**ADM 2304 – ASSIGNMENT 2**

***Due Date: Sunday, March 5, 2017 at midnight.***

***Instructions:***

1. *You may use Minitab or other software for any calculations. However,* ***you must show your manual calculations unless indicated otherwise****. You may paste your output onto your assignment to show your use of Minitab; however, this output does not replace any of the steps outlined below. This means that answers that are exclusively Minitab output may receive only one mark.*
2. *If you are performing a hypothesis test, make sure you state the hypotheses, the level of significance, the rejection region, the test statistic (and/or p-value, if requested), your decision (whether to reject or not to reject the null hypothesis), and a conclusion in managerial terms that answers the question posed. These steps must be completed in addition to any Minitab output.*
3. *The data in the Minitab data files can be found in the file* ***Assign2Data.xlsx****.*
4. *Remember to include your integrity statement. Your assignment should be uploaded to* ***Blackboard*** *in pdf format.*

**Question 1 – Real Estate**

The dataset *RealState* includes two random samples of 100 listings each that have been extracted from a much larger original dataset. One sample has houses with fireplaces and the other has houses without fireplaces. The spreadsheet also has: Price ($), Living Area (sq. ft.), and Age (years) for each listing.

a) Describe, in one short phrase each, the populations being examined.

b) Carry out a hypothesis test to see whether there is a real difference in the mean price of these two populations. Assume the two population variances are unequal.

Note: The level of significance is not specified. Draw your conclusions by referring to the strength of evidence indicated by the p-value. Remember the three alpha values most commonly used: 0.05 = statistically significant, 0.01 = highly statistically significant, and 0.001 = very highly statistically significant.

c) Find the corresponding 95% confidence interval for the difference in mean house prices.

d) Describe in one sentence how the hypothesis test in part b) and the confidence interval in part c) reflect the relationship between confidence intervals and hypothesis tests as introduced in class.

**Question 2 – Consumer Spending Patterns**

You are on the financial planning team for monitoring a high-spending segment of a credit card. You know that customers tend to spend more during December before the holidays, but you're not sure about the pattern of spending in the months after the holidays. The dataset *Spending* has the monthly credit card spending of 200 customers during the months December, January, February and March, and the average across January to March.

a) Carry out a hypothesis test to compare December with the average spending over the three following months (variable JantoMarAvg) to confirm the belief that customers spend more on average in December. Use a 5% level of significance.

b) Construct a 95% one-sided confidence interval to supplement the hypothesis test you carried out in part a). Do the confidence interval confirm your previous conclusion?

**Question 3 – Top 100 Novels**

Years ago a meme went around Facebook and other parts of the internet about the BBC's "Top 100 Books", with the statement that most people have read only six of the books listed. Your friend at Carleton suggests that Carleton students have read more of those books than either uOttawa students or UofT students. Determined to prove that uOttawa students are well read, you collect some data. Using appropriate sampling techniques, you poll uOttawa, Carleton, and UofT students to see how many of these books they've read. Here are the results of your poll:

| | |
|---|---|
| uOttawa students (148 total): | 84 have read 0 to 9 books on the list<br>43 have read 10 to 19<br>21 have read 20 or more |
| Carleton students (137 total): | 91 have read 0 to 9 books on the list<br>24 have read 10 to 19<br>22 have read 20 or more |
| UofT students (119 total): | 70 have read 0 to 9 books on the list<br>35 have read 10 to 19<br>14 have read 20 or more |

a) Put the data into a two-way table with the number of books categories in the rows and the university categories in the columns. The data in this table should be observed counts. Create another table showing the corresponding expected counts. Show no more than three decimal places in the second table and make sure your two tables show the totals for the rows, columns, and overall.

b) Perform a hypothesis test to check if the distributions for number of books read are the same across the three universities at the 0.01 significance level. That is, test the independence of the two categorical variables, number of books read and university attended.

c) Is the chi-squared approach appropriate here? Why or why not? Hint: Look at the "expected" table.

If you are interested, the BBC list can be found at this site:
http://www.listchallenges.com/kaunismina-bbc-6-books-challenge

**Question 4 - Consumer Perceptions of Food Safety**

Vendors of prepared food are very sensitive to the public's perception of the safety of the food they sell. Food sold at outdoor fairs and festivals may be less safe than food sold in restaurants because it is prepared in temporary locations and often by volunteer help. What do people who attend fairs think about the safety of the food served? One study asked the following question to people at a number of fairs:

*How often do you think people become sick because of food they consume prepared at outdoor fairs and festivals?*

The possible responses were:

1 = very rarely
2 = once in a while
3 = often
4 = more often than not
5 = always

In all, 303 people answered this question.

The dataset *FoodSafety* includes the following variables: subject, sfair, srest, and gender. The variable "sfair" contains the responses concerning safety of food served at outdoor fairs and festivals. The variable "srest" contains responses to the same question asked about food served in restaurants.

a) Is there good evidence that men and women differ in their perceptions about food safety at fairs? Is this true for restaurants? Select an appropriate nonparametric test and use Minitab to obtain your results. Explain your conclusions.

b) Now you wonder if people are more concerned about the safety of food served at fairs than they are about the safety of food served at restaurants. Explain carefully why we cannot answer this question by applying the Wilcoxon Rank-Sum Test (a.k.a. the Mann-Whitney Test) to the variables "sfair" and "srest".

c) We suspect that restaurant food will appear safer than food served outdoors at a fair. Do the data give good evidence for this suspicion? Use Minitab to perform an appropriate nonparametric test and report your findings. Provide a p-value and your conclusion.