STAT 502 - Homework 1

Due date: Thursday, October 14, 11:59PM. Submit your homework solutions to the course Canvas page. Total points: 16. **Late homework will not be accepted.**

1. **(5 Points)** Suppose we are studying the effect of diet on height of children, and we have two diets to compare: diet A (a well balanced diet with lots of broccoli) and diet B (a diet rich in potato chips and candy bars). We wish to find the diet that helps children grow (in height) fastest. We have decided to use 20 children in the experiment, and we are contemplating the following methods for matching children with diets:

   (a) **(1pt)** Let them choose.

   (b) **(1pt)** Take the first 10 for A, the second 10 for B.

   (c) **(1pt)** Alternate A, B, A, B.

   (d) **(1pt)**Toss a coin for each child in the study: heads corresponds to A, tails corresponds to B.

   (e) **(1pt)** Get 20 children; choose 10 at random for A, the rest for B.

   Describe the benefits and disadvantages of using these five methods.

   (Source: Gary W. Oehlert. *A first course in design and analysis of experiments.*, 2000, p. 29, Exercise 2.3.)

2. **(2 Points)** Generate a vector of length 20 filled with A's and B's corresponding to each of the methods in 1.(b) - 1.(e). For this exercise only, please submit your R code. (Each question is worth 0.5 points.)

3. **(5 Points)** Plant growth: Results from an experiment to compare yields (as measured by dried weight of plants) obtained under two different treatment conditions. The data are available on the course Canvas webpage. Read in the data with `data <- readRDS("Plants.RDS")`.

   (a) **(1pt)** Exploratory data analysis: Calculate the sample means, medians and standard deviations for the two treatment groups. Plot the CDFs for the two groups.

   (b) **(3pt)** Hypothesis tests:

      i. **(1pt)** The H_0 hypothesis is that there is no difference between the two treatments in terms of plant growth. For the given data, compute the the KS-statistic (see lecture slides) and the variance ration statistic: $\max\{s_1^2/s_2^2, s_2^2/s_1^2\}$, where s_j^2 is the variance of the j-th group.

      ii. **(1pt)** Plot histograms of the randomization distributions for each statistic, and indicate the observed value of the test statistics on the plots.
      Hint: Use Monte-Carlo sampling with replacement.

      iii. **(1pt)** Obtain the p-values for the randomization test based on each statistic. What are your conclusions about the differences between the two treatments?

   (c) **(1pt)** Multiple testing: If both tests in 3b are performed at the level-0.05, what is the probability of at least one of them rejecting H_0, if H_0 is true? Explain.

4. **(4 Points)** As part of a larger experiment, Dale (1992) looked at six samples of a wetland soil undergoing a simulated snowmelt. Three were randomly selected for treatment with a neutral pH snowmelt; the other three got a reduced pH snowmelt. The observed response was the number of Copepoda removed from each microcosm during the first 14 days of snowmelt.

Reduced pH			Neutral pH		
256	159	149	54	123	248

   (a) **(1pt)** Perform an exact randomization test of the hypothesis of interest using as a test statistic the absolute value of the difference of the group means. In notation from class, use:

$$g(\mathbf{y_A}, \mathbf{y_B}) = |\bar{\mathbf{y}}_\mathbf{B} - \bar{\mathbf{y}}_\mathbf{A}|.$$

   (b) **(1pt)** Redo the (exact) randomization test but now use the t-statistic

$$\begin{aligned} g_t(\mathbf{y_A}, \mathbf{y_B}) &= \frac{|\bar{\mathbf{y}}_\mathbf{B} - \bar{\mathbf{y}}_\mathbf{A}|}{s_p\sqrt{1/n_A + 1/n_B}}, \quad \text{where} \\ s_p^2 &= \frac{n_A - 1}{(n_A - 1) + (n_B - 1)} s_A^2 + \frac{n_B - 1}{(n_A - 1) + (n_B - 1)} s_B^2. \end{aligned}$$

   Compare your results.

   (c) **(2pt)** Give a mathematical explanation for what you observe when comparing the results of the two tests. Do it for general n_A and n_B.
   Hint: Start by writing s_p^2 in the form:

$$s_p^2 = \frac{1}{(n_A - 1) + (n_B - 1)} \left(h(\mathbf{y_A}, \mathbf{y_B}) - \frac{n_A n_B}{n_A + n_B} (\bar{\mathbf{y}}_\mathbf{B} - \bar{\mathbf{y}}_\mathbf{A})^2 \right).$$

   How does h change under randomization?

(Source: Gary W. Oehlert. *A first course in design and analysis of experiments.* 2000, p. 29, Exercise 2.4.)