

Group Project Information

Every student will be assigned to a small group for the project in week three of the course. Students may form their own group if they wish (group size 3-4 members). Each group will select three questions to analyze from an instructor provided list of project questions (available in Week 3). Groups will answer the questions through the application of statistical concepts covered in the course. The group will prepare and submit two reports about the questions, solutions, and conclusions. The first report, an informal report due in week 5, will only address Descriptive Statistics (Chapters 1-4). The second report, due near the end of the course, will be a formal report which includes both Descriptive and Inferential Statistics. Please note that although one group report is to be delivered for the assignment, all group members are expected to complete all of the analysis questions so that all members are fully engaged with each analysis.

Report 1: Descriptive Statistics & Graphical Displays (see the section with this same name under the “Assignment Details” section below)

- Produce a summary statistics table for each project question with relevant graphics and discuss briefly the data. Appendix 'A' provides details for what should be included in the report.
- Due in Week 5

Report 2: Inferential statistics and completed report (see “Statistical Inference” section below)

- Apply statistical methods used in class to answer the project questions.
- Produce a formal/professional report including descriptive statistics. Appendix ‘B’ provides some report guidelines.
- Due in Week 9.

Description of the Data and Problem:

You are a group of statisticians who provide consulting services to small biotechnology and life science companies. A client has performed an exploratory pilot study to investigate the relationships between two cholesterol lowering drugs, lifestyle information, and blood lipid levels. Data were collected from each patient at a screening visit and at a follow-up visit three months later. At the screening visit, demographics, lifestyle information, and blood lipid readings were collected. At the follow-up visit, weight and blood lipid readings were collected as well as information as to whether or not the patient experienced stomach pain while taking the cholesterol lowering drug.

These data are contained in the file **lipid2.xls**:

Column Name Description

Patient ID - unique number for each patient

Drug - cholesterol lowering drug taken: A or B

Stomach Pain - Did the patient experience stomach pain during the study? Yes or No

Sex - F or M

Age - age in years

Height - height in inches

Systolic BP - blood pressure - systolic

Diastolic BP - blood pressure - diastolic

Exercise - minutes of exercise per day

Coffee Consumption - cups per day

Alcohol Consumption - drinks per week: 0, <2, >2

Weight - weight at screening visit

Weight3 - weight at follow-up visit

Cholesterol - cholesterol at screening visit

Cholesterol3 - cholesterol at follow-up visit

Triglycerides - triglycerides at screening visit

Triglycerides3 - triglycerides at follow-up visit

HDL - HDL at screening visit

HDL3 - HDL at follow-up visit

LDL - LDL at screening visit

LDL3 - LDL at follow-up visit

The client has requested you to investigate any interesting or meaningful relationships that may exist in these data and create a written report of your results and findings. Your findings will then be used by the client in the planning of future clinical studies.

Assignment Details:

Your group will select three questions of interest from the instructor's list of assigned questions and then define an appropriate subset of the data to investigate these questions. Three statistical analyses will determine the questions you analyze, and they are:

- (1) A Hypothesis Test and Confidence Interval for Two Means (Chapter 11)
- (2) A Hypothesis Test using the Chi-Square Distribution (Chapter 12)
- (3) Analysis of the Relationship between Two Continuous Variables (Chapters 4 & 14)

Your report will contain the following sections:

(I) Introduction – Provide an overview of the project, the questions that were studied (and why you chose these particular questions), and analyses that were performed. Define the data that were used, details about any transformations performed on the data, and discussion on existing outliers.

(II) Statistical Analysis of Each of the Three Questions – Each question being analyzed should include descriptive and inferential data. Details of each are as follows:

· **Descriptive Statistics & Graphical Displays** (Chapters 2 and 3) – Provide some meaningful descriptive statistics about the data such as mean, median, standard deviation, range, quartiles, IQR, fences, outliers, etc. and put this data in a table. Include charts that will best display the data – usually bar graphs, histograms, regression plots, are effective, but other types of charts presented in the textbook may also be helpful. Make sure your tables and charts are appropriately labeled and clearly discussed in your text. Large data tables showing transformations, if necessary, should be placed in an appendix so as not to interfere with the flow of the report. Refer to Appendix 'A' for examples of what is expected.

· **Statistical Inference** – For each of the analyses state the question of interest which was investigated, analysis method that was used, mathematical/statistical details, your conclusion, and an interpretations statement. Each of these should be presented in the same way as the examples in the book showing the steps – hypotheses, critical values, formulas, test statistics, decisions/conclusions, and summary interpretation statement. **It is necessary to show intermittent calculations**, and show enough of your work so that numbers can be evaluated. Present both the “classical” and “p-value” approaches and provide an interpretation statement for the

confidence intervals. The p-values can be obtained using the tables (and estimating if necessary) or by using a calculator or software.

The question involving regression should include a scatter plot of the data and the least-squares regression line on a single graph, calculation of the correlation coefficient, and the equation of the least-squares regression line. Comment on the relationship that appears to exist based on the results of the scatter plot and correlation coefficient and coefficient of determination. Interpret the slope and y-intercept of the least-squares regression line. Test whether a linear relationship exists between the explanatory and response variables and calculate a confidence interval for the slope of the regression line. Remember to provide an interpretation statement for your tests.

(III) Conclusion – Summary of your analyses and recommendations for further study.

(IV) Lessons Learned – Thoughts and comments about your group project experience, and suggestions for improving the group project or online statistics course.

(V) Appendix – Place any large amounts of data, data transformations, or long statistical analyses in an appendix so that the report body can flow without interruption. Include summary statistics in the body of the report so the reader does not have to repeatedly refer to the appendix for such information. It is acceptable to use only certain columns of the data or to perform simple data transformations (changes, % change, etc.) as long as you provide the details in your report so that the results can be independently verified. The goal of this assignment is to simulate a real-world situation of investigating, analyzing, and interpreting data.

Listed below are some examples of previous project questions, the appropriate analysis method, and the data needed:

- Is there a significant difference between the two drug groups in mean LDL level at screening? Approach: Hypothesis Test and Confidence Interval for Two Means. Use the columns Drug and LDL to calculate the mean and standard deviation for each drug group.
- Is there a significant difference between males and females in mean weight change from screening to follow-up? Approach: Hypothesis Test and Confidence Interval for Two Means. This requires a data transformation. Create a new column which contains the difference between Weight and Weight3. Then use this new column and Sex to calculate the mean and standard deviation for males and females.
- Is there evidence to indicate that the proportion of patients who experienced stomach pain differs among the three alcohol consumption groups? Approach: A Hypothesis Test using the Chi-Square Distribution. There are three Alcohol Consumption groups and two Stomach Pain groups, so tabulate the number of patients in each of the six groups.
- Is there significant evidence that the proportion of people who did not exercise experienced more stomach pain? Approach: A Hypothesis Test using the Chi-Square Distribution. This requires a data transformation. Create a new column Exercise Status which is Yes if the patient has a value > 0 in the Exercise column, No if Exercise = 0. Now there are two Exercise Status groups and two Stomach Pain groups, so tabulate the number of patients in each of the four groups.

Tips for Productive Work Groups

1. The purpose of the group project is to simulate a real-world project through collaboration. Your exchanging ideas, sharing knowledge, and supporting each other will benefit each member of the group. You will meet some talented people with diverse backgrounds with a common interest, so have fun.
2. Groups will be formed based on the information provided by students in their “student introductions” in their first Discussion Board assignment. Students will be informed of their groups during week three.
3. Contact your group members as soon as possible, and identify a project leader or someone who will keep track of schedules. If you are having difficulty contacting any group member by the end of week 5 please make note of it in your Part 1 report email.
4. Establish a project schedule listing due dates for the completion of each section of the report.
5. Select the analysis questions from the list of project questions provided by the instructors. Examples of the types of questions to be analyzed can be found in their respective chapters in the textbook.
6. ***Each member of the group is expected to analyze and complete each of the three project questions*** in order for meaningful group discussions to occur. This greatly improves the learning process as students may come up with different ways to solve the problems.
7. Produce a rough draft of the report by Wednesday following Quiz 8 so that all group members may have several days to review the report and make any necessary revisions.
8. Save the final version of your report in an Adobe PDF file and email it to the instructors. Be sure to include a title page with the Group # and all participating group members.
9. Complete the Group Project Survey which includes a peer evaluation of group performance.
10. Enjoy the experience.

Recommended Report Outline

Introduction

- Project background
- Questions of interest and why you chose them?
- Approach to answering the questions
- Data discussion: transformations and/or outliers

Question #1 (Difference between two means)

- Descriptive statistics
 - Discuss the data
 - Include a table of summary statistics (mean, std dev, quartiles, fences, ...)

- Include a chart(s) that effectively displays the data
- Inferential statistics (follow the procedures in the textbook examples)
 - Null and alternative hypotheses
 - Level of significance (why this level?)
 - Critical values (explain how you got them)
 - Test statistics (classical method and p-value method)
 - Test requirements met?
 - Display formulas and intermittent calculations through final answer (so instructors can see where in the process an error was made, if any).
 - Test conclusion (compare test statistic with critical value)
 - Interpretation Statement (see textbook for examples)
 - Confidence intervals

Question #2 (chi-square)

- Descriptive statistics
 - Discuss the data
 - Include a table of summary statistics appropriate for the type of data (contingency table).
 - Include a chart(s) that effectively display the data
- Inferential statistics (follow the procedures in the textbook)
 - Null and alternative hypotheses
 - Level of significance (why this level?)
 - Critical values (explain how you got them)
 - Test statistics (classical method and p-value method)
 - Test requirements met?
 - Display formulas and intermittent calculations through final answer (so instructors can see where in the process an error was made, if any).
 - Test conclusion (compare test statistic with critical value)
 - Interpretation Statement (see textbook for examples)

Question #3 (Regression)

- Descriptive statistics
 - Discuss the data
 - Include a table of summary statistics
 - Include a chart(s) that effectively display the data
 - Show the regression line and equation in the chart of data
 - Explain the regression equation (slope and y-intercept)
 - Discuss the coefficient of correlation and determination
- Inferential statistics – test the significance of the least squares regression model (follow the procedures in the textbook)
 - Null and alternative hypotheses
 - Level of significance (why this level?)
 - Critical values (explain how you got them)
 - Test statistics (classical method and p-value method)
 - Test requirements met?
 - Display formulas and intermittent calculations through final answer (so instructors can see where in the process an error was made, if any).
 - Test conclusion (compare test statistic with critical value)

- Interpretation Statement (see textbook for examples)
- Confidence intervals

Conclusion

- Summary of the analyses and ideas for future tests or research.

Lessons Learned

- Provide comments on the group project (interest in the subject matter, interaction among members, value of project, ideas for improvement)
- Comment on any other part of the statistics course that you enjoyed or would like to see improved.

Testimonials from Former Students Regarding the Group Project

Usually there are some students that are concerned about participating in a group project in a fully online course. These testimonials are presented to help reduce concerns so that students can enjoy the experience.

I thought the group project was a great experience for learning how to crunch actual data and put the statistical analysis tools we've learned in the class to use. For me, it was the most beneficial.

I really enjoyed working with this group and learned a great deal about how to apply statistics to problems where the outcome is not clear cut. Thanks!

Overall this was a pleasant experience. Everybody in my group contributed to the project and it was much easier to do as a group than it would have been individually.

It was a very exciting experience with our group project. It provided a platform to apply the statistical concepts that we learned in this course. Provided an insight how things work in a live project and to interact with people from different background and gain from there experience.

...it was fun and useful to have some freedom to apply the principles we learned throughout the class. It tied things together, was relevant and valuable. Interesting to see how others approached the problems and the tools they used.

I really enjoyed this project, just like you all said I would. Honestly, this is a great way to end the class...helped me work through the use of the inferential tests we learned during the class, and it gave me confidence in my own statistical skills. Overall, we were a group that respected and appreciated each member's contributions. Project members brought a lot of energy and purpose to the project. There was a sharing of ideas and many decisions were taken based on consensus. It was fascinating using different communication tools to coordinate tasks and keep in touch and it was beautiful to see the project taking shape to its final conclusion. It truly was a global project with a global effort and can be a model for commercial application.

Overall, it was a great experience. There was a lot of talent in the group and everyone contributed something unique to the final product. The hardest part was trimming down the report so that it's not too tedious to read, but I think the overall product is very interesting and certainly hammered home the main goals of the class.

The group project was a great experience. I learned a lot. The project allowed me to utilize Excel like I had never done before and this was truly a great learning experience. My group's members were both equally great to work with. Our scheduled appointments to get the work done for our project was timely, a good use of time and very productive. We learned from each other and I think we put together a great project. I am quite happy to have had the chance to work in a group as it was just like a real-world experience where a challenge or goal is proposed and group interaction and individual skills are brought to the table to come up with a solution... The course has been very insightful to the world of statistics in everyday living. I believe that the course was

comprehensive and exactly what I was looking for in an introductory course.

Appendix 'A': Examples for Part 1 descriptive statistics tables are as follows:

Problem 1: Quantitative data (for example: Is there a significant difference in weight between male and female soccer players at the college level?)

Statistic	Male	Female
Sample Size		
Mean		
Standard Deviation		
Q0 Min		
Q1		
Q2 Median		
Q3		
Q4 Max		
Range		
IQR		
Upper Fence		
Lower Fence		
Outliers		

Provide histograms for each variable displaying data distribution, and comment on the distribution and any outliers. Box plots would also be a nice addition.

Problem 2: Qualitative Data

Produce a contingency table for the variables (Chapter 4.4, Table 9 is an example), and a bar chart for a visual display (refer to Figure 28 on page 230 in the textbook).

Problem 3: Quantitative data (for example is there a relationship between weight and height for male soccer players at the college level)

- present a summary statistics table similar to the example shown above
- provide histograms for each variable, and box plots if desired
- include a scatter plot showing the data and LSR line equation.

Appendix 'B': Example for Part 2 - Inferential Statistics

Prepare for the analysis by reviewing the test requirements and verifying that they are satisfied.

Step 1: Determine the null and alternative hypotheses.

Step 2: Select a level of significance.

- Step 3: Compute the test statistic (for both classical and p-value approaches).
- Step 4: Compare the critical values to the test statistics and draw a conclusion.
- Step 5: Provide an “interpretation statement” which formally summarizes the test.

Guidelines for producing the final report:

- Include a cover page with participating group members and your group number.
- Only one report file created in a word processing document and **saved as a PDF** file with the course name and group number in the file name (**example, “SP21 Biostats Group XX Project Report”**) should be sent to the instructor (do not send spreadsheets, or PPT presentations, or any other format).
- Font size should be legible in text, charts, tables, etc.
- Carry decimal places out to an appropriate number of places (not too many, not too few).
- Large amounts of raw data should be placed in an appendix so that it doesn't interrupt the flow of the body of the report. Only summary statistics tables and appropriate charts should be in the body of the report, in addition to test calculations, comparisons with critical values, and conclusions. Raw data tables should go in the appendix.
- Email the final PDF report file to the instructor via Canvas with your course name and group number in the subject line (**example, “SP21 Biostats Group XX Project Report”**)
- Following the report outline in the “Project Introduction File” is recommended. And review the Project Scorecard file to ensure you are providing the appropriate report content.