Exam 1: MN4912 AY23Q3

11 pages total, due April 19 at 23:55 California time.

This Exam assesses your ability to use data to learn about relationships among variables, and communicate your analysis.

Data set

All questions are based on the `loans_data.csv` data set (the .xlsx version is identical), with variables described in `loans_data_description.docx`, attached to this question and available in Resources/Data.

Research questions (RQ)

There are seven Research Questions (RQs) about the relationships among variables in the data set.

Student questions (SQ)

For each RQ, you need to respond to one or more Student questions (SQs), for a total of 16 questions. Many of the SQ's will ask you provide a chart, a table and/or a text summary responding to the RQ.

Chart When asked for a chart, it should be:

- **appropriate** to the RQ, including conditioning on the correct variable
- **complete**:
  - it should be clear from the chart what variables are represented and how they are measured (units if appropriate), and
  - the reader should be able to read axis ticks and labels
- **correct**, using the right data, with the right labels, e.g. don't label an axis "percent" if it's showing proportion.

Table When asked for a table, it should:

- show appropriate summary statistics, conditioned on each level of the explanatory variable,
- make it clear what variables are represented, including what the units are, and what each value in the table means, and
- it should be correct.

Text summary Unless otherwise specified, the text summary should:

- be one to three-ish complete, grammatical sentences,
- give a qualitative summary of the relationship, including strength and direction (if appropriate), and
- give a quantitative summary of the relationship, including at least one, and generally more than one appropriate summary statistics that illustrate the qualitative summary

Rules and recommendations

Rounding: round proportions to the nearest 1% or 0.001. When in doubt, round other values to the nearest 0.1.

You may use any other resources provided in this class, including spreadsheets provided before the exam. It **is allowed** to search for general information like definitions and how to do things in R or Excel.

It is **not allowed** to search for anything specific to the questions on the exam, or otherwise seek any help, except for clarification from professor. Please email me with any questions. I will try to respond quickly.

Please do not share the content of this exam with anyone. The only exception is you can discuss the exam with each other but not with future students.

RQ 1 Does default rate depend on the purpose of the loan?

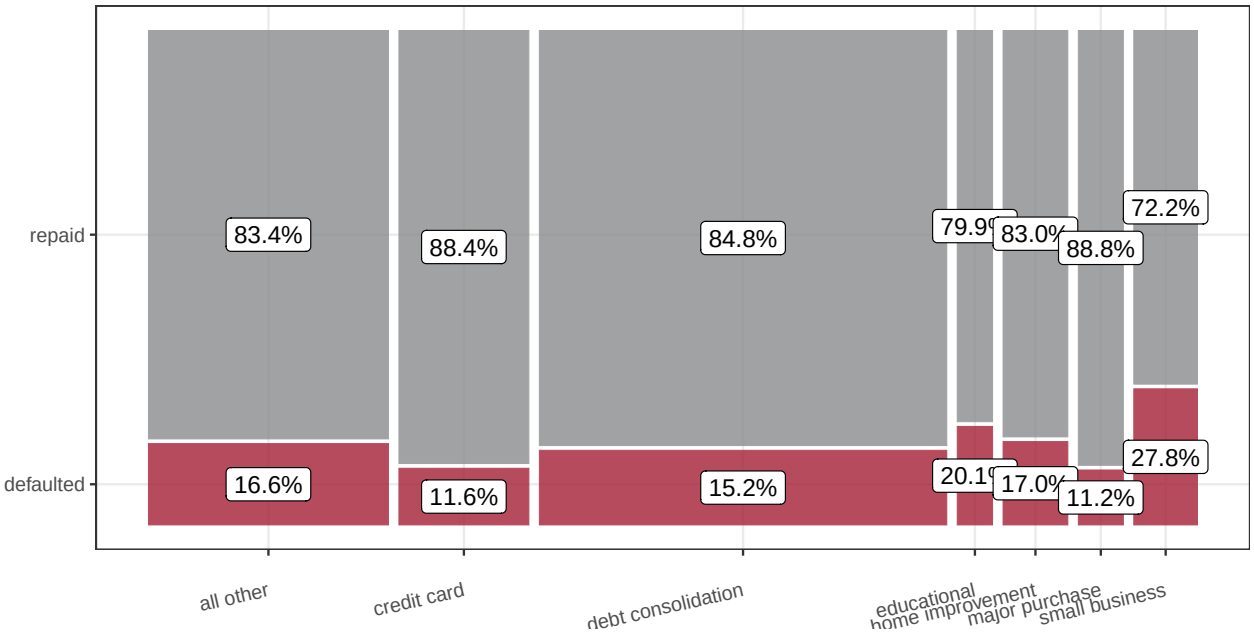

Figure 1: Default rate as a function of the loan's purpose.

SQ 1.1 Give the variables represented in each role in Figure 1:

explanatory: _____ response: _____

SQ 1.2 Text summary

Fill in the blanks and circle the correct options in parentheses to complete the text summary:

> Default rate differs somewhat by loan purpose. The highest default rate was for _____ loans: _____% of them defaulted. The lowest default rate was for _____ loans: _____% of them were repaid in full. Perhaps surprisingly, the default rate for credit card loans was _____ percentage points (higher / lower) than the default rate for educational loans. In other words, educational loans defaulted at a rate _____% (higher / lower) than credit card loans.

RQ 2 Does the debt-to-income ratio seem to influence the interest rate borrowers receive?

SQ 2.1 For this research question, what are:

a. the variables?

b. variable types?

c. variable roles?

d. appropriate type of chart?

e. appropriate summary statistic(s)?

SQ 2.2 Chart

SQ 2.3 Text summary—give a qualitative summary plus one summary statistic

RQ 3 How is interest rate related to FICO score?

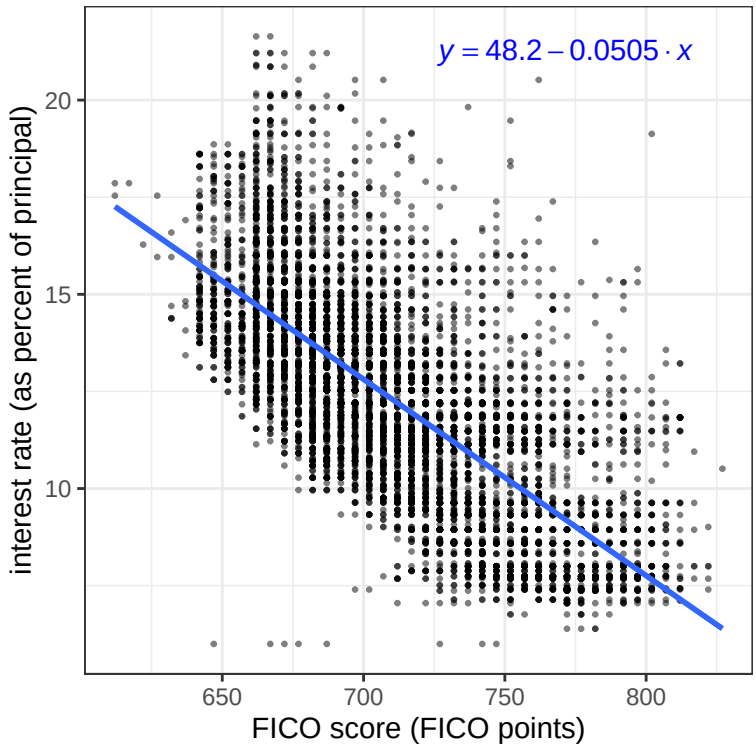

Figure 2: Chart visualizing the relationship between FICO score and interest rate on the loan. A linear model is shown in blue, together with its formula.

SQ 3.1 Text summary—with qualitative and quantitative summaries (use the slope of the line)

Hint: the term *percentage point* should appear in your answer

RQ 4 Can the borrower's annual income be used to predict a loan's monthly installment amount?

SQ 4.1 For this research question, what are:

a. the variables?

b. variable types?

c. variable roles?

d. appropriate type of chart?

e. appropriate summary statistic(s)?

To answer this research question, you could use either annual income or the natural log of annual income as the predictor. Assume you would like to use a linear model.

SQ 4.2 Which variable would make a better predictor and why?

Circle the best answer a. or b. and explain in the box below.

a. annual income

b. natural log of annual income

and why?

RQ 5 Does the number of times the borrower has been 30+ days late provide an indication of whether they will default?

SQ 5.1 For this research question, what are:

a. the variables?

b. variable types?

c. variable roles?

d. appropriate type of chart?

e. appropriate summary statistic(s)?

SQ 5.2 Chart

SQ 5.3 Table

SQ 5.4 Text summary

RQ 6 Does the purpose of a loan seem to affect the interest rate?

SQ 6.1 For this research question, what are:

a. the variables?

b. variable types?

c. variable roles?

d. appropriate type of chart?

e. appropriate summary statistic(s)?

SQ 6.2 Table

SQ 6.3 Text summary

RQ 7 Is this data set consistent with the conventional wisdom regarding interest rates and default risk?

The conventional wisdom is that lenders charge higher interest rates for higher-risk loans, i.e. those that are more likely to default.

SQ 7.1 Based on Figure 3, is this data set consistent with the conventional wisdom?

Why or why not—as a text summary, both qualitative and quantitative, and include metadata

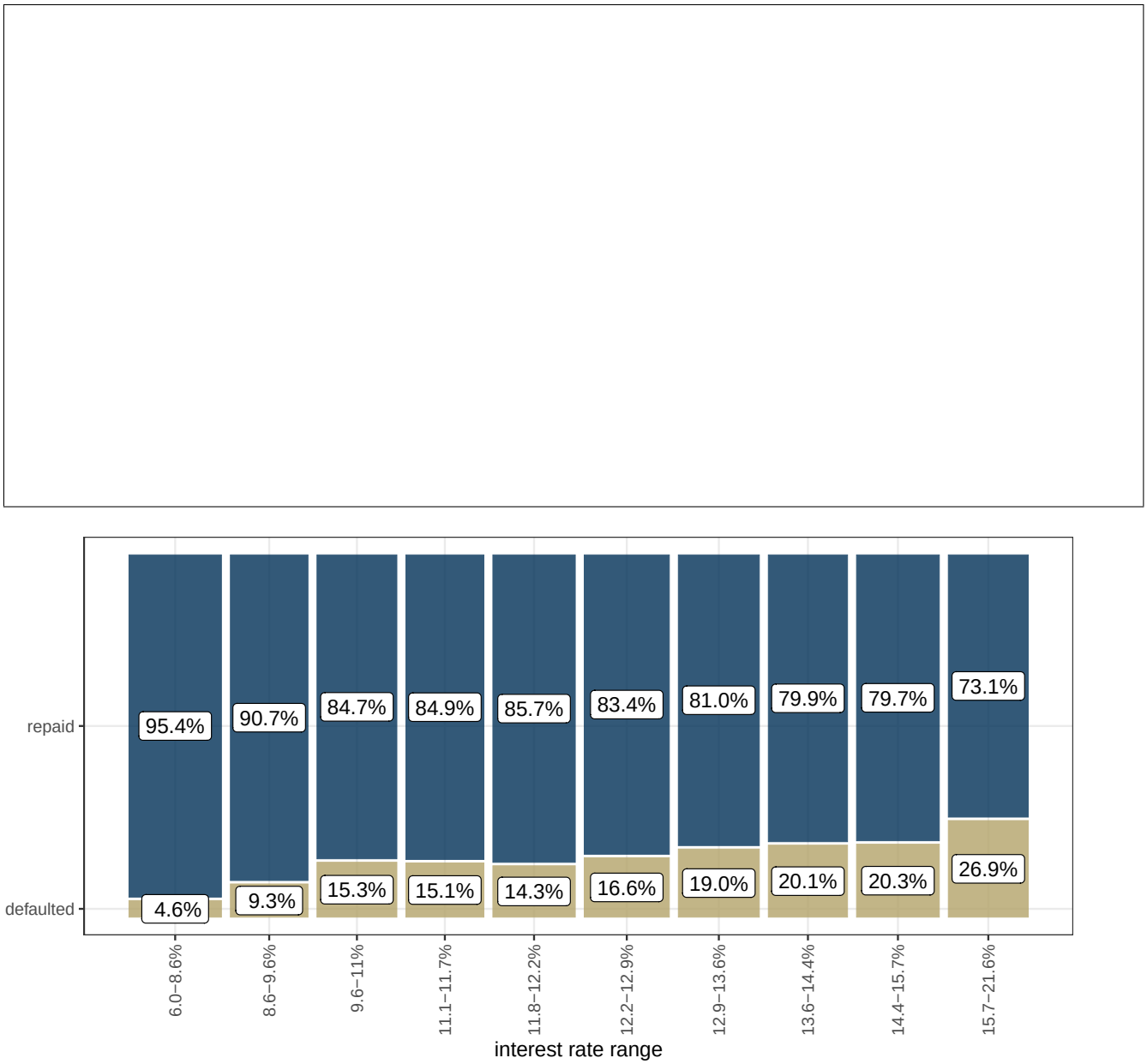

Figure 3: A chart visualizing the relationship between interest rate and default.