431 Quiz 1 for Fall 2022

Thomas E. Love, Ph.D.

Due 2022-10-10 at 9 PM: Version 2022-10-05 00:18:05

Table of contents

Instructions for Students

There are 24 questions on this Quiz, and this PDF is 26 pages long. Be sure you have all 26 pages. It is to your advantage to answer all 24 Questions. Your score is based on the number of correct responses, so there's no chance a blank response will be correct, while a guess might be, so you should definitely answer all of the questions.

0.1 The Google Form Answer Sheet

All of your answers should be placed in the Google Form Answer Sheet, which will be found (once the Quiz starts) at https://bit.ly/431-2022-quiz1-answer-sheet. All of your answers must be submitted through that Google Form by 9 PM on Monday 2022-10-10, and no extensions will be made available, so do not wait until Monday evening to submit. We will only accept responses through the Google Form.

The Google Form's final question requires you to type in your full name to affirm that you followed the rules for the Quiz. You must complete that affirmation before you can submit your responses. After submission (like a Minute Paper) you will be emailed a copy of your submission, with a link allowing you to edit your responses.

If you wish to work on some of the quiz and then return later, do so by [1] completing the final question (the affirmation) by typing in your full name, and then [2] submitting the quiz. You will then receive a link at your CWRU email which will let you return to the quiz as often as you like without losing your progress.

0.2 Getting Help During the Quiz

This is an open book, open notes quiz. You are welcome to consult the materials provided on the course website and that we've been reading in the class, but you are not allowed to discuss the questions on this quiz with anyone other than Professor Love and the teaching assistants. You will be required to complete a short affirmation that you have obeyed these rules as part of submitting the Quiz.

If you need clarification on a Quiz question, you have exactly two ways of getting help:

1. You can ask your question in a private question to the instructors on Campuswire, or
2. You can ask your question via email to **431-help at case dot edu**.

While the complete Quiz is available, we will not answer questions about the Quiz except through the two approaches listed above. We promise to respond to all questions received before 5 PM on 2022-10-10 at that time, if not sooner.

- Specific questions are more likely to get helpful answers.

- We will not review your code or your English for you.
- We will not tell you if your answer is correct, or if it is complete.
- We will email all students if we find an error in the Quiz that needs fixing.

0.2.1 When Should I ask for help?

We recommend the following process.

- If you encounter a tough question, skip it, and build up your confidence by tackling other questions.
- When you return to the tough question, spend no more than 10-15 minutes on it. If you still don't have it, take a break (not just to do other questions) but an actual break.
- When you return to the question, it may be much clearer to you. If so, great. If not, spend 5-10 minutes on it, at most, and if you are still stuck, ask us for help.
- This is not to say that you cannot ask us sooner than this, but you should **never, ever** spend more than 20 minutes on any question without asking for help.

0.3 Scoring and Timing of the Quiz

Question 7 is worth 8 points, and the other 23 questions are worth 4 points each, summing to 100 points. The questions are not in any particular order, and range in difficulty from "things I expect everyone to get right" to "things that are deliberately tricky".

The Quiz is meant to take 4-6 hours. I expect most students will take 3-8 hours, and some will take as little as 2 or as many as 10. It is not a good idea to spend a long time on any one question.

Dr. Love will grade all Quizzes, and you should have your result by 5 PM on Wednesday 2022-10-12.

0.4 Writing Code into the Google Form

Occasionally, we ask you to provide a single line of code. If not otherwise specified, a single line of code in response should contain no more than two pipes, although you may or may not need the pipe in any particular setting. Moreover, do not include the `library` command at any time for any of your code. Assume in all questions that all relevant packages have been loaded in R.

0.5 The Data Sets for this Quiz

We have provided **four** data sets that are mentioned in the Quiz. You will find them in our Shared Google Drive in the Quiz 1 folder. They may be helpful to you.

- `algae.csv`, first mentioned in Question 10
- `sleep.csv`, first mentioned in Question 13
- `newborn.csv`, first mentioned in Question 15
- `pcare.csv`, first mentioned in Question 21

I have also provided an R Markdown file and an HTML file, described in Question 07.

0.6 Packages used in Developing this Quiz

This doesn't mean you need to use all of these packages or even that I used them all, but it does mean that I did not add any additional packages to this list in building the quiz or the answer sketch. You will note that the `mosaic` package is not listed here, but I did use the `favstats` function from that package using the `mosaic::favstats()` approach in preparing this material.

```
library(broom)
library(Epi)
library(equatiomatic)
library(glue)
library(ggrepel)
library(ggridges)
library(gt)
library(kableExtra)
library(simputation)
library(janitor)
library(naniar)
library(patchwork)
library(tidyverse)

theme_set(theme_bw())
```

1 Question 01

The initial results of the Systolic Blood Pressure Intervention Trial (SPRINT) received a lot of attention.

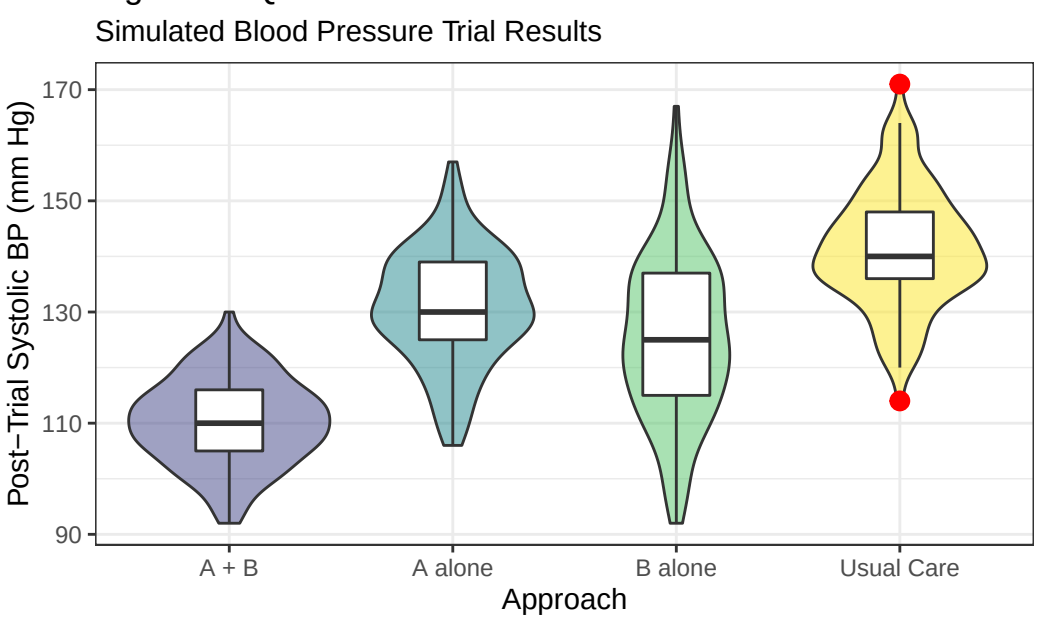
Quoting a press release from the National Institutes of Health:

> SPRINT evaluates the benefits of maintaining a new target for systolic blood pressure, the top number in a blood pressure reading, among a group of patients 50 years and older at increased risk for heart disease or who have kidney disease. A systolic pressure of 120 mm Hg, maintained by this more intensive blood pressure intervention, could ultimately help save lives among adults age 50 and older who have a combination of high blood pressure and at least one additional risk factor for heart disease, the investigators say.

Consider a hypothetical trial, where two different interventions are studied to see whether patients in another population besides that studied in SPRINT may have their blood pressure effectively managed to fall at the target level (120 mm Hg or lower).

500 patients were included in this trial, and were randomly allocated (125 to each intervention) so that we have 125 patients receiving both interventions A and B, 125 receiving A alone, 125 receiving B alone, and 125 receiving usual care (neither A nor B). The post-trial Systolic Blood Pressure results for all 500 patients are shown in the Figure for Question 01.

Figure for Question 01



Question 01 continues on the next page.

Question 01 (continued)

Consider the following statements:

I. The group of patients receiving usual care had the smallest number of patients with SBP at 120 or lower after the trial.

II. The group of patients receiving B alone had the largest spread in their distribution of post-trial systolic blood pressures.

III. The group of patients receiving both A and B had more than 90 patients with post-trial SBP at 120 or lower.

Which of these statements are true?

a. I only
b. II only
c. III only
d. I and II
e. I and III
f. II and III
g. All three statements
h. None of the three statements

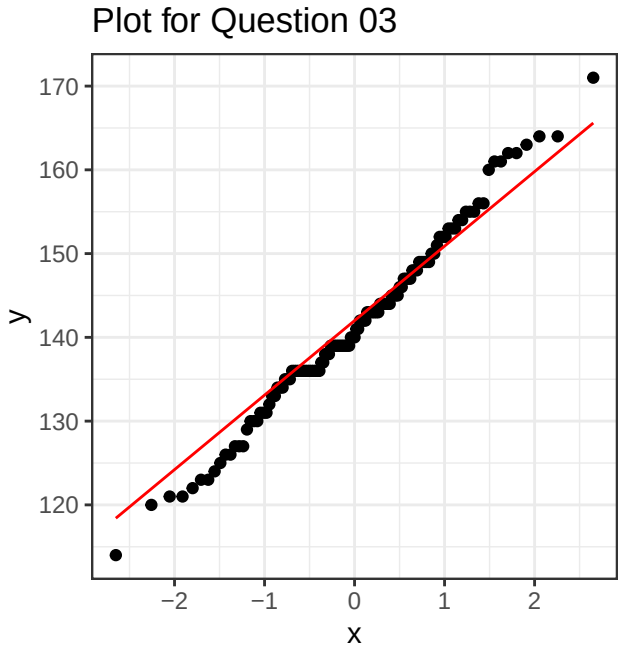
2 Question 02

Which of the four blood pressure trial groups discussed in Question 01 produced the individual subject with the lowest post-trial systolic blood pressure?

a. The group receiving both A and B
b. The group receiving A alone
c. The group receiving B alone
d. The group receiving usual care
e. It is impossible to tell from the information provided.

3 Question 03

The normal Q-Q plot shown as the Plot for Question 03 below is taken from one of the four blood pressure trial groups discussed in Questions 01 and 02. Which one?



a. The group receiving A alone
b. The group receiving B alone
c. The group receiving usual care
d. The group receiving both A and B

4 Question 04

Suppose you have access to a tibble, called `q04`, in R, which contains information on a cohort study. Specifically, you have collected data to look at the impact of exposure to an industrial solvent (which is stored in a four-level character variable called `solvent` which can be either none, modest, moderate or profound) on the probability of a bladder cancer diagnosis (stored as a three-level character variable called `diagnosis` which can be either definite, possible, or no).

Provide a single line of R code (you may use at most two pipes) to obtain an appropriate numerical summary describing the distribution of solvent exposure within each `diagnosis` group in the `q04` tibble.

5 Question 05

Continuing the study from Question 04, you now have more granular information on the exposure level to the solvent. You now have an `exposure` measure, which is the percentage of the Occupational Safety and Health Administration (OSHA) recommended exposure limit, so that 100 = the recommended exposure limit for this solvent, and values above 100 indicate exposures that exceed that limit, while values below 100 indicate exposures that are at least somewhat "safe". Suppose a new tibble called `q05` is available to you containing this `exposure` measure as well as the bladder cancer `diagnosis` variable from Question 04. Provide a single line of R code (using at most two pipes) to yield an appropriate numerical summary of the distribution of solvent exposure (using the new `exposure` variable) within each `diagnosis` group in the `q05` tibble.

6 Question 06

We built a model to predict birth weight (in ounces) for newborn animals growing up in one of two locations (one seems worse than the other at providing support for the mother) on the basis of location and length of gestation (in days). The data includes 80 newborns and the model's R-squared value is 37%. Here's the data for the first couple of newborns:

subject	birth_weight	location	gestation_length
S-001	122	worse	276
S-002	103	better	261

Our regression model yields this equation. Which of the interpretations listed below is correct? (CHECK ALL THAT APPLY.)

$$\begin{aligned}\widehat{\text{birth\_weight}} = &-122.7 + 0.9(\text{gestation\_length}) + 11.1(\text{location}_{\text{worse}}) \\ &- 0.08(\text{gestation\_length} \times \text{location}_{\text{worse}})\end{aligned}$$

a. Each additional day of gestation increases estimated birth weight by 0.9 ounce, regardless of location.
b. Subjects who grew up in the worse location had a larger estimated birth weight than subjects who had the same gestation length, but grew up in the better location, assuming both newborns had a gestational length of at least 200.
c. The correlation between predicted birth weight using this model and actual birth weight is 0.37.
d. The model's predicted birth weight for Subject S-001 is smaller than its predicted birth weight for subject S-002.
e. None of the statements above are correct.

7 Question 07 (8 points)

As part of the Quiz materials, you'll find an R Markdown file called `question7_initial.Rmd` and an HTML file called `question7_resultswewant.html`. In order to generate that `resultswewant` file, you need to solve four problems with the `initial` R Markdown file. In Question 7 (parts a, b, c and d), we ask you to identify (by the line number in the initial file) the four places where a change needs to be made, and specify what that change should be in order to produce the results you see in the `resultswewant` document.

As a hint, I'll remind you that the use of `echo = FALSE` in a code chunk means that the code is not shown in the final document, and thus, there are no problems with the use of `echo = FALSE` in Question 7.

7.1 Part 07a. (2 points)

There are four critical errors in the initial R Markdown file. List the first one by specifying the line number in which it occurs in the initial R Markdown file, then specify how it needs to be fixed.

7.2 Part 07b. (2 points)

List the second critical error by specifying its line number in the initial R Markdown file, then specify how it needs to be fixed.

7.3 Part 07c. (2 points)

List the third critical error by specifying its line number in the initial R Markdown file, then specify how it needs to be fixed.

7.4 Part 07d. (2 points)

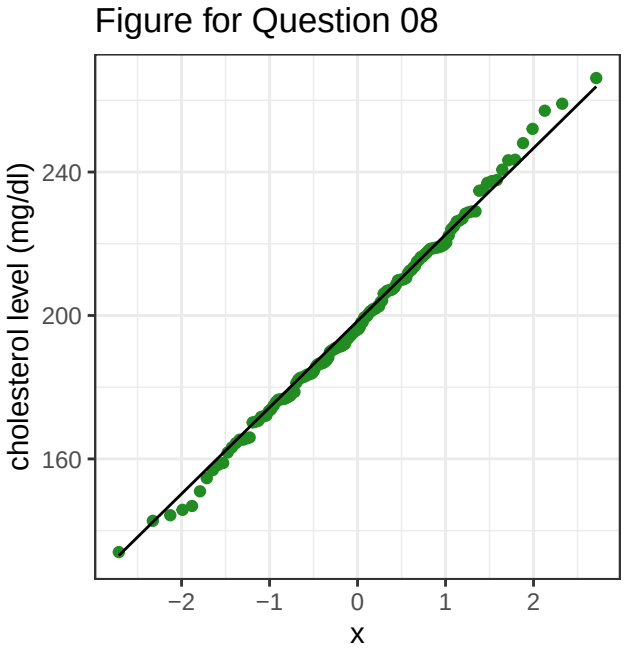
List the fourth critical error by specifying its line number in the initial R Markdown file, then specify how it needs to be fixed.

7.5 Part 07e. (optional: 2 points of extra credit)

There is actually a fifth problem in the initial R Markdown file, which doesn't need to be corrected to get the `resultswewant` HTML file, but which should be corrected anyway. Specify its line number in the initial file, and then specify how it should be fixed.

8 Question 08

The Figure for Question 08 is a Normal Q-Q plot of cholesterol levels (in mg/dl) for 150 adult American men.



Which of the following statements best describes the distribution of the cholesterol levels?

a. Symmetric, but substantially outlier-prone in comparison to what we would expect from a Normal distribution.
b. Approximately Normally distributed, with a mean of approximately 250 mg/dl and a standard deviation of approximately 25 mg/dl.
c. Approximately Normally distributed, with a mean of approximately 250 mg/dl and a standard deviation of approximately 10 mg/dl.
d. Approximately Normally distributed, with a mean of approximately 200 mg/dl and a standard deviation of approximately 25 mg/dl.
e. Approximately Normally distributed, with a mean of approximately 200 mg/dl and a standard deviation of approximately 10 mg/dl.
f. Not approximately Normally distributed, but instead substantially skewed to the left.
g. Not approximately Normally distributed, but instead substantially skewed to the right.

9 Question 09

In the introduction and Chapters 1-4 of *The Art of Statistics*, David Spiegelhalter describes several common features of a strong visualization of data, including some identified by Alberto Cairo. Which of the following features are described as being characteristic of high-quality work in this regard? (CHECK ALL THAT APPLY.)

a. The graph's design is chosen so that relevant patterns become noticeable.
b. The graph contains reliable information.
c. The graph's appearance is presented in an attractive way.
d. The graph is honest, clear, and contains deep insights.
e. The graph should never connect data points gathered at different times.
f. The graph is accompanied by a meaningful title, and clear labels and captions.
g. The graph is of the raw data only.
h. The graph helps to raise more questions and encourage the reader to explore.

10 Question 10

We have provided the `algae.csv` data set with the Quiz materials. This data file describes 200 water samples collected from the same river over a period of 3 months, of which 184 have complete data. Each of those 184 observations contains information on a series of chemical parameters measured in the water samples, one of which is the mean value of orthophosphate, contained in the variable `oPO4`.

Consider the histograms shown in the Figure for Question 10 on the next page, and suppose your goal is to approximate a Normal distribution with some transformation of the `oPO4` data.

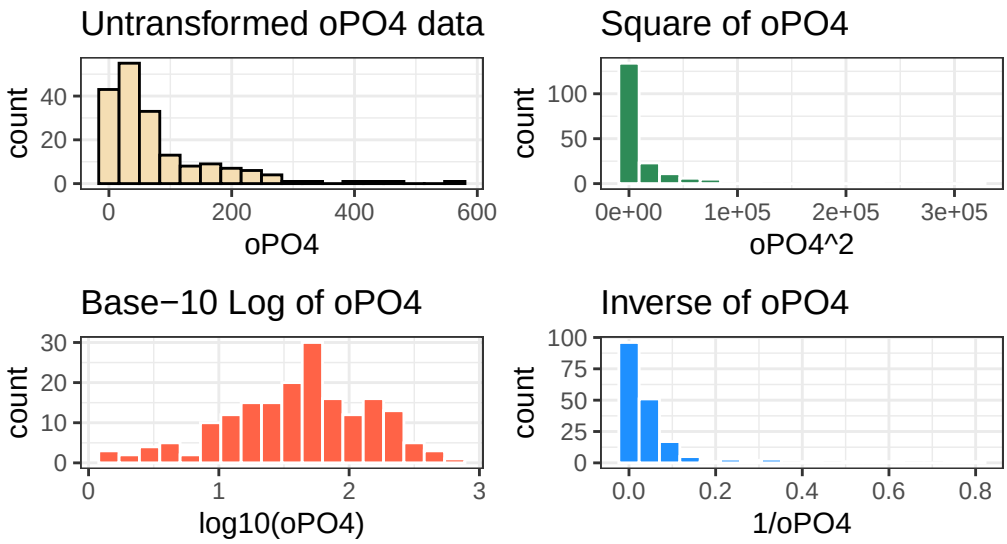
Which of the following options describes the most logical transformation to use in trying to accomplish the goal of approximating the transformed data's distribution with a Normal model?

a. The untransformed oPO4 data
b. The square of the oPO4 data
c. The base-10 logarithm of the oPO4 data
d. The inverse of the oPO4 data
e. It is impossible to tell from the information provided.

Question 10 continues on the next page.

Question 10 (continued)

Figure for Question 10



184 complete cases in algae

11 Question 11

Return to the `algae.csv` file I provided to you and discussed in Question 10, and fit a linear model to predict the **natural** logarithm (**not** the base-10 logarithm used in Question 10) of the algae frequency `a1` using the natural logarithm of the `oPO4` (orthophosphate) in the same water sample.

- Note that you will have to manage the data to use only those samples with (1) complete data on both variables in your model, and (2) which have values of `a1` that exceed zero.

How many water samples are included in your model?

12 Question 12

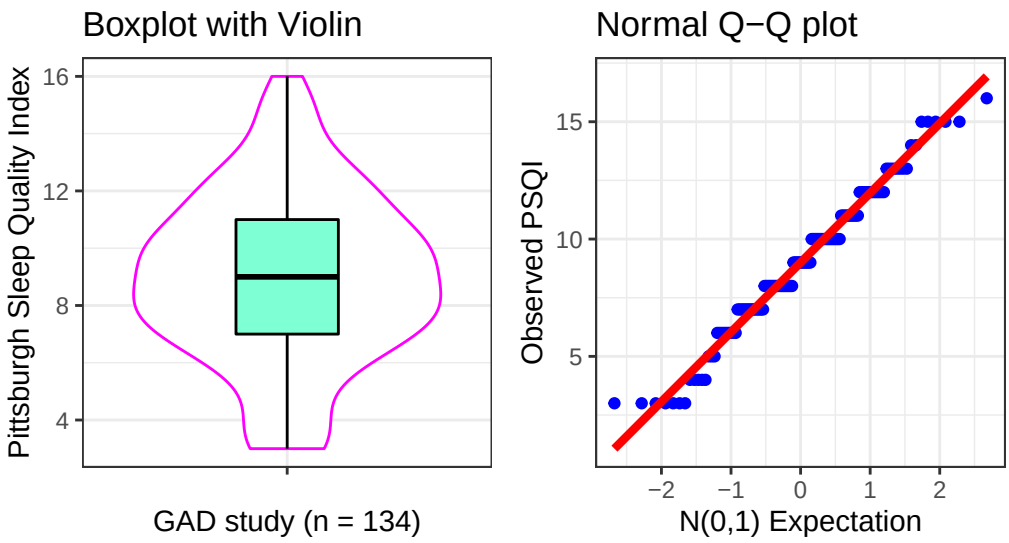
Based on your model developed in Question 11, what is the predicted value of the actual algae frequency `a1` for a sample with an `oPO4` value of 64?

- Round your response to a single decimal place.
- Think carefully about what your model predicts before responding.

13 Question 13

The Pittsburgh Sleep Quality Index (PSQI) is a self-rated 19-item questionnaire to assess sleep quality and disturbances over the past month. The PSQI yields a global score which ranges from 0 to 21, with higher scores indicating greater overall sleep disturbance. A study of older adults with a diagnosis of generalized anxiety disorder (GAD) yielded a sample of 134 subjects stored in a file provided to you called `sleep.csv`. I used the `sleep` data to produce the Figure for Question 13. Before building this plot, I loaded all necessary packages.

Figure for Question 13



Which of the following bits of R code were **NOT** used in generating the Figure for Question 13? (CHECK ALL THAT APPLY.)

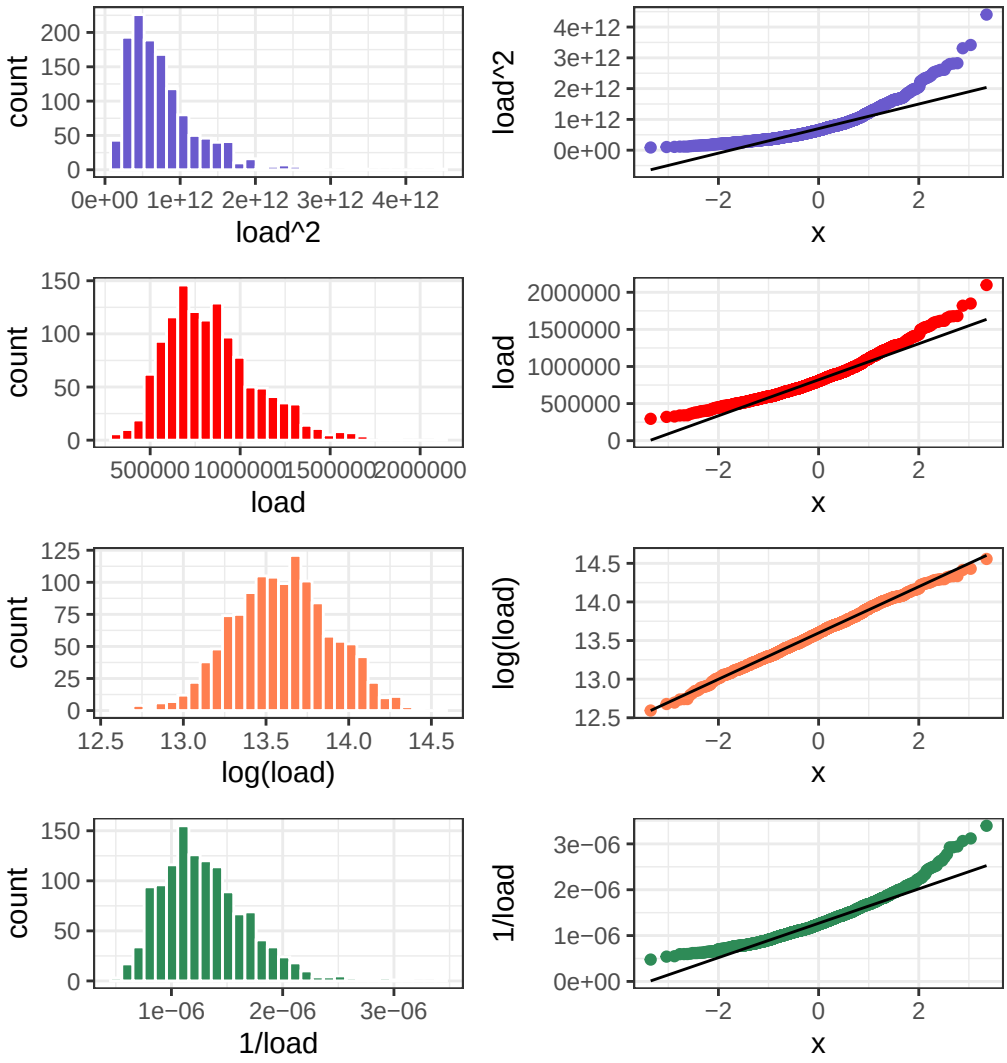
a. `labs(y = "Observed PSQI", x = "N(0,1) Expectation", title = "Normal Q-Q plot")`
b. `geom_boxplot(fill = "aquamarine", col = "black", width = 0.2)`
c. `(p1 / p2) + plot_annotate(title = "Figure for Question 13")`
d. `labs(x = "GAD study (n = 134)", y = "Pittsburgh Sleep Quality Index", title = "Boxplot with Violin")`
e. `p2 <- ggplot(data = sleep, aes(sample = psqi))`
f. `geom_qq(col = "blue") + geom_qq_line(col = "red", lwd = 1.5)`
g. `geom_violin(col = "magenta", fill = "white")`
h. `p1 <- ggplot(data = sleep, aes(x = "", y = psqi))`
i. Check this option only if all the bits of code listed in options a-h were used.

14 Question 14

1,251 subjects were given a hepatitis C RNA quantitative test which measured the amount of Hep C virus present in their blood, in IU/ml, and this is called the viral load. Anything over 800,000 is usually considered high, and anything under that is low. Those with low viral load have a better chance of responding to treatment. Consider the Figure for Question 14.

Figure for Question 14

Exploring Viral Load Power Transformations



Question 14 continues on the next page.

Question 14 (continued)

If our goal is to obtain a transformation of the data which is well fit by a Normal model, which of the following options appears to be our best choice?

a. Taking the square of the viral load.
b. Taking the viral load, untransformed.
c. Taking the natural logarithm of the viral load.
d. Taking the inverse of the viral load.
e. None of these options.

15 Question 15

I have provided you with a data set called `newborn.csv`. After you import that into R as a tibble called `newborn`, the result should contain a variable called `apgar5` that contains scores on the APGAR scale at five minutes for 130 infants, although 4 of the values are listed as NA.

You wish to obtain the standard deviation of the APGAR scores in the `newborn` tibble. If you need to know more about the APGAR score, visit https://goo.gl/9rxkVU. Your task is to mark the box next to EACH of the R commands listed below that produce the SAMPLE STANDARD DEVIATION of APGAR scores at five minutes for the 126 infants not marked as NA. (CHECK ALL THAT APPLY.)

a. `mosaic::favstats(~ apgar5, data = newborn)`
b. `summary(newborn)`
c. `sd(newborn$apgar5)`
d. `newborn |> summarize(sd(apgar5, na.rm = TRUE))`
e. `newborn |> filter(complete.cases(apgar5)) |> summarize(sd = sd(apgar5))`
f. `newborn |> select(complete.cases(apgar5)) |> summarize(sd = sd(apgar5))`
g. None of these will produce the correct value.

16 Question 16

The `starwars` data set is part of the `dplyr` package loaded by the tidyverse. You can learn more about it at https://dplyr.tidyverse.org/reference/starwars.html if you like. In that data, we find information on Star Wars characters, specifying 14 different variables, including each character's `homeworld` and `gender`. After loading the packages used in developing this Quiz (specified in the Instructions for Students), try running the following code:

```
starwars |>
  select(name, hair_color, birth_year) |>
  gt()|> tab_header("Some Characters in Star Wars")
```

Note that I'm not showing all of the rows in the table here, to save some space.

Some Characters in Star Wars

name	hair_color	birth_year
Luke Skywalker	blond	19.0
C-3PO	NA	112.0
R2-D2	NA	33.0
Darth Vader	none	41.9
Leia Organa	brown	19.0
Owen Lars	brown, grey	52.0
Beru Whitesun lars	brown	47.0
R5-D4	NA	NA
Biggs Darklighter	black	24.0
Obi-Wan Kenobi	auburn, white	57.0

Your job is to modify the code I've provided to produce a new table which includes only those characters in the `starwars` data set that have:

1. brown hair (do not include people with both brown and grey hair, for example)
2. a birth year of 19 or higher (years are measured here before the Battle of Yavin)
3. a known, non-missing, homeworld.

Now, review your new table, and answer these questions:

a. How many characters of feminine gender appear in your new table?
b. How many characters of masculine gender appear in your new table?

17 Question 17

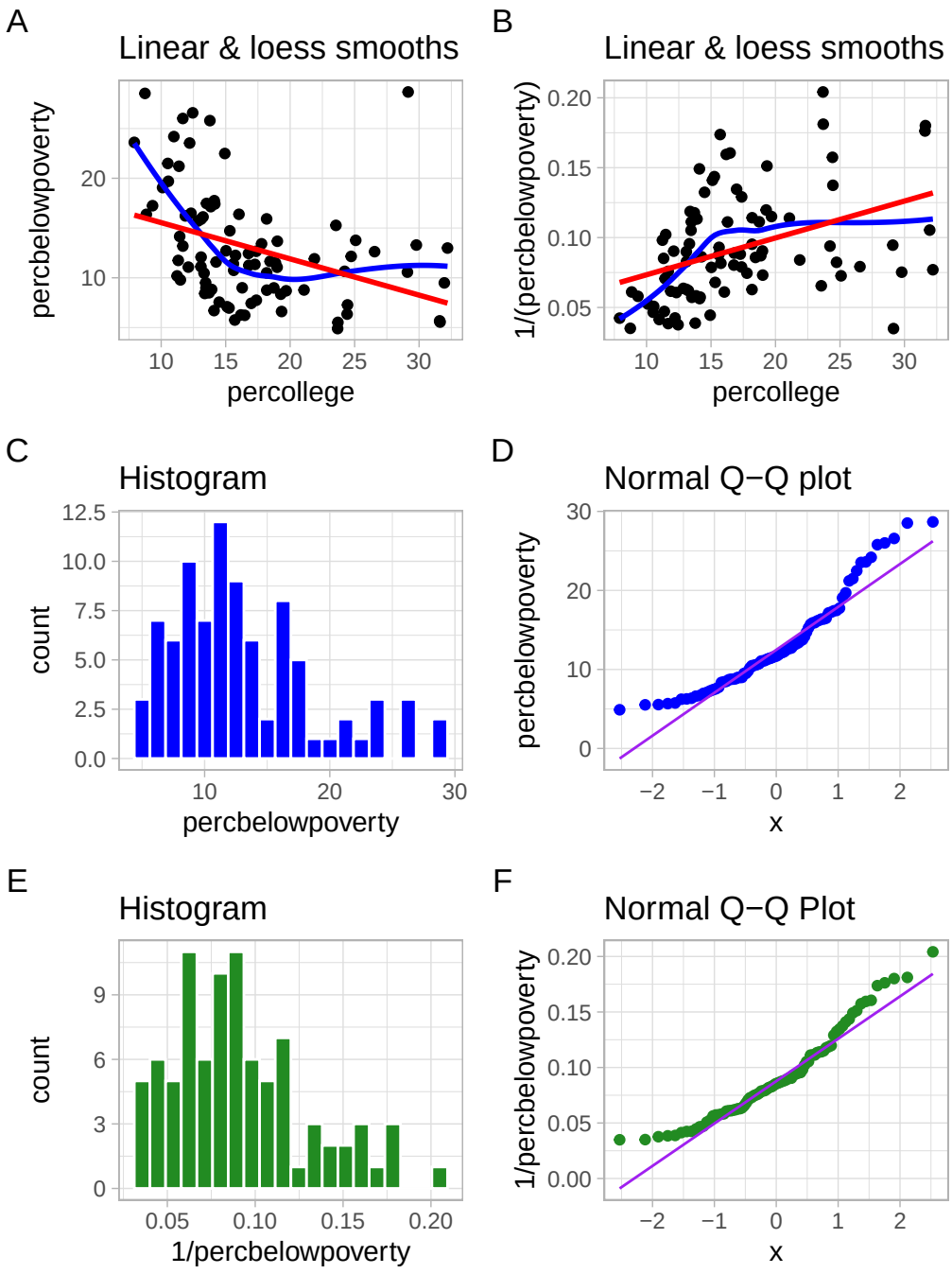
Suppose you are using a subset of the `midwest` data from the `ggplot2` package. You are trying to determine for this subset whether or not a transformation of the outcome (specifically, taking the inverse of the outcome) is necessary to fit a linear regression model to describe the relationship between `percollege` (the predictor, specifically the percent college educated) and `percbelowpoverty` (the outcome, specifically the percent below the poverty level). Which of the Plots shown in the Figure for Question 17 would be of the most help in assessing whether using this transformation would improve the assumption of linearity?

a. Plot A
b. Plot B
c. Plots C and D
d. Plots E and F
e. They would all be equally useful

The Figure for Question 17 is shown on the next page.

Question 17 (continued)

Figure for Question 17

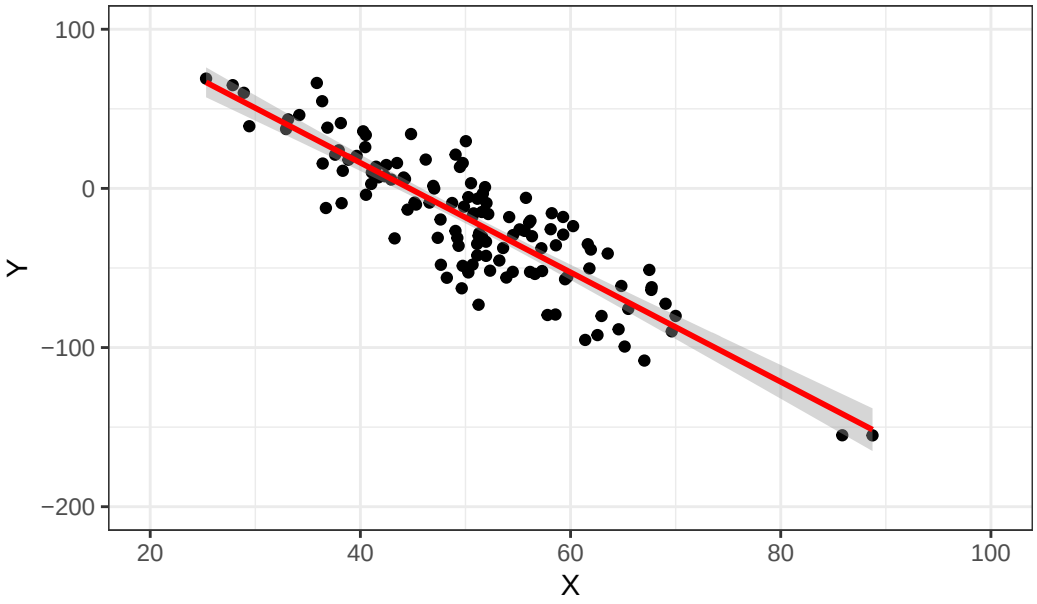


18 Question 18

Consider the following possible summaries of a linear model fit to predict Y from X, describing the scatterplot shown in the Figure for Question 18. Which of these summaries is correct?

a. Model: y = 3.4 + 154 x, with R-squared = -0.76
b. Model: y = 3.4 - 154 x, with R-squared = -0.26
c. Model: y = -3.4 + 154 x, with R-squared = 0.76
d. Model: y = -3.4 + 154 x, with R-squared = 0.26
e. Model: y = 3.4 + 154 x, with R-squared = 0.76
f. Model: y = 3.4 + 154 x, with R-squared = 0.26
g. Model: y = 154 - 3.4 x, with R-squared = -0.76
h. Model: y = 154 - 3.4 x, with R-squared = -0.26
i. Model: y = 154 + 3.4 x, with R-squared = 0.76
j. Model: y = 154 + 3.4 x, with R-squared = 0.26
k. Model: y = 154 - 3.4 x, with R-squared = 0.76
l. Model: y = 154 - 3.4 x, with R-squared = 0.26

Figure for Question 18



19 Question 19

Sir Austin Bradford Hill's criteria have been a building block for scientific work for 55 years. In *The Art of Statistics*, David Spiegelhalter outlines recent work by Jeremy Howick and colleagues to separate these criteria into direct, mechanistic and parallel evidence.

For each description below, identify whether what is provided is best categorized as direct, mechanistic or parallel evidence.

Columns:

1. direct
2. mechanistic
3. parallel

Rows:

a. The observed effect is consistent with what is already known.
b. The observed effect is consistent with a plausible biological mechanism.
c. The observed effect is seen again in a new, similar study.
d. The observed effect is smaller in subjects who are less exposed.
e. The observed effect is preceded in time by the presumed cause.

20 Question 20

Choose the five number summary (minimum, Q1, median, Q3 and maximum) that matches the stem-and-leaf plot of LDL cholesterol levels shown in the Figure for Question 20.

a. Min: 53 Q1: 100 Median: 135 Q3: 155 Max: 241
b. Min: 53 Q1: 100 Median: 131 Q3: 158 Max: 241
c. Min: 53 Q1: 100 Median: 131 Q3: 155 Max: 241
d. Min: 53 Q1: 100 Median: 122 Q3: 155 Max: 241
e. Min: 53 Q1: 100 Median: 135 Q3: 158 Max: 241
f. Min: 53 Q1: 100 Median: 122 Q3: 158 Max: 241

Figure for Question 20

```
The decimal point is 1 digit(s) to the right of the |

 5 | 33
 6 |
 7 | 1
 8 | 8
 9 | 39
10 | 078
11 | 07
12 | 2
13 | 15
14 | 1456
15 | 58
16 |
17 | 08
18 |
19 | 9
20 |
21 | 1
22 |
23 |
24 | 1
```

Setup for Questions 21-23

Questions 21-23 make use of the `pcare` data that describe 111 patients seen in primary care. Note that the `pcare.csv` data file is provided as part of the Quiz materials. The variables are:

- `subj_id` = subject ID
- `age` = subject's age (in years)
- `totChol` = subject's total cholesterol, in mg/dl
- `smoke` = current smoking status (yes or no)
- `BMICat` = category determined by body-mass index (Healthy, Overweight or Obese)
- `heartRate` = heart rate, in beats per minute
- `diaBP` = diastolic blood pressure, in mm Hg.

Tibble (with Code) for Questions 21-23

```
pcare <- read_csv("data/pcare.csv", show_col_types = FALSE)
pcare
```

```
# A tibble: 111 x 7
   subj_id   age smoke totChol BMIcat     heartRate diaBP
     <dbl> <dbl> <chr>   <dbl> <chr>          <dbl> <dbl>
 1      69    47 Yes       300 Healthy           76  60
 2      98    40 No        205 Obese             60  60
 3      90    41 No        274 Overweight        80  61.5
 4      46    49 No        208 Healthy           65  63
 5     105    42 Yes       173 Healthy           65  63
 6      15    39 Yes       226 Healthy           85  64
 7     103    45 Yes       268 Healthy           63  64
 8      52    51 No        216 Healthy           90  66
 9      70    52 No        302 Healthy           63  67.5
10      26    47 Yes       294 Healthy           62  68
# ... with 101 more rows
```

21 Question 21

How many of the subjects in `pcare` are not currently smoking, are either overweight or obese, and have a diastolic blood pressure above 100 mm Hg?

22 Question 22

Write a single line of R code that will fit a linear regression model to predict heart rate on the basis of age, as well as smoking status, using the `pcare` tibble. Your resulting model should allow both the slope and the intercept of the age-heart rate relationship to change based on the subject's smoking status. Be sure that your code will work, and in particular, that you haven't spelled anything incorrectly.

23 Question 23

The two versions of the Figure for Question 23 (labeled A and B, on the next page) each show the same data (from the `pcare` tibble), using two different plotting approaches.

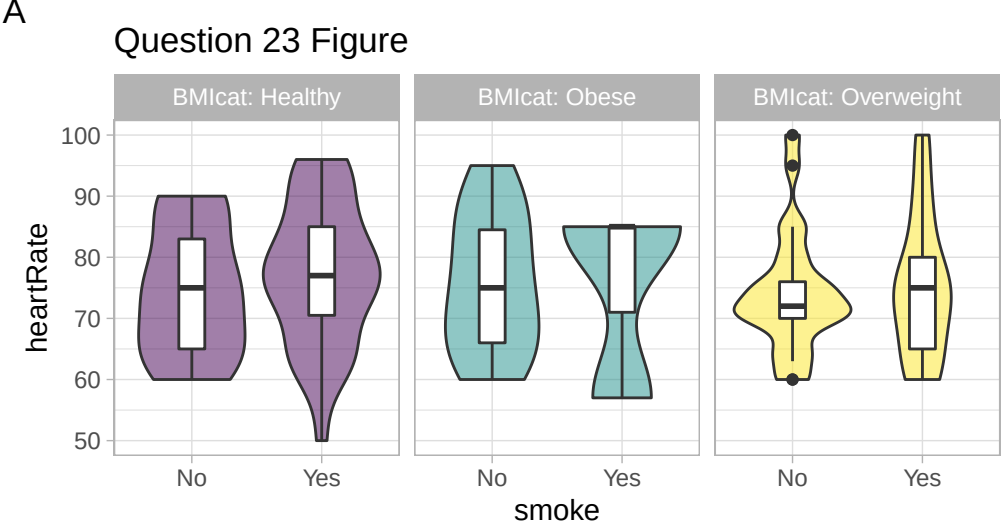
Note that the code for Plot A is provided below, accompanied by line numbers. I obtained Plot B by making two changes to the code for Plot A.

Specify the **two** changes to the code that must be made to Plot A in order to obtain Plot B. Do this by specifying what your new code would be only for the lines that require changes.

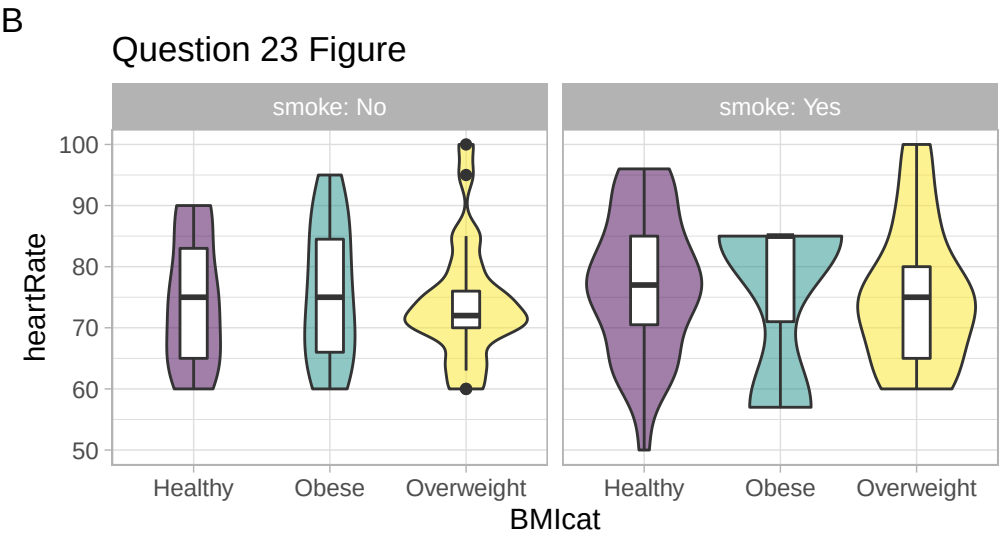
```
## Code for Plot A

[1] ggplot(pcare, aes(x = smoke, y = heartRate, fill = BMIcat)) +
[2]  geom_violin() +
[3]  geom_boxplot(fill = "white", width = 0.2) +
[4]  facet_wrap(~ BMIcat, labeller = "label_both") +
[5]  guides(fill = "none") +
[6]  scale_fill_viridis_d(alpha = 0.5) +
[7]  labs(title = "Question 23 Figure") +
[8]  theme_light()
```

A


Question 23 Figure
BMIcat: Healthy
BMIcat: Obese
BMIcat: Overweight
heartRate
100
90
80
70
60
50
No
Yes
smoke

B


Question 23 Figure
smoke: No
smoke: Yes
heartRate
100
90
80
70
60
50
Healthy
Obese
Overweight
BMIcat

24 Question 24

Classify each of the following variables by their type.

The rows are:

a. Highest level of education completed (grade school, high school, college, higher than college)
b. Cause of death (for instance, homicide, heart failure, etc.)
c. Total body calcium of a patient with osteoporosis (to the nearest gram)
d. Province of residence for a group of Canadian citizens.
e. Days between attacks for a patient diagnosed with relapsing-remitting multiple sclerosis.
f. Self-reported amount of learning completed, based on a four item scale with the following responses for each item: didn't learn anything, learned a little bit, learned enough to be comfortable with the topic, learned a great deal.

The columns are:

- Quantitative
- Ordinal categorical
- Nominal categorical
- It is impossible to tell

THIS IS THE END OF THE QUIZ