

Stat 311 Summer 2022 Quiz 1

Quiz 1 consists of three equally weighted problems, due uploaded to Gradescope by 11:30 PM PDT July 12th. Do not wait until the last minute to upload as **no late quizzes will be allowed**.

The quiz is written for about 1 hour and 45 minutes assuming you studied but you may complete the quiz anytime during the open window (since there is so much supplementary material for Problems 1 and 2, I decided to not have a timed quiz); just be sure to leave yourself time to upload your answers before the deadline.

This quiz is open Stat 311 notes, textbook, homework and posted supplementary materials only. All responses must be your own. Do not make be regret giving an untimed quiz—if I suspect that you collaborated with other people or put down answers that match something you found on the internet, your quiz score will be zero and I will file a report with the Student Conduct office. By uploading your quiz to Gradescope, you are acknowledging that you adhered to the rules and academic conduct standards set by the University of Washington.

Pay attention to sentence or word length requirements. We will not read more than the allowed limits. Also, always keep context in mind and report units when applicable.

If you have questions about any of the questions, you may post a private message on Ed Discussion. Do note that I can only guarantee responses during daytime hours on Monday and Tuesday. I will not answer any questions after 5 PM PDT on Tuesday, July 12th.

Problem 1 (10 points): Read the short UW News article for an overview of a study that showed that areas with historical redlining were associated with more air pollution. Then look at the published journal article that provides more details about the study. The news and journal articles can be found in the Quiz 1 assignment on Canvas. Use information in these articles to answer parts (a) – (g). Except for part (f), limit your responses to at most two sentences.

- a) Was this an experiment or observational study. Briefly explain. (0.5 point)
- b) What is meant by the term redlining that is used in the two articles? (1 point)
- c) What were the main sources of data for this study? (0.5 point)
- d) Race and ethnicity were combined into which aggregate groups for this study? Include the HOLC percentages of each group. (2 points)
- e) What are the main two pollutants that were investigated in this study and why did they focus on just these two? (1 point)
- f) Figure 1 of the journal article shows population weighted distributions for both pollutants. Summarize what you can glean from the top left plot [Unadjusted NO₂ national aggregation]. Be sure to speak to the HOLC grades and race/ethnicity. Limit your response to at most 150 words. (3 points)
- g) Figure 2 in the journal article looks at the interaction between racial/ethnic groups and historical redlining grades. What does the right plot (PM_{2.5} difference) tell you about the association between these two variables? (2 points)

Stat 311 Summer 2022 Quiz 1

Problem 2 (10 points): Five measurements were taken on two species of fish from a single lake ($n = 35$ Bream and $n = 33$ Perch). For this problem, fish weight is the response (y) and cross length, height and diagonal width are predictors of fish weight. This problem does not require any coding; rather, we are providing summary graphs, summary tables, and `lm` output in the supplementary handout, Quiz1Problem2.pdf, posted in the Quiz 1 assignment on Canvas. Use the information in this handout to answer parts (a) – (g).

- a) Look at the density plot in the last row of Figure 2 and the histograms in Figure 6 of the supplementary handout (pages 3 and 7) and describe the overall distribution of the observed sample widths. Also compare the sample distributions for widths individually for Bream and Perch. Use no more than four sentences. (1.5 points)
- b) Look at Figure 2 (page 3) of the supplementary handout and interpret the overall joint relationship for Weight on Width. Also comment on the relationship when considering species. [Hint: make clear, specific observations regarding the relationships]. Use no more than three sentences. (1.5 point)
- c) Using Figure 2 (page 3), what is the overall sample correlation between Height and Width. What are the correlations for Height and Width by species? How do correlations by species compare with the overall correlation? Use no more than two sentences. (2 points)
- d) Write out the regression equation for Weight on Height. In one sentence, interpret the estimated slope parameter for this regression in the context of the problem. (1 point)
- e) Report and in one sentence interpret the coefficient of determination for the regression of Weight on Length3 in the context of the problem. (1 point)
- f) We have provided the output for simple linear regressions of Weight on the three predictor variables we are considering. Of the three models, which model do you think is the best single predictor model? Use all the information (scatterplots, `lm` outputs, residual plots, and histograms of the residuals) to support your choice. We are looking for written answers that are in the context of the problem and that you support your choice with more than a single piece of information. Limit your answer to a maximum of 100 words. You may use bullet points if that helps you organize your answer. (2 points)
- g) On pages 12 and 13 of the supplementary handout, we provide regression output that includes Width and the additional categorical variable Species. Of the three models for Weight on Width (single regression all species, differing intercepts by species (parallel lines), or different slopes by species), which model do you think is best? Use the information from the regression outputs and Figure 7 to justify your answer. Limit your answer to at most three sentences. (1 point)

Stat 311 Summer 2022 Quiz 1

Problem 3 (10 points; 2 points each): True or False. If the statement is True, then indicate True. If the statement is False, then indicate False and carefully explain with no more than three sentences why the statement is false.

- a) A researcher is interested in administering a survey to gauge attitudes of Catholic Church members regarding the current Pope. A random sample of five Catholic Churches in King County are selected and all church members of those five churches are surveyed. This type of sample is called a stratified sample.
- b) A student is collecting data on movie preferences for a class project. The student stands outside a neighborhood theater and asks people exiting the theater if they would be willing to take a brief survey. The student can get 50 people willing to participate over the course of six hours. The 50 completed surveys are an example of a simple random sample.
- c) A study compared a group of men who had heart attacks with a similar group of controls. The proportion of men with male pattern baldness was compared between the two groups. This is an example of an observational study.
- d) A sample of households in a community is selected at random from the telephone directory. In this community, 4% of households have no telephone, 10% have only cell phones, and another 25% have unlisted telephone numbers. The largest issue with this sampling scenario is response bias.
- e) The contingency table shown below is from a Pew Research Center [study](#), published in May 2022, that looked at use of video conferencing services (Zoom, Webex or other) for three levels of work from home status.

Work from home status	Use of Video Conferencing for Work				Total
	Often	Sometimes	Hardly Ever	Never	
All or most of the time	66	17	7	11	101
Sometimes	49	28	16	7	100
Rarely or never	35	30	18	17	100
Total	150	75	41	35	301

The joint percentage of people that sometimes work from home and never use video conferencing for work is 2.3% (rounded to one decimal place), and among people that sometimes use video conferencing for work, 16.8% (rounded to one decimal place) work from home all or most of the time, 28% sometimes work from home and 30% rarely or never work from home.