Unit 1 homework instructions

DKU Stats 101 Spring 2025 Session 4

2025-03-17

Table of contents

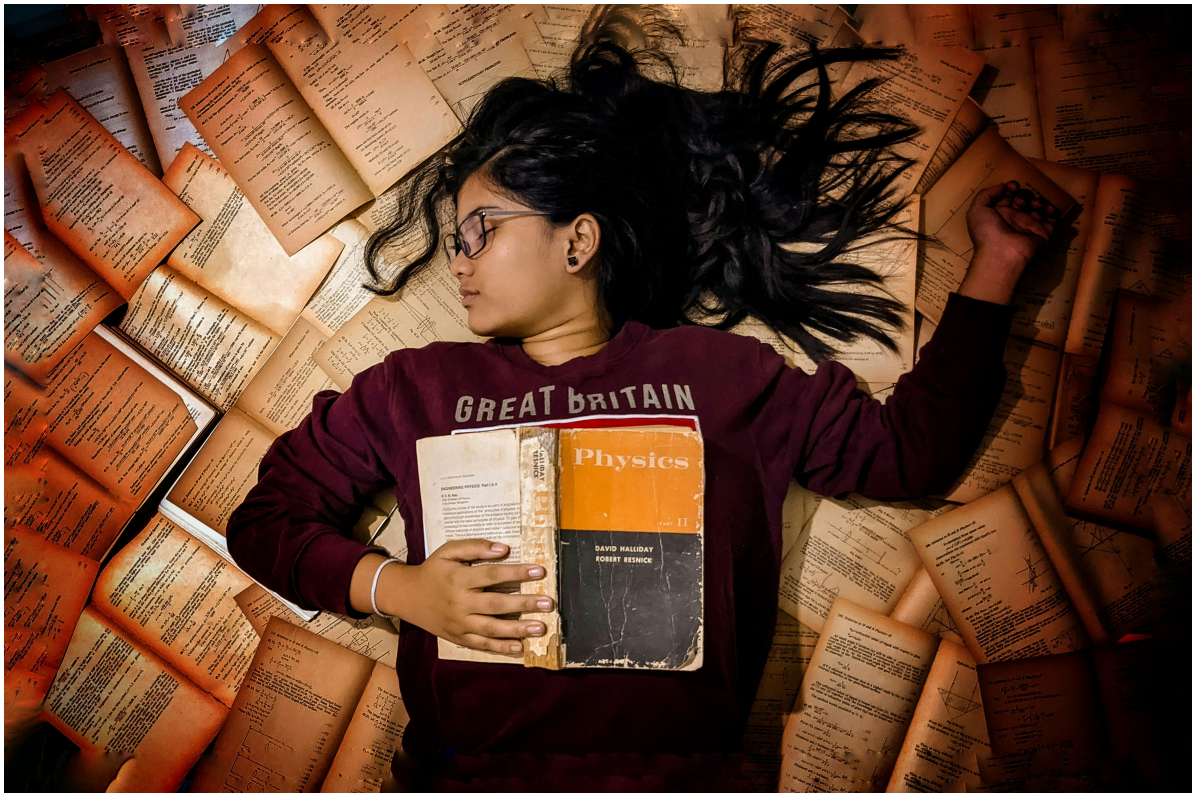
Scoring guide

Content

- Getting the right answer is only a small part of the grade
- Good quality interpretation of your results is the name of the game
- If you see something that looks unusual in your data (outlier, some unusual distribution type) - investigate it!
- When explaining your results, say something interesting about them. Did it match your expectations? Why or why not?
- Brief explanations that simply repeat what I can visually see myself will not receive a good score
- On the other hand, filling the homework with pages of not very interesting description is not valuable either. The goal isn't to write the most words, but find the most interesting things in the data.
- You do not need to be an expert in sleep for a good score, but I will expect you to look up basic information, such as "what is a normal score on the PSQI?" and "what is are"? and so on to help you understand and set expectations your data.
- The information requested in the question prompts are only a starting point, if you find other interesting information along the way, please report that. You don't need to look at the data forever but if there is obviously something else interesting in the data you should report it.
- You must have up to Question 3 completed for the homework check on March 23rd

Technical

- Make sure your graphs are produced using `ggplot()`, are well labeled, and are easy to read.
- Make sure your tables are produced with the `kable()` function from the `knitr` package, are well labeled, and are easy to read. You can make your tables prettier with the `kableExtra` package.
- Make sure you do not have anything rendered in your PDF file besides your results and, when asked for by a question, your code. That means no warnings, messages, or other output should appear in your final rendered PDF file.
- Make sure to accurately mark each page a question answer appears on when submitting on GradeScope.



Questions

Question 1: Describing your data (10 points)

1a. Where is this data from?

For this dataset, describe the data according to the five Ws & *how* defined in the textbook Chapter 1.2. What are some possible problems with the *who* and *what* of the dataset?

You can refer to the article PDF included in this `.zip` file for variable definitions.

1b. What are the variable types?

For the following variables, please make a table. The first column should be the variable name, the second the variable type (as defined in the textbook Chapter 1.3) and the third column should be the units.

- `id`

- `sex`
- `age`
- `race`
- `ethnicity`
- `totalepworth`
- `anxiety_clinical`

Question 2: Displaying and describing the data (15 points)

For the moment, we are going to focus on `totalepworth` of females and males. You can create a subset of your data using the `filter()` verb as you learned in the DataCamp lab.

2a. Filtering your data

Using the `filter()` verb as described in the DataCamp lab, make a subset of your data that only includes females and another that includes only males. Show the code you used to make the subset using the `#| echo: true` code block option.

2b. Investigating `psqi_continuousduration`

Using the Think-Show-Tell framework from the textbook (example on page 71), investigate the distribution of `totalepworth`, comparing males vs. females.

Note 1: for this question and all other Think sections in the homework, you do not need to report the W's of the data (you have already completed this in Q1)

Note 2: You are recommended to look up how this variable is defined

Think

For this section, please write down your expectations, why you expect it, the variable meaning, and, given the variable type, the best way to display the data

Show

For this section, please make an appropriate graph or table and briefly describe what you observe

Tell

Please interpret the meaning of your finding here, especially with respect to your expectation

2c. Investigating `totalepworth`

Using the Think-Show-Tell framework from the textbook, investigate the distribution of the width of the Chinese paintings

Note: You are recommended to look up how this variable is defined

Think

For this section, please write down your expectations, why you expect it, the variable meaning, and, given the variable type, the best way to display the data

Show

For this section, please make an appropriate graph or table and briefly describe what you observe

Tell

Please interpret the meaning of your finding here, especially with respect to your expectation

2d. Thinking about your results

Consider the results of 2b. and 2c. together. What can we understand about sleepiness of students from these two investigations?

Question 3: Relationships between categorical variables (10 points)

3a. Investigating the categorical relationship between `sex` vs. `psqi_category`

Investigate the relationship between `sex` vs. `psqi_category`

Hint 3: you can see an example of some ways to display this information here

Think

For this section, please write down your expectations, why you expect it, the variable meaning, and, given the variable type, the best way to display the data

Show

For this section, please make an appropriate graph or table and briefly describe what you observe

Tell

Please interpret the meaning of your finding here, especially with respect to your expectation

3b.Thinking about your results

Think deeply about what these results mean and what are the possible factor(s) that could be driving these results. Explain what additional information you would like to gather to test your interpretation.

Complete up to here by March 23 at 23:59:00

Question 4: Comparing groups (15 points)

Note: You are recommended to look up how this variable is defined

4a. Compare the groups of students on the variable `race` and `psqi_globalscore`

Think

For this section, please write down your expectations, why you expect it, the variable meaning, and, given the variable type, the best way to display the data

Show

For this section, please make an appropriate graph or table and briefly describe what you observe

Tell

Please interpret the meaning of your finding here, especially with respect to your expectation

4b. Compare the groups of students on the variable `age` and `psqi_globalscore` - please treat `age` as a categorical variable

Think

For this section, please write down your expectations, why you expect it, the variable meaning, and, given the variable type, the best way to display the data

Show

For this section, please make an appropriate graph or table and briefly describe what you observe

Tell

Please interpret the meaning of your finding here, especially with respect to your expectation

4d. Thinking about your results

Consider the results of 4b. and 4c. together. What can we learn about the differences in sleep across the different student categories? What do you think causes these differences or similarities? How would you confirm your guess as to the cause of the differences/similarities?

Question 5: Considering deviations (10 points)

5a. Selecting your data

Pick two types of clinical symptoms (such as aggressive or anxiety) and create subsets of only students who answered yes to those questions.

5b. Finding the average

Calculate the average `psqi_global` for each of the two groups. Show your code using the `#| echo:` true code block option.

5c. Normalizing the data

Find how many z units each of the averages for the score are away from the overall mean of of score and interpret your results.

Think

For this section, please write down your expectations, why you expect it, the variable meaning, and, given the variable type, the best way to display the data

Show

For this section, please make an appropriate graph or table and briefly describe what you observe

Tell

Please interpret the meaning of your finding here, especially with respect to your expectation

5d. Thinking about your results

What are some of the implications of your findings with regard to the motivation of this question? What are some of the limitations of this analysis? What other kind of analysis would you like to do to answer this question?

Question 6: Your own investigation (15 points)

6a. Selecting your own question

Similar to the previous questions, think of your own question that you would like to ask of the data. Use the Think-Show-Tell procedure to conduct your investigation. Think deeply about what your result means.

Think

For this section, please write down your expectations, why you expect it, the variable meaning, and, given the variable type, the best way to display the data

Show

For this section, please make an appropriate graph or table and briefly describe what you observe

Tell

Please interpret the meaning of your finding here, especially with respect to your expectation

6b. In summary

Sum up everything that you have learned from questions 1-6. Do not simply repeat/rephrase your previous results but try to say something larger that synthesizes the results together to draw a more meaningful general conclusion. Try to relate your findings to those present in the academic paper provided.