<u>BME 423 Fall 2021: Homework 2</u>

*Due on <u>Wednesday, October 13, 2021 at 5pm</u>. Solutions will be posted later that day to allow you to use them as study material for the exam, so this is a **<u>hard</u>** deadline. You will submit your work as a PDF uploaded to Blackboard.*

Please read the homework guidelines before starting your assignment. Statistical tables necessary for completing this assignment are posted on Blackboard.

All statistical tests are to be performed "by hand".

1. Hypothesis Testing (30 points)

(a) A group of medical researchers would like to see if certain interventions can improve cognitive function in subjects who have shown a decline in their cognitive abilities. Subjects over the age of 65 who showed some level of cognitive decline were divided into two groups: 22 subjects were assigned to the control group (no intervention) and 27 subjects were assigned to the treatment group (prescribed specific diet and physical exercise regimens, as well as "brain training" mental exercises). After six months, each subject took a scored cognitive assessment test. The control group had $\bar{x}_1 = 49$ and $s_1 = 8.4$; the treatment group had $\bar{x}_2 = 57$ and $s_2 = 9.5$. Can the researchers conclude that there is a difference in cognitive scores between the two treatment populations? Use the appropriate statistical test that is formulated for only two groups. You can assume that these interval data are Normally distributed.

(b) Based on the results of an annual health wellness screening program, the average low-density lipoprotein concentration (LDL) of USC faculty and staff is 115 mg/dL. (High levels of LDL are associated with an increased risk of developing coronary heart disease.) We want to determine whether the average LDL for USC undergraduate students is different than 115 mg/dL. To make this assessment, we randomly selected 16 USC undergraduates who have consented to participate in the study and measured their LDL. The students' average LDL was found to be 103 mg/dL with a standard deviation of 16 mg/dL. Based on the data from this sample, is there a statistically significant difference in LDL between USC undergraduate students and the average value obtained from USC faculty/staff?

(c) Most anticancer therapies can cause myelosuppression leading to a reduced concentration of neutrophils in the blood, which increases the risk that patients will develop life threatening infections. For this reason, it is not uncommon to also administer filgrastim, a granulocyte colony stimulating factor, to increase neutrophil production in the bone marrow, together with the anticancer therapy. A study was designed to examine the effect of three different doses of filgrastim. Patients were randomized into three different dose groups ($n = 15$ for each group), and after one month of therapy each patient's blood neutrophil concentration was measured. The results are summarized in the table below.

Dose Group	Mean neutrophil count (10^9 cells/L)	Standard Deviation (10^9 cells/L)
Group 1: Low Dose	3.7	1.1
Group 2: Medium Dose	4.6	1.4
Group 3: High Dose	5.2	2.1

An ANOVA test demonstrated that the F test-statistic is significant at α = 0.05 ($F = 3.3839$, $P < 0.05$), with a value of $s_{with}^2 = 2.5267$. Using the appropriate pairwise comparison method and the **Bonferroni** correction, determine which dose groups result in differences in neutrophil count.

You will need to calculate $t_{0.0167}(42)$, which you should do using the R command: `qt((0.05/2)/3, 42, lower.tail=FALSE)`.

2. Methods for Proportions and Nominal Variables (30 points)

(a) Dioxin is known to be toxic in humans (can cause cancers, damage the immune system, cause developmental problems) when present in the environment at concentrations as low as 15 parts per trillion. In one study on the genetic effects of dioxin using health department records from several US States bordering the Great Lakes, investigators examined the consequences of dioxin exposure in one of the two parents, on the proportion of their offspring that are male or female. The results are as follows:

- The proportion of female babies born to parents whose father was exposed to dioxin but not the mother was 0.565. Total number of babies born in this group was 186.
- The proportion of female babies born to parents whose mother was exposed to dioxin but not the father was 0.455. Total number of babies born in this group was 220.
- The total number of babies born was 406, of which 205 were female.

Use a z Test to determine if there is a difference in the proportion of female babies born when either only the father or only the mother was exposed to dioxin.

(b) The following data were taken from a study of 1000 expectant mothers. The expectant mothers were given questionnaires about their prenatal care, including whether or not they took folate supplements throughout their pregnancy. After birth, the infants were evaluated for neural tube defects (NTDs; birth defects of the brain, spine, or spinal cord), the following results:

Group	**No NTD**	**Yes NTD**
Took folate	471	4
No folate	498	27

i. Use a χ^2 test to determine if there is a difference in development of NTDs between the folate and no-folate populations.

ii. If there is a difference, quantify the degree of association between the mother's <u>not supplementing with folate</u> and the <u>development of NTD(s)</u> in the infant. Provide your result in a single complete sentence that would convey your conclusion in a way that could be understood by a non-expert in statistics or medicine (e.g., a typical reader of a local newspaper).

3. Methods for Proportions and Nominal Variables (30 points)
Patients with myocardial infarctions were assigned to one of four thrombolytic (clot-dissolving) treatments and then followed to determine 30-day mortality. The data are summarized in the table below:

	Survived	Died
Control Group: Standard Care	887	117
Group 1: t-PA and IV Heparin	967	79
Group 2: Streptokinase and IV Heparin	901	85
Group 3: t-PA, streptokinase, and IV Heparin	912	59

(a) By hand, perform the appropriate statistical test to determine if there is a difference in the survival among the four populations.

(b) The following results were obtained using the `chi-square` function in R to perform all six pair-wise χ^2 Tests.

```
------------------------------------------
> # Control vs. Group 1
> chisq.test(MI_Ctl_G1)

   Pearson's Chi-squared test with Yates' continuity correction

data:  MI_Ctl_G1
X-squared = 9.4944

------------------------------------------
> # Control vs. Group 2
> chisq.test(MI_Ctl_G2)

   Pearson's Chi-squared test with Yates' continuity correction

data:  MI_Ctl_G2
X-squared = 4.6895

------------------------------------------
> # Control vs. Group 3
> chisq.test(MI_Ctl_G3)

   Pearson's Chi-squared test with Yates' continuity correction

data:  MI_Ctl_G3
X-squared = 18.234

------------------------------------------
> # Group 1 vs. Group 2
> chisq.test(MI_G1_G2)

   Pearson's Chi-squared test with Yates' continuity correction

data:  MI_G1_G2
X-squared = 0.64312

------------------------------------------
> # Group 1 vs. Group 3
> chisq.test(MI_G1_G3)

   Pearson's Chi-squared test with Yates' continuity correction

data:  MI_G1_G3
X-squared = 1.4982

------------------------------------------
> # Group 2 vs. Group 3
> chisq.test(MI_G2_G3)

   Pearson's Chi-squared test with Yates' continuity correction

data:  MI_G2_G3
X-squared = 4.2807
```

Perform the appropriate multiple pairwise comparisons using the Holm correction to determine which treatments, if any, differ from the control (the Standard Care group). Present your final/overall conclusion on which treatment groups are different from the control (and which treatment groups are not different) using an overall level of significance of 0.05. Use the needed information from the R results that have been provided to do your calculations and assessments. (You may assume that all implementations in R have been done correctly.)