Long Assignment for
General Statistics
Mathematics 112

Ayesha Delpish
Mathematics and Statistics Department

Introduction to Faculty Colleagues

Course
MTH 112 - General Statistics is used by many students to satisfy the first-year core. These students are from a variety of majors and may not be quantitatively advanced.

Assignment Fit into Course
This assignment will be used throughout the course. Each of the first 2 phases are linked to a specific unit.

Aims for Assignment
The aim of the assignment is for students to apply statistics in real-world contexts and to use it as evidence to support their conclusions.

Course Learning Outcomes
- Consume: Use statistical results appropriately to make life and academic decisions
- Create: Create models and compute appropriate measures to assist in decision making
- Communicate: Create an oral or written report on the results of a statistical study and interpret for those who have little or no statistical background

Specific Learning Outcome
As a result of working on this assignment the student will be able to generate and analyze appropriate statistics in order to draw conclusions. Students will also be able to present their findings in writing and orally to a non-statistical audience.

Important Features of the Assignment
There are 3 phases to this course project. A few notes:

1. Students are randomly assigned data sets. They receive individual files with specific prompts for their data set. They use the same data set throughout but it (and the question) is expanded as we progress through each phase. The scenario and audience is included in the data file.
2. At the end of each phase, students rate their group member and submit a group scoring sheet which they use to evaluate their team members.

3. There are dates built into the syllabus for peer and instructor review of drafts. Students use established rubrics (which we call coversheets) to provide feedback to their peers. This is the same sheet that I use to grade the final assignment—having students use the rubric for peer review creates an opportunity for clarification of the required components.

4. Each phase builds on and extends some elements of the previous phases and I collect all phases at the end of the course.

**Modifications Made**

1. I re-did the Phase I and Phase II instructions to specify the purpose, clarify the intent and place the phase in the larger context of the overall project

2. I clarified some elements of the rubric and project information sheet

3. I updated all data files to change the prompts from a general question, to a focused task reportable to a specific person. So I’ve identified a person to whom the report should be written.

---

**The Assignment I Distribute to Students**

**MTH 112 – General Statistics**

**Data Analysis Project**

**Term Project & Presentation**

There will be a group project due at the end of the semester. You will work in groups of 3-5 students to describe and analyze data to answer questions using techniques covered in this course. This will allow you to apply the skills you learn in this course to the real world which in turn will enhance your appreciation and retention of the material. There will be *mini-projects* due during the semester—guidelines and due dates are provided in a separate announcement on Moodle. The goal of the project reports is to slowly build the project pieces as the term progresses. During the final week of classes, each group will present their work. Each person’s grade for the term project will be an individual grade. Individual grades will be determined by the instructor and team member evaluations.
Sample Data File I Give to Students

BAT Mobile Data
Albuquerque Police Department began a special enforcement program aimed at countering DWI (driving while intoxicated) accidents. The program was composed of a squad of police officers, Breath Alcohol Testing (BAT) devices, and a van which houses a BAT device and is used as a mobile BAT station (Batmobile). These data were collected by the Division of Governmental Research of the University of New Mexico under a contract with the National Highway Traffic Safety Administration of the Department of Transportation to evaluate the Batmobile program. The first 29 observations are for a control period and the next 23 observations are for the experimental (Batmobile) period.

Variable Names
1. QTR: Quarter (Quarters 1-29 = Control, Quarters 30-52 = Experimental)
2. ACC: Injuries and fatalities from Wednesday to Saturday nighttime accidents

Your Task
You have been hired by the Albuquerque Chief of Police to investigate whether there are fewer injuries and fatalities during the experimental quarters than during the control quarters.

Raw data appears here.

Part 1
T-test

No late assignments will be accepted

In this assignment you will work in groups of 3-5 to describe and analyze data to answer questions using techniques covered in this course thus far. This will allow you to apply the skills you learn in this course to the real world which in turn will enhance your appreciation and retention of the material.

You will submit a typed report addressed to your client with all appropriate computer output incorporated into the body of the report. A large portion of your grade will be determined by how effectively you communicate your results. You should also select the most relevant parts of the analysis, not turning in gobs of output with no “story” or explanation. Your discussion should use language understandable by a non-statistician, but you may use standard technical terms and notation.
Your project must include the following sections:

- **Introduction:** Briefly describe the investigation and why it may be of interest. Describe the source of the data and any information on the data collection plan that you have. You must also clearly identify the test that you intend to use (1-sample, 2-sample independent or matched pairs) here and justify your choice.

- **Descriptive Statistics:** Produce and describe all necessary graphs and statistics. Provide a detailed explanation of what these statistics tell you about the data and the assumptions.

- **Statistical Inference:** Be sure to show all steps of the HT as done in class.

- **Conclusion:** Summarize your results. Comment on anything of interest that occurred to you during the project. Did the data behave roughly as you expected or did some of the results surprise you? Point out any unusual data values or interesting phenomenon. What other questions would you like to ask about the data?

**Things to Remember**
Check the model assumptions!!! **Comment on whether you think the assumptions are sufficiently met.**

You may find the following materials on the Moodle site:

1. Coversheet/rubric that must be placed at front of your project before turning it in
2. Data set descriptions with question of interest for individual data set
3. Sample project write-up

**Grading**
This mini-project is worth 6% of your overall course grade.

- Your project grade will depend on more than the “correctness” of your final model; after all, more than one model might be quite reasonable for your data.

- I will grade you on several criteria: first, on meeting the minimum requirements as outlined above; second, on the organization and professional presentation of your complete report; third, on the initiative and creativity shown by your team.
That’s right – show initiative! What’s described above is the **minimum**, so you should feel free dazzle me with other demonstrations of your knowledge😊

**Group Scoring Sheet**

*(Each student submits this after each Part and it is used to adjust individual scores as appropriate. If scores are to be adjusted down, I always call in the student for a conference.)*

Please rate each group members’ contribution *(include yourself)* to the specified assignment *(Phase I, Phase I, Phase III)* using the following three categories:

1-Below Average  
2- Average  
3-Above Average

**Group Member Scoring Sheet**

This information will only be made available to the instructor and will be used to determine individual scores on the assignment.

<table>
<thead>
<tr>
<th>Student Full Name <em>(Include yourself)</em></th>
<th>Score</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Overall comments about completion of the assignment *(Use back side of sheet if necessary)*

Each student must submit a folded completed form to Dr. Delpish on the due date of each group assignment.

---

**Part 2**

**Regression**

No late assignments will be accepted.
This project extends the analysis that you began in Phase II by incorporating regression models. **You must work with your group and data from your first project unless you've received permission from the instructor to do otherwise.** You will submit a typed report addressed to your client with all appropriate computer output incorporated into the body of the report. In your report you will describe and analyze data to answer questions using techniques covered in this course thus far. This will allow you to apply the skills you learn in this course to the real world which in turn will enhance your appreciation and retention of the material.

A large portion of your grade will be determined by how effectively you communicate your results. You should also select the most relevant parts of the analysis, not turning in gobs of output with no “story” or explanation. Your discussion should use language understandable by a non-statistician, but you may use standard technical terms and notation.

Your project MUST include the following sections:

- **Introduction**
  Extend your introduction from Phase I to briefly describe your investigation and why it may be of interest. You must also clearly identify the regression test that you intend to use here and justify your choice. Be sure that your response and predictor variables are clearly stated.

- **Descriptive Statistics**
  - *Basic graph and statistic:* Produce and describe/comment on the necessary graph and statistic relevant for correlation analysis.
  - *Assumptions check:* Identify the regression assumptions and then do the appropriate checks. If an assumption is not met, include and justify appropriate remedial measures.
    - *Note:* If a transformation is suggested, you must perform the said transformation and use that model from this point onward!!!

- **Parameter Estimation**
  - Start with a short paragraph describing what you plan to do and then include the formulae and working for your parameter estimates.
  - Your final model (predicted line) should be clearly identified using the correct notation.
  - End this section with interpretations of the regression coefficients and your coefficient of determination (if appropriate) in context and with sensible units.
• **Statistical Inference**
  o Perform the hypothesis test to answer the question provided in your data file for your final model.
  
o Be sure to show all steps of the HT as done in class.
    - You need NOT include working for the TS—instead, copy and paste the relevant SAS output into your document in step 6 and highlight the appropriate value in the output.
    - You must perform both the RR and the P-value methods for the test as done in class.
  
o After completing your HT, construct the relevant confidence interval (you may use the SE from technology) and comment on whether the interval supports the initial conclusions found in step 7 of the HT.

• **Conclusion:** Summarize your results.
  o Comment on anything of interest that occurred to you during the project.
  
o Did the data behave roughly as you expected or did some of the results surprise you?
  
o Point out any unusual data values or interesting phenomenon.
  
o What other questions would you like to ask about the data?

**Things to Remember**
• Clearly explain your process.
  
o So if you start with the basic model but then decide to do a transformation (IF YOU DECIDE ONE IS INFACT NEEDED), you must present/discuss all of the information for the basic model (including the statistics—scatter plot, correlation and coefficient of determination, equation of the line and graphs checking assumptions), clearly explain what assumptions are not met and your strategy for correcting it.
  
o If you do come up with a better line, you must present all of the statistics, graphs etc for that model also.
  
o It is the **final model** that you should use for inference.
• Your project grade will depend on more than the “correctness” of your final model; after all, more than one model might be quite reasonable for your data.

• This phase is worth 6% of your overall course grade.

---

**Part 3**

**Presentation**

This phase is the *presentation of your project to your clients*.

**Basic Requirements for Presentation**

1. For each question (you had 2 questions) clearly identify
   a) research question
   b) analysis method used
   c) Findings—at a minimum be sure to provide the TS, and P-value with your conclusions; for regression, the line for your final model and appropriate correlation statistics must be reported as well

2. Usefulness of statistics
   a) Identify 4 research questions of your own that you may be able to use the 2 methods highlighted in the project to answer (2 questions per method). These questions must be relevant to the group members’ field of study.

3. Comment on overall quality of your project/group experience.
   a) Identify at least 2 positives of your experience
   b) Identify at least 2 negatives of your experience

**Format**

• I expect that each group will prepare appropriate visual aids (i.e., PowerPoint slides, posters or typed transparency films)

• The presentation should be approximately 15-20 minutes.

**For Audience**

• You are expected to ask questions and provide comments after the presentation
Grading
In addition to instructor grading,

- This presentation is worth 3% of your overall grade.
- Each student in the audience will grade the group’s presentation (see Group Rubric on Moodle site)
- Each group member will grade their team members’ overall contribution to the project (see Group Member Scoring Sheet on Moodle site)

You will be graded on style and against every other group, so having the basics are not enough to ensure a top notch score.

Cover Sheet
(Used for Review and Final Assessment)

Team Member Names (Print):

__________________________________________  _______________________________________

Content (72)

<table>
<thead>
<tr>
<th></th>
<th>6</th>
<th>4</th>
<th>2</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction: Questions of interest</td>
<td>Good</td>
<td>Fair</td>
<td>Poor</td>
<td>Missing</td>
</tr>
<tr>
<td>2. Description of data collection/source</td>
<td>Good</td>
<td>Fair</td>
<td>Poor</td>
<td>Missing</td>
</tr>
<tr>
<td>3. Analysis Method correctly identified</td>
<td>Good</td>
<td>Fair</td>
<td>Poor</td>
<td>Missing</td>
</tr>
<tr>
<td>4. Numerical summary of data with support</td>
<td>Good</td>
<td>Fair</td>
<td>Poor</td>
<td>Missing</td>
</tr>
<tr>
<td>5. Discussion of robustness to any violations</td>
<td>Good</td>
<td>Fair</td>
<td>Poor</td>
<td>Missing</td>
</tr>
<tr>
<td>6. Graphical description of data with support</td>
<td>Good</td>
<td>Fair</td>
<td>Poor</td>
<td>Missing</td>
</tr>
<tr>
<td>7. Assumptions identified and checked</td>
<td>Good</td>
<td>Fair</td>
<td>Poor</td>
<td>Missing</td>
</tr>
<tr>
<td>8. Hypothesis test conducted with support</td>
<td>Good</td>
<td>Fair</td>
<td>Poor</td>
<td>Missing</td>
</tr>
<tr>
<td>9. Relevant Output incorporated and identified</td>
<td>Good</td>
<td>Fair</td>
<td>Poor</td>
<td>Missing</td>
</tr>
<tr>
<td>10. Conclusion: Results related to questions</td>
<td>Good</td>
<td>Fair</td>
<td>Poor</td>
<td>Missing</td>
</tr>
<tr>
<td>11. Conclusion: Additional comments provided</td>
<td>Good</td>
<td>Fair</td>
<td>Poor</td>
<td>Missing</td>
</tr>
<tr>
<td>12. Conclusion: Future work/ additional questions</td>
<td>Good</td>
<td>Fair</td>
<td>Poor</td>
<td>Missing</td>
</tr>
</tbody>
</table>

Organization (18)

<table>
<thead>
<tr>
<th></th>
<th>6</th>
<th>4</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Presentation (outline, typed, length)</td>
<td>Good</td>
<td>Fair</td>
<td>Poor</td>
</tr>
<tr>
<td>2. Logical flow of ideas</td>
<td>Good</td>
<td>Fair</td>
<td>Poor</td>
</tr>
<tr>
<td>3. Spelling, punctuation, grammar</td>
<td>Good</td>
<td>Fair</td>
<td>Poor</td>
</tr>
</tbody>
</table>

Additional Comments:
<table>
<thead>
<tr>
<th>Score</th>
<th>Content</th>
<th>Organization</th>
<th>Presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Is well thought out and supports the solution to the question Reflects application of critical thinking Has clear goal that is related to the analytic topic Is well supported Is accurate</td>
<td>Information is clearly focused in an organized and thoughtful manner. Information is constructed in a logical pattern to support the conclusion.</td>
<td>Multimedia is used to clarify and illustrate the main points. Format enhances the content. Presentation captures audience attention. Presentation is organized and well laid out.</td>
</tr>
<tr>
<td>3</td>
<td>Is well thought out and supports the solution to the question Has application of critical thinking that is apparent Has clear goal that is related to the topic Is pulled from several sources Is accurate</td>
<td>Information supports the solution to the question.</td>
<td>Multimedia is used to illustrate the main points. Format is appropriate for the content. Presentation captures audience attention. Presentation is well organized.</td>
</tr>
<tr>
<td>2</td>
<td>Supports the solution to the question Has application of critical thinking that is apparent Has no clear goal Is pulled from a limited number of sources Has some factual errors or inconsistencies</td>
<td>Project has a focus but strayed from it at times. Information appears to have a pattern, but the pattern is not consistently carried out in the project. Information loosely supports the conclusion.</td>
<td>Multimedia loosely illustrates the main points. Format does not suit the content. Presentation does not capture audience attention. Presentation is loosely organized.</td>
</tr>
<tr>
<td>Score</td>
<td>Content</td>
<td>Organization</td>
<td>Presentation</td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>--------------</td>
<td>--------------</td>
</tr>
</tbody>
</table>
| 1     | Provides inconsistent information for solution of problem  
      Has no apparent application of critical thinking  
      Has no clear goal  
      Is pulled from few sources  
      Has significant factual errors, misconceptions, or misinterpretations | Content is unfocused and haphazard.  
      Information does not support the conclusion to the question.  
      Information has no apparent pattern. | Presentation appears sloppy and/or unfinished.  
      Multimedia is overused or underused.  
      Format does not enhance content.  
      Presentation has no clear organization. |