431 Lab 01 Sketch and Grading Rubric

Instructor: Dr. Thomas E. Love   Lab Author: Mr. Wyatt P. Bensken

Due: 2021-09-06 | Last Edit: 2021-09-07 13:15:12

Contents

1 Loading Packages

```
library(patchwork); library(tidyverse)

-- Attaching packages --------------------------------------- tidyverse 1.3.1 --

v ggplot2 3.3.5     v purrr   0.3.4
v tibble  3.1.3     v dplyr   1.0.7
v tidyr   1.1.3     v stringr 1.4.0
v readr   2.0.0     v forcats 0.5.1

-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
```

2 Learning Objectives

1. Be comfortable interacting with R to gain additional information about an embedded data set.
2. Describe a provided visualization, including commenting on its shape and the relationship observed.
3. Contextualize the information presented in the figure, by using additional information about the data and building upon your description.
4. Critically consider, and apply, the PPDAC approach as laid out by David Spiegelhalter in the Introduction of *The Art of Statistics: Learning from Data.*

3   Question 1. Video

In this question, you will create and submit a short (30 seconds or less, please) video of yourself (recording yourself using Zoom is a solid approach.) In this film, we should be able to see your face clearly and also hear you clearly, so make sure that is the case. The resulting video should ideally be sent to us as an **.mp4** file.

3.1   You will be doing two things in the video:

1. First, say hello, and then state your name, loudly and clearly, so that the viewer can learn to pronounce it correctly. Please use both your given name and your family name. If you prefer to be called by a nickname, please specify that, too.

2. Then, tell us something about you that we don't know, and might be interested to learn. It can be anything at all. We're hoping to get to know you a little better, and this can be something about your professional or private life, or whatever you feel you would like to share with us. We're hoping to facilitate connections here.

3.1.1   Some tips for the video. . .

- Expect that your fellow students (as well as the Teaching Assistants and Dr.Love) will see your video before the semester is over, so keep that in mind as you decide what to tell us.
- Do not worry about the quality of the video, or adding elaborate theatricality, props or scenery. If you'd like to do so, that's fine, but just make sure that we can see you and hear what you are saying clearly.
- We're not kidding about the 30-second time limit. Please don't let your video exceed 30 seconds.

3.2   Grading Rubric

To grade Question 1, the TAs need to be able to save the video and watch it. The videos need to meet three criteria for full credit on part 1.

- They clearly state their given name and family name, so that the viewer can learn to pronounce it correctly.
- They tell us something about themselves.
- We asked them to keep this to a maximum of 30 seconds, but as long as they're less than 45 seconds that's fine.

If you can save and watch the video and they accomplish all three of these things, 25 points. I expect that almost every student will get the full 25 points here.

- If they don't manage to do any one of these things, they should receive 20 points
- If they don't manage to do two of these things, they should receive 15 points.
- If they don't do any of these things, they should receive 0 points.

4 Question 2. Interpret a Visualization Built in R

We used R and the `tidyverse` (specifically the `ggplot2` part of the `tidyverse`) to build two plots using the `DNase` data set from the `datasets` package automatically loaded by R. Below, you'll find the two plots we built (labeled Plot A and Plot B) and the code we used to build the plots and then display them in a single Figure with an appropriate title (for which we used the `patchwork` package.)

Use the Help window in R to learn about the `DNase` data set, and in particular, about the two variables displayed in the plots below, and their scientific context.

Then write a paragraph (100-150 words is the appropriate length) which explains what the Figure's two plots indicate about the relationship between the two variables, and (more generally) what you have learned about the data (or science) from the Figure. Be sure to discuss how the two plots work together specifically addressing the question of what you can learn from Plot A that you can't learn from Plot B and vice versa.

4.1 Plot A

```
plotA <- ggplot(DNase, aes(x = conc, y = density)) +
    geom_point() +
    geom_smooth(method = "loess", formula = y ~ x) +
    theme_bw() +
    labs(title = "Plot A")
```
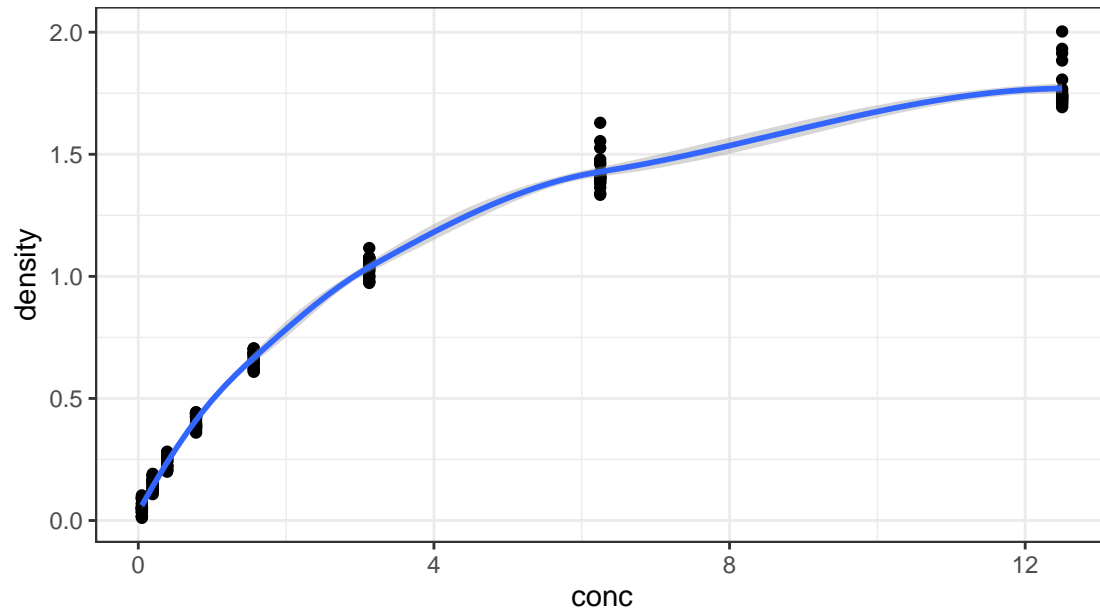
4.2 Plot B

```
plotB <- ggplot(DNase, aes(x = factor(conc), y = density)) +
    geom_boxplot() +
    theme_bw() +
    labs(title = "Plot B")
```

4.3 Code to combine Plots A and B into one Figure and display the result
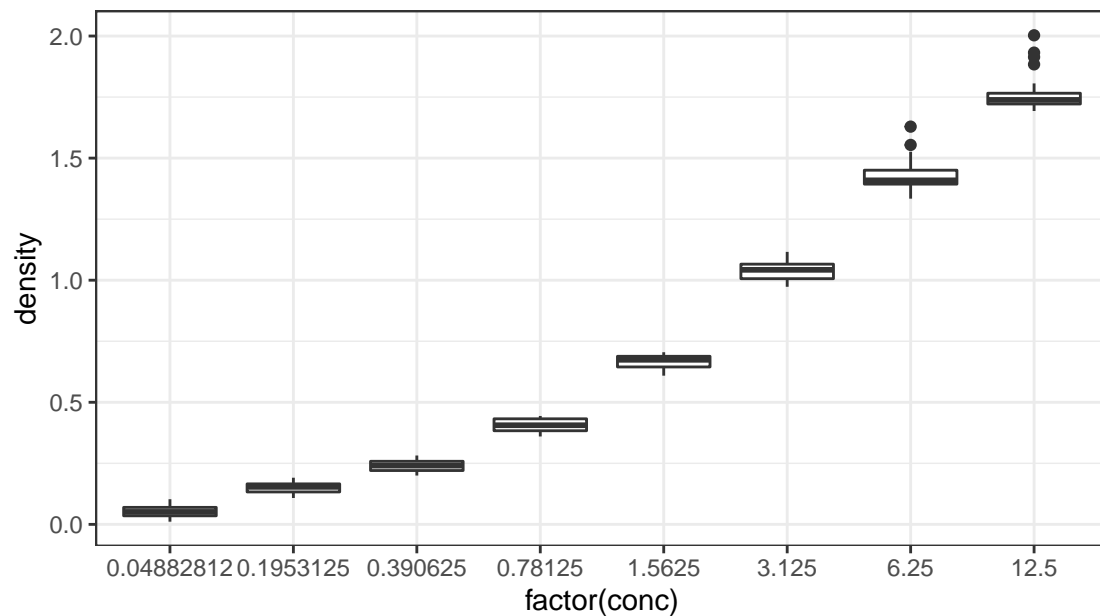
```
plotA / plotB +
    plot_annotation(
        title = "Question 2. Association of `density` and `conc` in the `DNase` data")
```

Question 2. Association of 'density' and 'conc' in the 'DNase' data

Plot A

Plot B

4.4 The `DNase` data set: Help file

- *Description*: The `DNase` data frame has 176 rows and 3 columns of data obtained during development of an ELISA assay for the recombinant protein DNase in rat serum.
- The variable `conc` is a numeric vector giving the known concentration of the protein.
- The variable `density` is a numeric vector giving the measured optical density (dimensionless) in the assay. Duplicate optical density measurements were obtained.

4.5 What Were We Looking For?

We don't write answer sketches for essay questions. We'll likely share some excerpts written by students in the class (anonymously) later, but we can tell you what we were hoping to see.

1. We want you to write in complete, grammatically correct English sentences. We want you to make your points as clearly as possible, in your own words, not, for instance, just copy-and-pasting what's in the help file.
2. We want you to accurately describe what the graph indicates about the relationship between `conc` and `density` as shown by the data, specifically that higher concentrations of the DNase protein are associated with higher values of measured optical density in the assay.
3. We wanted you also to describe the shape of the relationship, specifically that it appeared somewhat non-linear. It appears that the impact of changing the `conc` level is a bit more substantial on `density` at lower `conc` levels than at higher levels.
4. We wanted you to remark on the nature of the experiment, that several `density` measures were taken at each `conc` level, and perhaps to suggest that the blue smooth curve follows fairly closely to the average of those `density` measures at each observed `conc` level. This explains why the points in the plot fall in vertical lines at certain `conc` levels, and do not appear at other levels - that is the design of this study.
5. It also would have been helpful to avoid suggesting any sort of causal relationship. We don't know enough about the study to even suggest that higher `conc` *caused* larger values of `density` or anything like that. Among other things, we don't know what else might influence this relationship, and we don't know what else might have been controlled for in this study.
6. We also wanted you to note how Figure A was far more useful to visualize the relationship between the two variables than Figure B.

The TAs will provide a few comments, centered around these ideas, in reaction to your paragraph. We hope this is helpful to you, as you think through future work.

4.6 Grading Rubric

Students will receive 25 points if they've written a paragraph which explains what they learned from the plot and puts it in the context of what the two variables mean without factual inaccuracies. Again, I expect most students to get the full 25 points here.

- If they don't define what density and conc mean in context from looking at the DNase information, then they should lose 5 points for that, and should be told about it in TA comments.
- There are lots of ways in which they could describe the association in the plot. If they write something down which seems vaguely reasonable, then OK for now. If they use causal language (like writing "rising conc causes increased density") point out in TA comments that that's a stronger conclusion that can be justified here, but they shouldn't lose any points. A more tempered response like "higher levels of conc were associated with higher density" is sort of what we're aiming for. - It would be fine if they described the problems of interpolating when only a few "conc" values were studied, but if they don't talk about that, don't worry about it.
- If they write something (or more than one thing) that is definitely inaccurate, then they should be told about that in TA comments, and lose 5 points for factual inaccuracy (even if they have more than one such problem.)
- Comments should be provided about English grammar and syntax (I recognize that some TAs may be better at this than others) and in reaction to what they've discussed, but in this Lab, we won't drop any points for that.
- The paragraph was meant to be 100 words or less, but we won't count. If they've written something that looks to have exceeded that by quite a bit (200 words or more, perhaps) then they should be told about this in TA comments.
- They have at least some comment relating to Figure A being more useful than Figure B to show the relationship between the two variables. If they do not, they lose 5 points and please note this.

5 Question 3. React to the Introduction of *The Art of Statistics*

Read the introduction to David Spiegelhalter's book *The Art of Statistics*. Then write a short essay (of 100-150 words, please) describing how the PPDAC problem solving cycle might be helpful to you in the context of some "problem" you are interested in solving. Please feel free to draw on your own experience solving problems in a systematic way, and don't feel obliged to write about a "problem" that is related to biology or medicine or health or science. Anything you can explain briefly and that you are interested in could work well here.

Again, your response to this question (and to the other Questions) should be written using clear and complete English sentences and minimizing jargon.

We don't write answer sketches for essay questions.

5.1 Grading Rubric

Students will receive 25 points if they've written a paragraph which (a) describes a problem they are interested in solving and (b) shows some indication that they're thinking about how PPDAC (Problem-Plan-Data-Analysis-Conclusion) might be useful in the context of solving that problem. I expect, again, most students to receive 25 points.

- If they write an essay, but don't describe their problem at all, they should lose 10 points for that.
- If they write an essay, but don't indicate how PPDAC might help at all, they should lose 10 points. They certainly don't need to indicate how each piece of PPDAC might help, but at least some of the PPDAC stuff should be indicated for them to get credit.
- Detailed comments should be provided about English grammar and syntax (I recognize that some TAs may be better at this than others) and in reaction to what they've discussed, but in this Lab, we won't drop any points for that.
- The paragraph was meant to be 100 - 150 words, but we won't count. If they've written something that seems to substantially exceed this (250 words or more, perhaps) then they should be told about this in TA comments.

6   Question 4. Connecting Spiegelhalter to Question 2

In Question 2 above, you've been given data (D) and a very rough analysis (A), and from your own write-up have drawn some conclusions (C). The PPDAC approach is notably a cycle, and so after we create a conclusion we return to another problem. Given what you've learned about the data, what you've seen in the analysis, and the conclusions you've drawn in Question 2, briefly (again, in 100-150 words) describe your thoughts as to what the next cycle of the PPDAC approach for a study of these data could be.

In your response, clearly specify what you see as the Problem, Plan, Data, Analysis, and Conclusion while also highlighting how this builds upon the results already presented.

We don't write answer sketches for essay questions.

6.1   Grading Rubric

Students will receive 25 points if they've written a paragraph which clearly lays out the next steps in a PPDAC approach and they can tie it to their interpretation presented in Question 2.

- If they write an essay, but it doesn't seem to build upon what they did in Question 2 they should lose 10 points.
- If they don't specify the next Problem, Plan, Analysis, and Conclusion they should lose 10 points.
- If they don't highlight how this directly builds upon the previous work they should lose 5 points.
- If there is an essay, but it is not written in complete sentences (e.g., just a bulleted list) they should lose 10 points.
- If there is no essay, they should lose all 25 points.