

Data Exam Rubric

36-402, Spring 2021

Each row is one requirement; you should strive to meet the description in the Excellent column. We will grade by adapting this rubric to the specific data exam questions.

Overall Requirements

Contents

	Excellent	Needs Revision
Length is appropriate	Report is in 12-point font with 1.5 line spacing and 1-inch margins. Report is ≤ 12 pages, including figures	Report is too long. Pages after 15 will not be graded
No R code	No R code is presented in the report, only relevant results	Chunks of R code are shown directly in the report
No R output	No R output is presented in the report; results are formatted as tables or plots	Raw R output is pasted directly into the report
No math	Report explains results in words, without mathematical formulae or derivations	Report contains mathematical derivations

Figures and Tables

These criteria apply to all figures and tables you present in your report.

	Excellent	Needs revision
Legibility and design	Figures are clear, simple, legible, and attractive	Figures are hard to read; fonts are too small; some graphics are blurry or squished
Labeling	All legends and axes are clearly labeled, including units	Labels missing or use raw variable names instead of descriptive text
Captions	Figures have captions explaining their purpose	Figures have no captions or their purpose is unclear

Writing and Style

	Excellent	Needs revision
Grammar and style	Grammar is correct and style is appropriate to audience	Grammatical or spelling errors make text hard to read
Precision	Text is clear and	Text is unclear or excessively

	unambiguous	wordy
Uncertainty always given	Results presented with clear measures of uncertainty, such as confidence intervals or standard errors	Many results given without any uncertainty
Sourcing (if applicable)	If information is used from outside sources, references are clearly given; quotations clearly mark any verbatim text from sources	Text or figures used without attribution. Warning! Using material from other sources without <i>clear</i> attribution is an academic integrity violation.

Requirements for Specific Sections

Your report must contain sections with the following headings, as shown in the template we provide. Requirements for each section are shown below. **You must mark these sections in Gradescope**, just like you marked each problem on homework.

Introduction

	Excellent	Needs revision
Describes research questions	Introduction sets out questions to be answered	Unclear what questions you intend to answer
Briefly summarizes results	Introduction briefly mentions the results of the analysis	No results mentioned or results are unclear

Exploratory Data Analysis

	Excellent	Needs revision
Data is clear	Meaning of relevant variables given with units	Important variables not explained or have no units
Univariate checks	Distributions of variables checked (when they are); outliers mentioned	Important aspects of variables not checked
Multivariate checks	Relationships between predictors and response checked with appropriate tables/graphs; any trends described	EDA does not explore the relationships that will be included in the model
Only necessary graphics	Only the necessary figures are shown in the report	Many figures don't show anything relevant

Initial Modeling & Diagnostics

	Excellent	Needs revision
Models constructed	Report shows all models	Some or all models missing

	requested in the prompt	
Models explained	Report explains choices made in model building, such as how variables are coded or transformed	Unclear why or how the models were made
Variable selection	Choice of variables to include in the models is explained, e.g. by saying what confounders or interactions had to be included	Text does not say why variables were chosen
Diagnostics checked	Diagnostics and goodness-of-fit checked for each model	Unclear if model assumptions are checked and how well they fit
Choice justified	Text clearly explains why final models was chosen over alternatives, e.g. with a test or through diagnostics	Unclear why chosen model is better
Problems noted	Caveats and problems are noted and their potential effect on results explained	Problems are hidden and their potential effect on results ignored

Model Inference and Results

	Excellent	Needs Revision
Results answer questions	Statistical results answer the questions asked	Statistical results do not answer the substantive questions
Statistical results clearly presented	Tests and estimates are presented clearly and accurately	Explanation misstates what the results show or misinterprets the statistics
Statistical results correct	Statistical methods correctly implemented	Errors in code or math mean the results are incorrect

Conclusions

	Excellent	Needs Revision
Conclusion presents results	Summarizes conclusions presented in report	Describes conclusions not justified or described in the body of the report, or doesn't say conclusions
Conclusion notes limitations	Conclusion notes any limitations on conclusions and what could be done to address these	Conclusion ignores limitations of results and is too confident

