ASDS - Applied Statistics with R

Fall 2018, YSU

Homework No. 09

Due time/date: No Due Date

Note: Please use R only in the case the statement of the problem contains (R) at the beginning. Otherwise, show your calculations on the paper. Supplementary Problems will not be graded, but you are very advised to solve them and to discuss later with TA or Instructor.

Problem 1. State a hypothesis (i.e., state the Null and Alternative Hypo's) to:

a. Check if a die is loaded;
b. Check if the data comes from the $\mathcal{N}(2,4)$ or it comes from $\mathcal{N}(1,3)$. Supposedly, the latter is true;
c. Check if the average Armenian female height (for persons of the age between 25 and 49) is 158.1 cm[1]
d. Check if the number of female students is larger than the number of male students at YSU;

For each case, write what it means that we will have Type I Error and Type II Error.

(Supplementary) In each case suggest a method to test the hypothesis, if possible.

Problem 2. Assume $x_1, ..., x_{100}$ is an observation from the model $\mathcal{N}(\mu, 4)$. We want to test the hypothesis

$$H_0: \mu = 0 \qquad vs \qquad H_1: \mu = 1.$$

The Rejection Region is $RR = \{\overline{x} > 0.5\}$. The sample mean of our observations x_k is $\overline{x} = 0.32$.

a. What is the decision of the test?
b. Calculate the Type I Error;
c. Calculate the Type II Error;
d. Calculate the power of the test.

Problem 3. I believe that in our Department, 50% of undergrad students are from FinMath Major, 40% are from Math Major and the rest are from Mechanics. Assume I am randomly asking 40 students about their Major, and get that 18 of them are from FinMath, 10 are from Math and 12 are from Mechanics. Is this data supporting my claim? Test at the 5% significance level.

[1]See https://en.wikipedia.org/wiki/List_of_average_human_height_worldwide

Problem 4. (**R**) Generate 50 random numbers using **R** from the Normal Distribution using the Mean $\mu = 0.6$ and Variance $\sigma^2 = 4$. Round that numbers up to two decimal points (i.e., if you will have 0.12334565, then you need to round it to 0.12).

Now assume that our rounded numbers are also from the Normal Distribution (maybe with another mean and variance). Test the hypothesis, at the 5% significance, that the rounded numbers are from the Normal with the mean 0.5 versus they are not.

Problem 5. (**R**) From some internet source download some Company Stock historical prices. Plot the histogram of weekly returns.

a. Construct a 95% CI for the mean (weakly) return, assuming that returns are Normally Distributed;

b. Test the hypothesis that the mean (weakly) return is 0.02 at 95% confidence level, again assuming that returns are Normally Distributed;

c. Estimate the probability that the (weakly) return will be $>= 0$. Construct a 90% CI for that probability.

Problem 6. In this problem, we will design the famous t-test for one sample. The have the following observation: $x_1, x_2, ..., x_n$ coming from the Normal Distribution. We want to test the Hypothesis about the Mean:

$$H_0 : \mu = \mu_0 \qquad \text{vs} \qquad H_1 : \mu \neq \mu_0.$$

We model this by a random sample: we assume that we have a random sample

$$X_1, X_2, ..., X_n \overset{IID}{\sim} \mathcal{N}(\mu, \sigma^2),$$

where we do not know the value of σ^2. We want to test the hypothesis

$$H_0 : \mu = \mu_0 \qquad \text{vs} \qquad H_1 : \mu \neq \mu_0.$$

So we need to design the test.

a. The Model is written above. We choose the following Test Statistics:

$$t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}},$$

where $s^2 = \frac{\sum_{k=1}^{n}(X_k - \overline{X})^2}{n-1}$. What is the distribution of t under the hypothesis H_0?

b. We take a Rejection Region of the form $|t| > c$. Find the critical value c such that the significance level of the test will be α.

c. Describe the test: in which case we will reject the hypothesis H_0, and in which case we will not.

d. I have generated the following data from a Normal Distribution, using **R** (I am truncating my data for the sake of simplicity):

$$1.88, \ -0.16, \ 1.95, \ 0.30, \ -0.41, \ 1.49, \ -0.30, \ -0.74, \ 2.34, \ 0.28$$

Check the hypothesis that I was using the mean 1.2 vs. I was using other mean.

e. I am changing the hypothesis to be

$$H_0: \ \mu = \mu_0 \qquad \text{vs} \qquad H_1: \ \mu > \mu_0$$

We use the same Test Statistics. Choose a reasonable Rejection Region, find the critical value and describe the test in this case.

Problem 7. On a new highway, the speed limit is set to 70 km/h. The police department is claiming that the average speed is higher than 70 km/h on that highway. In support, the department is providing a data from aragachaphs for randomly chosen 10 cars: the recorded speeds are (in km/h):

$$66, 79, 80, 74, 81, 79, 65, 78, 77, 69.$$

Is this data providing a sufficient evidence against the hypothesis that the average speed is 70 km/h? Use 1% significance level. You can assume that the car speeds are normally distributed.

Problem 8. Historically, 70% of the students of one of the Stat Professors are passing the Stat course. Assume, on a given year, only 54 of 74 students are passing the course. Perform a test of significance level 0.05 that the passing rate is 70% versus it is less.

Problem 9. Here are some facts from the 2017 RA Parliament elections. According to Wikipedia (see https://en.wikipedia.org/wiki/Armenian_parliamentary_election,_2017), the opinion poll conducted from 1827 March 2017 among 1005 persons resulted in the following distribution of percentage of votes:

Party	TsD	HHK	HVK	Yelq	HYD	AD	C-HJK	ORO	HKK	ND/NA
	29	30	5	9	7	3	4	4	3	

According to the same source, the actual results were[2]

Party	TsD	HHK	HVK	Yelq	HYD	AD	C-HJK	ORO	HKK
	27.35	49.15	3.72	7.78	6.58	0.94	1.66	2.07	0.75

Is the sample data from the poll consistent with the actual results? Us the significance level 0.05.

Note 1: You can add the above table category ND/NA (Not Decided yet/Not Available) to any of the parties you want, or you can distribute evenly between the parties.

[2] In fact, there is an inconsistency between Armenian and English language pages at Wikipedia. In the English language page, the HHK percentage votes are 49.17, and it is 49.15 in the Armenian language page. And the sum of percents in the English-language page is a little bit higher than 100%, so I am using the Armenian page results.

Note 2: No politics, please! Only pure Statistics (and a χ^2-test) $\ddot\smile$

Problem 10. Assume we want to test a hypothesis about the mean of the Normal Distribution, when the variance is known. In particular, we want to design a test in the following way: we want to have both the significance level $\alpha = 0.05$ (the confidence level 95%), and we want to have a power not less than 95%. Our test is ($\mu_0 < \mu_1$ are given, fixed):

$$H_0: \ \mu = \mu_0 \qquad vs \qquad H_1: \ \mu = \mu_1.$$

If the number of the sample is fixed, it is possible that we will not be able to control both Type I and II Errors probabilities. On the other hand, we can choose the number of observations, n, to obtain the necessary bounds for error probabilities. So assume

$$X_1, ..., X_n \sim \mathcal{N}(\mu, \sigma^2)$$

is our Random Sample, and σ^2 is known. We will test the above hypothesis based on the Test Statistics

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}.$$

Our Rejection Region will be

$$RR = \{Z > c\}.$$

a. Choose the Critical Value c such that the significance level of our test is α;

b. Calculate the Probability of the Type II Error under the obtained above Rejection Region;

c. Choose the number n so that the power of the test will be not less than 95%;

d. What will happen with n, if μ_1 will approach μ_0? Explain.

Problem 11. (from Ross's textbook) A previous sample of fish in Lake Michigan indicated that the mean polychlorinated biphenyl (PCB) concentration per fish was 11.2 parts per million with a standard deviation of 2 parts per million. Suppose a new random sample of 10 fish has the following concentrations:

$$11.5, 12.0, 11.6, 11.8, 10.4, 10.8, 12.2, 11.9, 12.4, 12.6$$

Assume that the standard deviation has remained equal to 2 parts per million, and test the hypothesis that the mean PCB concentration has also remained unchanged at 11.2 parts per million. Use the 5 percent level of significance.

Problem 12. (from Ross's textbook) A farmer claims to be able to produce larger tomatoes. To test this claim, a tomato variety that has a mean diameter size of 8.2 centimeters with a standard deviation of 2.4 centimeters is used. If a sample of 36 tomatoes yielded a sample mean of 9.1 centimeters, does this prove that the mean size is indeed larger? Assume that the population standard deviation remains equal to 2.4, and use the 5 percent level of significance.

Problem 13. (from Ross's textbook) A fast-food establishment has been averaging about \$2000 of business per weekday. To see whether business is changing due to a deteriorating economy (which may or may not be good for the fast-food industry), management has decided to carefully study the figures for the next 8 days. Suppose the figures are

$$2050, 2212, 1880, 2121, 2205, 2018, 1980, 2188$$

a. What are the null and the alternative hypotheses?

b. Are the data significant enough, at the 5 percent level, to prove that a change has occurred?

c. What about at the 1 percent level?

Problem 14. (from Ross's textbook) A recently published study claimed that the average academic year salary of full professors at colleges and universities in the United States is $\$87,800$. Students at a certain private school guess that the average salary of their professors is higher than this figure and so have decided to test the null hypothesis

$$H_0 : \mu \leq 87,800 \qquad \text{against} \qquad H_1 : \mu > 87,800$$

where μ is the average salary of full professors at their school. A random sample of 10 professors elicited the following salaries (in units of $1000):

$$91.0, 79.8, 102.0, 93.5, 82.0, 88.6, 90.0, 98.6, 101.0, 84.0$$

a. Is the null hypothesis rejected at the 10 percent level of significance?

b. What about at the 5 percent level?

Problem 15. (from Ross's textbook) It has been common wisdom for some time that 22 percent of the population have a firearm at home. In a recently concluded poll, 54 out of 200 randomly chosen people were found to have a firearm in their homes. Is this strong enough evidence, at the 5 percent level of significance, to disprove common wisdom?