

MTH 133
MIDTERM PROJECT, CHAPTERS 1-4
First submission due: June 23rd

This is the mid-term project. You will be exploring the relationships between smoking and/or drinking and various types of cancers. Your task will be to use the things that we have learned up until this point and apply them intelligently. Your assignment is to explore the relationship of the variables in your data set, draw meaningful conclusions and construct a professional looking document demonstrating your findings.

Your paper should be structured in the way that most scientific papers are:

Title

Author(s)

Abstract: Give a short (50-200 words) summary of what the purpose of the paper was, what you did, what you found and what your conclusions were. This is usually the last thing you write.

Introduction and Data: Here, explain the motivation behind your study. Give some history (cite your sources!) and explain why we should even be worried about cancer at all. Why should we be worried about smoking? If nobody smokes, who cares? The next section should provide support for why we should care about smoking, drinking and cancer.

Data: The section is where you DESCRIBE your data set. That is, you should present tables and graphs similar to what we did for the school project. You should present tables of descriptive statistics, plots to see how cancers stack up against one another and plots of smoking and/or beer consumption. The point here is to explain to the reader EXACTLY what it is that you're working with. It's an accounting of the data at hand, and also lends support for why the reader should even be concerned with the subject matter.

Methods: Next, you should explain your methodology. What did you do? What techniques did you use? Assume your reader knows absolutely nothing. Explain terms and ideas here. You cannot ever explain too much. Remember, in the real world people just are not that scientifically knowledgeable. Explaining things to them in simple terms (without being condescending) will get you far and can make the difference between your work being noticed or ignored. This is obviously a regression project. Therefore, you should include a discussion of regression, what it is and why it's used.

If you had to do anything to the data to get it to work in your project, be sure to mention it here.

Results: What did you find? Here, you should present tables and graphs of your results with detailed explanations. You are doing a regression project. Therefore, you will need to present scatter plots and regression stats. Try to make any tables and graphs as efficient as you possibly can. That is, make the whole thing as readable as possible.

Conclusions: What can you conclude from your results? How certain are you? What did you learn about the link between smoking/drinking and the various cancers? How could this study have been improved? What were its limitations? Where could have it been inaccurate? What kinds of problems were there in the methodologies? What kinds of problems existed in the data or in the data collection methods? Remember, science is not absolute nor is it ever perfect. There is always room for improvement.

Sources: Please try to limit your use of outside sources. Spend your time on the project itself.

Data: Include your raw data set.

You are allowed to work on this in teams (max 2-3 people) and turn your final papers in as a team. Of course, you can work on your own if you like.

A couple of notes:

1. Avoid stating absolutes. It is easy to say that “smoking causes lung cancer” but the truth is that we do not for absolute certain that smoking causes lung cancer. We know that people who smoke have higher rates of lung cancer. Therefore, we conclude that it is highly probable that something about smoking cigarettes raises ones risk for lung cancer. Is it the smoke? The filter? The lack of oxygen when one inhales? The art on the package? Do smokers generally engage in other kinds of risky behavior? There are many possibilities. Thus, in science, we avoid stating things in black and white terms, since the reality is often much more complex. Therefore, we look to likelihood and probable events. “It is very likely, given these observations, that smoking cigarettes raises one’s risk of developing lung cancer”.
2. Before you start, examine your data. What does it look like? Are there some potential problems? Weird values that are likely incorrect? Missing values? Make a box plot. Determine the outliers. Remember, you want to present as realistic a picture as you can. Outliers will

influence your results negatively. You may want to compare your results with the outliers in and with them left out and see how different they are. If you remove outliers, state this in your paper and let the reader know which ones you removed and why. Also, try to theorize why these particular points are unique and state this. Do not ever expect data to be “ready to go”. 99.9% of the time you have to do something to make it work.

3. I have included a paper that a student wrote a couple of terms ago to give you an idea of what is expected of you.
4. Remember, if I get your paper and I write more than you do, something is seriously wrong. This assignment is GRADED. That is, good work will be rewarded. That other kind of work will not.