**QTM 100 – Final Project Submission**

**Due date: Monday, May 4th, before 11:59 pm.**

**Introduction**

The final project submission is intended to function as a short, polished research paper (in some fields this is known as a research letter). Be aware that the final submission forms 90% of your project grade, which in turn forms 20% of your final QTM 100 grade. For this reason, we expect extremely high-quality work and will grade accordingly.

**Instructions**

Formulate 3 research questions (you should use the research questions from your preliminary submission as a guide). They may, in fact, be the same as in your preliminary submission so long as they conform to the following rules:

1) One research question must address the potential association between your **dichotomous** response variable and one of the **categorical or dichotomous** explanatory variables.

2) One research question must address the potential association between your **numerical** response variable and a **categorical or dichotomous** explanatory variable.

3) One research question must address the potential association between your **numerical** response variable and one of the **numerical** explanatory variables (an ordinal variable may be acceptable – check with your TA).

Write a **650 word** research letter and submit your final project individually on Canvas as a MS Word document or PDF (do not use Pages). Your research letter should include (1) an introduction which states the significance of your topic, your research questions, and your hypotheses; (2) a methods section which details the data cleaning and recoding steps as well as the statistical tests you used; (3) a results section which plainly states your findings; (4) a discussion section where you discuss whether your research questions were answered, hypotheses supported (or not), and the importance of your findings; and (5) appendices which contain your tables of results, bivariate plots, and R code. The rubric and more detail for each section are provided below.

**Your reader**

Research letters are not written for TAs or graders. Research letters are written for people who care about the topic. You should assume that your reader is aware of the dataset (but not an expert) and is well-versed in statistics (but not necessarily R). Do not assume that your reader has the level of knowledge of the dataset or R that your TA does.

**Late policy**

You will be deducted 10% for each calendar day after the due date. Thus, submissions received at 12:00 AM on Tuesday, May 5th will receive a maximum score of 90/100 points. Submissions received on Wednesday will receive a maximum score of 80/100. And so on.

Please read below the detailed grading rubric and detailed explanation.

**QTM-100 Project Grading Rubric**

|  |  |
| --- | --- |
|  | **Points** |
| Formatting (see next page) | 3 points |
| Title (be descriptive but succinct) | 1 point |
| Overall quality | 3 points |
| Holistic evaluation | 4 points |
| **Introduction** | **17 points** |
| Introduce your topic and state why it’s important to study | 2 points |
| State your research questions, rationale, and hypotheses | 15 points |
| **Methods** | **18 points** |
| Introduce your dataset and variables used | 2 points |
| Describe all data cleaning and recoding procedures used, including one significant recoding | 6 points |
| A description for each statistical test used | 10 points |
| **Results** | **4 points** |
| A few sentences summarizing results | 4 points |
| **Discussion** | **15 points** |
| A discussion of your findings | 10 points |
| Study limitations and brief conclusion | 5 points |
| **Appendix** | **35 points** |
| Appendix A: Tables of results | 20 points |
| Appendix B: Bi-variate plots of your explanatory vs response variable(s) | 9 points |
| Appendix C: Submit well-commented R code | 6 points |
| **Total** | **100 points** |

**Rubric explanation**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Overall**

**Formatting [3 points]:** Your final project should be typed in 12-point Times New Roman font, double-spaced, 1-inch margins.

The main text of your research letter (i.e. from introduction to discussion/conclusion, not including appendices) must not exceed 650 words, one standard for research letters. ***Nothing past 650 words will be read or graded.***

Appendices begin on new pages; tables belong on new pages (i.e. not on the same page as the last page of your text). **Use the separate citation guide to format all citations.**

At no point in the main body of the paper should R code or R variable names be given.

**Title [1 point]:** Be descriptive.

**Overall quality [3 points]:** You should write in full sentences – no bullet point type sentence fragments. You should write clearly and concisely. Anything that may be omitted without the loss of clarity, must be. Other things equal, the less lines of text, the better. You may, but are not required to, use the 1st person (I/We). Your paper should be readable and free of grammar, spelling, and punctuation mistakes. If you use abbreviations, define them first: “… used the 2012 Current Population Survey (CPS).”

**Holistic evaluation [4 points]:** Your research letter must be well-structured, cohesive, and readable. This means that your reader should be able to read through your research letter once (twice maximum) and understand what you did, what you found, and why it matters. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Introduction**

**State the topic and why it’s important [2 points]**: There is no point in finding uninteresting or useless associations between variables. Your research should matter – explain why it matters in the introduction.

**State your research questions + rationale + hypotheses [15 points]**: State each research question with sufficiently compelling or intuitive rationale for why you think the two variables are associated. You will be deducted points if you:

1. have no rationale.
2. if your rationale is unsupported by previous studies or, indeed, common sense. (e.g. age is the response variable for the explanatory variable household income)
3. if your rationales are circular. This, for instance, is a bad rationale: hospital revenue is determined by hospital type because hospital type determines hospital revenue.

For your hypotheses,a precise verbal description is fine.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Methods**

**Methods [18 points]:** Describe your dataset and the variables you used. Do not give mysterious R variable names – give the name of the concept or scale. Again: at no point in the research letter should the reader see R code or R variable names. Describe all data cleaning steps and recoding procedures you used (e.g. categories you combined). If you recoded unreasonable values, say so; also state why these values are unreasonable. It is not enough to say “I cleaned the variable” without describing *how* you cleaned it.

Just like in the preliminary submission, one of your variables should have required significant recoding (such as making a numerical variable dichotomous or categorical). Make sure that your discussion of this recoding is clear and concise. It should, at a minimum, tell the reader what the variable was like before, *and* how you modified it. Justifying your choice with a source is a great way to strengthen your argument.

Name each statistical test you used for each research question. If you have many levels of a factor variable, you will likely want to group some of them. If you did so, explain why; if you did not, explain why.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Results**

**Results [4 points]:** Summarize your results in clear, concise sentences. Include p-values and test statistics in parentheses after each result. Example: (χ2 = 2.35, p = 0.04). You may refer the reader to your table in a parenthetical note (e.g. see Table 1 in Appendix A). Include all information that would be relevant to a reader who is interested in understanding your results. For instance, if you run a t-test and find significant results, tell the reader which mean was higher and by how much (or give both means).

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Discussion**

**Discussion [15 points total]:** Do not merely restate your results. Discuss the importance of your findings, e.g. whether your findings agree with your hypotheses, whether the findings are practically significant (as opposed to statistically significant), the implications of your results, etc. Synthesize your results – show how any of your results are related to your other results. Discuss the limitations of your study. Mention Type I or Type II errors and how they may apply to your findings.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Appendix A**

**[20 points]:** Tables of results

You should produce 4 tables:

1. a table with descriptive statistics (e.g. mean (sd)/N (%) and n size of ALL variables you used)
2. A table for the results of research question 1
3. A table for the results of research question 2
4. A table for the results of research question 3

Round all results to 2 or 3 decimal places (pick 2 or 3 and be consistent)

Each table should have a title (e.g. Table 1. Association between X,Y, and Z. XX dataset, n = XXX) and be sequentially numbered.

Cute, creative, or stylistic tables are discouraged (no color-coded columns, irregular fonts, etc.). Use black/white tables with minimal horizontal and vertical lines. Construct the tables in Word or in Excel.

Each appendix begins on a new page. Do not start an appendix on the same page as information from a previous section.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Appendix B:**

**[9 points]** Bi-variate plots

This appendix begins on a new page. You will present three plots for the three research questions.

Format your plots and upload **at most** 2 per page (1 per page is fine). As always, label axes and factor levels. Where applicable, include the unit of measurement in the axis label (e.g. Household income in USD). Do **NOT** include the title in the plot itself – as with your tables, write out the title in your document (not in your code) and number them sequentially. This, in particular, solves the problem of insufficient space for a title. If there is anything in a plot or table that needs to be explained, make a note below the object.

When you use factor variables, ensure that they are ordered correctly so that a trend is easily discernable. You will be deducted points if your plots are not easily understandable.

Use marginal proportions for barplots. This makes clearer the association between variables which have significant differences in the sizes of groups. Showing barplots of raw frequency adds nothing that your table doesn’t already do.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Appendix C**

**[6 points]:** R code

This appendix begins on a new page. Your R code should be clearly labeled with your name and lab number. It should be well-commented and formatted – at a glance, a reader should be able to locate statistical analyses/recoding code. Ensure your code is readable (e.g. put spaces between chunks of code). Your comments are as much for your reader as it is for you – this means that the standard of ‘readability’ depends on whether a reader finds it easy to read as opposed to what YOU find easy to read.

*You should also submit your R code as a separate R file.*

**Tips from TAs**

**General**

Write your final submission early – you will find it difficult at first to present everything within the 650 word limit. By “early” I mean you should have a first draft a week ahead of the deadline.

If you have too few observations for a particular category in a factor variable, recode the variable so that you have sufficient observations.

Recheck the coding/cleaning of variables (do not skip the basics).

Ensure you are using proportions rather than frequency for plots.

Run your script one last time before submitting the project to ensure that you have not inadvertently mangled your code.