descriptive statistics is large and consistent enough to represent a true or real difference (and not due to sampling error) in the larger population from which the sample is drawn is what **inferential statistics** answers for us. This can also be extended to the observed relationship or association between variables obtained from descriptive statistics such as scatter plot.

Stages involved in inferential statistics

Formulate Research Questions; Identify Independent and Dependent Variables; Formulate Hypotheses; Perform Inferential Test; and Interpret and Contextualize the outputs.

Categories of inferential tests

1. **Parametric tests:** These are used when parametric assumptions are satisfied. Parametric assumptions include normal distribution (checked with Shapiro-Wilk test or Q-Q plot or histogram density plot); absence of significant outliers (checked with Box plot); homogeneity of variance (checked with Levene's Test); etc.
2. **Non-parametric tests:** These are used when parametric assumptions are not satisfied.

How to interpret Parametric assumptions

If the **p-value** is **greater than 0.05**, that **parametric assumption** (e.g., normal distribution) **is satisfied**.

Types of inferential tests based on the purpose or research question

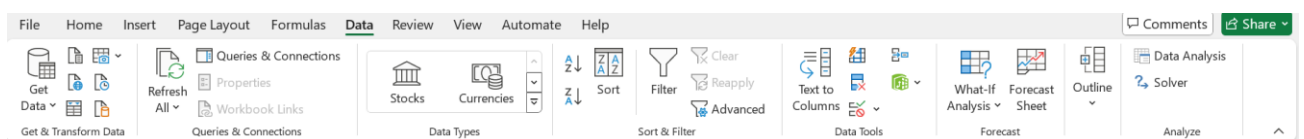
1. **Tests of Comparison:** These are used to compare a dependent variable across groups or conditions or across combination of groups or conditions. This helps to answer the research question of whether significant difference exists in a dependent variable across groups or conditions or across combination of groups or conditions.
2. **Tests of Association:** These are used to test if there is significant relationship or correlation or association or connection between two or more variables. If significant association exists between variables, these tests can also be used to predict the dependent variable (outcome) using the independent variables (predictors).

How to interpret results of inferential tests

If the **p-value** is **less than 0.05**, there is **significant difference** in the dependent variable across groups or conditions or across combination of groups or conditions; or there is **significant relationship or association** between variables.

How to perform inferential test in Microsoft Excel

Click the **Data** ribbon, then click **Data Analysis** as shown in the figure below and select the appropriate inferential test. Microsoft Excel offers only parametric inferential tests by default, but you can write formulas for the non-parametric inferential tests yourself.



If **Data Analysis** option is missing under the **Data** ribbon, click on **File**, then **More**, then **Options**, then **Add-ins**, select **Analysis ToolPak**, then click **Go**, ensure you tick **Analysis ToolPak** button, then click **OK**. Once this is done, it will show the **Data Analysis** option under the **Data** ribbon. In some versions of Microsoft Excel, after clicking **File**, it might show **Options** directly without the need to click **More** first.

Selection of inferential tests for comparison

Parameters being compared:	Dependent Variable Type	Independent Variable Type	Parametric test	Non-parametric test
Averages of two INDEPENDENT groups	Continuous	Nominal (Binary)	Independent T-test	Mann-Whitney test or Wilcoxon Rank Sum test
Averages of 3 or more INDEPENDENT groups	Continuous	Nominal	One-Way ANOVA*	Kruskal Wallis test
Averages of two paired/matched samples or conditions (e.g., pain level before and after a treatment)	Continuous	Time or Condition variable	Paired T-test	Wilcoxon Signed Rank test
Averages of three or more matched/repeated samples or conditions (e.g., pain level before, 6 months after, and 1 year after treatment)	Continuous	Time or Condition variable	Repeated Measures ANOVA	Friedman test

*ANOVA means Analysis of Variance

Selection of inferential tests for relationship or association

Parameters being assessed:	Dependent Variable Type	Independent Variable Type	Parametric test	Non-parametric test
Relationship between two continuous (or ordinal for non-parametric) variables	Continuous	Continuous	Pearson's Correlation Coefficient	Spearman's Correlation Coefficient or Kendall's Tau
Relationship between two variables	Continuous	Nominal (Binary)	Point-Biserial Correlation	Rank-Biserial Correlation
Relationship between two categorical variables	Categorical	Categorical		Chi-Square test
Predicting one variable using another predictor variable	Continuous	Continuous or Ordinal	Simple Linear Regression	Transform the data or use non-parametric regression

Predicting one variable using another predictor variable	Categorical (Nominal or Ordinal)	Continuous or Nominal or Ordinal	Logistic regression	Classification models
--	----------------------------------	----------------------------------	---------------------	-----------------------

Selection of inferential tests when several independent variables are involved

Parameters being assessed:	First Independent Variable Type	Second/Subsequent Independent Variable Type	Inferential Test
Predicting a continuous dependent variable using two or more predictor variables	Continuous	Continuous or nominal (Binary)	Multiple Linear Regression
Comparing averages of a continuous dependent variable	Nominal (Independent groups)	Nominal (Independent groups)	Between-Subject Factorial ANOVA
Comparing averages of a continuous dependent variable	Nominal (Paired/Repeated)	Nominal (Paired/Repeated)	Within-Subject Factorial ANOVA (or Repeated Measures Factorial ANOVA)
Comparing averages of a continuous dependent variable	Nominal (Independent groups)	Nominal (Paired/Repeated)	Mixed Factorial ANOVA
Comparing averages of a continuous dependent variable	Nominal	Continuous	Analysis of Covariance (ANCOVA)

Need help with this topic?

Click or scan this code to book an Academic Skills Tutor appointment.

