**Statistics 500 - Homework #1, Fall 2020**
**Due by 12pm on Friday 8/28/2020**

Reading Assignment: Statistical Sleuth, Chapters 1 and 2.

Please access the course on your canvas page and read the class syllabus. Check that you can download homework assignments and data files.

1. For each of the following experiments, answer questions (a)-(b).
   (a) Identify (i) experimental units, (ii) observational units, (iii) treatments, and (iv) response variable.
   (b) Does the experiment utilize the three principles of experimental design: (i) replication, (ii) blocking, (iii) randomization? Provide a yes/no answer to each of the principles, and briefly justify your answer.

   **Experiment I:** Some researchers were interested in studying the effects of different fertilizer amount (Low Nitrogen and High Nitrogen) and different genotypes (one energy line and one grain line) of sorghum on biomass. For each genotype, six pots of one-week-old seedlings were available, and each pot held one seedling. For each genotype, the researchers randomly assigned three pots to high nitrogen treatment (H) and the remaining three pots to low nitrogen (L) treatment. After 2 weeks, the fresh weight for each seedling was measured. In total, we have 12 observations.

   **Experiment II:** A statistics teacher wanted to determine if having business students use clickers to respond to questions posed in a business statistics class would improve student learning. The teacher decided to have students use clickers in one class of introductory business statistics 226. She did not have students use clickers in a second class of business statistics 226 that she taught during the same semester. She tossed a coin to select the class to use the clickers. She used the same book and the same lectures in both classes and gave the same assignments and same exams to both classes. There were 90 students in each class. At the end of the semester, she compared the final exam scores for the students in the class that used clickers to the final exam scores for the students in the class that did not use clickers.

2. A researcher would like to investigate the effect of feeding high-oil corn to one-year-old steers. Five diets are selected for use in the experiment. The diets are (A) a traditional corn diet, (B) a diet with 25% high-oil corn, (C) a diet with 50% high-oil corn, (D) a diet with 75% high-oil corn, and (E) a diet with 100% high-oil corn. Five farms, each with 5 steers raised in individual pens, are available for use in the experiment. The researcher is planning to measure the ribeye area of each steer at slaughter. There are two proposed experimental designs:

   *Design 1*: One farm (including all 5 steers) is randomly assigned to each of the five diets: A, B, C, D, and E.

   *Design 2*: Within each farm, one steer was randomly assigned to each of the five diet treatments.

   Which design is better based on the design principles? Why?

3. For a study about a new treatment for pancreatic cancer, 100 pancreatic cancer patients are recruited. For ethical reasons every patient in the study is given the highest current standard of care which consists of surgery to remove cancerous tumors from the pancreas followed by a period of radiation therapy. After receiving the standard treatment, each patient is given the opportunity to have a

second experimental treatment that consists of coating some of the cancerous cells removed from the patient during surgery with an antigen that the human body will reject and injecting the coated cells back into the pancreas of the same patient. The basic idea is that this second treatment will help prevent recurrence of cancer tumors because the patient's immune system will attack the coated cells and also attack other cancerous cells in the pancreas that the surgery and radiation treatments may have missed. The 57 patients who volunteer for the second treatment and the 43 patients who refuse the second treatment are followed for three years to determine how long each patient survives.

(a) Is this study an experiment or not? Justify your answer.
(b) Describe whether or not the principles of designing experiments (including control of extraneous variation, randomization, and replication) were used in this study.

4. A farmer wants to make an informed decision about which variety of corn to use. Which variety of corn will produce higher yields, A or B, or will the two varieties produce equal yields on average? To answer this question, she decides to run an experiment, a field trial, to compare two corn varieties, A and B. She has a field with 36 plots of equal size for the experiment. She randomly selects 18 plots to plant variety A and on the other 18 plots she plants variety B. The yields for the 36 plots are given in the file **farming.csv**.

   (a) Use SAS to evaluate the following summary statistics for each treatment group: Q1, Q3, IQR, Median, Sample Mean, and Sample Std. Deviation. Summarize these values in a table similar to the one below.

| Statistic | Variety A | Variety B |
|---|---|---|
| Q1 | | |
| Q3 | | |
| IQR | | |
| Median | | |
| Sample Mean | | |
| Sample Std. Dev. | | |

   (b) Use the SGPLOT procedure in SAS to construct side-by-side box plots of yields for the two varieties of corn and include it with this assignment.
   (c) Use the box plots from part (b) and the summary statistics from part (a) to describe and compare features of the distributions of yields for the two varieties of corn.
   (d) Use SAS to conduct a randomization test to test whether the mean yield of the two varieties are the same or different. Make sure to give the null and alternative hypotheses, observed test statistic, p-value, decision and conclusion for the test at the significance level of 0.05. The conclusion should be stated in the context of the problem.

5. Suppose researchers conducted a randomized two-treatment experiment with three experimental units per treatment. The observed data is given below. The researchers want to test whether the means of the response variable are the same for both treatments or not. To answer this question, we will use randomization test.

| Treatment 1 | | | Treatment 2 | | |
|---|---|---|---|---|---|
| 4.8 | 5.2 | 5.0 | 7.7 | 8.2 | 8.1 |

   (a) Based on the observed dataset, what is the sample mean difference between the two treatments?

(b) How many possible permuted datasets can you get? (How many different ways to assign treatment labels to the six experimental units for this experiment?)
(c) List all possible permuted datasets, and calculate the difference between two sample means for each permuted dataset.
(d) Based on your results in (a) and (c), what is the p-value for this randomization test?

6. This question is intended to help you get familiar with the results for linear combination of random variables.

a. Suppose $Y_i$ is a random variable with mean $\mu_i$ and $\mu_i$ is known for all $i$. Calculate:

i. $E(2Y_1 + 3)$.
ii. $E(Y_1 + Y_2 + Y_3)$.

b. A constant real number $c$ can be viewed as a random variable that always takes the value $c$. Calculate
i. $E(c)$.
ii. $Var(c)$.

c. If two independent random variables $Y_1$ and $Y_2$ have the same variance $\sigma^2$, calculate
i. $Var(Y_1 - Y_2)$
ii. $Var[(Y_1 + Y_2)/2]$

d. $\text{Var}(Y_i) = \sigma^2$ for $i = 1, 2, \ldots, n$, that is, the random variables $Y_i$ have the same variance; and $\sigma^2$ is a known constant. Suppose $Y_i$ and $Y_j$ are independent for all pairs of $i, j$ $(i \neq j)$. Calculate:

i. $Var(\sum_{i=1}^{n} Y_i)$
ii. $Var(\frac{1}{n}\sum_{i=1}^{n} Y_i) = Var(\bar{Y})$

***In case you need some resources for scanning written work to PDF. Here I list some tools that you can use:***

Natively in iOS – iPhone, iPad
https://www.cnet.com/news/how-to-use-ios-11s-notes-app-as-a-document-scanner/

Or via Microsoft Office Lens|PDF Scan App
https://apps.apple.com/us/app/microsoft-office-lens-pdf-scan/id975925059

Android
https://www.howtogeek.com/166610/who-needs-a-scanner-scan-a-document-to-pdf-with-your-android-phone/

Or via Microsoft Office Lens|PDF Scan App
https://play.google.com/store/apps/details?id=com.microsoft.office.officelens&hl=en_US