Design and Analysis of Experiments, HW 1

Due Tuesday, December 9

Instructions: This homework is due Tuesday, December 9, at the beginning of class. (Late hw's/e-mailed hw's are not accepted.) Be sure to include all generated Minitab output in your hw, especially graphics. **NOTE: All material for a given problem - including your work plus corresponding Minitab output and graphics - is to be on one page (or consecutive pages). For example, putting all your graphs at the end of the hw is not acceptable.** In other words, I should be able to grade your hw by reading it front to back without flipping back and forth. Also, be sure to put your mail box number on your hw. If you have any questions about the preceding or the questions below, please contact me.

0: Be sure to do the following: read chapter 1 (plus Canavos DOE handout), skim sections 2.1-2.3, read section 2.4 (skip 2.4.3 on sample size determination), and the probability and hypothesis testing review handouts.

1: (Based on consult with Rose-Hulman human-powered vehicle designer.) A human-powered vehicle designer wants to determine which of three seat heights - low, medium, or high - maximizes the power output of the vehicle's cyclist. The designer measures the power output of the cyclist by having him pedal all out for two minutes on a dynamometer. Because this task is demanding, the cyclist can only do three runs on a given day. Since the designer wants three power measurements for each seat height, this means the cyclist must do runs on three different days. With respect to this scenario, answer the following questions:

i. What is the goal of this experiment? What is the response variable and what is/are the factor(s) of interest? Are there any nuisance factors?

ii. Keeping in mind the basic principles discussed in setion 1-3, write out how the vehicle designer should conduct the experiment, i.e., tell him *exactly* what he should have the cyclist do on each of the three days.

iii. Describe/discuss your experimental design in part ii with respect to randomization, replication, and blocking.

2: Clay tiles are fired in a kiln. Unfortunately, some of the tiles crack during the firing process. A manufacturer of clay roofing tiles would like to investigate the effect of clay type on the proportion of cracked tiles. Two different types of clay are to be considered. One hundred tiles can be placed in the kiln at any one time. There will be slight variations in firing temperature at different locations in the kiln, and firing temperature may also effect cracking.

i. What is the response variable in this investigation? What is the factor of interest? Are there any extraneous (nuisance) factors?

ii. Suppose you are asked to do the experiment. Construct a design/experimental procedure which minimizes the effect of temperature variation within the kiln.

3: The continuous random variable X has the following density:

$$f_X(x) = \begin{cases} 0, & x < 0 \\ \frac{2}{9}x, & 0 \leq x \leq 3 \\ 0, & x > 3 \end{cases}$$

Determine the following:

i. Determine P($X < 1$).

ii. Compute the mean of X, μ_X.

iii. Compute the variance and standard deviation of X, σ_X^2 and σ_X, respectively.

4: The following numbers represent an IID sample of size 10 from the RV X in problem 4 above:

`2.91, 1.65, 2.25, 2.18, 2.64, 0.68, 1.12, 2.52, 1.58, 1.63`

i. Compute the sample mean (manually or using Minitab's `Stat -> Basic Statistics -> Display Descriptive Statistics` routine.)

ii. Compute the sample variance and standard deviation similary.

iii. How well do the above estimates agree with the actual values of μ_x, σ_X^2, and σ_X you computed in problem 4. Why don't the estimates agree with these values?

6: Suppose the yield strength (ksi) of specimens of A36 grade steel is normally distributed with $\mu = 43$ and $\sigma = 4.5$. Answer the following:

i. What proportion of specimens have a yield strength of at most 40 ksi? At least 60 ksi?

ii. What proportion of specimens have a yield strength of between 40 and 50 ksi?

iii. What yield strength value separates the strongest 75% of all specimens from the others?

7: Recall that the IID assumption is required by essentially all inference procedures used in MA223/MA382 and almost all procedures we will use in this class. Do the following (note that the data sets are provided on the course website):

i. Are the `viscosity` data IID? Check by

1. constructing a time series plot (`Graph -> Time Series Plot`) and checking if the data are identically distributed by determining if there are any systematic trends in location or spread then
2. use the `autotest` macro to determine if the data are independent.

ii. Are the `resistivity` data IID? Check using via the same procedure used in part i.

8: In this problem you will compare the average reaction times of subject 2 under the "Think, Act" (TA) and "Act" (A) protocols by testing H_0: $\mu_{TA} = \mu_A$ vs. H_1: $\mu_{TA} \neq \mu_A$ using a two-sample t-test as follows:

i. The first step in any data analysis is to look at the data using appropriate graphics. Construct a time series plot containing both samples (columns `TA2` and `A2`) using `Graph -> Time Series Plot -> Multiple`. (Recall that data is in `rtimes.MTW` data set on course website.) Based on this graph of the data, do you think H_0 is true?

ii. In order to use a two-sample t-test, the two samples must be IID. Determine if the samples are IID by constructing and interpreting time series plots of each sample and using the `autotest` macro. What do you conclude?

iii. Assume, irrespective of your conclusion in part ii, that both samples are IID. Since the sample sizes are small, in order to use a two-sample t-test the samples must be from normal populations/processes. Test that each sample is from a normal population/process using Minitab's normality test procedure (`Stat -> Basic Statistics -> Normality Test` - use the default (Anderson-Darling) test. What do you conclude at $\alpha = 0.05$?

iv. Irrespective of your conclusion in parts ii and iii, assume the two-samples are IID and from normal populations/processes so that you can do a two-sample t-test using Minitab's `Stat -> Basic Statisticts -> 2-sample t`. Using $\alpha = 0.05$, what do you conclude?

v. Manually verify the value of the t-test statistic reported by Minitab using the following formula:

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}.$$

(Hint: You can easily compute the means and standard deviation using Minitab's `Stat -> Basic Statistics` procedure.)

vi. In part iii you should have concluded that at least one of the samples was non-normal. Since your samples sizes are small, you should use the two-sample bootstrap macro `boot2mean` to compare the two means. Do so. What do you conclude at $\alpha = 0.05$?

9: Analyze the data of problem 2-26, page 61, as follows:

i. Input the data into Minitab.

ii. Recalling that the first step in any data analysis is to look at the data, construct side-by-side boxplots of the data as follows:

1. Select *Graph -> Boxplot*
2. Select *Multiple Y's Simple*, click OK
3. Select both columns for *Graph variable*

Compare the two samples with respect to location and spread using this graphic. Recall that the boxplot measure of location is the median (center line in the box) and its measure of spread is the IQR (*Interquartile Range*) the distance between the 1st and third quartiles of the data (bottom and top edges of the box).

iii. Assume that the data are IID so that you can test whether or not the two samples are from normal populations. What do you conclude at $\alpha = 0.05$?

iv. Using your result from part iii, test $\mu_1 = \mu_2$ vs. $\mu_1 \neq \mu_2$ using an appropriate procedure. What do you conclude at $\alpha = 0.05$?

10: Suppose you are testing H$_o$: $\mu_1 = \mu_2$ using a two-sample t-test. The value of your test statistic is t = 1.753 with 15 degrees of freedom. Compute the p-values below using table II in your book or else using Minitab's `Graph -> Probability Distribution Plot` as follows:

1. Click on `View Probability` then click `OK`.
2. Under the `Distibution` tab, select `t` for distribution and enter `15` for the degrees of freedom.
3. Under the `Shaded Area` tab, click on the `X Value` radio button and then enter the value of your test statistic for the `X value`.
4. Finally, select the appropriate tail area to use for computing your p-value then click `OK`.

i. If the alternative hypothesis is H$_1$: $\mu_1 > \mu_2$, what is the p-value?

ii. If the alternative hypothesis is H$_1$: $\mu_1 < \mu_2$, what is the p-value?

iii. If the alternative hypothesis is H$_1$: $\mu_1 \neq \mu_2$, what is the p-value?

11: Short answer questions.

i. Give the two important benefits/properties of randomization.

ii. What does the acronym "MVT" stand for?

iii. "MVT" is "business speak" for what type of experimental design? (Hint: Read section 1.5.)