MSAN 631: Design and Analysis of Experiments

Cumulative-Rolling-Take-Home-Final Examination

Due Date: Thursday June 29th by 5:00pm

Description:

The list of problems below is intended to assess your competency with both technical and non-technical experimental design concepts. These problems also test your ability to analyze experiment data both by hand and by R. You may complete these problems at your own pace, so long as they are completed and submitted by the due date indicated above. You are expected to complete this examination independently and submit your own work.

Problems:

1. Discuss the merits and drawbacks of an experiment.

2. Discuss the importance of replication, randomization, and blocking in the context of an experiment.

3. Briefly explain the philosophy and mechanics of an "A/B" test.

4. Suppose that an online retailer is interested in determining whether the average purchase price is larger among customers in "condition A" than among customers in "condition B". Formally:

$$H_0: \mu_1 \leq \mu_2 \text{ versus } H_A: \mu_1 > \mu_2$$

   To test this hypothesis the retailer randomizes $n = 1000$ customers to each condition and records their purchase prices. Data for this experiment can be found in the file `retail.txt`. Use R to analyze this data and hence conclude whether condition "A" results, on average, in significantly larger purchase prices than condition "B", at a 5% level of significance. Be sure to assess the validity of any assumptions your analysis makes.

   Assuming an effect size of $\delta = \mu_1 - \mu_2 = 1$, and a common standard deviation of 5, calculate the power of this hypothesis test, and determine the sample size that would be sufficient to achieve 90% power.

5. Suppose your colleague wishes to compare two groups but they are unwilling to use parametric hypothesis tests such as t-tests for means or z-tests for proportions. Instead they wish to use a non-parametric randomization test, and they ask you for your help. Unfortunately you are busy, so you only have time to write pseudocode for such a test. Provide that pseudocode here.

6. Suppose we are interested in testing the following hypothesis:

$$H_0: \pi_1 = \pi_2 \text{ versus } H_A: \pi_1 \neq \pi_2$$

where π_i is the proportion of individuals that conform to some condition in population $i = 1, 2$. Derive a formula for the sample size n that each sample would need to have to be able to this hypothesis a $100 \times \alpha\%$ significance level and a power of $1 - \beta$.

Using this formula, calculate the sample size necessary for testing this hypothesis assuming Type I and II error rates of 5%, and assuming $p_1 = 0.78$ and $p_2 = 0.84$. Repeat this calculation using the `power.prop.test()` function in R.

Note that more problems will be assigned as the course progresses.