

Short Course: Outline

At the completion of this short course you will be able to:

- Perform statistical analyses, report their findings and interpret their results.
- Critically appraise the statistical aspects of research publications in bioscience.

You will be confident and enjoy statistical analysis! 😊

Day 1 – Building up the Background

Module 1: Some Basic Concepts: Measurement scales/types of data, sample selection; randomized clinical trial (RCT), case-control study, cross-sectional study, cohort study, matched case control study. Does the choice of a statistical method for data analysis depend on the measurement scale? Yes, largely! Want to know in details? Next modules.

Module 2: Summary Statistics: Summarizing data using simple statistical methods (e.g., mean, median, quartiles, sd, etc) and graphical presentation (e.g., box-plot, bar-diagram, phi-chart, histogram, scatter plot, etc.). At the completion of this module you will be able to analysis the data using some basic statistical methods and report the results.

Module 3: Distribution of Data: Quantifying uncertainty in data analysis, and the normal, t and chi-squared distribution. No study results are perfect!!! Numerous errors occurred from the beginning to the end in all studies that causes some uncertainty in the statistical analysis results. Shouldn't you quantify and report it? Certainly! In this module you will learn the technique of quantifying uncertainty in the analysis results.

Module 4: Comparing two Groups/Methods: Comparing two groups/methods using hypothesis test or p-value. The p-value gives the strength of evidence that the two groups are the same (different)? **Example:** Consider a study where you want to compare a new drug with a standard drug. Randomly select two groups of patients where one group is treated by the new drug and another group is treated by the standard drug. Study Question: Is the new drug is effective? In this module you will know the method to assess the effectiveness of the new drug.

Day 2 – Common Data Analysis Methods

Module 5: Comparing two Groups/Methods: Comparing two groups/methods using 95% confidence interval. The 95% confidence interval compares the two groups with certain level of confidence and also gives the uncertainty in the study results. **Example:** Consider a study where you want to compare development of ovarian cancer b/w parous and nulliparous women. Suppose the difference of cancer rate b/w them is 5%. Is this difference is statistically significant? What is the uncertainty in this estimated difference? In this module you will learn statistical methods to address these questions.

Module 6: Comparing more than two groups – Analysis of Variance (ANOVA): If the exposure variable has more than two categories, we often wish to compare the mean outcomes from each of the groups defined by these categories. Analysis of variance (ANOVA) is a statistical method useful to compare more than two groups. **Example:** Consider a study where the clinicians are interested in whether the mean haemoglobin levels of patients with various sickle cell diseases are the same. This can be determined using the ANOVA method which is based on the assessing how much of the overall variation in the outcome is attributable to differences between the exposure group means – hence the name ANOVA.

Module 7: Association of Outcome and Exposure: Assessing the association between outcome and exposure using chi-squared test. Outcome is the event of interest in a study and exposure is the cause to the outcome. **Example:** Consider a study where you want to investigate the blood pressure and social status. We may think that people with very high social status will have healthy blood pressure level! However, we cannot be sure about this without having appropriate statistical analysis. How do you evaluate the association between blood pressure and social status? You will learn this in this module.

Module 8: Risk Comparison: Risk comparison is very common in biomedical research. We use either relative risk (RR) or odds ratio (OR) to compare risk of developing the disease among exposed and unexposed groups. Can we use RR or OR for the same data? In 20% of the published articles authors made a wrong choice! **Example:** Consider a study where you want to evaluate the risk of developing lung cancer among smokers compared to non-smokers. There are two methods to compare the risk, RR and OR. Which method should you use? Can you use either of RR or OR? My answer is NO! Why? You will learn it in this module.

Module 9: Non-parametric Methods: Often the data we collect for analysis gives us hard time by not following the normal distribution. In such situation we cannot use the statistical methods we discussed in Modules 4 – 5 for data analysis. Statisticians are very smart! They developed alternative statistical techniques known as non-parametric methods. In this module you will learn some commonly used non-parametric methods in medical research.

Day 3 – Prediction of Risk/Survival & Power Analysis

Module 10: Prediction of an Outcome/Event: Prediction of adverse event or identifying risk factors for some event of interest – e.g., what are the risk factors for cardiac surgery mortality?; how many glasses should you drink on Friday night? – do you know the risk factors of your blood alcohol level? In this module you will learn statistical methods that are used to identify risk factors and then use these risk factors to predict adverse outcome or outcome of interest ((linear and logistic regression analysis).

Module 11: Survival Analysis: Survival analysis is appropriate for time to event data, e.g., consider a cancer study where you want to know the chance that a cancer patient who survived today will also survive tomorrow. Consider another study where you are interested in comparing the survival time of two groups of patients underwent cardiac surgery - one group was operated by the trainee surgeons and another group by the consultants. Here the outcome/event of interest is mortality of patients after having cardiac surgery. The study results will tell you whether patients operated by consultants have longer life compared to those operated by trainees. Kaplan-Meier survival curve, Long-rank statistic and Cox proportional-hazards ratio are the statistical methods we use to analyze time to event data – you will learn them in this module.

Module 12: Sample Size Calculation/Power Analysis: In any study the question arises how many cases we should include in the study. How do we select this magic number? In fact, the number should not be too small (under powered) or too big (huge cost/time involved). We should select an optimum number of cases. Learn the statistical methods in this module for determining this magic number.

Case Studies: Several case studies will be discussed that will help you to elevate your understandings in Modules 1 – 12. You will be provided a flow chart that will guide to choose correct statistical methods to evaluate the hypotheses on your study. You will never be scared of statistical analysis! Be a great BioSTATistician!