**BST 140.652 Midterm Exam**

Notes:

- You may use your one 8.5 by 11 formula sheet.
- Please use only the basic mathematical functions on your calculator.
- Show your work on all questions. Simple "yes" or "no" answers will be graded as if blank.
- Please be neat and write legibly. Use the back of the pages if necessary.
- There are 5 questions one per page
- Good luck!

____________________

signature and **printed name**

1. In a review of the author Jane Austen's work, scholars found the following relative frequencies of the words "an", "that" and "this"

| an | that | this |
|---|---|---|
| .53 | .35 | .12 |

An new story claimed to be Austen's was discovered with the word "an" 140 times, the word "that" 100 times and the word "this" 50 times. Are these counts consistent with Austen's traditional frequencies? Perform an appropriate test and interpret.

**Comments** Use a chi-squared goodness of fit test. It has two degrees of freedom.

2. Two drugs, $A$ and $B$ , are being investigated in a randomized trial with the data are given below. Investigators would like to know if the Drug A and Drug B are equivalent in the terms of side effects.

| | | None | Nausea | Nausea and Vomiting | $N$ |
|---|---|---|---|---|---|
| Drug A | Observed Counts | 80 | 15 | 5 | 100 |
| | Expected Counts | | | | |
| Drug B | Observed Counts | 60 | 20 | 10 | 90 |
| | Expected Counts | | | | |

State relevant null and alternative hypotheses and fill in the expected cell counts. How many degrees of freedom would the Chi-squared statistic have? (Do not perform the test.)

**Comments** The hypotheses of interest is

$$H_0 : p_{A1} = p_{B1}, p_{A2} = p_{B2}, p_{A3} = p_{B3}$$

versus

$$H_a : \text{At least one pair is not equal}$$

where $p_{ij}$ is the probability of side effect $j = 1, 2, 3$ (non, nausea and nause and vomiting) for drug $i$ (A, B). Note writing $H_0$ : as "the probability of side effects for Drug A is equal to that of Drug B" is not correct (as it is imprecise). The degrees of freedom of the test would be $3$. Use the notes from the class to calculate the expected cell counts. Note that there is no reason to round the expected cell counts to be integers.

3. A colleague is going to study change in forced expiratory volume (FEV) over two years. In this study she is going to take individual subjects and subtract their FEV at follow-up from their FEV at baseline. She would like to test whether or not there is a decline in FEV. She needs to know how many subjects are required to have an appropriately power the study.

   *Give a list of the quantities that you will need and assumptions that you will make to perform the calculation.* Do not use any symbols; state the quantities in English.

   **Comments**

   - The desired probability of detecting a decline in FEV should one exist (power).
   - The desired probability of incorrectly declaring a decline in FEV when one does not exist (alpha)
   - A scientifically meaningful change in FEV to be detected and an estimate of the standard deviation for the change in FEV (the ratio of these two numbers would suffice)
   - Likely the power calculation would assume independence and identically distributed data
   - In addition, some degree of normality or applicability of the CLT would likely be assumed.

4. You are given three sample mean across a stratifying variable. They are (in order) $5.0$, $7.0$, $8.0$. You are also given their standard errors: $2.0$, $1.0$ and $3.0$ respectively. Give a sensible estimate of the common mean across strata. **Extra credit 5 points** What would be an appropriate standard error for the resulting common mean?

   **Comments** Let $X_i$ be each sample mean $(i = 1, 2, 3)$. Let $W_i = 1/SE_i^2$; therefore, $W_1 = 1/2.0^2 = .25$. Then the estimate is

$$\frac{W_1X_1 + W_2X_2 + W_3X_3}{W_1 + W_2 + W3}$$

   (plug in the numbers yourself). The rational behind this procedure is that the three sample means can be considered independent normally distributed with sample variances $SE_i^2$. We know from the lecture that the most appropriate way to combine such data uses inverse variance weights.

5. In a study of aquaporins, 12 frog eggs were randomized, 6 to receive a protein treatment and 6 controls. If the treatment of the protein is effective, the frog eggs would implode. The resulting data was

| | Imploded | Did not | Total |
|---|---|---|---|
| Treated | 5 | 1 | 6 |
| Control | 2 | 4 | 6 |
| Totals | 7 | 5 | 12 |

State the appropriate hypotheses and set up the calculation of a P-value. *Do not solve for the final number, just leave your answer in an equation format.*

**comment** Most everyone got this. Simply use Fisher's exact test. For the one sided test of $H_a : P_T > P_C$ add the probability of the two tables with 5 and 6 in the upper left hand cell.