MIDTERM PROJECT: LIFESTYLE DATA

You should now have three sets of numbers describing three different aspects of your lifestyle. Use the steps below to guide your data analysis and reporting. Be sure to label each step in your write up. Make sure that all graphs you produce are done on graph paper and all writing is neat and legible. All work must be shown to get full credit. NOTE: if you have very bad handwriting, it is in your best interest to type the report.

Create a new file in JMP IN and enter your lifestyle data. You will have three columns of data, one for each variable. Use the ANALYZE commands in JMP IN to analyze your data. While JMP IN can provide most of the statistics and graphs that are required for this project, you will have to calculate some statistics such as the range and interquartile range using information provided by JMP IN. You will also have to expand and contract stem-and-leaf plots by hand.

In addition to completing each of the 11 items listed below, your grade will depend on how well you address each of the following:

• Correct use of statistical language
• Correct calculations (for any that you perform)
• Correct and complete graphs (for any that you create)
• Accurate summaries of information and correct conclusions
• Clarity and legibility of your writing

Here are the items you must address in your paper. Please number the items 1 through 11.

1. For EACH of the three variables you measured, describe:
   • why you selected the variable (why were you interested in it?),
   • how you measured the variable (inches, miles, ounces, cups, hours, minutes, etc.),
   • how you collected the data for the variable (how did you make accurate measurements or keep accurate records)
   • what you expected to find out about the variable, including whether or not you expect two or more of the variables to be related and why.

2. a. Use JMP IN to produce descriptive statistics, a box plot, a histogram, and a stem-and-leaf plot for the first variable (Use ANALYZE -> DISTRIBUTION from JMP IN to obtain statistics and graphs).
   b. Report the mean and median for the distribution (report them from JMP IN).
   c. Report the range, interquartile range, and standard deviation. (You will have to calculate the range and interquartile range).
   d. Use the graphs AND statistics to describe and interpret the distribution for the first variable. Determine the best measures of center and variability to use for this distribution.

3. Repeat steps 2a through 2d for your second variable.

4. Repeat steps 2a through 2d for your third variable.

5. Choose one of your variables and either contract or expand the stem and leaf plot in order to get a better picture of the distribution. YOU HAVE TO DO THIS BY HAND. State whether or not you think the new stem and leaf plot is a better representation of the distribution, and WHY.

6. REPEAT step 5 above for another one of your variables.
7. For one of your variables, compare the values for weekdays (Monday through Friday) to the values for the weekend (Saturday and Sunday). Create a back-to-back stem-and-leaf plot to make the comparison (Weekdays on the left side and Weekend on the right side). YOU HAVE TO DO THIS BY HAND. Describe any similarities and differences between the distributions with respect to their centers, spread, and shapes.

8. a. Choose TWO of the variables that you think might be related. Use JMP IN to make a scatterplot of these variables (Use ANALYZE -> FIT Y BY X).
   b. Use JMP IN to report the correlation (click the red triangle and select Density Ellipse; set it to .95) and the equation for the line of regression (click the red triangle and select Fit Line). Click the blue diamond next to Correlation to display the correlation value. Interpret what the correlation tells you about the relationship between the two variables. Use the scatterplot, the line of regression, and the correlation to describe the relationship between the variables that you see in the scatterplot (comment on both the direction and the strength of the relationship).
   c. Use the RESIDUAL PLOT to determine if the regression line provides a good description of the relationship between the two variables (click the red triangle next to Line of Fit and select Plot Residuals). See the discussion on pages 552-556 of the textbook.
   d. Use JMP IN to determine if there is a bivariate outlier. Use the Density Ellipse, Residual Plot, and a change in R-Square of more than .05 to identify a potential bivariate outlier. Make a print out of BOTH scatterplots and statistics (before and after the point is excluded). You must use the density ellipse, residual plot, and change in R-Square to argue for why a point is or is not a bivariate outlier.

9. Go to the ROW menu and choose CLEAR ROW STATES. Then, REPEAT steps 8a through 8f for a SECOND PAIR of variables.

10. Select one of the variables and determine if it has a normal distribution. Use the procedures presented in the Variability Classwork activity (you must test for all 5 characteristics to receive full credit. You must do calculations for any statistics that are not provided by JMP IN). Based on the information, DISCUSS whether or not the variable has a normal distribution.

11. Write a few paragraphs to summarize your lifestyle data and state what you learned by analyzing the data. Your discussion should be related to the expectations you described in your answer to the first question.

12. Attach a copy of your Lifestyle Data Collection Sheet to the report.
LIFESTYLE DATA COLLECTION SHEET

Over the next 5 weeks you will be asked to record some data about your personal life. Pick three QUANTITATIVE variables that you can measure on a daily basis, and fill in your data on the chart below. Try to choose at least two variables that you think are related to (or depend on) each other. At the end of five weeks you will analyze the data and write up the results as a student project. Here are some general ideas of things you can measure about your daily life:

- **leisure** (e.g., minutes you read for pleasure, watch TV, play video games, talk on the phone)
- **study habits** (e.g., minutes you study each day, number of study breaks for each hour of study)
- **sleep habits** (e.g., how many times you get up at night, hours of sleep you get each night)
- **exercise habits** (e.g., minutes of exercise you get each day, miles you walk or jog each day)
- **eating habits** (e.g., how much money you spend on fast food each day)

You may choose variables from other lifestyle categories as well.

**IMPORTANT**: For each variable, mark the days that represent Saturday (Sa) and Sunday (Su). For example, if Day 1 is a Tuesday, then days 5, 12, 19, 26, and 33 are Saturdays, whereas days 6, 13, 20, 27 and 34 are Sundays.

Variable 1:

<table>
<thead>
<tr>
<th>Day</th>
<th>Day</th>
<th>Day</th>
<th>Day</th>
<th>Day</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>15</td>
<td>22</td>
<td>29</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>16</td>
<td>23</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>17</td>
<td>24</td>
<td>31</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>18</td>
<td>25</td>
<td>32</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>19</td>
<td>26</td>
<td>33</td>
</tr>
<tr>
<td>6</td>
<td>13</td>
<td>20</td>
<td>27</td>
<td>34</td>
</tr>
<tr>
<td>7</td>
<td>14</td>
<td>21</td>
<td>28</td>
<td>35</td>
</tr>
</tbody>
</table>

Variable 2:

<table>
<thead>
<tr>
<th>Day</th>
<th>Day</th>
<th>Day</th>
<th>Day</th>
<th>Day</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>15</td>
<td>22</td>
<td>29</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>16</td>
<td>23</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>17</td>
<td>24</td>
<td>31</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>18</td>
<td>25</td>
<td>32</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>19</td>
<td>26</td>
<td>33</td>
</tr>
<tr>
<td>6</td>
<td>13</td>
<td>20</td>
<td>27</td>
<td>34</td>
</tr>
<tr>
<td>7</td>
<td>14</td>
<td>21</td>
<td>28</td>
<td>35</td>
</tr>
</tbody>
</table>

Variable 3:

<table>
<thead>
<tr>
<th>Day</th>
<th>Day</th>
<th>Day</th>
<th>Day</th>
<th>Day</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>15</td>
<td>22</td>
<td>29</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>16</td>
<td>23</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>17</td>
<td>24</td>
<td>31</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>18</td>
<td>25</td>
<td>32</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>19</td>
<td>26</td>
<td>33</td>
</tr>
<tr>
<td>6</td>
<td>13</td>
<td>20</td>
<td>27</td>
<td>34</td>
</tr>
<tr>
<td>7</td>
<td>14</td>
<td>21</td>
<td>28</td>
<td>35</td>
</tr>
</tbody>
</table>
Final Project Design: Research Questions and Survey Design

For the final project for this course, you need to plan, design, administer, and write an analysis of a research project of your own choosing. This activity is designed to help you get started on the survey.

Part 1: Choose Your Research Questions

1. **Brainstorm** a set of beliefs (no less than three) that you want to—and reasonably can—test.
   - Example: I believe that people whose children live at home eat out less often each week than people who have no children at home.
   - Example: I believe that men drink more alcoholic beverages per week than do women.
2. Decide on your target population. Who are the people you want to study?
3. Think about how you will select your sample. What type of people do you need to answer your research questions? How will you make sure your sample is representative of the target population? What method will you use to obtain your sample?

Part 2: Design the Questionnaire

4. Draft an outline for a five-to-ten question survey.
   - **BACKGROUND QUESTIONS:** Some of the questions must ask respondents’ about their backgrounds (e.g., age, gender, ethnicity, etc.). Background questions are used to describe your sample. They may also be used to divide your sample into two groups in order to test for a significant difference (e.g., males vs. females, or older vs. younger people). Background questions are usually placed at the beginning of the questionnaire.
   - **BELIEF QUESTIONS:** Your questions can be based on only one or all of the beliefs you brainstormed for item 1 above. You can write more than one question for each belief stated in item 1 above. Check to make sure each question is necessary—that it will gather data needed to test your hypotheses or estimate your specified parameters.

Write and rewrite each survey question so that it is:
   a. simply and clearly worded;
   b. impossible to misinterpret the question;
   c. avoid giving an answer.

Do not use response categories when the person can give an exact estimate. For example, do not use age ranges when asking for age, or do not give ranges as response options when asking for how much or how often a person does something. Simply ask for an estimate (“On average, how often do you …”) and then leave a blank space for the response. You can calculate means and standard deviations for exact estimates, but you can only get counts and percentages for response categories or ranges.

5. Design the survey plan.
   a. Decide if it will be in written or interview format.
   b. If it will be in interview format, how will you record the responses?
   c. Determine when and how you will carry out the survey.

6. Write a brief introduction to begin each interview or survey that is something like the following:
   - I am conducting this survey to fulfill a requirement for my statistics course at the University of Minnesota. Please take a few minutes to fill it out. Thank you.

You must turn in a summary of your research questions and a rough draft of your survey by the class period indicated on the syllabus. The instructor will give you feedback and suggest changes before you actually administer the survey and collect data.
Part 3: Pilot and Administer the Questionnaire

8. Pilot your survey with a small group of people (family, friends, class-mates, etc.). See if they have any problems or useful suggestions on how to improve the questions. Make any necessary revisions before administering the survey for the actual study.

9. Administer the surveys or interview questions to twenty to thirty people.

Part 4: Analysis and Written Report

You need to carry out the procedures for at least THREE hypothesis tests. No matter what you choose to do, at least one of the hypothesis tests must test for a difference between two samples.

10. The final product for this project is a written report. Each student must produce his or her own report written in his or her own words. Use the following format to write up the research results:

Your report will have four sections. Use the following headings for each of the four sections:

a. **Introduction:**
   What were the overall research questions? Why were you interested in these research questions? What did you expect to find out? What kind of differences did you expect?

b. **Methods:**
   Describe how the research was conducted. Give as much detail as possible. Describe how you designed your survey, including the piloting of the questions (see Part 3 above).
   Describe how you obtained your sample and how you administered the research (when, how, how many, etc.). Be clear and precise (for example time of day, day of the week, location on campus).
   You want to describe the participants in the sample so that a reader can determine the larger population of people they represent. Use statistics to do this. You only need to report statistics for the variables you are using to describe the sample (e.g., averages, standard deviations, medians, interquartile ranges for each measure; graphs from JMP IN showing the distribution of each measure including histograms and outlier box plots).

c. **Analyses.**
   Describe how you analyzed the data.
   State the type of hypothesis test you are conducting (e.g., a one-sample hypothesis test for a proportion, or a two-sample hypothesis test for a difference between means, or a Chi-square analysis). **For each hypothesis test, state the null and alternative hypotheses.** Report the statistics on which you base your decisions (e.g., in a two-sample hypothesis test, report the means and standard deviation for both groups, the confidence interval, and any other information necessary to draw a conclusion about the null hypothesis).
   Interpret the result of each hypothesis test (i.e., what do the results of the hypothesis tests imply about your beliefs or research questions?).

d. **Conclusions:** Address the following questions:
   What did you learn about the original research questions? Were your beliefs supported by the results of the hypothesis tests? State how your beliefs were or were not supported.
   Reflect on what you learned about the process of designing, administering, and analyzing data from a research project? Are there any ways that you can see to improve the research? If so, state what they are and how they would improve the research.

e. **Attachments:** Attach all relevant printouts that contain the graphs and statistics for descriptions of distributions and the hypothesis tests. Also attach one blank copy of the survey.