BM5033 Statistical Inference Methods in Bioengineering

Problem Set 3

Instructions

1. For each scenario below, specific questions are asked. However, to get the best outcome from these problems, try analyzing each situation by asking more questions yourself about the topics covered in the course. For instance, for each situation below, you should describe the data using central tendency and dispersion measures, plot the data, state the null and alternate hypotheses, perform the power analysis to get an idea of the sample size, check assumptions of the hypothesis tests to be used, identify the test, run the test, and interpret the test result.
2. The datasets for the questions in this problem set can be obtained at this link.

Questions

1. Suppose you wish to study the effect of different food supplements (`feed`) on the weights of the chicken (`weight`).
   (a) Use boxplots and a plot of means to visualize the difference between feed types.
   (b) Test for the assumptions required for the analysis of variance.
   (c) Perform an appropriate version of the ANOVA to compare the effectiveness of different feeds.
   (d) If ANOVA shows a difference in the group means, perform post-hoc analysis to identify the feed that results in the highest weight gain for the chicken.
2. Cushing's syndrome is a hypertensive disorder that has been shown to be associated with a higher amount of cortisol secretion by the adrenal gland. In the data for this question, the observations are urinary excretion rates (mg/24hr) of two steroid metabolites `Pregnanetriol` and `Tetrahydrocortisone`. Based on some experiments it has been hypothesized that the excretion rate of `Pregnanetriol` changes based on the underlying type (coded `a` (adenoma), `b` (bilateral hyperplasia), `c` (carcinoma) or `u` for unknown) of Cushing's syndrome.
   (a) Use boxplots and a plot of means to visualize the difference between excretion rates of the two steroid metabolites for different Cushing's type.
   (b) Test for the assumptions required for the analysis of variance.
   (c) Perform an appropriate version of the ANOVA to compare `Pregnanetriol` secretion rates in different syndrome types.
   (d) Perform an appropriate version of the ANOVA to compare `Tetrahydrocortisone` secretion rates in different syndrome types.
   (e) If ANOVA shows a difference in the group means in the last two questions, perform a post-hoc analysis for each.
3. `Anorexia` is characterized by an unwarranted fear of being overweight. Its symptoms include attempts to maintain a below-normal weight through various means. In the data for this question, the weights (in lbs) of the subjects are listed before (`PreWt`) and after (`PostWt`) a couple of treatment methodologies (Cognitive Behavioural treatment `CBT` and family treatment `FT`) along with a control (`Cont`) group where no treatment was given.
   (a) Use boxplots and a plot of means to visualize the difference between different groups.
   (b) Calculate the difference in body weight for each subject.
   (c) Test for the assumptions required for the analysis of variance.
   (d) Perform ANOVA and post-hoc analysis and comment on the effectiveness of the therapies.

4. Suppose you are designing an experiment to study the effect of a nano-particle-based drug delivery method. For this, you have planned to synthesize Poly (lactic-co-glycolic acid) (PLGA) nanoparticles, Poly-lactic acid nanoparticles, and gold nanoparticles. For a comparison of the effectiveness of the delivery method, you also plan to administer the drug without loading it into the nanoparticles. In experiments, you intend to measure the number of cells after 4 hours of drug administration using different delivery modes. From some preliminary observations, you have seen that the variance in the cell numbers is 1000. What should be your sample sizes to achieve a statistical power of 80% at 0.05 level of significance if you want to detect a difference of 100 cells?

5. In a cardiac ward at a hospital, a team of researchers sought to test a novel drug, `CardioNova`, against the established treatment, `BetaBlock`, for chronic heart failure. They recruited patients, carefully dividing them into two groups. One group received `CardioNova`, while the other continued with `BetaBlock`. The scientists observed the patients' progress, measuring their ejection fraction (a measure of heart function) and overall quality of life. Their goal was to uncover if `CardioNova` offered a significant improvement over `BetaBlock` in managing chronic heart failure. The data are given in the dataset. Analyze.

6. A medical device company has developed a new surgical instrument designed to improve the efficiency and accuracy of a specific surgical procedure. They want to evaluate the effectiveness of this new instrument compared to the current standard of care. The company gave the new and standard instrument to several clinicians and observed their procedure time and post-operative complications. Analyze.

7. Sometimes in some experimental situations involving multiple groups, we are not interested in comparing all groups against each other but against one specific group. This special group is often labeled *control group*. For such a situation, the regular post-hoc analysis after ANOVA approach is not going to work owing to its inherent design. Instead, we use `Dunnett's test`. It performs many-to-one comparisons. You can see the details of this method at this link on R documentation webpage.

   Consider one such situation- assume there are four groups in one study where the first group is a control group, while the rest are the test groups involving different treatments. We want to know if the test groups differ significantly from the control group. Use `Dunnett's test` to analyze the experimental data.

8. A researcher is planning an experiment to investigate the effects of three different teaching methods (A, B, and C) on student test scores. Based on previous research, the researcher estimates that the average test score for Method A will be 85, Method B will be 80, and Method C will be 75. Assuming a standard deviation of 10 points, the researcher wants to determine the minimum sample size needed to detect a significant difference among the three methods with a power of 0.80 at an α-level of 0.05.

   (a) calculate the effect size based on the estimated mean differences and standard deviation.

   (b) perform a power analysis for a one-way ANOVA.

   (c) Determine the required sample size per group.

9. For the situation described in the previous question, suppose 9 schools have been selected randomly to run the experiment.

   (a) How many factors are there in this experiment?

   (b) What is the nature (fixed vs random effects) of the factor/factors involved in the experiment?

   (c) Suppose each classroom in any school in this study requires a fixed monetary resource to run the experiment. Therefore, the experiment can be performed on a fixed number of classrooms given a finite amount of money available for the study. Imagine, the experimenter has the resources to run experiments in 180 classes (distributed over multiple schools). Consider two possibilities - (i) 10 schools with 18 classrooms in each school, (ii) 20 schools with 9 classrooms in each school. Analyze the pros and cons of both experimental setups from a statistical viewpoint.

   (d) Will it have any influence if the money is only to run experiment on 160 classes.