COMM611_assignment_2_hypothesis_testing

Xuanjun Gong

2025-02-12

Assignment 2

Q1

Toss a fair coin ($p_{head} = 0.5$) 10 times, and record each outcome as X_i. What is the expected value and variance of the statistic $\bar{X}$ for $\bar{X} = \frac{1}{10}\sum_{i=1}^{10} X_i$?

Q2

For a normal distributed random variable $X \sim N(2, 100)$. Suppose we have collected $i.i.d$ random samples X_i of sample size $n = 25$ from the population X. What is the sampling distribution for the sample mean $\bar{X}$?

Q3

A health communication researcher is studying the effectiveness of a new public health campaign aimed at increasing awareness of the dangers of excessive sugar consumption. Previous national surveys indicate that, on average, people score 50 points on a nutrition knowledge scale (ranging from 0 to 100).

To assess whether the campaign has improved nutrition knowledge, the researcher surveys a random sample of 10 individuals who were exposed to the campaign. Their scores on the nutrition knowledge scale are as follows:

55,55,45,60,55,55,50,60,55,60

1. Calculate the sample mean and sample variance for nutrition knowledge.
2. Now you want to conduct a two-tailed t test to test if the campaign has improved people's nutrition knowledge from the general nutrition knowledge in the previous national survey. State your null hypothesis and alternative hypothesis.
3. Construct the t statistic and state the sampling distribution for this t statistic. $t = \frac{\bar{X}-\mu}{\sqrt{S^2/n}}$.
4. What is the rejection region for your sample mean? Given that $P(t_9 < -2.262) = 0.025$.
5. What is the p value for your t test? Write down the R code to compute this p value.
6. Construct the 95% confidence interval for people's average nutrition knowledge after viewing the promotional message.
7. What is your conclusion of the hypothesis testing? Does the health campaign improve people's nutrition knowledge?

Q4

In a tech company, your team want to test a new feature in your product (a social media app). In order to test whether or not this new feature can improve your customers' satisfaction of using the product. Your team conducted a A-B testing experiment in following ways:

You randomly sampled a group of $n = 10000$ of existing customers and assigned them into control group. Then you randomly sampled another group of $m = 100$ of existing customers and assigned them into experiment group.

You updated the app in the experimental group with the new feature. One month later, you measured customers' satisfaction rate in both control and experiment group.

You noticed that the data scientist in your team wanted to conduct a independent sample t test with equal variance assumed. But you are not sure if the equal variance is a valid assumption. Thus you plan to conduct a F test to test for equal variance.

Given that for the control group, you measured people's satisfaction $\bar{X} = 5$, and sample variance $S_x^2 = 1$. For the experiment group, you measured people's satisfaction $\bar{Y} = 2$, and sample variance $S_y^2 = 1.2$.

1. State your null hypothesis and alternative hypothesis.
2. Construct the F statistic, and gives the distribution of this F statistic.
3. Compute p values for your hypothesis testing, and make your conclusions. You can just write down the R code to calculate the p value.
4. Suppose your team decided to conduct the independent sample t test without equal variance assumed. Construct the t statistic, and determine its distribution (degree of freedom) using the formula below.
5. Write down the R code to compute the p value and write the p value down. What is your conclusion?

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} \sim t_k \qquad k = \frac{(\frac{S_x^2}{n} + \frac{S_y^2}{m})^2}{\frac{1}{n-1}(\frac{S_x^2}{n})^2 + \frac{1}{m-1}(\frac{S_y^2}{m})^2}$$

Q5

Suppose you plan to conduct a survey before the 2028 presidential election to know which candidate (Democratic vs. Republican) will win the 2028 election. You plan to randomly sample 1000 participants and ask them which candidate they will vote. You want to test whether or not the Democratic candidate will win the election. Since you have a large sample size, we can assume the sampling distribution for the proportion of people vote for Democratic candidate $\bar{X}$ follows a normal distribution $N(\mu, \sigma^2)$, and conduct a one-tailed Z test with significance level $\alpha = 0.05$.

1. State your null hypothesis and alternative hypothesis.
2. What is the sampling variance for $\bar{X}$ if null hypothesis is true?
3. Suppose you have observed that $\bar{X} = 0.55$, construct your Z statistic.
4. What is your rejection region for the observed statistic $\bar{X}$?
5. What is your p value for the one-tailed Z test? You can just write down the R code to calculate the p value.
6. You want to know if your study have sufficient power to detect the effect, so you conducted a post-hoc power analysis to calculate your power with the assumption that $H_1 : p = 0.55$. Compute the power. Note that the sampling variance will change depending on p.