Minute Paper after Class 03 Feedback

Thomas E. Love, Ph.D.

2023-01-25

Table of contents

1 R Setup

```
knitr::opts_chunk$set(comment = NA)

library(tidyverse)

theme_set(theme_bw())
```

2 Sample

n = 55 of 56 students responded by the NOON 2023-01-25 deadline. Thanks!

3 Where are you in Jeff Leek's small book "How To Be A Modern Scientist"?

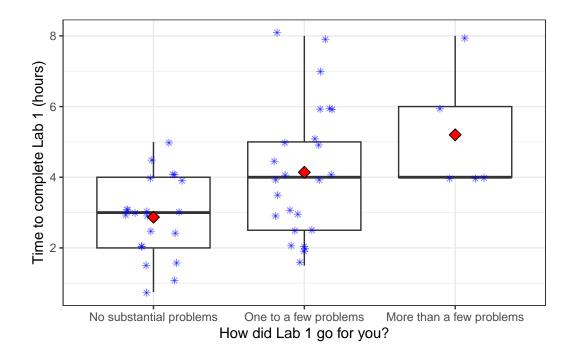| Students	Response
10	I've not yet purchased the book. I need to do that soon.
22	I've purchased the book, but not started reading it yet.
20	I've read some, but not all, of the book.
3	I've read all of the book.

We want you to have finished reading the book by class time on 2023-02-02.

4 How did Lab 1 go for you?

Students	Response
4	I was unable to complete the Lab on time.
5	I had more than a few substantial problems but eventually completed the Lab on time.
25	I had between one and a few substantial problems but eventually completed the Lab on time.
21	I didn't have substantial problems with the Lab, and completed it on time.

5 About how long (in hours) did it take you to complete Lab 1?

omitting students who did not complete the Lab

```
ggplot(min03, aes(x = lab1_went, y = lab1_time)) +
  geom_boxplot() +
  geom_jitter(width = 0.2, shape = 8, col = "blue", alpha = 0.5) +
  stat_summary(fun = "mean", geom = "point",
               shape = 23, size = 3, fill = "red") +
  labs(x = "How did Lab 1 go for you?", y = "Time to complete Lab 1 (hours)")
```

6 Did you seek out help from us for Lab 1?

Students	Hours	Campuswire	Slides/Videos	Notes
3	No	No	No	No
8	No	No	No	Yes
5	No	No	Yes	No
19	No	No	Yes	Yes
2	No	Yes	No	No
1	No	Yes	No	Yes

Students	Hours	Campuswire	Slides/Videos	Notes
5	Yes	No	No	No
7	Yes	No	No	Yes
2	Yes	No	Yes	Yes
1	Yes	Yes	No	No
2	Yes	Yes	Yes	Yes
Total	**17**	**6**	**28**	**39**

Available responses (check all that apply) were:

- (**Hours**) I went to TA office hours.
- (**Campuswire**) I asked for help on Campuswire.
- (**Slides/Videos**) I rewatched part or all of Dr. Love's class recordings and/or reviewed his slides.
- (**Notes**) I read part of Dr. Love's Course Notes to help me answer Lab 1's questions.
- I didn't do any of these things.

7 What was the most important thing you've learned during 432 so far?

(*edited and grouped by TEL; some responses dropped, others arranged from more to less common*)

- Troubleshooting / Using Quarto

  - Some of the differences between R Markdown and Quarto.

- The use of survey weights, including how to use the **survey** package to accomplish a few tasks with NHANES data.
- Refresher/reminder of material from 431, including

  - data ingest and cleaning,
  - data visualization development and interpretation,
  - inference on means and rates with confidence intervals, and
  - linear regression models.

- Validating a predictive model by splitting into training and testing samples

  - Different summary statistics for in-sample (AIC, BIC, raw and adjusted R^2, σ) and out-of-sample (validated R^2, mean absolute prediction error, median APE and maximum absolute error) model comparison/evaluation

- Logistics of the Course

- Update R and RStudio to the latest version
- Keep up with the Calendar
- I do not need to be intimated by this class, we are set up for success!
- Resources for Getting Help

- The difference between clinical significance and statistical significance
- Reminder of how to use gt() effectively
- Honestly - nothing yet, but i know that will change soon.
- More efficient ways to transform variables with functions like starts_with.
- Using the ~ one function for assigning numbers

8 Your Questions: My Answers about 432

I edit these questions a bit for clarity, and answer most, but not all, that are asked. Some receive individual answers via email, and others I don't really know how to answer, so I don't. If I didn't answer your question, you are welcome to try again, perhaps on Campuswire, in person or during TA office hours, or just with a rephrased version in the next Minute Paper. Those of you without questions this week, try not to make a habit of never asking a question here, but don't worry about not asking anything if you really don't have any questions.

8.1 About Statistics / Data Science Issues

- When analyzing these big data sets, is it more appropriate to ask an interesting question, even if the model that comes of it isn't very effective, or to explore the data to identify potential relationships that would be interesting to model?

  - It's a false choice. Both approaches can be helpful, and complimentary. I don't know why you wouldn't try to do both.

- Is there a percent difference in AIC or BIC between models that you find significant?

  - Smaller values of these measures are better. Don't attribute a "Yes/No" categorization in the form of "Significantly better" if you don't have to do so. AIC and BIC are just two of many tools to evaluate the quality of a model. Use several such tools.

- I am confused about the rationale of weighted analyses in the first place. Is this specific to surveys?

  - Our topic in class 3 was **survey weighting**. We haven't yet talked about other sorts of weighting.
  - By survey weighting here, I mean survey sampling weights which are meant to help us use our sample to more effectively represent the population as a whole.

- You use weights in a survey whenever you deliberately oversample certain groups within the population of interest, and you want to describe not your sample, but rather make inferences about that population.
- The references in the Class 3 README should be of some help.
- In particular, I recommend the opening section of https://stats.oarc.ucla.edu/r/seminars/survey-data-analysis-with-r/ as a good way to get a little more background.

- When using survey weights in practice, is it typical to build models both with and without the weights for comparison?

  - Not especially, no.

- What is the difference between adding a variable into a linear model (with +) or adding the same variable as an interaction term (with *)?

  - See the discussion of interaction in [Chapter 10 of the 432 Course Notes](#).

8.2 About the Course

- Are we at a reasonable place to start project A? Or should we wait to reach a particular class. We review the details and deadlines for project A? Would like to get started soon.

  - You can start Project A immediately, if you like. I don't plan to spend meaningful time in class discussing Project A before the Plan is due in mid-February. If you have specific questions about the instructions, please let us know on Campuswire.

- In 432 we are using more advanced data science tools to build data sets like joins and the %in% operator. Other than spending time reading the course notes to parse out why the tool was used, and targeted reading of R for Data Science, can you share your advice for quickly understanding use of these tools that we didn't (commonly) use in 431?

  - Those are excellent choices. The Sources page at https://github.com/THOMASELOVE/432-classes-2023/tree/main/sources has others.

- Is 432 easier or harder than 431?

  - It isn't intended to be meaningfully different from 431 in terms of difficulty level.

- Is it all right to ask TAs about Labs that are due later than this week, so that I can get ahead of things?

  - Sure. Campuswire, too.

- Is the expectation that we download the quarto file for each class if we want to utilize the code. I think last semester you mentioned that sometimes cutting/pasting from pdf rather than html can create issues with code.

– That was when I was providing you with HTML only, back in 431. Not true this term.

- What's so important about Jeff Leek's book?

  – I suggest you invest the time in reading it.

- Will we have the chance to work with 3d visualizations (e.g. 3d scatterplots) during this course?

  – Not as part of our general work, no. They are fun sometimes to look at, but they don't actually convey information particularly effectively, in my view.

- Will we learn how to build Shiny apps in this course?

  – No. If you want to learn about building interactive web apps with Shiny straight from R, a good place to start is the Shiny page at https://shiny.rstudio.com/. But this isn't something we'll do here.

- Does this course cover medical image segmentation?

  – No. I know nothing about the subject.

- Will we cover mixed effects models in 432 or is that in a different course?

  – While we will discuss mixed effects models a little bit, this isn't a course in that subject at all.

8.3 About R/RStudio/Quarto/Coding

- I am struggling with (some relatively fundamental issue in getting Quarto/R to do something) - how can I get help?

  – This is the ideal scenario to visit TA office hours.

- Why Quarto rather than R Markdown?

  – I'll let the folks at Posit handle this issue at https://quarto.org/docs/faq/rmarkdown.html

- What situations would you use R scripts for, when performing statistical analysis?

  – I never do. I always use Quarto (and in the past R Markdown) rather than an R script.

- Why use the |> pipe rather than the %>% pipe?

  – %>% is a predecessor to |>. For more details, read the pipes chapter in R for Data Science (2nd Edition) at https://r4ds.hadley.nz/workflow-pipes.html

- How do your report an RStudio bug to the company so that they fix it in the next update?

    – You might want to ensure it's actually a bug first. Ask about it on the Posit Community Forum at https://community.rstudio.com/ would be my first suggestion to get it in front of the folks at Posit.

9 Your Comments and Non-432 Questions

9.1 Thanks and Comments

- Appreciate all the TAs!
- enjoying the class and on-line resources immensely
- I really appreciate you taking time to create the class notes! They are super informative and probably the best course material that I have ever encountered.
- I'm liking 432 so far and hope to continue growing on it
- The PDF notes are great for annotating during class.
- Maybe it's just me, but I felt like we went a bit too fast through all the code, it was hard for me to take notes and also keep up today. Thank you!

    – I encourage you to look at the Course Notes a bit, and think about whether taking notes on code is the best use of your time in class, but it's up to you.

9.2 Other Questions

- Do we get extra credit if we see your play?

    – **Absolutely not**. I'd love to see you, but coming to see the show carries no effect on your grade in this course in any way. That would be completely unfair to people who cannot come, for whatever reason, or who simply aren't interested.

- See you in the show!

    – Great! Please feel encouraged to stick around for a few minutes after the curtain call in the theater, and I'll come out (once I'm back in street clothes) to accept any and all well-wishes or complaints.

- Do you like chocolate?

    – I do not, no. There is a modest-sized list of things I either cannot or will not eat, despite my size, and chocolate, coffee, garlic and alcohol head the list.

- How do you manage to get so much done all by yourself? What's the hidden secret behind it?

- Lots of people have helped me over the years, and continue to do so, not least our current TAs and students.
- The positive answer is that I've got a number of systems in place to keep me on track, I'm enthusiastic about people learning this stuff, and I've been at this long enough to feel like I know what's important to do, and what can be postponed (this is a double-edged sword, of course.)
- The more negative answer is that I eat poorly (most of the time), sleep a good deal less than I should, and fight some issues with letting go of things that I don't think are sufficiently polished.

- If I have questions about a project outside of the class, am I able to ask you for guidance?

  - If you're willing to accept the answers "I don't know" or "I don't have time to talk about this today" then sure. But wait until after February 1, if that's OK.