**Some Project Writing Hints**

**Keep in mind the purpose of this paper** The purpose of this paper is NOT to dazzle me with your knowledge of statistical methodology – the meaning of a p-value or how to write down the null and alternative hypotheses (do this on the exams!) The purpose of this paper is to show me you know how to USE statistics to make a persuasive quantitative argument and to communicate results to other scientists with a similar level of statistics background.

The audience for the paper is a scientist interested in your research question. This scientist has knowledge of statistical techniques and interpretation.

**Remember the BIG picture of your paper:**

The report should include the following LABELED sections:

**Introduction**: What are we doing? Why are we doing this?

**Methods**: Where do the data come from? What did we do to it?

**Results**: What did we learn from the data? Tell me a story *supported* with statistics, not *about* statistics. This includes figures and tables.

**Discussion**: Why do we care about what we found? How generalizable are the results? What other questions would we like to answer?

**Appendix:** (a separate document) Include all your R code and output with notes so I can follow along. Any other work you did that isn’t directly applicable to the final report goes in here too. (extra graphs, more details about missing data or evaluating whether assumptions are satisfied).

Here is **a results writing strategy**:

For the first paragraph, start with simple descriptions: *50% of the subjects are male, 20% smoke cigarettes, and 40% are overweight.* Often, studies will include demographic or other baseline characteristics about the subjects in the study, even if they are not the main variables of interest. This baseline descriptive information can be helpful for determining generalizability of the study (for instance, if a random sample is not used what population is this study representative of?) Sometimes these baseline characteristics are just included in a table (e.g. mean/SD or proportion for each variable) and only a couple important ones are discussed in text.

In a second paragraph, discuss the relationships: *Males are twice as likely as females to smoke cigarettes and 3 times as likely to be overweight.* Try to make it flow together nicely: report things in an order that makes sense and that emphasizes your point.

Think about tables or figures that support this story. If you’re making the same comparison across many categories (e.g. the proportion of people with heart disease for 6 categories of BMI), consider reporting the general pattern, or pointing out a few important proportions in the text, and putting the rest in a table or graph. Insert references (parenthetically, or “as seen in Figure 1, ....) into the text.

THEN add p-values and/or confidence intervals in parentheses. If you’ve put results in a table, consider adding the p-value or confidence interval to the table as well, or perhaps instead of in the text.

Other common issues tend to be in **the discussion section**, particularly

* Identify any limitations of your study. Discuss the potential impact of such limitations on the conclusions.
* Identify strengths and weaknesses of your analysis.

These are two separate issues. *Limitations of the study* often have to do with study design... is it generalizable? If not, why not? Are there potential measurement errors or other biases? (e.g. students feel social pressure to lie about drug use, sex etc when asked in survey). Are you trying to measure something that isn’t easily defined (e.g. “Very happy”, “Somewhat happy”, etc)?

*Strengths and weaknesses of the analysis* are statistical (or possibly study design) decisions that you made which may strengthen or weaken your potential conclusions.

Example of a strength: Very large sample size means very high power (especially interesting if you fail to reject null hypothesis... this means it has very low probability of a type II error... and even though we can’t officially *accept Ho,* this is stronger evidence that Ho may really be true.)

Example weaknesses:

* Had to combine two or more categories of responses into one in order to use your chosen statistical method (e.g. had to make the explanatory variable binary for two-prop-z-test or t-test), which means you now can’t make inferences about those groups separately.
* We treated a variable as continuous even when it was recorded as (1, 2, 3, 4, 5, 6 or more). The numeric value of the last group is questionable (would analysis change if we had replace it with the numeric value of 7 instead of 6? 6.5? 8? which of these is most representative of the answers “6 or more”... we really don’t and can’t know).

These limitations and weaknesses are natural suggestions for further research. When suggesting further research, feel free to make suggestions of things you don’t know how to do, or you can’t do because you don’t have the necessary data/information.