431 Quiz 2 for Fall 2025

Deadline: Wednesday 2025-12-03 at Noon

Thomas E. Love

2025-11-20 10:04 am

Table of contents

Instructions for Students

There are **28** questions on this Quiz and this PDF is **43** pages long. Be sure you have all **43** pages[1]. It is to your advantage to answer all **28** questions. Your score is based on the number of correct responses, so there's no chance a blank response will be correct, and a guess might be, so you should definitely answer all of the questions.

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 post the questions online, use some sort of AI to help you, or discuss the questions on this quiz with anyone other than me (not even with the teaching assistants.) You will be required to complete a short affirmation that you have obeyed these rules when submitting the Quiz.

0.1 The Google Form Answer Sheet

All of your answers should be placed in the Google Form Answer Sheet, located at…

- https://bit.ly/431-2025-quiz2-answer-form

All of your answers must be submitted through the Google Form by noon on Wednesday 2025-12-03, without exception. The form will close at 12:30 PM on that date, and no extensions will be available, so do not wait until late in the morning on Wednesday to submit your work. We will only accept responses through the Google Form.

The Google Form contains places to provide your responses to each question, and a final affirmation where you'll type in your name to tell us that you followed the rules for the Quiz. You must complete that affirmation before you can submit your responses. When you submit your results (in the same way you submit a Minute Paper) you will receive an email copy of your submission, with a link that will allow you to edit your work.

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

0.2 The Data Sets

I have provided **five** data sets to you (called `hosp650.csv`, `kid_scr.rds`, `rheuma.xlsx`, `wtchg1.xpt` and `wtchg2.rds`) that are mentioned in the Quiz.

[1] The session information ends the document, but I encourage you to read **everything** that precedes that section.

0.3 Getting Help

If you need clarification on a Quiz question, you have exactly one way of getting help:

- Ask your quiz question via email to me at **Thomas dot Love at case dot edu**.

During the Quiz period (3 PM 2025-11-20 through 12:30 PM 2025-12-03) we will not answer questions about the Quiz in TA office hours or in person.

All questions sent before 9 AM on 2025-12-03 will receive a response from me. Some tips:

- Specific questions are more likely to get helpful answers.
- I will not review your code or your English for you.
- I will not tell you if your answer is correct, or if it is complete.
- I will email all students if we find an error in the Quiz that needs fixing.

> WARNING! I will be away for a week during the break!
>
> I will check my email on Thursday 2025-11-20, Friday 2025-11-21, Saturday 2025-11-22 **and then not again until** Sunday 2025-11-30.
> This means that from Sunday 2025-11-23 through Saturday 2025-11-29, your emailed questions about the Quiz will sit patiently and await my return.

0.4 When Should I Ask For Help?

- 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 me (Dr. Love) a question via email.
- 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.

> If you are stuck, ASK A QUESTION
>
> You should **NEVER** spend more than 20 minutes on any question without asking me for help. Just email me at **THOMAS DOT LOVE AT CASE DOT EDU**.

0.5 Scoring the Quiz

The 28 questions are worth between 3 and 8 points, yielding a total possible score of 108 points. The questions are not in any particular order, except that Questions 1 and 2 are longer (and worth more points) than most of the rest. In addition, Questions 6, 8, 18-20, 23 and 27 each require you to write one to three sentences. The Questions range in difficulty from “things I expect everyone to get right” to “things that are deliberately tricky”.

0.6 How long should this take?

The Quiz is meant to take 6 hours. I expect most students will take 4-8 hours, and some will take as little as 3 or as many as 10.

Some questions will unquestionably take more time than others.

0.7 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 can contain **at most** two pipes, although you may or may not need the pipe in any particular setting. Note that I exclusively used the `|>` pipe, and not the `%>%` pipe, in developing this Quiz, but you may use either in your responses.

You should not include the `library` command at any time for your responses on the Quiz that ask for code. Assume in all questions that all of the packages listed below have been loaded in R.

0.8 Setting Parameters

1. If you need to **set a seed** in any Quiz 2 situation, please use `43123` as that seed.
2. If you need to run a set of bootstrap replications, use **1000** replications.
3. For each question requiring the specification of a **confidence level**, I have specified that level in the question. In most (but not *all*) questions, I am using **90%** as the confidence level.

0.9 R Packages and Love-431.R script

Here is a list of packages in R. This doesn't mean you need to use all of these packages, and it also doesn't mean that I actually used all of them in building the Quiz and its answer sketch. It means only that I didn't use any other packages, other than **xfun** for session information, so you can do the entire Quiz without using any other packages.

I will provide this listing to you at https://github.com/THOMASELOVE/431-quizzes-2025/tree/main/quiz2 so you can easily copy and paste it into your own RStudio session.

```r
library(broom)
library(car)
library(DescTools)
library(Epi)
library(GGally)
library(glue)
library(gt)
library(haven)
library(here)
library(infer)
library(janitor)
library(knitr)
library(mice)
library(MKinfer)
library(mosaic)
library(naniar)
library(olsrr)
library(patchwork)
library(readxl)
library(rstanarm)
library(easystats)
library(tidyverse)

source(here("data/Love-431.R"))

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

The actual questions start on the next page. Good luck!

Tip

Statistics is a "getting the details right" business. Read each question carefully, to be sure you get the details right in your response.

1 Question 1 (6 points)

Write a clear and well-composed essay (at least 150 words) describing an important idea from David Spiegelhalter's *The Art of Statistics* about doing statistical science well that Dr. Love **did not discuss** in Classes 1-23. Your essay should state the idea in your own words, and should indicate why you feel it is important.

If you quote Spiegelhalter (and we prefer that you do), specify the Chapter containing your quote. If I discussed your idea in class, you'll lose 1 of 6 available points. If your essay is unclear, or if you miss Spiegelhalter's point, that will have a bigger impact on your score. Each Chapter in Spiegelhalter includes a summary of key points. Feel free to use these to help spark ideas, but do not quote the summaries.

As you'll see in the Google Form Answer Sheet, you can either type your response directly into the form, or upload a Word, PDF or Google Doc containing your response to this Question.

The instructions for Question 1, including this sentence, are 175 words long.

2 Question 2 (8 points)

Suppose we want to estimate the change in weight (in pounds) for 50 overweight male adult subjects who enter into a strict nutritional regimen. In that regimen, subjects drink nothing other than water, and eat nothing but a variety of potatoes for two weeks, then spend four weeks eating only high-nutrition vegetables, and still drinking only water.

Each subject is measured before beginning the six-week regimen, and then again after the regimen is complete. Subjects were selected at random from a large group of overweight male adults who were interested in the study and met initial screening criteria.

An analyst prepares the output for Question 2 to follow, using both the `wtchg1.xpt` and `wtchg2.rds` data files. Note that `wtchg1.xpt` is a SAS transport file that can be read into R thanks to the **haven** package, while `wtchg2.rds` is an R tibble.

The analyst used each of these files to help build the Question 2 output to follow.

Tip

The `wtchg1` file shows the data in one orientation, while the `wtchg2` file shows another orientation for exactly the same data.
The first three rows from each file are shown below.

```r
head(wtchg1, 3)
```

```
# A tibble: 3 x 4
  subject before after  diff
    <dbl>  <dbl> <dbl> <dbl>
1       1    263   253    10
2       2    257   268   -11
3       3    223   227    -4
```

```r
head(wtchg2, 3)
```

```
# A tibble: 3 x 3
  subject phase  weight
    <dbl> <fct>   <dbl>
1       1 before    263
2       1 after     253
3       2 before    257
```

Question 2 has three parts.

a. (2 points) First, we are interested in building appropriate point and 98% uncertainty interval estimates for the mean weight change during the regimen. Specify which of the five Results prepared by the analyst and shown below (Result 1, 2, 3, 4 or 5) is most appropriate in this setting, given the information available to you.

b. (3 points) Write a complete English sentence or two to tell us why your selected result in Question 2a. is the most appropriate choice.

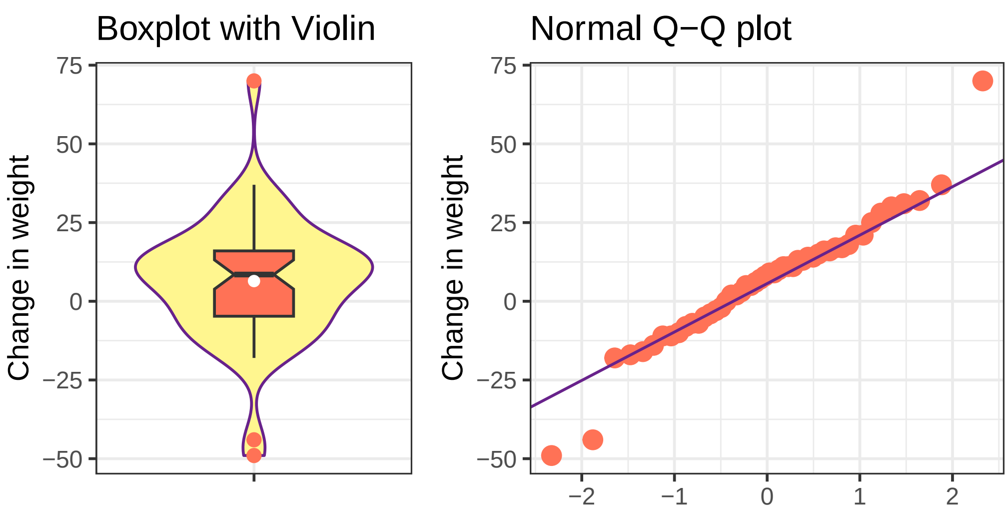
c. (3 points) For the Result you selected in Question 2a., write 2-4 English sentences stating and carefully interpreting the point and 98% interval estimates in their appropriate context for this study.

As you've noted, we are using a **98% confidence level** for Question 2.

The analyst's output follows on **the next 5 pages**.

Some Question 2 data summaries

Plot 2A. Weight Change (Before − After diet) in pounds



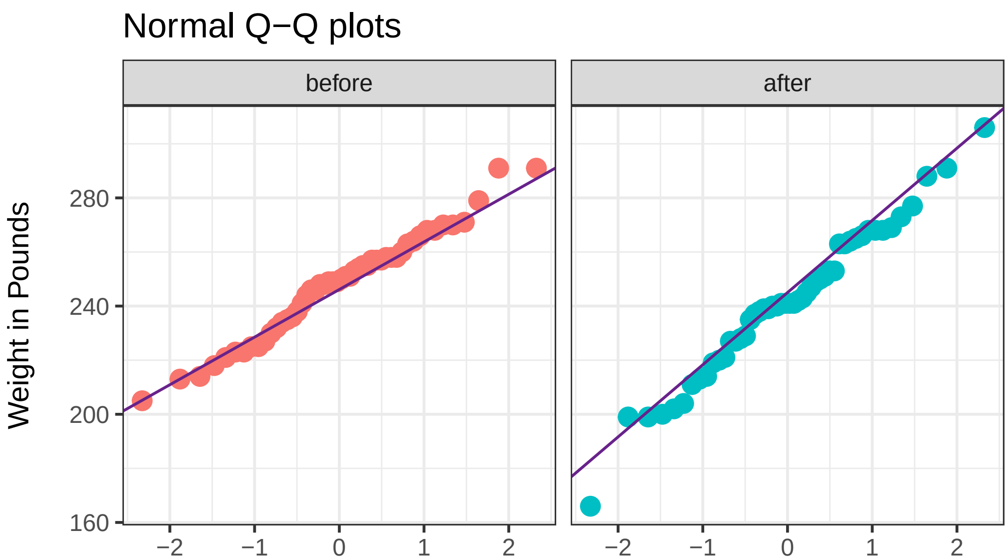
White dot indicates sample mean.

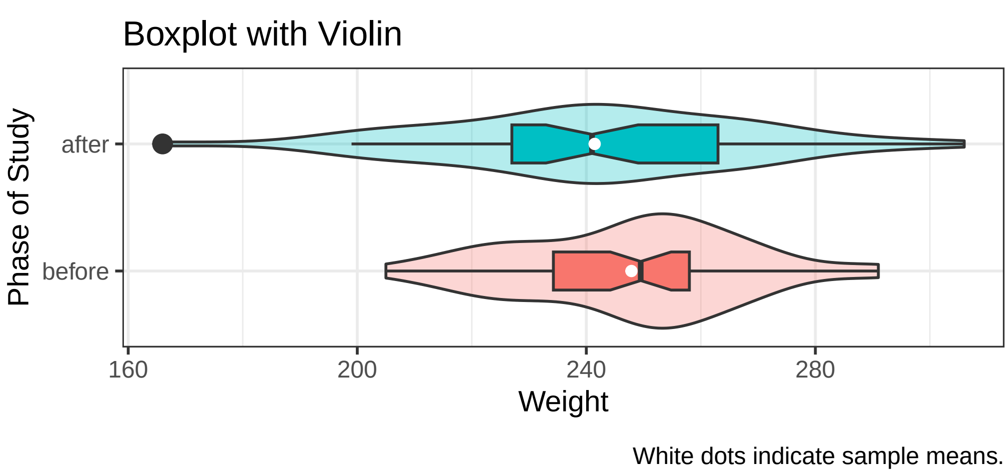
```
# Correlation Matrix (pearson-method)

Parameter1 | Parameter2 |    r |        98% CI | t(48) |         p
-----------------------------------------------------------------
before     |      after | 0.69 | [0.47, 0.83] |  6.68 | < .001***

p-value adjustment method: Holm (1979)
Observations: 50
```

Results from lovedist() for Question 2

variable	n	miss	mean	sd	med	mad	min	q25	q75	max
before	50	0	247.88	19.43	249.5	20.0	205.0	234.2	258.0	291.0
after	50	0	241.46	27.01	241.0	30.4	166.0	227.0	263.0	306.0
diff	50	0	6.42	19.46	8.5	14.8	-49.0	-4.8	16.0	70.0

Plot 2B. Weights Before and After Diet in pounds





Result 1 for Question 2

```r
result1 <- lm(weight ~ phase, data = wtchg2)
model_parameters(result1, ci = 0.98, pretty_names = FALSE, include_info = TRUE)
```

```
Parameter   | Coefficient |   SE |           98% CI | t(98) |      p
--------------------------------------------------------------------
(Intercept) |      247.88 | 3.33 | [240.01, 255.75] | 74.50 | < .001
phaseafter  |       -6.42 | 4.71 | [-17.55,   4.71] | -1.36 | 0.176

Model: weight ~ phase (100 Observations)
Sigma: 23.526 (df = 98)
RMSE : 23.290
R2: 0.019; adjusted R2: 0.009

Uncertainty intervals (equal-tailed) and p-values (two-tailed) computed
  using a Wald t-distribution approximation.
```

Result 2 for Question 2

```r
result2 <- lm(diff ~ 1, data = wtchg1)
model_parameters(result2, ci = 0.98, pretty_names = FALSE, include_info = TRUE)
```

```
Parameter   | Coefficient |   SE |         98% CI | t(49) |     p
-----------------------------------------------------------------
(Intercept) |        6.42 | 2.75 | [-0.20, 13.04] |  2.33 | 0.024

Model: diff ~ 1 (50 Observations)
Sigma: 19.459 (df = 49)
RMSE : 19.264
R2: 0.000; adjusted R2: 0.000

Uncertainty intervals (equal-tailed) and p-values (two-tailed) computed
  using a Wald t-distribution approximation.
```

Result 3 for Question 2

```
set.seed(43123)

result3 <- boot.t.test(weight ~ phase, data = wtchg2,
                       var.equal = FALSE,
                       conf.level = 0.98, R = 1000)

result3
```

```
Bootstrap Welch Two Sample t-test

data:  weight by phase
number of bootstrap samples:  1000
bootstrap p-value = 0.17
bootstrap difference of means (SE) = 6.42156 (4.652475)
98 percent bootstrap percentile confidence interval:
 -3.7206 16.8014
```

> **Note**
>
> I have edited out the second part of `result3`.

Result 4 for Question 2

```
set.seed(43123)

x_bar <- wtchg1 |> observe(response = diff, stat = "mean")

result4 <- wtchg1 |>
  specify(response = diff) |>
  generate(reps = 1000, type = "bootstrap") |>
  calculate(stat = "mean") |>
  get_confidence_interval(level = 0.98, type = "percentile")

result4 <- result4 |> mutate(point_est = x_bar$stat) |>
  relocate(point_est)
```

Result 4 for Question 2 (continued)

```
result4 |> kable(digits = 2)
```

point_est	lower_ci	upper_ci
6.42	0.52	12.36

Result 5 for Question 2

```
result5 <- t.test(weight ~ phase, data = wtchg2,
                  var.equal = FALSE, conf.level = 0.98)
model_parameters(result5)
```

```
Welch Two Sample t-test

Parameter | Group | phase = before | phase = after | Difference
---------------------------------------------------------------
weight    | phase |         247.88 |        241.46 |       6.42

Parameter |         95% CI | t(89.01) |     p
---------------------------------------------
weight    | [-4.73, 17.57] |     1.36 | 0.176

Alternative hypothesis: true difference in means between group before and group after is not
```

> **Note**
>
> The end of Result 5 for Question 2 reads:
>
> ```
> Alternative hypothesis: true difference in means between group
> before and group after is not equal to 0
> ```

This is the end of Question 2

Setup for Questions 3-16

Questions 3 - 16 each use the `survey15_2014_to_2025.csv` file which is found at this link. The file contains responses to 15 items that I have asked of students in the first week of 431, since 2014. A PDF of the survey itself is found on our Class 2 README.

The `survey15_2014_to_2025.csv` file includes a subject identification code called `student`, and the `year` (2014 through 2025) for their response.

Ingest the data into R (producing the tibble `surv15`) using the following code[2].

```
url1 <- "https://raw.githubusercontent.com/THOMASELOVE/431-data/refs/heads/main/data/survey15

surv15 <- read_csv(url1, show_col_types = FALSE) |>
  janitor::clean_names() |>
  select(student, height_in, statfuture, english, haircut,
         smoke, hand_span, lecture, alone, lastsleep,
         pulse, year) |>
  mutate(across(where(is.character), as_factor)) |>
  mutate(student = as.character(student),
         year = as_factor(year),
         smoke = fct_recode(factor(smoke), "Never" = "1",
                            "Former" = "2", "Current" = "3"))
```

Here are the first six rows of the resulting tibble.

```
surv15 |> head()
```

```
# A tibble: 6 x 12
  student height_in statfuture english haircut smoke   hand_span lecture alone
  <chr>       <dbl>      <dbl> <fct>     <dbl> <fct>       <dbl>   <dbl> <dbl>
1 202501         63          7 n             0 Never        19         2     3
2 202502         68          6 n            20 Current      20         4     3
3 202503         75          7 y            20 Never        24         2     2
4 202504         62          6 n             0 Never        20.1       3     3
5 202505         64          6 n             0 Never        23         3     3
6 202506         71          6 y             0 Never        23.5       2     2
# i 3 more variables: lastsleep <dbl>, pulse <dbl>, year <fct>
```

[2]This code is available at https://github.com/THOMASELOVE/431-quizzes-2025/tree/main/quiz2 so you can copy and paste it into RStudio.

3 Question 3 (3 points)

The `hand_span` variable in the **`surv15`** data stores hand span lengths, in centimeters.

How many of the students who have a non-missing response to `hand_span` gave responses that fell within two standard deviations of the sample mean of that measure?

4 Question 4 (3 points)

What *percentage* of students with a response to `hand_span` would we expect to see in your count in Question 3 if the Normal distribution was a *perfect* model for the distribution?

Round your answer to zero decimal places.

5 Question 5 (3 points)

The `smoke` variable in **`surv15`** divides subjects into Never, Former or Current smokers.

Among the subjects identified as Current smokers, identify the subject with the largest number of missing values in the `surv15` data. For that subject, which variables show an observed data value (i.e. are not missing)?

CHOOSE ALL VARIABLES WITH AN OBSERVED VALUE FOR THIS SUBJECT.

a. height_in
b. statfuture
c. english
d. haircut
e. hand_span
f. lecture
g. alone
h. lastsleep
i. pulse

6 Question 6 (5 points)

Ava and Ivy are two analysts, working with the **surv15** data to compare mean `haircut` prices (in US dollars) across groups defined by whether or not the subject's `height_in` is more than 67 inches, using a **90% confidence level**[3]. Each analyst assumes all missing `haircut` prices and `height_in` values are missing completely at random, and thus uses only the subjects with complete data on both variables. They produce the output shown below.

```
q06 <- surv15 |> filter(complete.cases(haircut, height_in)) |>
  mutate(height_tall = factor(height_in > 67))

q06 |> tabyl(height_tall) |> adorn_pct_formatting()
```

```
 height_tall   n percent
       FALSE 347   54.0%
        TRUE 295   46.0%
```

```
set.seed(43123)
boot.t.test(haircut ~ height_tall, conf.level = 0.90, R = 1000, data = q06)
```

```

    Bootstrap Welch Two Sample t-test

data:  haircut by height_tall
number of bootstrap samples:  1000
bootstrap p-value < 0.001
bootstrap difference of means (SE) = 12.15301 (2.622742)
90 percent bootstrap percentile confidence interval:
  7.817259 16.387134

Results without bootstrap:
t = 4.63, df = 567.83, p-value = 4.536e-06
alternative hypothesis: true difference in means is not equal to 0
90 percent confidence interval:
  7.886726 16.600120
sample estimates:
mean in group FALSE  mean in group TRUE
           37.69597            25.45254
```

[3] 46% of the subjects included had heights taller than 67 inches.

Question 6 (continued)

Ava assumes that each year's `haircut` prices follow a Normal distribution but with meaningfully different variance across the two `height` groups.

Ivy, on the other hand, agrees that the `haircut` prices in each `height` group have meaningfully different variances, but believes that they do not follow a Normal distribution.

a. (2 points) Which of the two analysts has selected the more appropriate approach? What is the point estimate and 90% confidence interval for that more appropriate approach? Round your estimates to two decimal places.

b. (3 points) In a sentence or two, explain why you think that the choice you made in Question 6a. is the more appropriate analysis.

7 Question 7 (3 points)

The `lecture` variable describes agreement (1 = Strongly Disagree to 5 = Strongly Agree) with the statement "I prefer to learn from lectures over learning from activities." The `alone` variable describes agreement (on the same 1-5 scale) with the statement "I prefer to work on projects alone over working on a team."

Create a 5x5 table with `lecture` in the rows and `alone` in the columns for the 647 subjects in **surv15** with complete information on both the `lecture` and `alone` variables. We will use this table in both Questions 7 and 8.

Your table should show one cell containing exactly three subjects. Identify the values of `lecture` and `alone` that apply to that cell.

8 Question 8 (3 points)

We performed a Pearson chi-square test using the table produced in Question 7, which yielded the following results.

```
Warning in stats::chisq.test(x, y, ...) :
  Chi-squared approximation may be incorrect

Pearson's Chi-squared test
X-squared = 112.19, df = 16, p-value < 2.2e-16
```

Based on the warning message, there is a clear reason to be concerned about the accuracy of the chi-square approximation. In a sentence or two, specify what that reason is.

9 Question 9 (3 points)

Consider the `english` variable, which specifies for each subject whether their most comfortable language is English (`english` = `y`), or some other language (`english` = `n`), and the `smoke` variable, which divides subjects into group 1 (non-smokers), or group 2 (former smokers), or group 3 (current smokers).

In Question 9, we will exclude the 8 subjects in **surv15** who have missing data on either `english` or `smoke`. Collapse the `smoke` responses of 2 and 3 into a single group.

Now, estimate the sample odds ratio for being a non-smoker for students who are most comfortable speaking in English, compared to students who are most comfortable speaking in a language other than English.

Use a Bayesian adjustment in this question (adding two to each count) **only if** any of the observed counts in the 2×2 table you build are below 20.

Provide a point estimate and the endpoints of a **90%** confidence interval for the odds ratio specified above, each rounded to two decimal places.

10 Question 10 (4 points)

Consider the responses to the item "How important do you think statistics will be in your future career?" as gathered in the `statfuture` variable. There have been seven years so far where the class had no missing responses on this item. **Use the data from only those seven years for Question 10.**

Now, estimate the proportion of subjects who responded with the value 7 (meaning extremely important), using the `saifs_ci()` approach in our `Love-431.R` script.

a. (2 points) Rounding to three decimal places, specify a 90% confidence interval (using the SAIFS approach) for the true value of the proportion of "7" responses.
b. (2 points) Is the sample proportion of "7" values from the 2025 survey, ignoring any missing responses to the `statfuture` variable, inside the interval you reported in Question 10a.?

Available responses for Question 10b are:

a. Yes, it is inside the interval
b. No, the sample proportion is lower than all values in the interval.
c. No, the sample proportion is higher than all values in the interval.

11 Question 11 (4 points)

Build a linear model (with ordinary least squares) to predict `haircut` price (in $) on the basis of `height_in` (height in inches) using the 635 subjects in **surv15** who have complete data on those two variables as well as complete data on the `english` variable. We'll call that model **fit11**.

a. (2 points) What does **fit11** predict will be the `haircut` price for a single new subject who is 64 inches tall? Round your response to two decimal places, and specify the units of measurement.

b. (2 points) Provide a **95%** prediction interval using **fit11** for the `haircut` price of the subject mentioned in Question 11a. Round each endpoint to two decimal places, but we will assume you are using the same units of measurement as you specified in Question 11a, so you don't need to specify those again here.

12 Question 12 (4 points)

Fit a new regression model (using ordinary least squares) to predict `haircut` price (in US $) on the basis of both `height_in` and `english`, again using the 635 subjects with complete data on each of those three variables. We'll call this model **fit12**.

a. (2 points) According to the (uncorrected) AIC and BIC, which of the two models (**fit11** or **fit12**) shows stronger performance?

> **Note**
>
> Options for Question 12a are:
>
> a. fit11 shows stronger performance on both AIC and BIC
> b. fit12 shows stronger performance on both AIC and BIC
> c. fit11 is better on AIC, but fit12 is better on BIC
> d. fit12 is better on BIC, but fit11 is better on AIC

b. (2 points) Specify the `student` code for the subject with the largest residual (in absolute value) for model **fit12**.

13 Question 13 (3 points)

For the model **fit12** you fit in Question 12, which of the following statements are true?

CHECK ALL OF THE TRUE STATEMENTS.

a. The fit12 model has a serious problem with its posterior predictive check.
b. The fit12 model has a serious problem with the assumption of linearity.
c. The fit12 model has a serious problem with the assumption of constant variance.
d. The fit12 model has at least one highly influential point.
e. The fit12 model has a serious problem with the assumption of Normality.
f. None of the above statements are true.

14 Question 14 (4 points)

Fit a Bayesian linear model to predict `haircut` price (in US $) on the basis of both `height_in` and `english`, again using the 635 subjects with complete data on each of those three variables, as we did in model **fit12**. Use the default weakly informative priors and set the random seed to `43123` just before fitting the model, which we'll call **fit14**. In Question 14, we will compare **fit12** to **fit14**.

Which of the following statements are true?

CHECK ALL OF THE TRUE STATEMENTS.

a. The 90% credible interval in model fit14 for the slope of `height_in` is narrower than the 90% confidence interval for the same slope in model fit12.
b. The residual standard deviation (σ) is smaller for model fit14 than it is for model fit12.
c. If we believe the assumptions (including the prior) for model fit14, then the probability is less than 0.80 that the true slope of `height_in` in the model is negative.
d. None of the statements above are true.

15 Question 15 (4 points)

The `lastsleep` variable in **surv15** contains responses to the item "How many hours did you sleep last night?"

a. (2 points) Which year has the smallest median `lastsleep` value?

b. (2 points) Build a boxplot showing the `lastsleep` data specifically for the year you identified in Question 15a. Specify the `lastsleep` values (including the units of measurement) for any points indicated by your boxplot to be outliers.

16 Question 16 (3 points)

Consider again the **surv15** data[4], but this time, restrict your analysis to students who

- responded in one of these four years: 2015, 2018, 2023 or 2025, **and**
- had a non-missing `pulse` rate

Perform an appropriate comparison of the mean pulse rates across these years, assuming (even if it's not true) that each year has a Normal distribution of pulse rates, and that the design is close enough to balanced that we will pretend it is.

Next, estimate then build a plot of pre-planned Tukey HSD confidence intervals comparing each of the six possible pairs of years on `pulse` rate that maintain 90% family-wise confidence.

Which of the following differences between pairs of means has a confidence interval that indicates that "no difference between the means" is **NOT** a reasonable conclusion?

SELECT EACH COMPARISON WHERE "NO DIFFERENCE" IS **NOT** A REASONABLE CONCLUSION

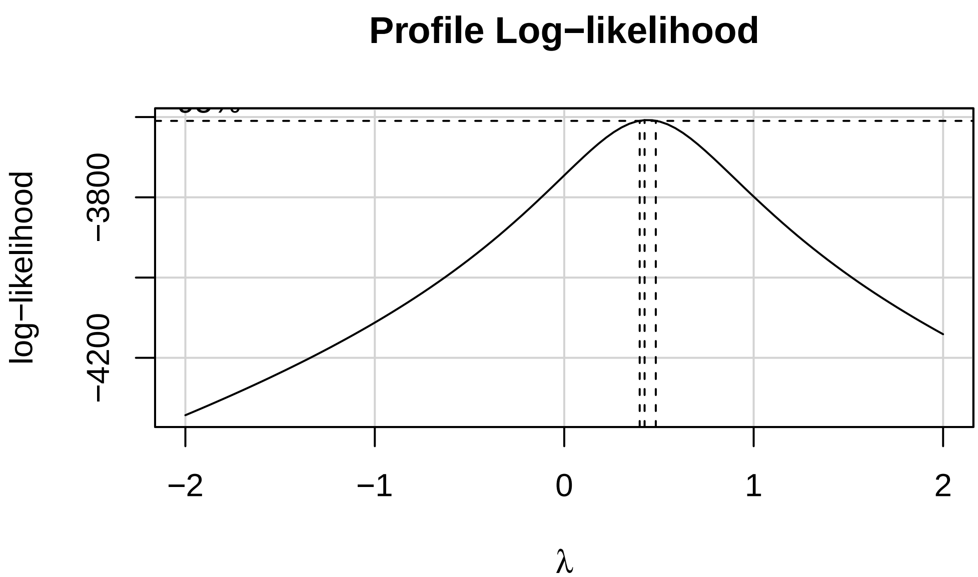
a. 2015 vs. 2018
b. 2015 vs. 2023
c. 2015 vs. 2025
d. 2018 vs. 2023
e. 2018 vs. 2025
f. 2023 vs. 2025
g. None of the above

> **Note**
>
> Question 16 is the last question that has anything to do with the **surv15** data.

[4] Don't worry. This is the last time we will use the **surv15** data in this Quiz.

17 Question 17 (3 points)

Consider the Box-Cox plot below, which was developed using a model to predict CD4 count (CD4 cells are the cells that the HIV virus kills; a normal range is about 500 - 1,500) using several predictors related to genetic makeup and several exposures of interest for a study involving 400 young men.



Which of the following is the most promising strategy for fitting a linear regression model to describe the relationship between the CD4 counts and the predictors of interest?

a. Model the inverse of CD4 count: 1/CD4 count.
b. Model the logarithm of CD4 count: $log(\text{CD4 count})$.
c. Model the square root of CD4 count: $\sqrt{\text{CD4 count}}$.
d. Model the CD4 count without transformation.
e. Model the square of CD4 count: $(\text{CD4 count})^2$.
f. None of the above.
g. We cannot tell from the information provided.

Setup for Questions 18-20

For Questions 18-20, the data in the `hosp650.csv` file describe 650 patients seen for primary care at one of two clinic locations affiliated with a metropolitan hospital. They are simulated, and include these five variables.

- `subject` = Subject Identification Number (not a meaningful code)
- `age` = subject's age, in years (all subjects are between 21 and 75)
- `insurance` = subject's primary type of insurance (4 levels: MEDICARE, COMMERCIAL, MEDICAID, UNINSURED)
- `ldl` = subject's LDL cholesterol level (in mg/dl)
- `clinic` = whether the subject receives primary care at the hospital's newly built clinic (NEW) or a specific older clinic (OLD)

18 Question 18 (3 points)

We plan to use the `hosp650` data to build a model using the main effects of `clinic` and `insurance` to predict LDL cholesterol level.

Build an appropriate interaction plot to help assess whether or not there is a large interaction apparent here.

Looking at your plot, tell us in a few sentences what you learn from it about the interaction between insurance type and clinic type on the LDL cholesterol outcome?

19 Question 19 (5 points)

Here are results from two models relevant to our discussion in Question 18.

```r
anova(fit19a)
```

```
Analysis of Variance Table

Response: ldl
                 Df Sum Sq Mean Sq F value    Pr(>F)    
insurance         3  23651  7883.8  6.4073 0.0002794 ***
clinic            1   7405  7405.4  6.0185 0.0144213 *  
insurance:clinic  3   4293  1430.8  1.1629 0.3231183    
Residuals       642 789939  1230.4                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(fit19b)
```

```
Analysis of Variance Table

Response: ldl
           Df Sum Sq Mean Sq F value    Pr(>F)    
insurance   3  23651  7883.8  6.4025 0.0002811 ***
clinic      1   7405  7405.4  6.0140 0.0144570 *  
Residuals 645 794231  1231.4                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

a. (2 points) Which model (fit19a or fit19b) is the one suggested by your interpretation of the interaction plot you built in Question 18?

b. (3 points) In a sentence or two, specify the difference in eta-square (η^2) between models fit19a and fit19b, and tell us what that difference suggests to you about which model to choose.

20 Question 20 (5 points)

Now consider adding a covariate, the subject's age, to the model you prefer (fit19a or fit19b) given what you have seen so far, to create a new model called **fit20**. Use a confidence level of 89%.

a. (2 points) Specify the resulting **fit20** regression equation, rounding all coefficients to two decimal places.

b. (3 points) Interpret the point estimate of the model's intercept term in the context of this study in a sentence or two.

Setup for Questions 21-24

The `kid_scr.rds` data below describe 1,000 children whose health was screened.

Variable	Description
`child_code`	subject (child) identifier
`waist`	child's waist circumference (cm)
`age`	child's age at screening (years)
`arm_circ`	child's arm circumference (cm)
`phys_health`	child's general health condition (4 levels)
`asthma`	parent/guardian ever told child has asthma? (Yes, No)
`plank_time`	# of seconds child holds plank position
`protein`	total protein consumed by child yesterday (g)
`water`	total plain water child drank yesterday (g)

```
data_codebook(kid_scr |> select(-child_code))
```

```
select(kid_scr, -child_code) (1000 rows and 8 variables, 8 shown)

ID | Name        | Type        | Missings |       Values |           N
---+-------------+-------------+----------+--------------+------------
1  | waist       | numeric     | 0 (0.0%) | [48, 144.7]  |        1000
---+-------------+-------------+----------+--------------+------------
2  | age         | numeric     | 0 (0.0%) |      [7, 15] |        1000
---+-------------+-------------+----------+--------------+------------
3  | arm_circ    | numeric     | 15 (1.5%) |  [15, 46.8] |         985
---+-------------+-------------+----------+--------------+------------
4  | phys_health | categorical | 0 (0.0%) |  1_Excellent | 467 (46.7%)
   |             |             |          |   2_VeryGood | 267 (26.7%)
   |             |             |          |       3_Good | 233 (23.3%)
   |             |             |          | 4_FairorPoor |  33 ( 3.3%)
---+-------------+-------------+----------+--------------+------------
5  | asthma      | categorical | 11 (1.1%) |         Yes | 186 (18.8%)
   |             |             |          |           No | 803 (81.2%)
---+-------------+-------------+----------+--------------+------------
6  | plank_time  | numeric     | 13 (1.3%) |    [1, 450] |         987
---+-------------+-------------+----------+--------------+------------
7  | protein     | numeric     | 10 (1.0%) | [4.2, 241.8] |         990
---+-------------+-------------+----------+--------------+------------
8  | water       | numeric     | 16 (1.6%) |    [0, 8591] |         984
----------------------------------------------------------------------
```

21 Question 21 (3 points)

Below is some partial code (with one missing line), designed to create a single imputation for the data in `kid_scr`, and place that imputed data set in the `kid_scr_si` tibble.

```
set.seed(43123)

kid_scr_imp_set <- mice(kid_scr, m = 1, printFlag = FALSE)

## ---- a single line of code is omitted here ----

prop_miss_case(kid_scr_si)
```

```
[1] 0
```

Provide the correct line of R code that is missing so as to complete the necessary imputation, store it in the appropriate place, and thus yield the result shown.

22 Question 22 (4 points)

Build a linear model using ordinary least squares to predict `waist` circumference using the main effects of the seven other available variables (excluding `child_code`.)

Call the model **fit22** and use the imputed data you developed in Question 21 to fit it. Do not transform the outcome or any of the predictors in fitting this model. Use a **94%** confidence level in Questions 22-24.

Which of the following statements are true about this model?

CHECK ALL OF THE TRUE STATEMENTS.

a. If Harry and Sally have the same values of the other six predictors in model fit22, but Harry's health is Fair or Poor, and Sally's is Good, fit22 predicts that Sally will have a larger waist circumference.
b. Model fit22 explains at least 75% of the variation in the outcome.
c. Model fit22 has a serious problem with collinearity.
d. A backwards elimination stepwise regression procedure using R's default settings applied to fit22 settles on a reduction of that model containing four of the original seven predictors.
e. None of these statements are true.

23 Question 23 (5 points)

We now plan to complete a multiple imputation on these data, to create a new model **fit23** containing all of the predictors in our model **fit22**.

a. (2 points) At a minimum, how many imputations should we perform in this situation, if we want to use a number of imputations that is divisible by 5?
b. (3 points) Suppose we use the number of imputations you specified in Question 23a, then fit model **fit23** across them. Estimate the R-square and adjusted R-square for this new model. Express these values as percentages, rounded to one decimal place. In a sentence or two, what do those values suggest about model **fit23**?

24 Question 24 (3 points)

Again using the model **fit23** from Question 23, suppose we find a point estimate and 94% confidence interval for the coefficient of `arm_circ` and want to carefully interpret its meaning in context. Which of the following sentences would be appropriate to include?

CHECK ALL OF THE SENTENCES YOU WANT TO INCLUDE.

a. A 94% confidence interval of the effect of increasing arm circumference by 1 cm on waist circumference, accounting for age, physical health status, asthma history, plank time, and recent protein and water consumption is consistent with values of that effect that could be positive or negative.
b. Our analysis assumes that the missing values in the kid_scr data are missing completely at random.
c. When we generalize beyond the participants in this screening to the population of children that they represent, and assuming that our linear model fit23 is correct, then our data are compatible at the 94% confidence level with asthma being associated with a reduction in waist circumference.
d. None of these statements should be included.

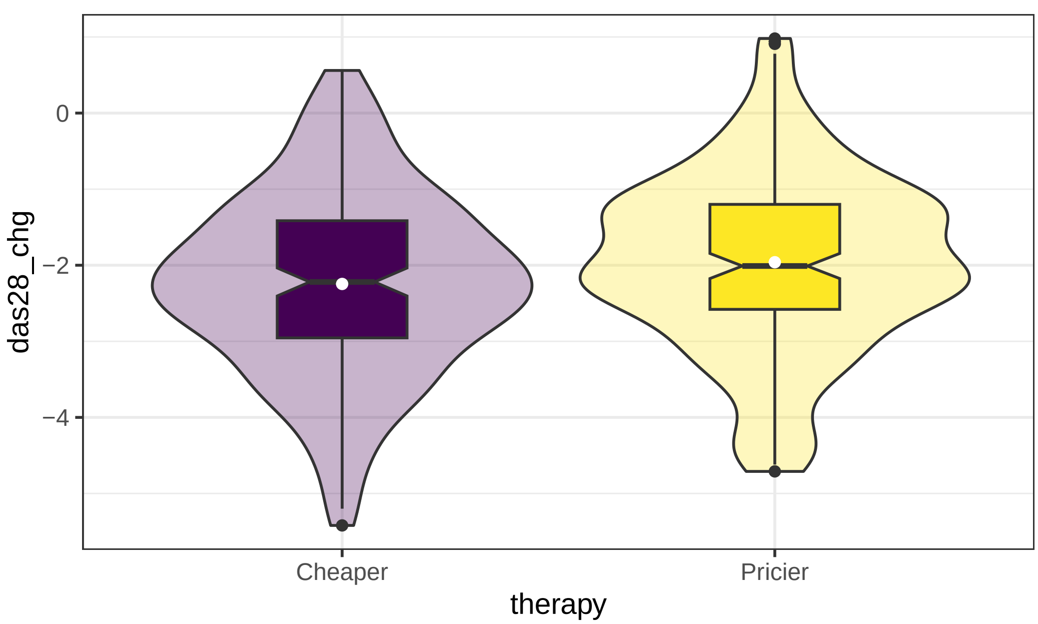
Setup for Questions 25-26

In a double-blind trial, 350 patients with active rheumatoid arthritis were randomly assigned to receive one of two therapy types: a cheaper one, or a pricier one, and went on to participate in the trial.

The primary outcome was the change in DAS28 at 48 weeks as compared to study entry. The DAS28 is a composite index of the number of swollen and tender joints, the erythrocyte sedimentation rate, and a visual-analogue scale of patient-reported disease activity. A decrease in the DAS28 of 1.2 or more (so a change of -1.2 or below) was considered to be a clinically meaningful improvement. I have provided the data in the `rheuma.xlsx` file. Here are a few potentially relevant summaries.

```
rheuma |>
  group_by(therapy) |>
  reframe(lovedist(das28_chg))
```

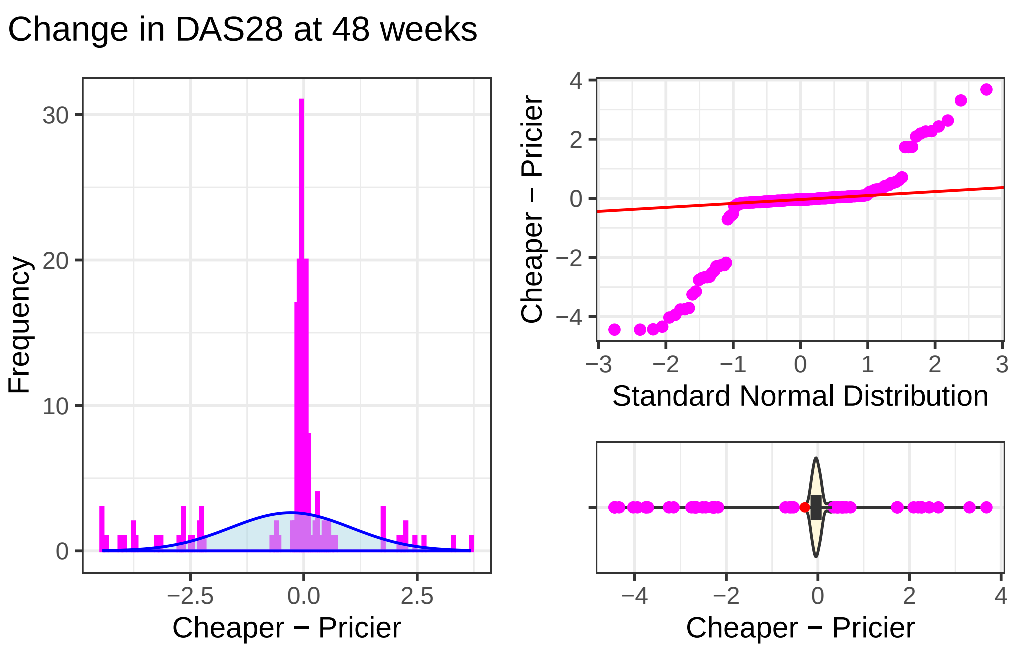
```
# A tibble: 2 x 11
  therapy     n  miss  mean    sd   med   mad   min   q25   q75   max
  <fct>   <int> <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 Cheaper   175     0 -2.25  1.19 -2.22  1.14 -5.42 -2.96 -1.42  0.56
2 Pricier   175     0 -1.96  1.17 -2.01  1.14 -4.71 -2.58 -1.2   0.98
```



White dots indicate sample means.

```r
rheuma2 <- tibble(
  Cheaper =rheuma$das28_chg[rheuma$therapy=="Cheaper"],
  Pricier =rheuma$das28_chg[rheuma$therapy=="Pricier"],
  Diff = Cheaper - Pricier)

rheuma2 |> reframe(lovedist(Diff))
```

```
# A tibble: 1 x 10
      n  miss   mean    sd     med   mad   min    q25    q75   max
  <int> <int>  <dbl> <dbl>   <dbl> <dbl> <dbl>  <dbl>  <dbl> <dbl>
1   175     0 -0.286  1.33 -0.0400 0.133 -4.44 -0.130 0.0500  3.68
```



25 Question 25 (3 points)

A student completed four analyses of the data described in the Setup for Questions 25-26, as shown below. Which of the following Analyses produces a 90% confidence interval for the change in DAS28 at 48 weeks which most appropriately compares the pricier therapy to the cheaper one?

d. Analysis D
e. Analysis E
f. Analysis F
g. Analysis G

Analysis D for Question 25

```
fit_d <- t.test(rheuma2$Pricier, rheuma2$Cheaper, paired = TRUE)
model_parameters(fit_d, ci = 0.90, pretty_names = FALSE)
```

```
Paired t-test

Parameter |   Group | Difference | t(174) |     p |       90% CI
----------------------------------------------------------------
Pricier   | Cheaper |       0.29 |   2.84 | 0.005 | [0.09, 0.48]

Alternative hypothesis: true mean difference is not equal to 0
```

Analysis E for Question 25

```
fit_e <- wilcox.test(das28_chg ~ therapy, data = rheuma,
                     conf.int = TRUE, conf.level = 0.90)
tidy(fit_e, conf.int = TRUE, conf.level = 0.90) |>
  select(-statistic, -p.value) |> kable(digits = 2)
```

estimate	conf.low	conf.high	method	alternative
-0.29	-0.5	-0.08	Wilcoxon rank sum test with continuity correction	two.sided

Analysis F for Question 25

```r
fit_f <- wilcox.test(rheuma2$Cheaper - rheuma2$Pricier,
                     conf.int = TRUE, conf.level = 0.90)
tidy(fit_f, conf.int = TRUE, conf.level = 0.90) |>
  select(-statistic, -p.value) |> kable(digits = 2)
```

estimate	conf.low	conf.high	method	alternative
-0.05	-0.07	-0.03	Wilcoxon signed rank test with continuity correction	two.sided

Analysis G for Question 25

```r
fit_g <- lm(das28_chg ~ therapy, data = rheuma)
model_parameters(fit_g, ci = 0.90, pretty_names = FALSE, include_info = TRUE)
```

```
Parameter      | Coefficient |   SE |         90% CI | t(348) |      p
---------------------------------------------------------------------
(Intercept)    |       -2.25 | 0.09 | [-2.39, -2.10] | -25.09 | < .001
therapyPricier |        0.29 | 0.13 | [ 0.08,  0.49] |   2.26 | 0.025

Model: das28_chg ~ therapy (350 Observations)
Sigma: 1.185 (df = 348)
RMSE : 1.181
R2: 0.014; adjusted R2: 0.012

Uncertainty intervals (equal-tailed) and p-values (two-tailed) computed
  using a Wald t-distribution approximation.
```

26 Question 26 (3 points)

Referring again to the study initially described in the Setup for Questions 25 and 26, which of the following analyses provides an appropriate 90% confidence interval for the difference (cheaper - pricier) in the proportion of participants who had a clinically meaningful improvement (DAS28 change of -1.2 or below) at 48 weeks?

j. Analysis J
k. Analysis K
l. Analysis L
m. Analysis M
n. None of the above.

The output for these analyses is shown on the next four pages.

Analysis J for Question 26

```
rheuma_j <- rheuma |>
  mutate(improved = das28_chg < -1.2) |>
  mutate(improved = fct_relevel(factor(improved), "FALSE"))
rheuma_j |> tabyl(therapy, improved) |> adorn_title(placement = "combined")
```

```
 therapy/improved FALSE TRUE
          Cheaper    32  143
          Pricier    45  130
```

```
twobytwo(32, 143, 45, 130,
         "cheaper", "pricier", "improved", "didn't improve")
```

```
2 by 2 table analysis:
------------------------------------------------------
Outcome   : improved
Comparing : cheaper vs. pricier

        improved didn't improve    P(improved) 95% conf. interval
cheaper       32            143         0.1829    0.1323   0.2472
pricier       45            130         0.2571    0.1978   0.3270

                                    95% conf. interval
             Relative Risk:  0.7111    0.4758   1.0629
         Sample Odds Ratio:  0.6465    0.3875   1.0784
Conditional MLE Odds Ratio:  0.6473    0.3737   1.1112
    Probability difference: -0.0743   -0.1600   0.0127

             Exact P-value: 0.1212
        Asymptotic P-value: 0.0947
------------------------------------------------------
```

Analysis K for Question 26

```r
rheuma_k <- rheuma |>
  mutate(improved = das28_chg <= -1.2) |>
  mutate(improved = fct_relevel(factor(improved), "TRUE"))
rheuma_k |> tabyl(therapy, improved) |> adorn_title(placement = "combined")
```

```
 therapy/improved TRUE FALSE
          Cheaper  145    30
          Pricier  132    43
```

```r
twobytwo(145, 30, 132, 43,
         "cheaper", "pricier", "improved", "didn't improve")
```

```
2 by 2 table analysis:
------------------------------------------------------
Outcome   : improved
Comparing : cheaper vs. pricier

        improved didn't improve    P(improved) 95% conf. interval
cheaper      145             30         0.8286    0.7654   0.8775
pricier      132             43         0.7543    0.6851   0.8124

                                   95% conf. interval
             Relative Risk: 1.0985    0.9859   1.2239
         Sample Odds Ratio: 1.5745    0.9338   2.6549
Conditional MLE Odds Ratio: 1.5724    0.9053   2.7585
    Probability difference: 0.0743   -0.0111   0.1585

             Exact P-value: 0.1140
        Asymptotic P-value: 0.0886
------------------------------------------------------
```

Analysis L for Question 26

```
rheuma_l <- rheuma |>
  mutate(improved = das28_chg <= -1.2) |>
  mutate(improved = fct_relevel(factor(improved), "TRUE"))
rheuma_l |> tabyl(therapy, improved) |> adorn_title(placement = "combined")
```

```
 therapy/improved TRUE FALSE
          Cheaper  145    30
          Pricier  132    43
```

```
twobytwo(145, 30, 134, 41, conf.level = 0.90,
         "cheaper", "pricier", "improved", "not improved")
```

```
2 by 2 table analysis:
------------------------------------------------------
Outcome   : improved
Comparing : cheaper vs. pricier

        improved not improved    P(improved) 90% conf. interval
cheaper      145           30         0.8286    0.7765   0.8705
pricier      134           41         0.7657    0.7090   0.8142

                                   90% conf. interval
             Relative Risk: 1.0821    0.9899   1.1829
         Sample Odds Ratio: 1.4789    0.9509   2.2999
Conditional MLE Odds Ratio: 1.4772    0.9200   2.3856
    Probability difference: 0.0629   -0.0080   0.1331

             Exact P-value: 0.1835
        Asymptotic P-value: 0.1450
------------------------------------------------------
```

Analysis M for Question 26

```r
rheuma_m <- rheuma |>
  mutate(improved = das28_chg < -1.2) |>
  mutate(improved = fct_relevel(factor(improved), "TRUE"))
rheuma_m |> tabyl(therapy, improved) |> adorn_title(placement = "combined")
```

```
 therapy/improved TRUE FALSE
          Cheaper  143    32
          Pricier  130    45
```

```r
twobytwo(144, 31, 131, 44, conf.level = 0.90,
         "cheaper", "pricier", "improved", "didn't improve")
```

```
2 by 2 table analysis:
------------------------------------------------------
Outcome   : improved
Comparing : cheaper vs. pricier

        improved didn't improve    P(improved) 90% conf. interval
cheaper      144             31         0.8229    0.7703   0.8655
pricier      131             44         0.7486    0.6909   0.7986

                                   90% conf. interval
             Relative Risk: 1.0992    1.0023   1.2055
         Sample Odds Ratio: 1.5602    1.0111   2.4076
Conditional MLE Odds Ratio: 1.5582    0.9794   2.4942
    Probability difference: 0.0743    0.0020   0.1457

             Exact P-value: 0.1176
        Asymptotic P-value: 0.0917
------------------------------------------------------
```

27 Question 27 (5 points)

On 2019-09-25, Maggie Koerth-Baker at FiveThirtyEight published "We've Been Fighting the Vaping Crisis Since 1937." In that article, she quotes a 2019-09-06 article at the *New England Journal of Medicine* by Jennifer E. Layden et al. entitled "Pulmonary Illness Related to E-Cigarette Use in Illinois and Wisconsin: A Preliminary Report." Quoting that report:

> E-cigarettes are battery-operated devices that heat a liquid and deliver an aerosolized product to the user. … In July 2019, the Wisconsin Department of Health Services and the Illinois Department of Public Health received reports of pulmonary disease associated with the use of e-cigarettes (also called vaping) and launched a coordinated public health investigation…. We defined case patients as persons who reported use of e-cigarette devices and related products in the 90 days before symptom onset and had pulmonary infiltrates on imaging and whose illnesses were not attributed to other causes.

In the study, 53 case patients were identified, but some patients gave no response to the question of whether or not "they had used THC (tetrahydrocannabinol) products in e-cigarette devices in the past 90 days." 33 of the 41 reported THC use. Assume those 41 subjects are a random sample of all case patients that will appear in Wisconsin and Illinois in 2019.

Estimate an appropriate 90% confidence interval for the **PERCENTAGE** of case patients in Illinois and Wisconsin in 2019 that used THC in the 90 days prior to symptom onset using the Agresti-Coull method as demonstrated in class using a function from the **mosaic** package. Note that I've emphasized the word **PERCENTAGE** here, so as to stop you from instead presenting a proportion.

a. (2 points) Specify your point estimate of this **PERCENTAGE**, and then the lower and upper bound for your 90% confidence interval, in each case rounded to a single decimal place.
b. (3 points) Carefully explain the meaning of your 90% confidence interval in the context of this study in a few sentences.

28 Question 28 (3 points)

Suppose we are considering six potential sets of predictors in regression models for the same outcome, and obtain the following results in our test sample. We will assume all six models show comparable performance and adherence to regression assumptions in the model development sample.

Model	RMSPE	Mean APE	Maximum APE	Validated R^2
1	10	8	15	0.63
2	14	10	14	0.57
3	11	10	16	0.61
4	7	9	13	0.59

Consider model X to be dominated by model Y when model Y shows better results on all four summary measures shown in the table above. Which of these models are dominated across these summaries by other models, so that we should no longer consider them in light of this analysis?

(CHECK ALL OF THE MODELS THAT ARE DOMINATED.)

a. Model 1
b. Model 2
c. Model 3
d. Model 4
e. None of these models are dominated.

This is the end of the Quiz. Congratulations!

Extra Note

If you've gotten this far and are still reading, it may pay off. Perhaps you'd like an opportunity to earn a few more points. To do so, send me an email (at `Thomas` dot `Love` at `case` dot `edu`) before Thursday **2025-12-04 at 9 AM** with the subject line "My Best Visualization" containing these two things:

(1) an attached screenshot or other small file containing a clear image of the single best visualization that **you** built using R this semester, for any purpose, inside this class or not. Be sure that there is only one image in your attached file.
(2) In one to three clear, complete and grammatically correct English sentences, tell me why you selected that visualization, and provide a little context for why you built it.

Anything that you think counts as a "visualization" (a table or a plot is what I have in mind) can be the choice.

This activity will be worth double the points that our typical "hidden" activities are worth, and, **in addition**, completing it to my satisfaction will earn you an extra 3 points on this Quiz.

This is the third and final opportunity for additional credit I have posted since Quiz 1. The others (which are not in this Quiz) are also due shortly after the Thanksgiving Break.

Session Information

```r
xfun::session_info()
```

```
R version 4.5.2 (2025-10-31 ucrt)
Platform: x86_64-w64-mingw32/x64
Running under: Windows 11 x64 (build 26200)

Locale:
  LC_COLLATE=English_United States.utf8
  LC_CTYPE=English_United States.utf8
  LC_MONETARY=English_United States.utf8
  LC_NUMERIC=C
  LC_TIME=English_United States.utf8

Package version:
```

abind_1.4-8	arrangements_1.1.9	askpass_1.2.1
backports_1.5.0	base_4.5.2	base64enc_0.1-3
bayesplot_1.14.0	bayestestR_0.17.0	BH_1.87.0.1
bigD_0.3.1	bit_4.6.0	bit64_4.6.0-1
bitops_1.0.9	blob_1.2.4	boot_1.3-32
broom_1.0.10	bslib_0.9.0	cachem_1.1.0
callr_3.7.6	car_3.1-3	carData_3.0-5
cellranger_1.1.0	checkmate_2.3.3	class_7.3-23
cli_3.6.5	clipr_0.8.0	cmprsk_2.2-12
coda_0.19-4.1	codetools_0.2-20	colourpicker_1.3.0
commonmark_2.0.0	compiler_4.5.2	conflicted_1.2.0
correlation_0.8.8	cowplot_1.2.0	cpp11_0.5.2
crayon_1.5.3	crosstalk_1.2.2	curl_7.0.0
data.table_1.17.8	datasets_4.5.2	datawizard_1.3.0
DBI_1.2.3	dbplyr_2.5.1	Deriv_4.2.0
desc_1.4.3	DescTools_0.99.60	digest_0.6.38
distributional_0.5.0	doBy_4.7.0	dplyr_1.1.4
DT_0.34.0	dtplyr_1.3.2	dygraphs_1.1.1.6
e1071_1.7-16	easystats_0.7.5	effectsize_1.0.1
emmeans_2.0.0	Epi_2.61	estimability_1.5.1
etm_1.1.2	evaluate_1.0.5	Exact_3.3
exactRankTests_0.8-35	expm_1.0-0	farver_2.1.2
fastmap_1.2.0	fontawesome_0.5.3	fontBitstreamVera_0.1.1
fontLiberation_0.1.0	fontquiver_0.2.1	forcats_1.0.1
foreach_1.5.2	Formula_1.2-5	fs_1.6.6
gargle_1.6.0	gdtools_0.4.4	generics_0.1.4
GGally_2.4.0	ggformula_1.0.0	ggiraph_0.9.2
ggplot2_4.0.1	ggridges_0.5.7	ggstats_0.11.0
gld_2.6.8	glmnet_4.1-10	glue_1.8.0
gmp_0.7-5	goftest_1.2-3	googledrive_2.1.2
googlesheets4_1.1.2	graphics_4.5.2	grDevices_4.5.2
grid_4.5.2	gridExtra_2.3	gt_1.1.0
gtable_0.3.6	gtools_3.9.5	haven_2.5.5
here_1.0.2	highr_0.11	hms_1.1.4
htmltools_0.5.8.1	htmlwidgets_1.6.4	httpuv_1.6.16
httr_1.4.7	ids_1.0.1	igraph_2.2.1
infer_1.0.9	inline_0.3.21	insight_1.4.2
isoband_0.2.7	iterators_1.0.14	janitor_2.2.1
jomo_2.7-6	jquerylib_0.1.4	jsonlite_2.0.0
juicyjuice_0.1.0	knitr_1.50	labeling_0.4.3
labelled_2.16.0	later_1.4.4	lattice_0.22-7
lazyeval_0.2.2	lifecycle_1.0.4	litedown_0.8
lme4_1.1-37	lmom_3.2	loo_2.8.0

lubridate_1.9.4
magrittr_2.0.4
markdown_2.0
MASS_7.3-65
Matrix_1.7-4
MatrixModels_0.5.4
matrixStats_1.5.0
memoise_2.0.1
methods_4.5.2
mgcv_1.9-4
mice_3.18.0
miceadds_3.18-36
microbenchmark_1.5.0
mime_0.13
miniUI_0.1.2
minqa_1.2.8
mitml_0.4-5
mitools_2.4
MKdescr_0.9
MKinfer_1.2
modelbased_0.13.0
modelr_0.1.11
mosaic_1.9.2
mosaicCore_0.9.5
mosaicData_0.20.4
multcomp_1.4-29
mvtnorm_1.3-3
naniar_1.1.0
nlme_3.1-168
nloptr_2.2.1
nnet_7.3-20
norm_1.0.11.1
nortest_1.0-4
numDeriv_2016.8-1.1
olsrr_0.6.1
openssl_2.3.4
ordinal_2023.12.4.1
otel_0.2.0
pan_1.9
parallel_4.5.2
parameters_0.28.2
patchwork_1.3.2
pbkrtest_0.5.5
performance_0.15.2
pillar_1.11.1
pkgbuild_1.4.8
pkgconfig_2.0.3
plyr_1.8.9
posterior_1.6.1
prettyunits_1.2.0
processx_3.8.6
progress_1.2.3
promises_1.5.0
proxy_0.4-27
ps_1.9.1
purrr_1.2.0
quantreg_6.1
QuickJSR_1.8.1
R6_2.6.1
ragg_1.5.0
rappdirs_0.3.3
rbibutils_2.4
RColorBrewer_1.1-3
Rcpp_1.1.0
RcppArmadillo_15.0.2.2
RcppEigen_0.3.4.0.2
RcppParallel_5.1.11-1
Rdpack_2.6.4
reactable_0.4.4
reactR_0.6.1
readr_2.1.6
readxl_1.4.5
reformulas_0.4.2
rematch_2.0.0
rematch2_2.1.2
report_0.6.2
reprex_2.1.1
reshape2_1.4.5
rlang_1.1.6
rmarkdown_2.30
rootSolve_1.8.2.4
rpart_4.1.24
rprojroot_2.1.1
rstan_2.32.7
rstanarm_2.32.2
rstantools_2.5.0
rstudioapi_0.17.1
rvest_1.0.5
S7_0.2.1
sandwich_3.1-1
sass_0.4.10
scales_1.4.0
see_0.12.0
selectr_0.4.2
shape_1.4.6.1
shiny_1.11.1
shinyjs_2.1.0
shinystan_2.6.0
shinythemes_1.2.0
snakecase_0.11.1
sourcetools_0.1.7.1
SparseM_1.84.2
splines_4.5.2
StanHeaders_2.32.10
stats_4.5.2
stats4_4.5.2
stringi_1.8.7
stringr_1.6.0
survival_3.8-3
sys_3.4.3
systemfonts_1.3.1
tensorA_0.36.2.1
textshaping_1.0.4
TH.data_1.1-5
threejs_0.3.4
tibble_3.3.0
tidyr_1.3.1
tidyselect_1.2.1
tidyverse_2.0.0
timechange_0.3.0
tinytex_0.57
tools_4.5.2
tzdb_0.5.0
ucminf_1.2.2
UpSetR_1.4.0
utf8_1.2.6
utils_4.5.2
uuid_1.2.1
V8_8.0.1

vctrs_0.6.5 viridis_0.6.5 viridisLite_0.4.2
visdat_0.6.0 vroom_1.6.6 withr_3.0.2
xfun_0.54 xml2_1.4.1 xplorerr_0.2.0
xtable_1.8-4 xts_0.14.1 yaml_2.3.10
zoo_1.8-14