Rev: Sep 3, 2006

# STAT 230 Project Description and FAQ

## Term Project

Why is there a term project? Understanding why may help you get a better grade. There is a lot of math in this statistics course but it really is more of a management and business course. You are tested on your knowledge of "number crunching" with quizzes and the final exam. However the most important elements are the business aspects and this is tested through the project which is very similar to a research paper but with statistics tests instead of references and footnotes. The term project will provide the opportunity to apply the statistical techniques you learn in this course to real-life business situations. Remember this is a management/business course not a math course and I will grade accordingly. You'll work on the project throughout the course and provide progress reports and a draft on each specific section (deliverable) of the project. Look at each part as a separate assignment and then near the end of the course you put all the parts together and your are done! See the course schedule for due dates. You must complete and submit your project by the last day of course. You may work alone or with one other student but you are responsible for turning in the complete project on time if the your team member does not complete his or her part of the project on time. If you work with one other student you must notify the instructor no later than the end of the third week of the course. Always put both names in any of the work if you do have two working on the project. You can decide to work alone if you are dissatisfied with your partner but you must still submit your project on time.

Note: For this and all parts of the project you will be provided a general template. You are required to use this template (in MS Word).

You are required to submit progress reports in the conference **Submit Project Progress Report Here!** Submit in MS Word and a spreadsheet for the calculations. The Word document should be the draft of the completed work on the assigned portion of the project. Think of progress reports as pre-grading which provides you a way to fix any problems before your submission. Your progress report is part of your participation grade. You are encouraged to submit as much of that portion (the specific deliverable) of the actual project for the progress report but do not submit past portions of the report as that will only confuse me. The more you submit for that portion the more I can comment. It is better that I give you a chance to improve on your work before I grade it at the end of the course. Please remember: this is NOT a math class but a management course and you should submit a management document! I want to see you apply statistics in a practical manner. Just don't "crunch" numbers or this will hurt your grade. Give me a clear, easy to read document that makes management sense! I will grade on clarity of writing so make an effort to communicate! Usually it is best to supply a narrative (like in Word or directly in the conference) with highlights of your data and calculations and then in Excel for all of the calculations. Although you can include Excel for your calculations they should be transferred to an appendix as part of the final project (upload one document). **If you do not submit on time then I will not comment later on that portion of the project so make it a point to meet the deadlines!**

Cross-sectional versus Time series data: Your data should be cross-sectional **NOT Time series**. Cross-sectional are data collected at approximately the same point in time while time series data are data collected over several time periods. Since we are interested in inferential statistics do not use time series (axes by days, months, years). For each deliverable explain the data in terms of it being a sample how you derived the sample and what is the population it represents. **Do not show me a stock prices over a period of days, weeks, months, years or other similar items.** To help you in your analysis of the type of sample you have, look on the lower right side of figure 1.2 (p. 10) as well as review pages 10-14. If you have for example, one stock that you analyze over weeks or months this will not be appropriate for you project. There are other and better analytical tools not covered in this than inferential statistics. If you are interested in the stock market then the type of appropriate data would be one 30 or more stocks taken at one point in time. Also gasoline prices over time is not acceptable.

At the beginning you select a theme (subject) for your paper and you must stay with that theme for all of the parts of the paper. You do not have to use the same set of data for all parts but they should be generally related to each other so your management report is coherent. You are encouraged to use data from your profession or future profession. If the project helps you in your job that is even better. There are many sources for data including your organization, government agencies, the Internet and the library. Your librarian at Maryland is here to help you. A common concern of students is that there is no data. There is data and perhaps too much data! If in doubt, contact

your librarian or get data at work. You may also be able to conduct simple surveys but they must have management relevance.

Do not include additional data or statistical analysis other than what is required below although in "real life" you would certainly be encouraged to do more than what is required in the project. Be sure to label your project with the same bold face headings as below so I understand what you are doing. Here are the instructions for each part of the project. Note: Style, grammar, clarity, relevance, insight, and organization are very important. A poorly written project will result in a lower grade.

**Deliverable 1: Descriptive Statistics**

(10 points) Our main emphasis in the course is "inferential" statistics which means taking samples and drawing an inference or conclusions about the population. Access raw data or a database from government, business, health, and similar official Web sites pertaining to your area of interest. Collect at least 30 pieces of numerical (quantitative) metric data (see p.15) but no more than an n of 50 (30-50 observations and only one theme). If you have a sample larger than 50 randomly select a subset so your n (sample size) is no more than 50. Explain where these data came from and why they are of interest to you. Describe the population and the variable. From the data, plot a histogram, a stem-and-leaf diagram and an ogive (polygon). Also calculate the mean, median, mode, range, standard deviation, and quartiles of the data. Create a boxplot. Explain what this analysis tells you.

In a separate appendix (or spreadsheet), list all 30-50 observations labeled from 1 to 30 (up to 50 if n=50) so I can duplicate your work if necessary. Since you will be doing a histogram you will need to select a sample that consists of numerical (quantitative) data not categorical (qualitative) data (see p.15). A bar graph is not the same as a histogram so in Excel, click bars (right click properties) /format data series/options/gap width (should equal to zero) and this should get rid of the gaps (histograms do not contain gaps...only bar graphs are used for categorical data. Your textbook describes other methods to provide an acceptable histogram and other graphics.

If you have a category/class in the data with zero observations then try to get rid of the gap by extending the width of the class interval or at the very least explain it in your comments. Histograms and other descriptive statistics should not add to the confusion or generate more questions but should answer and explain the data. Look at your descriptive statistics and ask if there are any questions that would be asked and can you answer them by modifying the descriptive statistics or adding a comment or label. See also p. 81, exhibits 2.1 and 2.2.

Note: *For this and all parts of the project you will be provided a general template. You are required to use this template (in MS Word).*

**Deliverable 2: Probability and Estimation**

(10 points) Calculate a binomial probability distribution relating to data of interest to you. Explain why the binomial distribution is applicable. You will also address the following areas:

- **Management Problem:** *Explain in a few sentences the purpose of using the binomial experiment. What do you hope to achieve by this experiment? What is the scenario? This is a key part of your grade for this section.*

- **Source of probability (P=?):** *Where did you get this probability? If it was from a sample, describe the same and include sample size. Sample size must be at least n>30. You can also use a mathematical probability or from an article that has stated a probability.*

- **Number of Trials** (n=?): *Explain why this number was picked.*

- **Successes to compute** (x=?): *Explain why you picked either x>? or x<? Do not use x=? since that is not usually too useful.*

Summary of results: In a few sentences explain your results and management or decision making implications. This is an important part of the Binomial deliverable.

*The provided template can help you organize this part of the project.*

## Statistical Inference (Deliverables 3 to 5)

*Perform each of the steps below using Excel. Explain in each case where these data came from and why they are of interest to you. Use your collected data as a sample of the population. Describe the population and variable that the data represents. Interpret each of your results with a conclusion and decision. For hypothesis testing use as appropriate the steps and format on page 344 (8.3). You do not need to include all the raw data in your narrative but do include the standard statistics and parameters such as n,* X, u, etc.

**Deliverable 3: One-Sample Hypothesis Test Confidence Intervals and (20 points of project grade). 2 parts**

3a. (10 points) Take a random sample and perform a *t*-test against a population mean (no time series for any parts of this project.) Test a hypothesis about a population mean. Besides the instructions above, explain what assumptions you made. Note: n must be between 15 to 20! Use no other sample sizes.

3b. (10 points) Use the same data as 1a and develop a 90% confidence interval and a 95% confidence interval for the population mean. Contrast this to 1a. What additional information or insight did this provide you over the traditional hypothesis testing.

**Deliverable 4: Two-Sample Hypothesis Tests & ANOVA (30 points of project grade). 3 parts**

4a. (10 points) Test a hypothesis comparing two population means. Collect two independent samples. Perform a *t*-test using two independent samples. Note: n must at least 10 for each sample but no more than 30 for each sample. Since independent both samples do not have to have the exact value of n.

4b. (10 points) Test a hypothesis comparing two population means using paired (matched) observations. Collect two dependent samples. Perform a *t* test for paired observations. In addition to the other requirements listed above explain why they are dependent samples. Note: n must at least n must at least 10 but no more than 30.

4c. (10 points) Collect data. Select three independent samples. Perform a one-way test for analysis of variance (ANOVA). In addition to the other requirements listed above be sure each sample is at least 10 or greater. Also include n, sigma, mean, for each sample.

**Deliverable 5: Chi-Square Tests (20 points of project grade). 2 parts**

5a. (10 points) Collect observed and expected data. Perform a chi-square goodness-of-fit test (chi-square test for differences in more than two proportions). In addition to the other requirements listed above include a table with your observed and expected data in your narrative. Most of the other data can be moved into the appendix.

5b. (10 points) Collect observed data in the form of a contingency table. Perform a chi-square independence test. In addition to the other requirements listed above include the contingency table in your narrative. Most of the other data can be moved into the appendix.

**Deliverable 6: Simple Linear Regression (10 points of project grade )**

Collect bivariate data. Perform a linear regression and correlation study, including creating a scatter diagram and a residual plot, determining the regression line, the coefficient of determination, and the correlation coefficient, testing for the significance of the linear relationship, and interpreting your results. Be sure to state your dependent variable and one independent variable and why you think there might be a relationship. Do this before running your regression. Note: must have at least 15 observations but no more than 30!

## Write-up

Your write-up or project will consist of all deliverables put together. You will use the Word template provided-- no other will be accepted! All these deliverables will be approximately ten to fifteen pages single-space (not including the appendix). (Excel outputs should be in an appendix in Word). Assume you are publishing your research in a trade journal or a report to the CEO. Include your analyses, conclusions, and decisions that are required (see above). Do not use first person narrative. The summary should consist of the following parts (deliverables) and should be clearly labeled as shown below (the template will do that for you):

You will upload to the assignment area:

1. One (only one) Word document which will contain the narrative and the appendices. You can also upload the Excel files but you still need the appendices in Word. You may also zip the files if you like. I do not want more than one Word document. Do include the appendices with the project.
2. Once you have uploaded verify that it is in fact there by downloading again and making sure the documents upload properly. You are responsible for insuring the documents are uploaded. If in doubt, contact the Help Desk.

---

## Additional Information

You should strive to make your write-up as clear as possible. Include enough information in the summary (p, z, t, n, LoS,x bar, u, Ho, Ha, etc.) so I understand your various conclusions but your calculations should be put in an appendix. The instructor should be able to reproduce your results!

2. Excel outputs should be converted to the appendix and in the write-up you should clearly reference where they are in the appendix! During the progress reports you can use an Excel instead of the appendix. However, your narrative should be in Word and NOT in Excel.

3. Consult UMUC's *Guide to Writing and Research* for correct style. Do not copy and paste what you write in conference and put in your paper.

## Progress Reports

You will submit progress reports (see the schedule) on how you are doing with all of the project assignments. Use the template provided (delete all red text) but only that portion of the template for the specific deliverable for that project. If you want the instructor to comment and offer advice on how to improve a specific part of your project please submit in the conference called "Submit Your Progress Report!" (not by email) by the deadline date listed in the schedule. Submit in Word and do not put your narrative in Excel! Do include the Excel spreadsheet but to show your raw data and calculations. Failure to post on time will not only hurt your participation grade but probably result in a project that could have been better. The instructor will not actually grade the specific part of the assignment (you will get participation credit) but tell you on how to improve on it for your final submission (above). The instructor will be "brutally honest" so do not take the constructive criticism personally. This will help you get a better grade. The more you post the more the instructor can comment on it. I will not comment on progress reports submitted late.

# FAQ

### What does the project consist of?

It consists of the Write-up of each deliverable and the appendices. You will also submit progress reports as listed in the schedule. You will use the Word template. At the end of course you will upload one Word file containing the write-up and appendices.

### Is there a sample I can see?

I will make available one or two examples. They are not in the exact format required but it will give you an idea.

**What am I to submit, where do I submit, and when do I submit my Project progress reports?**

> ***What***: The best is to post the complete draft or as much as possible so I can comment and you can improve that part of the project.
> ***Where:*** Post in the conference, "Submit Your Progress Report Here." Under this topic, will be a topic for each assignment and the deadline.
> ***When:*** It will be the last day listed in the deadline Due Date. For example, if the first week Due Date lists 10/23-10/31, the due date for the first project assignment would be 10/31. Also the topic will tell you when it is due.

**How do I send the project?**

You can post in the assignment area. Be sure you "submit for grading" and not as a draft. Once you have submitted go to your assignment folder, download the assignment and verify everything is complete. Once or twice students a semester a student doesn't check and has given me an incomplete assignment. Save the tears and verify by downloading and looking at your assignment. You will submit ONE Word document (the project and appendices) and an Excel file with your calculations.

You can optionally mail it to me but it must be post marked before no later than the deadline. Do not send it Express or Priority as I have problems with my post office.

**When is this project due?**

Look in the syllabus for the deadline.

**What is the format for submission?**

You are required to use the Word template provided (delete all red text). I will not accept any submission that is not the Word template provided.

**Grading criteria?**

The most important element I grade on the paper is your language and communications skills not just the number crunching. If you have a confusing, "messy" paper, did not use the template, disorganized with misspelled words, you will lose points. An A paper is simple to read and not a difficult one marked by excessive Excel copy and pastes (instead of in the appendix), missing standard statistics and parameters, tables and graphs with vague or missing titles and labels, numbers that do not state units of measure (feet, pounds, age, etc) that use vague terms such as "average," or incorrect statements such as "accept the null hypothesis." An A paper is a literate, insightful research paper that uses statistics to illustrate not confuse. An A paper respects the reader and tries to make it easy for the reader/manager to understand the key "messages."

Statistics is a communications tool and you better make your input clear and precise. If I don't understand what you are doing, what your conclusions are, then you will get a lower grade. Keep it as simple as possible while using the correct statistical terms and labeling everything. Generally do not use the term "average" if you mean mean, medium, or mode or say you "accept the null hypothesis (you can only fail to reject not accept)...you insist in not do so, you will lose points. I will see if you are making the right calculations and using the statistical tools appropriately. For example, on your binomial assignment, I will see if you are using the tool in a correct meaningful way. It is your job to use the statistical (whatever it is) in a meaningful way. For your project I am more interested in your communications skills than your math skills. Of course you still need to appropriately use the statistics.

Take care in explaining your data and terms. Do not be sloppy with your terms such as labeling average when you should label mean, medium, or mode. Clearly communicate and use proper English or I will grade you down if I don't understand where you are going. Label everything clearly and don't refer to items such as "it" or "they" but be specific. Think of your Project as a research paper where English and clarity do also count.

Be sure to read my comments to you and to all of the students in the "Progress Report." If I mentioned to you in the conference to change something and you fail to do so then I do take off extra points! This also includes what I say

to the class in general and to individual students if it applies to you. This is why progress reports are public because you should look at each submission and my comments to each student and learn from them.

**What if I can't find any data?**
I get this question every time from some students. You will have to work and spend some effort. It might take you about 6 hours to get the data you need. This is called research. There are many, many resources in the library and on the web. Do check out Maryland's online library which have data sources that might be able to apply to your project. Maryland's librarians might be able to help you search for data to use in your project. You may be able to get data from work that you can use and perhaps be able to convince your boss to do a project which you can adapt to this assignment!

You can also talk to your other instructors (past and present) and they may be able to point you to databases and journals that might provide you the data. You may be able to use some of the work in your other courses and apply to this project. Not only is this allowed by encouraged.

If you need to call me then that is ok. I will not do your research for you but I will try to lead you. Part of your grade is based on your ability to find the data.

**What do I submit in the conference?**
Be sure to submit your report when due in the Conference (topic: Submit your progress report for week x" If you missed one project report then it is too late for me to comment on the previous deliverables. Only include your work on the deliverable due for comment. Use that part of the template that relates to the deliverable in question. Don't submit the whole report each time.

Once submitted, I will comment on your submission. I do this via the conference because we all learn from each other's postings. Once the topic is closed we can continue the discussion via email.