Your Name: _____

Section: _____

36-201 INTRODUCTION TO STATISTICAL REASONING
Computer Lab Exercise – Lab #7
(there was no Lab #6)
Treatment for Depression: A Randomized Controlled Clinical Trial

Objectives:

1. To review research methods and the principles of experimental design that are typically used in an experiment.

2. To use exploratory data analysis techniques to analyze the results of an actual randomized controlled clinical trial for the prevention of the recurrence of depression. Specifically, to develop methods for the analysis of contingency tables.

3. To apply a general data-analysis approach in tackling this problem. Specifically, to practice the three main steps of that approach: (1) identifying the question, (2) analyzing the data, and (3) drawing conclusions from the data.

4. To develop further skills in using *Minitab* for data analysis.

Getting Started

For today's lab you will need to copy a file using the file server. **The data file is not a Minitab formatted file but rather a text file called** *depression.dat.*

Get the file now and start Minitab. Use Open Worksheet to open the file. Since this is a text file, you must select "Text" from "List Files of Type".

General Data-Analysis Approach

Even though different statistical problems may look quite distinct on the surface, they usually can be tackled using a single data-analysis approach. This approach has three main steps:

1. **Identify the Question**
2. **Analyze the Data**
3. **Draw Conclusions from the Data**

These main steps are all important parts of a thorough data analysis, and each one may require the completion of a series of sub-steps. For example, the first main step **Identify the Question** involves the following sub-steps: describe the problem, state the question in your own words, check the data format, and reflect on the study design. Keep in mind that each sub-step may, in turn, require several "sub-sub-steps" to be completed.

For a given data-analysis problem, the way a step is performed may be different, but the general sequence of steps is the same. This sequence can be viewed as a general template for how to solve data analysis problems. The outline below shows the main steps and their sub-steps. You should always consider these when solving a data-analysis problem.

1. **Identify the Question**
    A. **Describe the problem**
    B. **State the question(s)**
    C. **Check the data format**
    D. **Reflect on the study design**
2. **Analyze the Data**
    A. **Identify the relevant variables**
    B. **Determine the appropriate analysis**
    C. **Conduct the analysis**
    D. **Interpret the results**
    E. **Consider whether additional analyses are necessary**
3. **Draw Conclusions from the Data**
    A. **Re-state the question(s)**
    B. **Answer the question(s) based on analyses.**
    C. **Evaluate the strengths and weaknesses**

This week's lab is structured according to the above outline template. In fact, the above steps and sub-steps appear in bold as headings in this lab handout to structure the material. You will notice, however, that the details under each heading are specific to the data set you are analyzing. This organization is designed to help you see what is common across different statistics problems as well as how the details can vary.

1. Identify the Question

The idea behind the first main step **Identify the Question** *is that you should always think about the question(s) being asked before you dive into data analyses. This way, you can make sure that you understand how the data were collected and why they were collected. This can make later steps easier.*

1A. Describe the problem

Clinical depression is a recurrent illness requiring treatment and often hospitalization. Nearly 50% of people who have an episode of major depression will have a recurrence within 2-3 years. To be able to prevent the recurrence of depression in people who are at risk for the disease would not only go a long way to alleviate the pain and suffering of the individual patient but would also save society many thousands of dollars per patient in medical expenses and lost wages due to an inability to work.

During the 1980's the Federal government, through the National Institutes of Health (NIH), sponsored a multi-centered randomized controlled clinical trial to evaluate two drugs to prevent the recurrence of depression in patients who have had at least one previous episode of the illness (Prien et al., *Archives of General Psychiatry*, 1984). The following gives the basic design of the study:

The study was **multi-centered**. There were 5 medical clinics across the country that participated in this trial. Using many clinics enables the investigators to enroll many more patients into the study and allows for a diversity of patients to participate.

There were 3 **treatment** groups. Patients received either *Imipramine* (Imip), *Lithium* (Li), or a *Placebo* (Pl). Imip and Li are active drugs.

Patients were **randomized** to one of the 3 treatment groups, using a random device (like rolling a 3 sided die).

Patients were followed from 2-4 years to see whether or not they had a recurrence of depression. If they did not have a recurrence within this time frame, then their treatment was considered a *Success.* If they did have a recurrence, it was considered a *Failure*.

The study was **double-blinded**.

A number of additional background variables were measured for each patient.

1B. State the question(s)

Today you will analyze the results of this trial to answer the following questions:

(I) Which of the drugs (if either) was more successful in preventing the recurrence of depression relative to the placebo? and
(II) Which of the drugs (if either) delayed the recurrence of depression longer relative to the placebo?

1C. Check the data format

The variables in the data set are:

Variable Name	Description
HOSPT	Which hospital the patient was from: Labeled 1, 2, 3, 5 or 6.
TREAT	0=*Lithium*; 1=*Imipramine*; 2=*Placebo.*
OUTCOME	Response: 0=*Success* 1=*Failure* (recurrence of depression).
TIME	How long the patient was followed in the study, measured in weeks. For patients who had a recurrence of depression, TIME equals the number of weeks from their inclusion in the study until depression recurred. For patients who did not have a recurrence of depression, TIME equals the total number of weeks of the study.
GAS	*Global Assessment Scale* - a measure of social functioning from 0 to 100, taken at the beginning of the study. Small values are bad, large values are good.
ACUTET	How long the patient was depressed before the start of the current study, measured in days.
AGE	Age in years.
GENDER	Gender: 1=*Female* 2=*Male.*

Question #1: From the data worksheet fill in the blanks below for the first patient:

Which hospital was the patient from? _____
What treatment did the patient receive? _____
What was the patient's outcome? _____
How long was the patient followed? _____
What was the patient's GAS score? _____
How old was the patient? _____
Was the patient male or female? _____

1D. Reflect on the study design

Question #2: Is this an experiment or an observational study? Explain why.

In this study, patients were randomly assigned to the three treatment groups, so the final number of patients receiving each treatment is random. This should produce three groups of approximately equal size, with approximately equal distributions of possible lurking variables.

Question #3: How many and what percent of the patients received each treatment? How many total patients are in the study? You can use **Cross Tabulation** (which can be found in the **Stats→Tables→Cross Tabulation** menu) to get a frequency breakdown of TREAT.

♣ **Question #4: Was the randomization effective in assigning an approximately equal number of patients to each treatment group? Explain.**

Randomization is also supposed to "balance" variables that we didn't control by blocking. For example, the randomization should have assigned approximately the same distribution of ACUTET times to each of the three groups.

From the **Graph** menu, choose **Boxplot**. Under column Y, "Graph variables: Y vs. X," in the first cell in row number 1, type ACUTET. Under column X in row number 1, type TREAT. *Click* **OK**.

Question #5: Was the randomization successful in assigning approximately the same distribution of ACUTET scores to each treatment group? Explain.

Question #6: List other variables, from the table at the top of p. 4, that we hope are distributed equally among the treatment groups by the randomization.

2. Analyze the Data

After you are sure you understand the problem, the question(s) being asked, and the study design, it is time to analyze the data. In the second main step **Analyze the Data** *you will address each of the questions identified in main step 1 by doing the following: identify the relevant variables, determine the appropriate analysis, conduct the analysis, interpret the results, and consider whether additional analyses are necessary. Since this lab involves two questions (see part 1B), you will complete this sequence twice.*

I. WHICH OF THE DRUGS (IF EITHER) WAS MORE SUCCESSFUL IN PREVENTING THE RECURRENCE OF DEPRESSION RELATIVE TO THE PLACEBO?

2A. Identify the relevant variables

To assess the relative effectiveness of the treatments, we will examine the relationship between TREAT and OUTCOME.

2B. Determine the appropriate analysis

Question #7:

- Is OUTCOME an explanatory variable or a response? Is it qualitative or quantitative?
- Is TREAT an explanatory variable or a response? Is it qualitative or quantitative?

Question #8: Given your answers, what type of plot or table is most appropriate (choose: scatter plot, side-by-side boxplots, contingency table, etc.)?

2C. Conduct the analysis

From the **Stat** menu, go to the **Tables** sub-menu, and choose **Cross Tabulation** (this makes a contingency table). Type *Treat Outcome* in the box under "Classification variables". Make sure to put a space between the variable names and that the "Counts" and "Row Percents" check boxes (and no other checkboxes) are checked.

Question #9: Copy the table of observed values; label the rows and columns (not with numerical labels but with the appropriate words); include the row percents. *Be sure you know which variable is on the rows and which is on the columns and what each of the codes for the variables stand for.*

2D. Interpret the results

Question #10: Based on your analysis of the row %, is the drug "Imip" more effective, less effective or no different than "Li" in preventing the recurrence of depression? Why?

♣ **Question #11: Based on your analysis of the row %'s, is the drug "Li" more effective, less effective or no different than "Placebo" in preventing the recurrence of depression? Why?**

2E. Consider whether additional analyses are necessary

An additional analysis we can do is to look at **Standardized Residuals**. From the **Edit** menu, choose **Edit Last Command Dialog**. *Uncheck* the checkboxes next to **Counts** and **Row percents**, and *check* the checkbox next to **Chisquare analysis**. Then *check* the radio button *"Above and std. residual"* which adds to the contingency table the *Expected values* and *standardized residuals*, as well as the χ^2 value (sum of squares of standardized residuals). Click **OK**. (To figure out what each of the values in the Session window are, look at the "Cell Contents" table underneath the contingency table. *Click* the Zoom box if you cannot see the entire table in the Session window.)

Recall that each standardized residual in the table says how far "out of line" from the hypothesis of independence the corresponding observed count is:

- If the standardized residual is positive, the observed count was bigger than the expected count;
- If the standardized residual is negative, the observed count was smaller than the expected count;
- Standardized residuals between about -1.5 and 1.5 indicate cells that agree with independence;
- Standardized residuals between about -2.0 and -1.5 or between 1.5 and 2.0 indicate cells that give *mild* evidence against independence;
- Standardized residuals less than -2.0 or greater than 2.0 indicate cells that give *strong* evidence against independence.

Question #12:

i) Which cell of the table has the largest standardized residual? Interpret what this means in the context of the problem.

ii) Which cell of the table has the smallest standardized residual? Interpret what this means in the context of the problem.

II. WHICH OF THE DRUGS (IF EITHER) DELAYED THE RECURRENCE OF DEPRESSION LONGER, RELATIVE TO THE PLACEBO

2A. Identify the relevant variables

The amount of time in the study, TIME, is another way of evaluating how effective a treatment was. This is because once a patient had a recurrence of depression (i.e., a failure) they were no longer followed in the study: the longer a patient was in the study, the more effective the treatment must have been. Therefore, TIME and TREAT are the relevant variables for addressing the question os which drug delayed the recurrence of depression longer.

2B. Determine the appropriate analysis

Question #13:

- Is TIME an explanatory variable or a response? Is it qualitative or quantitative?
- Is TREAT an explanatory variable or a response? Is it qualitative or quantitative?

Question #14: Given your answers, what type of plot or table is most appropriate (choose: scatter plot, side-by-side boxplots, contingency table, etc.)?

2C. Conduct the analysis

From the **Graph** menu, choose **Boxplot**. Choose TIME as the Y-variable and TREAT as the X-variable.

2D. Interpret the results

♣ **Question #15: What are the approximate median** TIMES **for each treatment group? Based on this analysis which treatment or treatments appear to be better in preventing the recurrence of depression?**

2E. Consider whether additional analyses are necessary

Question #16: Can you think of any other ways to compare the lengths of time patients on each of the three treatments stayed in the study, or are you happy with the boxplot analysis?

3. Draw Conclusions from the Data

3A. Re-state the Question(s)

(I) Which of the drugs (if either) was more successful in preventing the recurrence of depression relative to the placebo? and (II) Which of the drugs (if either) delayed the recurrence of depression longer relative to the placebo?

3B. Answer the Question(s) Based on Analyses

Question #17: Given all the analyses above, how would you answer each question:

(I) Which of the drugs (if either) was more successful in preventing the recurrence of depression relative to the placebo?

(II) Which of the drugs (if either) delayed the recurrence of depression longer relative to the placebo?

♣ **Question #18: Which treatment would you recommend to prevent the recurrence of depression, based on this study? Explain why.**

3C. Evaluate Strengths and Weaknesses

Question #19: Are there other alternative hypotheses or explanations for these results that you are concerned about that might make you qualify your answer? If so, name one or two.

Remember to **delete** *files and folders that you might have created.*