[C]

Fatemeh Nadi
810101285
fatemehnadi@ut.ac.ir

Homework 1
Statistical Inference , Fall 1401

November 3, 2022

1

If we want to find a relationship and limit ourselves to an observational study, we can take a survey from the students; divide our population into three groups; how many of them study with music that has lyrics and study with music that has no lyrics, and how many studies without music, and finally by comparing the distribution of scores of these three groups or the average of each Check how effective the methods are.
However, to design an experimental experiment, we randomly divided the students into three groups and placed them in three different classes. Before these, we measure the average or the distribution of the scores of each class. We will repeat this test for several months for each class, and finally, we will evaluate the grades of each category, the level of progress in each lesson, etc. In the first class, we follow the same teaching method as before In the second class, while the teacher is teaching, music with lyrics should be played, and encourage students to study with music that has lyrics at home. In the third class, while the teacher is teaching, music without lyrics should be played and assigned to students to study with music without lyrics at home. Finally, each group must evaluate which method is more efficient

2

- a
  There may be no confounding factors, and their sunglasses are less effective, but proof of this must control some conditions further experimentation should be done rather than assuming cause and effect.
  But if we assume that there is, the following things are possible.
  1. An increase in prices in 2018 may have led to decreased sales according to the web.
  2. If the same people lived in the city during 2017 and 2018, people may already have umbrellas in 2017 and might not need to buy them.

- b
  maternal age. As seen in the figure, the main factor affecting Down syndrome is the mental age of the mother, not the birth order. The frequency of Down syndrome in young mothers is low regardless of birth order, and the frequency of Down syndrome in older mothers is high regardless of birth order. On the other hand, there is a strong relationship between the mental age of mothers and the order of birth, so naturally, the mother's age at the birth of the first child is lower than her age at the birth of the fifth child. However, when stratified by birth order and maternal age, we

can see that birth order did not have an independent effect. The apparent association with birth order resulted from confounding and overestimation caused by maternal age.

- c
  The students who gave more importance to the lesson tried to use that software and followed the advice of their teacher to get the maximum learning and participated in this case. Volunteer sampling. On the other hand, the students who use the software spend more than an hour studying, which is a confounding factor.

- d
  We have a confounding factor, possibly because qualified individuals performed better at their previous companies. As a result, their managers have better and more detailed recommendation letters to express satisfaction, and those who provide better work have more details in their work to report. Not because they have a better recommendation letter, so they are hired; here, a positive association can be seen.

3

- a
  Cluster Sampling: All flights on every day are the clusters(homogeneous clusters), and choose five flights on a particularly selected day, and we are sampling all data within only 5 of these clusters.

- b
  Simple Random Sampling: All employees have the same chance to be sampled.

- c
  Stratified Random Sample: Divide the population into subgroups based on professionalism and ensure that every subset is correctly represented in the sample.

- d
  Multi-Stage Sampling: We consider each page as a homogeneous cluster and divide the population clusters, randomly sampling a few clusters, then randomly sample within these clusters when choosing the last number.

4

- a
  explanatory variable : Study hours/week.
  response variable : GPA.

- b
  Cannot find a connection, maybe to some extent semi-positive association; the more study hours, the higher the GPA. Furthermore, in the range of 5 to 35, we have all scores, so we can conclude that the time of studying is not just a factor in the score. We can consider IQ, focus, facility, and so on Some unusual observations are: there is a person with a GPA above 4, but 4 should be the limit. Two people claimed to study more than 60 hours per week.

- c
  Observational study. We do have no treatment and controlled group.

- d
  We can't because it is an observational study, we can just point to an association (or correlation).

5

- a
  In the histogram, the two modes are more visible and skewness. In the boxplot, the outliers and median value are more apparent, skewness and percentile.

- b
  Men and women on average have very different finishing times. One mode is for men while the other is for women.

- c
  The women's finishing times are higher and more variable than the men's finishing times.

- d
  The women's time is always higher than the men's time (proof the bimodal distribution in the previous section), and they are both decreasing over time. They decrease rapidly at first, then at a slower pace over time.

6

- a
  No. The woman group has much higher cured rate (almost twice).

- b
  the null hypothesis (H0): cured is independent of gender or The effect of cure on men and women is equal.
  the alternative hypothesis (H1): cured is dependent of gender or The effect of cure on men and women is not equal.

- c
  In this experiment, we have 28 cured and 32 not cured people. We repeat this test 200 times or 1000 times regardless of gender; in this way, we select 60 people who suffered from high blood pressure and were given specific treatment, and in each test, we measure and show the difference between the percentage of cured and not cured.

- d
  According to the plot, the distribution of the difference in the percentage of men and women treated is approximately Gaussian On average, men and women cured equally, indicating that this treatment is independent of gender.

7 R

Please see this file: "Q7-R.Rmd" and "Q7-R.html"
Code and explanation are provided.

8 R

Please see this file: "Q8-R.Rmd" and "Q8-R.html"
Code and explanation are provided.