Statistics 500 - Homework #2, Fall 2020
Due by 12pm on Friday 09/04/2020

Reading Assignment: Statistical Sleuth, Chapters 2 and 3

1. Suppose Z is a random variable with a standard normal distribution, i.e., $Z \sim N(0,1)$.

   a. We know that $\Pr(Z \le 0.66) = 0.745$. Using only this probability and properties of standard normal distribution we discuss in class, evaluate the following probabilities:

   i) $\Pr(Z \le 0.66)$
   ii) $\Pr(|Z| \le 0.66)$
   iii) $\Pr(Z = 0.66)$

   b. You may use normal distribution tables or computer tools such as the applet we use in the lab 3 to answer the following questions:
   i)   Find the 0.2 quantile of Z, i.e., fine the value of a such that $\Pr(Z \le a) = 0.2$.
   ii) Find the value of b such that $\Pr(|Z| \le b) = 0.8$.

2. Based on a WHO report, the mean hemoglobin level for a population of children is 112g/L, and the standard deviation is 9 g/L. Suppose that the hemoglobin level approximately follow a normal distribution.

   a. What percentage of children in this population have hemoglobin level below 103 g/L?
   b. What percentage of children have hemoglobin level between 100 and 124 g/L?
   c. What is the median hemoglobin level for this population of children?
   d. What is the 20$^{\text{th}}$ percentile of this population?
   e. What are endpoints of the shortest interval (a, b) such that 80% of the population of children have hemoglobin between a and b?

3. Suppose that we have independent random samples from two normal distributions:
   $Y_{11}, Y_{12}, \dots, Y_{1n} \sim N(\mu_1, \sigma_1^2)$ and $Y_{21}, Y_{22}, \dots, Y_{2n} \sim N(\mu_2, \sigma_2^2)$
   In a hypothetical situation, suppose that we know the population variances σ_1^2 and σ_2^2, but we don't know the means. We will use the sampling distribution of $\bar{Y}_1 - \bar{Y}_2$ to conduct a hypothesis test:
   $$H_0: \quad \mu_1 = \mu_2$$
   $$H_A: \quad \mu_1 \ne \mu_2$$

   a. Derive the sampling distribution of $\bar{Y}_1 - \bar{Y}_2$.
   b. Give the expression of a test statistic for this hypothesis test.
   c. Provide the formula to calculate the p-value based on your statistic in (b).

4. Medical researchers randomly assigned 30 hypertensive subjects to one of two treatment groups. The subjects assigned to one group received a treatment regimen based on a "gold standard" drug that is commonly prescribed to treat hypertension. The subjects in the second group received a treatment regime based on a new drug. One question of interest is "Do the new drug treatment produce a smaller mean systolic blood pressure level in hypertensive patients than the gold standard treatment?"

   a. Is this study an experiment or an observational study? Justify your answer.

If you think this is an experiment, answer part b below. If you think this is an observational study, answer part c.

b. Identify the following components of this experiment.
experimental units:
response:
treatments:

c. Identify the following components for this observational study.
Sampling units:
response:
conditions under study.:

5. The following table shows systolic blood pressure measurements for the thirty subjects after receiving the assigned treatments for six weeks.

New Drug				Gold Standard			
130	138	130	153	139	153	132	145
110	145	131	118	119	148	167	129
141	121	132	133	139	127	143	123
137	111	127		149	131	146	

The data are posted as **systolich.txt**. There is one line for each subject on this data file. Each line contains three values separated by spaces: patient identification number, treatment group (coded 1 for new drug and 2 for gold standard), systolic blood pressure. Still, the research interest is to answer the question: "Do the new drug treatment produce <mark>a smaller mean</mark> systolic blood pressure level in hypertensive patients than the gold standard treatment?"

a. Compute the value of a two-sample t-statistic for testing that the mean systolic blood pressure levels are the same for the two drugs against the alternative that mean systolic blood pressure lower <mark>levels differ</mark>. Assume homogeneous variances and use a pooled variance estimate and report

$t =$ _____   d.f. = _____ p-value = _____

State your conclusion in the context of this study.

b. Using the pooled variance estimate, construct a 95% confidence interval for the difference between the mean systolic blood pressures (new drug – gold standard) for the two treatments. Carefully state your interpretation of this confidence interval.

6. Perform a randomization test of the null hypothesis that mean systolic blood pressure levels are the same for the two drugs used in problem 5. Still, the research interest is to answer the question: "Do the new drug treatment produce a smaller mean systolic blood pressure level in hypertensive patients than the gold standard treatment?"

a. Obtain an approximate p-value for a randomization test by computing the two-sample t-statistic for each of 25000 data sets created from new random assignments of 15 patients to each of the two drugs. Report the p-value and state your conclusion. Is it consistent with the conclusion you reached in part (a) of problem 5? Does the histogram for the values of the t-statistics appear to be well approximated by a t-distribution?

b. Obtain an approximate p-value for a randomization test by computing the difference in the sample means for each of 25000 data sets created from new random assignments of 15 patients to each of the two drugs. Report the p-value and state your conclusion. Is it consistent with the conclusion you reached in part (a) of this problem (problem 6)?