**Question 1a, 2a, 3a. Report 2**

Question 1 (Difference Between Means):

a) Is there a significant difference in Cholesterol change from screening to follow up for patients who have taken Drug A compared to Drug B for patients aged 22-25? (Create a new data column for the change in cholesterol, for example CHOL3-CHOL.)

Question 2 (Chi-Square):

a) Is there an association between patient BMI and cholesterol level at screening? Consider overweight, normal and underweight categories for BMI

Question 3 (Regression):

1. Is there a relationship between weight and LDL level at screening for male patients aged 22-25?

**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.

**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.

**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.

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.

