2028: Basic Statistical Methods
Homework 5

This homework is due **Thursday, Nov. 5** in class BEFORE class starts. Late papers will not be accepted. Please do not turn in any papers to any mailbox.

- Please remember to staple if you turn in more than one page.
- Please make sure to SHOW ALL WORK in order to receive full credit.

1. **Hypothesis Testing for the Population Mean**

   (a) Textbook 9-62 page 318: baseball coefficient of detection. You may use data baseball.txt. Please skip part (b), and assume the data is normally distributed. Also, since the true variance is unknown, you may use sample variance.

   9-62. Consider the baseball coefficient of restitution data first presented in Exercise 8-92.
   (a) Do the data support the claim that the mean coefficient of restitution of baseballs exceeds 0.635? Use $\alpha = 0.05$. Find the P-value.
   (b) Check the normality assumption.
   (c) Compute the power of the test if the true mean coefficient of restitution is as high as 0.64.
   (d) What sample size would be required to detect a true mean coefficient of restitution as high as 0.64 if we wanted the power of the test to be at least 0.75?
   (e) Explain how the question in part (a) could be answered with a confidence interval.

   (b) The drying time for a certain type of paint under specified test conditions is known to be normally distributed with mean value 75 min and standard deviation σ 9 min. Chemists have proposed a new additive designed to *decrease* average drying time. It is believed that the drying times with this additive will remain normally distributed with $\sigma = 9$. Because of the expense associated with the additive, evidence should strongly suggest an improvement in average drying time before such a conclusion is adopted. Experimental data consist of drying times from $n = 25$ test specimens.

      i. State the hypotheses to be tested. Comment on your choice of test hypotheses?
      ii. Find the type I error for the rejection region $\bar{X} \leq 70.8$.
      iii. Compute the type II error of the test based on the rejection region $\bar{X} \leq 70.8$ when the alternative H_A is true and $\mu = 72$.

2. **Hypothesis Testing for the Proportion Parameter**

   (a) Textbook Problem 9.89 page 329.

9-89. An article in the *British Medical Journal* ["Comparison of Treatment of Renal Calculi by Operative Surgery, Percutaneous Nephrolithotomy, and Extra-Corporeal Shock Wave Lithotrips," (1986, Vol. 292, pp. 879–882)] found that percutaneous nephrolithotomy (PN) had a success rate in removing kidney stones of 289 out of 350 patients. The traditional method was 78% effective.

(a) Is there evidence that the success rate for PN is greater than the historical success rate? Find the P-value.

(b) Explain how the question in part (a) could be answered with a confidence interval.

(b) Textbook Problem 9.91 on page 329.

9-91. A researcher claims that at least 10% of all football helmets have manufacturing flaws that could potentially cause injury to the wearer. A sample of 200 helmets revealed that 16 helmets contained such defects.

(a) Does this finding support the researcher's claim? Use $\alpha = 0.01$. Find the P-value.

(b) Explain how the question in part (a) could be answered with a confidence interval.

3. **Computer Problem**

Cloud seeding has been studied for many decades as a weather modification procedure (for an interesting study of this subject, see the article in *Technometrics*, "A Bayesian Analysis of a Multiplicative Treatment Effect in Weather Modification," Vol. 17, pp. 161 - 166). The rainfall in acre-feet from 23 clouds that were selected at random and seeded with silver nitrate is given the the quiz data file.

Instructions for reading the data. The data is in the file 'rain.txt'. In these data, there is only one column that specifies the rainfall data.

To read the data in R, save the file in your working directory (make sure you have changed the directory if different from the R working directory) and read the data using the R function `read.table` as follows:

```
data = read.table("rain.txt")
```

Then, convert the data to numeric, using:

```
Rain= as.numeric(data[,1])
```

Now, answer the following questions:

(a) Does the rain fall seem to follow an approximate normal distribution? Use a histogram.

(b) What is the two-sided 95% confidence interval for the mean rain fall? Does that support the claim that the mean rain fall is not equal to 35?

(c) Now, we want to test the claim that the mean rain fall is above 30. What is the null and alternative hypotheses that we should use to test this claim?

(d) Perform a t-test to test the claim in question 3. Use a confidence interval of 95% What are your conclusions?

You may use:

```
t.test(data,alternative,mu,conf.level)
Where,
       data should be Rain,

       alternative is either alternative=c("greater"), or alternative=c("less")
       or alternative=c("two.sided") depending on your alternative hypothesis,

       mu should be mu=30,

       conf.level should be conf.level=0.95
```

(e) What was the p-value for the hypothesis test in Question 4? Based on this p-value, would you reject the null hypothesis at 99% confidence level? How about at 90% confidence level?