431 Quiz 2

Thomas E. Love

Version: 2021-10-21 11:16:01

Instructions for Students

There are 22 questions on this Quiz, and this PDF is 18 pages long. Be sure you have all 18 pages.

It is to your advantage to answer all of the 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.

The Google Form Answer Sheet

All of your answers should be placed in the Google Form Answer Sheet, located at https://bit.ly/431-2021-quiz2-answer-sheet. All of your answers must be submitted through the Google Form by 11 PM on Monday 2021-10-25, without exception. The form will close at that time, and no extensions will be made available, so do not wait until late Monday evening to submit. We will not accept any responses except through the Google Form.

The Google Form will contain 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 and then submit your results. 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 results.

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.

Scoring and Timing

All questions are worth 4, 5 or 6 points as indicated, 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 class time on Thursday 2021-10-28.

IMPORTANT NOTES: On Writing Code

1. 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.
2. You need not include the `library` command at any time in your responses on the Google Form. Assume in all questions that all relevant packages have been loaded in R.
3. If you are asked to complete a bootstrap method, use the default number of bootstrap replications (this is `B = 1000` for the Hmisc package's `smean.cl.boot` and `B.reps = 2000` for Dr. Love's `bootdif` within `Love-boost.R`.) Use these defaults by simply not setting a value for `B` or `B.reps` in calling the relevant function. Be sure in either case that you have set a seed properly immediately before running the bootstrap procedure.
4. When completing any procedure that requires random sampling, use the command `set.seed(4312021)` to set your random seed, and use `4312021` as that random seed. Do this at the start of the chunk of R code where you use the procedure that requires a set of random numbers, and do it again if you need a new set of random numbers later in the Quiz.

Getting Help

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 post on Piazza to the instructors. (This is the only kind of post you will be able to make on Piazza during the Quiz.)
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 2021-10-25 at that time, if not sooner.

A few cautions:

- 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.

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.

The Data Sets

We have provided four data sets that are mentioned in the Quiz. You will find them in our Shared Google Drive in the Quiz 2 folder. They may be helpful to you.

- `quiz_wtchg`, first mentioned in Question 08
- `quiz_beds1` and `quiz_beds2`, first mentioned in Question 09
- `quiz_hosp`, first mentioned in Question 21

Packages used in building this document

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.

```
library(broom)
library(car)
library(Epi)
library(equatiomatic)
library(glue)
library(ggrepel)
library(Hmisc)
library(janitor)
library(knitr)
library(magrittr)
library(mosaic)
library(naniar)
library(patchwork)
library(pwr)
library(tidyverse)

source("data/Love-boost.R")

theme_set(theme_bw())
```

1 Question 01 (4 points)

Once a confidence interval is calculated, several design changes may be used by a researcher to make a confidence interval wider or narrower. For the changes listed in each of the rows below, indicate the impact of that change the width of the confidence interval by selecting the correct column.

Rows:

1. Increase the level of confidence.
2. Increase the sample size.
3. Increase the standard error of the estimate.
4. Use a bootstrap instead of a t-based approach to estimate the CI.

Columns:

a. CI will become wider
b. CI will become narrower
c. CI width will not change
d. It is impossible to tell

2 Question 02 (4 points)

Suppose you have a tibble with two variables. One is a factor called Exposure with levels High, Low and Medium, arranged in that order, and the other is a quantitative outcome. You want to rearrange the order of the Exposure variable so that you can then use it to identify for ggplot2 a way to split histograms of outcomes up into a series of smaller plots, each containing the histogram for subjects with a particular level of exposure (Low then Medium then High.)

Which of the pairs of `tidyverse` functions identified below has Dr. Love used to accomplish such a plot?

a. `fct_reorder` and `facet_wrap`
b. `fct_relevel` and `facet_wrap`
c. `fct_collapse` and `facet_wrap`
d. `fct_reorder` and `group_by`
e. `fct_collapse` and `group_by`

3 Question 03 (5 points)

Suppose that 95 of 155 applicants from students at private undergraduate institutions to a graduate school are accepted, while 160 of 345 from students at public undergraduate institutions are accepted. Estimate a two-sided 95% confidence interval for the relative risk of acceptance into graduate school for a "private undergrad" applicant as compared to a "public undergrad" applicant. Round your response to two decimal places, and provide both the point estimate and confidence interval.

4 Question 04 (4 points)

Each of the 500 applicants described in Question 03 applied to exactly one program at the graduate school: either Program A, B or C. Breaking down the applications, we find a few additional facts in addition to those specified in Question 03. In particular, we now also know that:

- Program A received 200 applications, and accepted 75.
- Program A accepted 35 of its 60 applicants who came from private schools.
- Program B received 150 applications in total.
- Program B accepted exactly half of its 20 applicants from private schools.
- Program C accepted 40 of its 75 applicants from public schools.
- Program C rejected 25 applicants from private schools.

Which of the following statements are true? (**CHECK ALL THAT APPLY.**)

a. Students from private schools have lower odds of being accepted into Program A than do students from public schools.
b. Students from private schools have lower odds of being accepted into Program B than do students from public schools.
c. Students from private schools have lower odds of being accepted into Program C than do students from public schools.
d. None of these statements are true.
e. There is insufficient information to decide which of statements a-c are true.

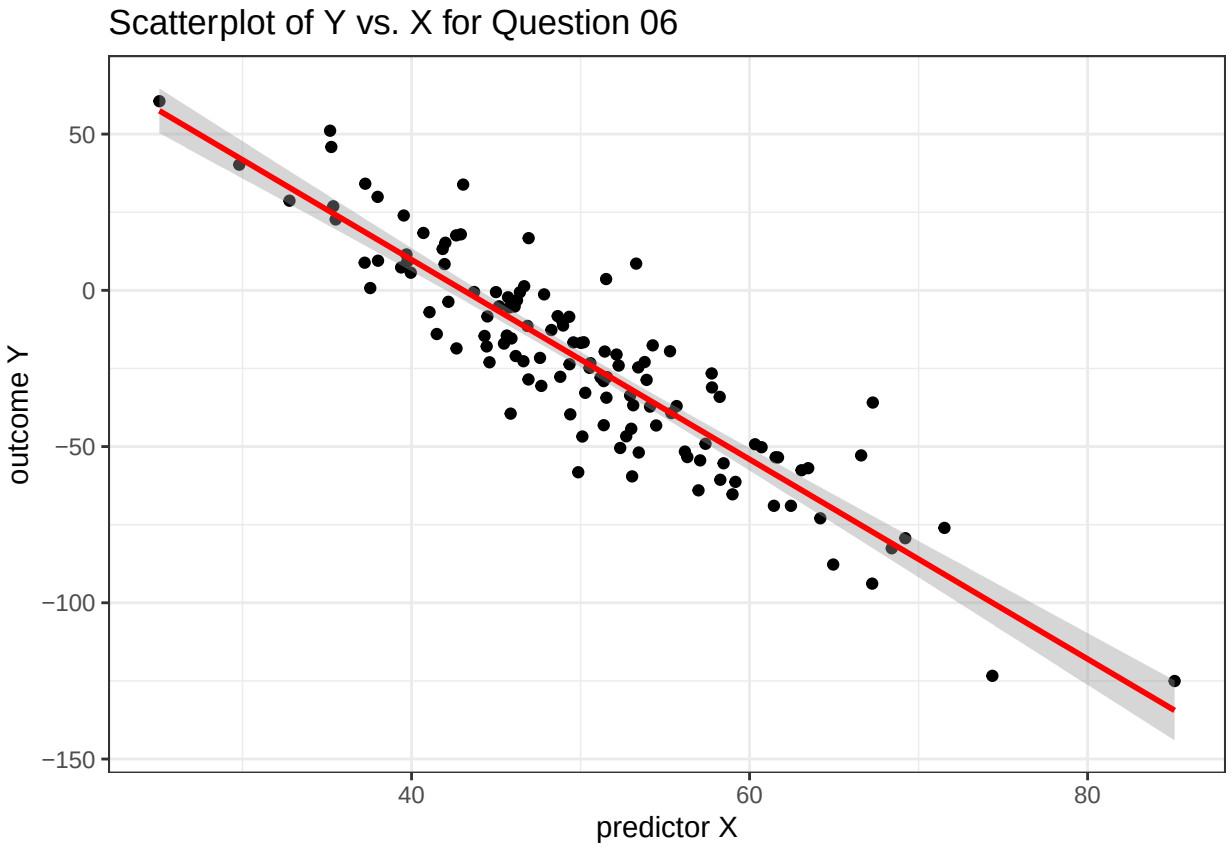
5 Question 05 (4 points)

A study reports that the sensitivity of a particular type of mammogram as a screening test for detecting breast cancer is 0.85, while its specificity is 0.80.

In a population of 1,000,000 women in which 2,500 women actually have breast cancer, what is the probability that a woman has breast cancer given that her mammogram is positive? Round your answer to three decimal places.

6 Question 06 (4 points)

Figure for Question 06



Consider the following possible summaries of a linear model fit to predict Y from X using 125 observations, describing the scatterplot shown in the Figure for Question 06. Which of these summaries is correct?

a. Model: y = 138 - 3.2 x, with R-squared = -0.82
b. Model: y = 138 - 3.2 x, with R-squared = -0.32
c. Model: y = 138 + 3.2 x, with R-squared = 0.82
d. Model: y = 138 + 3.2 x, with R-squared = 0.32
e. Model: y = 138 - 3.2 x, with R-squared = 0.82
f. Model: y = 138 - 3.2 x, with R-squared = 0.32
g. Model: y = 3.2 + 138 x, with R-squared = -0.82
h. Model: y = 3.2 - 138 x, with R-squared = -0.32
i. Model: y = -3.2 + 138 x, with R-squared = 0.82
j. Model: y = -3.2 + 138 x, with R-squared = 0.32
k. Model: y = 3.2 + 138 x, with R-squared = 0.82
l. Model: y = 3.2 + 138 x, with R-squared = 0.32

7 Question 07 (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.

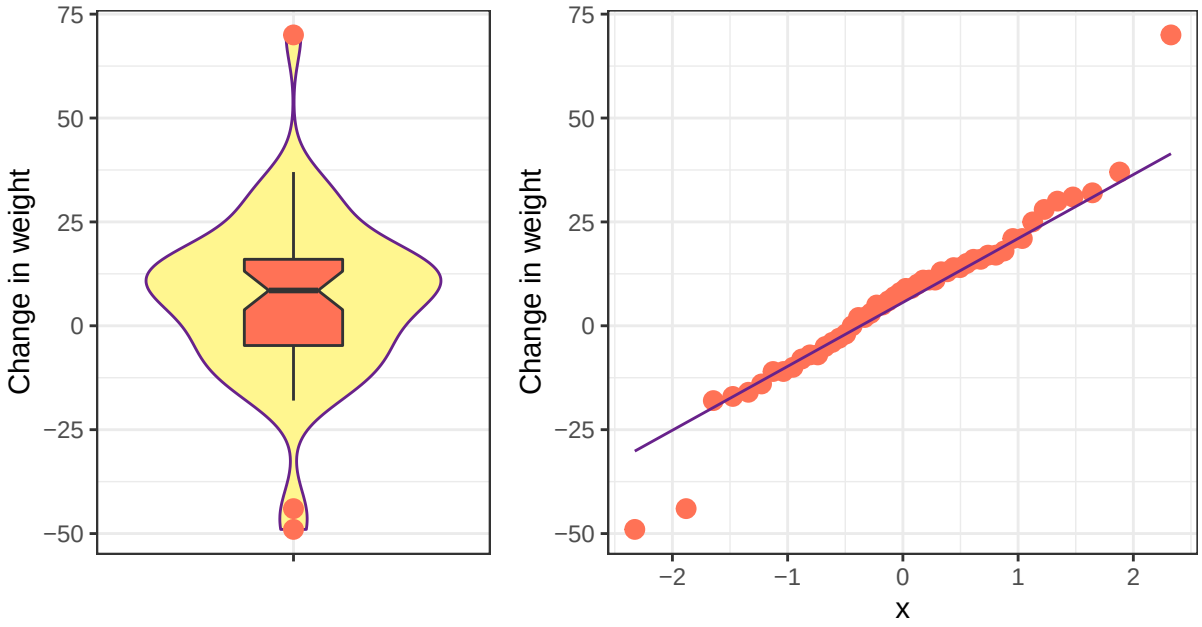
The entire report is available at https://www.nejm.org/doi/full/10.1056/NEJMoa1911614. In the study, 53 case patients were identified, but some patients gave no response to the question of whether or not "they had used THC (tetrahydrocannabinol) products in e-cigarette devices in the past 90 days." 33 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.

Use a SAIFS procedure to 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. Note that I've emphasized the word **PERCENTAGE** here, so as to stop you from instead presenting a proportion. Specify your point estimate of this **PERCENTAGE**, and then the lower and upper bound for your confidence interval, in each case rounded to a single decimal place.

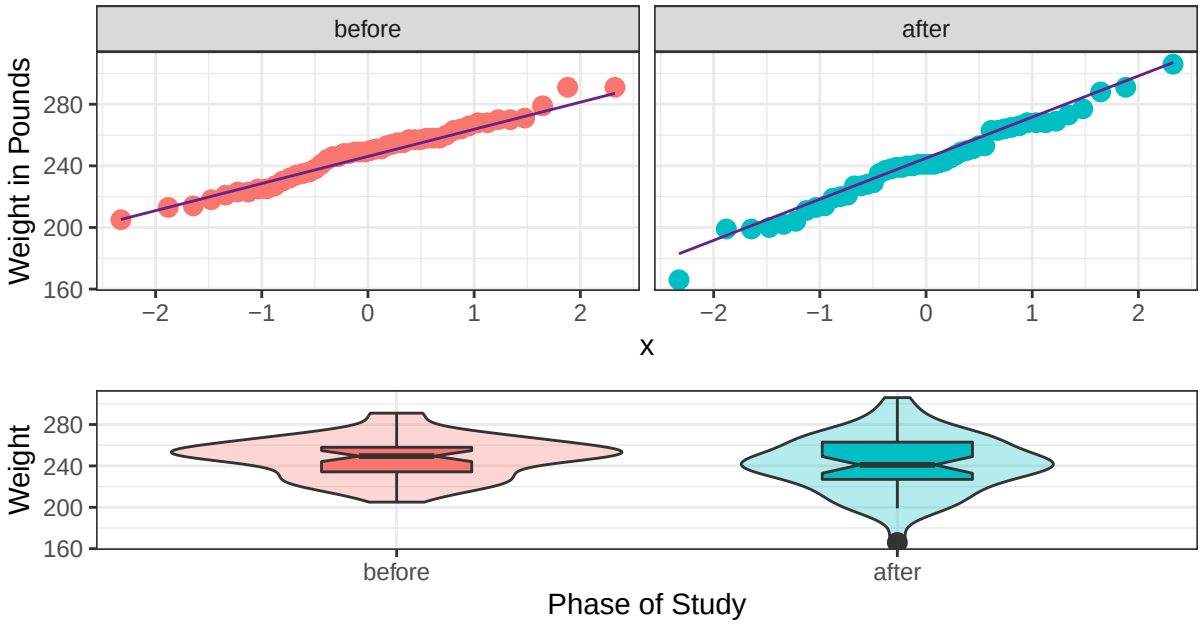
8 Question 08 (6 points)

Suppose we compare the change in weight (before - after, in pounds) for 50 overweight male adult subjects who enter into a rather strict nutritional regimen. Specifically, the 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 nothing but water. The team's statistical analyst prepares the following output, and I have provided you with the `quiz_wtchg` file containing the data, as well.

Plot 8A. Weight Change (Before − After diet) in pounds for n = 50 subjects



Plot 8B. Weights Before and After Diet in pounds for n = 50 subjects



Question 08 (continued)

Result 1 for Question 08

method	estimate	p.value	conf.low	conf.high	conf.level
One Sample t-test on `diffs`	6.80	0.0116	-0.149	13.749	0.99

Result 2 for Question 08

method	mean	conf.low	conf.high	conf.level
`Hmisc::smean.cl.boot` on `diffs`	6.800	0.059	13.402	0.99

Result 3 for Question 08

method	estimate	p.value	conf.low	conf.high	conf.level
Two Sample t-test for `before - after`	6.80	0.134	-5.034	18.634	0.99

Result 4 for Question 08

method	estimate	p.value	conf.low	conf.high	conf.level
Welch Two Sample t-test for `before - after`	6.80	0.134	-5.036	18.637	0.99

Result 5 for Question 08

method	mean	conf.low	conf.high	conf.level
`bootdif` from `Love-boost.R` for `before - after`	6.80	-3.94	18.46	0.99

Write a short essay to respond to Question 08 Three complete sentences should be sufficient. In that essay, you should:

a. specify which of the five Results above (1, 2, 3, 4 or 5) is most appropriate, and be specific about why it is the most appropriate choice
b. tell us what that result indicates about what can be said based on the available confidence intervals about the mean weight loss in a population represented by this sample, and in particular that population value's relationship to zero, justifying your decision by referring to the output provided.

9 Question 09 (6 points)

Consider the number of community hospital beds per 1000 population that are available in each of the 50 US states as well as the District of Columbia. The data for both 2010 and 2014 are provided for you in the `quiz_beds1` data set (where the number of beds per 1000 population are saved under the variable name `beds_index` and the indicator of year under the name `year`), and again in the `quiz_beds2` data set, using a different format, where the values for 2010 are saved under the variable name `beds_2010` and the values for 2014 are saved under the variable name `beds_2014`.

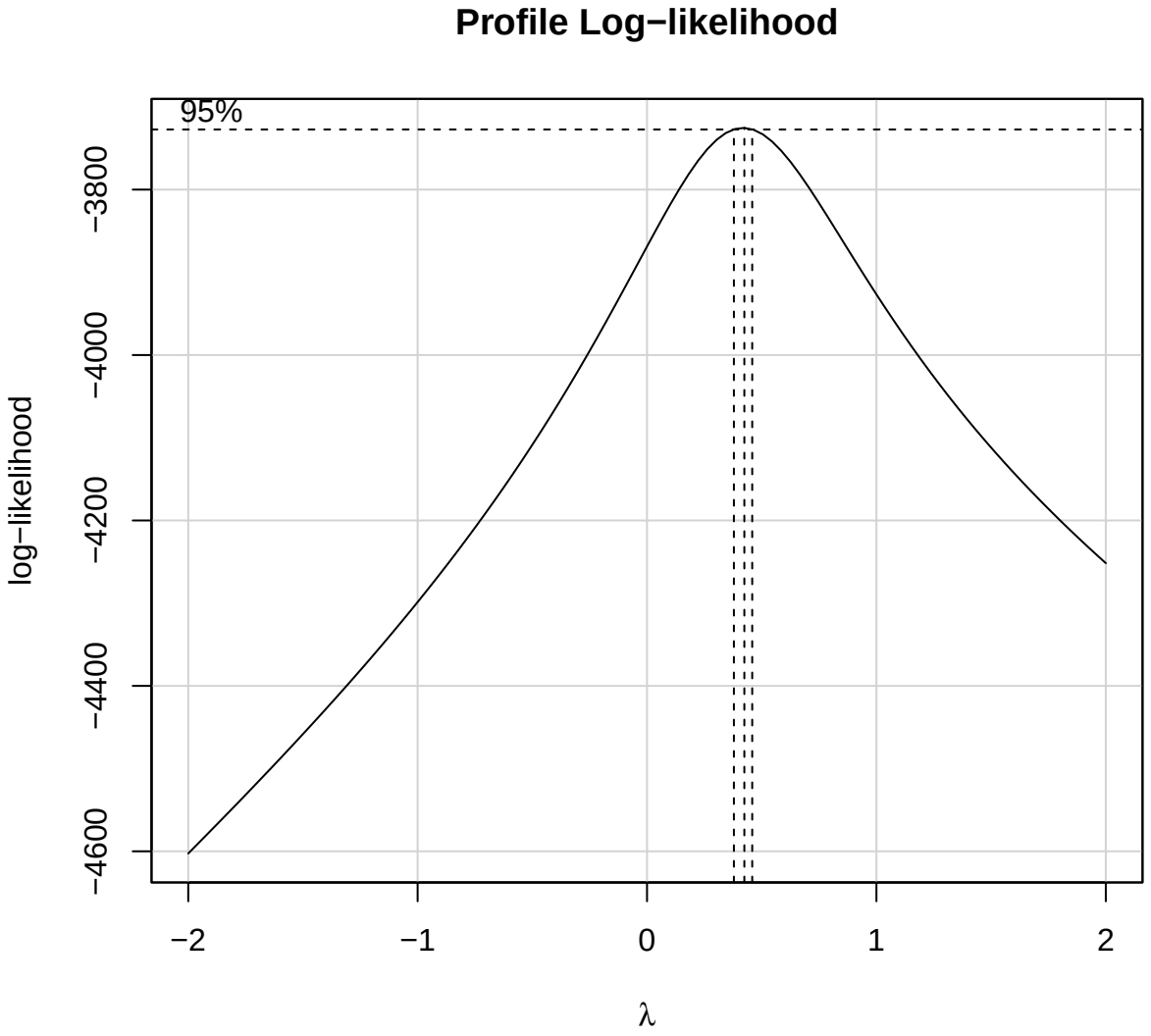
Use your understanding of the data to establish an appropriate 90% confidence interval for the true difference in the mean number of hospital beds comparing 2010 and 2014. Note that the appropriate point estimate for the mean difference (2010 - 2014) is 0.161 beds per 1000 residents.

Which of the following approaches is an appropriate way to obtain the desired confidence interval? (CHECK ALL THAT APPLY.)

a. `quiz_beds1 %$% t.test(beds_index ~ year, var.equal = TRUE, conf.level = 0.90)` which yields (-0.116, 0.438).
b. `quiz_beds1 %$% t.test(beds_index ~ year, var.equal = TRUE, alpha = 0.10)` which yields (-0.170, 0.492).
c. `set.seed(4312021); quiz_beds1 %$% bootdif(beds_index, year, conf.level = 0.90)`, which yields (-0.106, 0.422).
d. `quiz_beds2 %$% t.test(beds_2010 - beds_2014, conf.level = 0.90)`, which yields (0.126, 0.196).
e. `quiz_beds2 %$% t.test(beds_2010 - beds_2014, sig.level = 0.10)`, which yields (0.119, 0.203).
f. `set.seed(4312021); quiz_beds2 %$% smean.cl.boot(beds_2010 - beds_2014, conf = 0.90)`, which yields (0.127, 0.194).
g. `set.seed(4312021); quiz_beds2 %$% smean.cl.boot(beds_2010 - beds_2014)`, which yields (0.122, 0.202).
h. None of these responses are appropriate.

10 Question 10 (4 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 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.

11 Question 11 (5 points)

Consider the `starwars` tibble that is part of the `dplyr` package in the tidyverse. Filter the data file to focus on individuals who are of the Human species, who also have complete data on both their height and mass. Then use a t-based approach to estimate an appropriate 90% confidence interval for the difference between the mean body-mass index of Human males minus the mean body-mass index of Human females, without assuming that the population variances of males and females are the same. The data provides `height` in centimeters and `mass` in kilograms. You'll need to calculate the body-mass index (BMI) values - the appropriate formula to obtain BMI in our usual units of $\frac{kg}{m^2}$ is:

$$\text{BMI} = \frac{10,000 * \text{mass in kg}}{(\text{height in cm})^2}$$

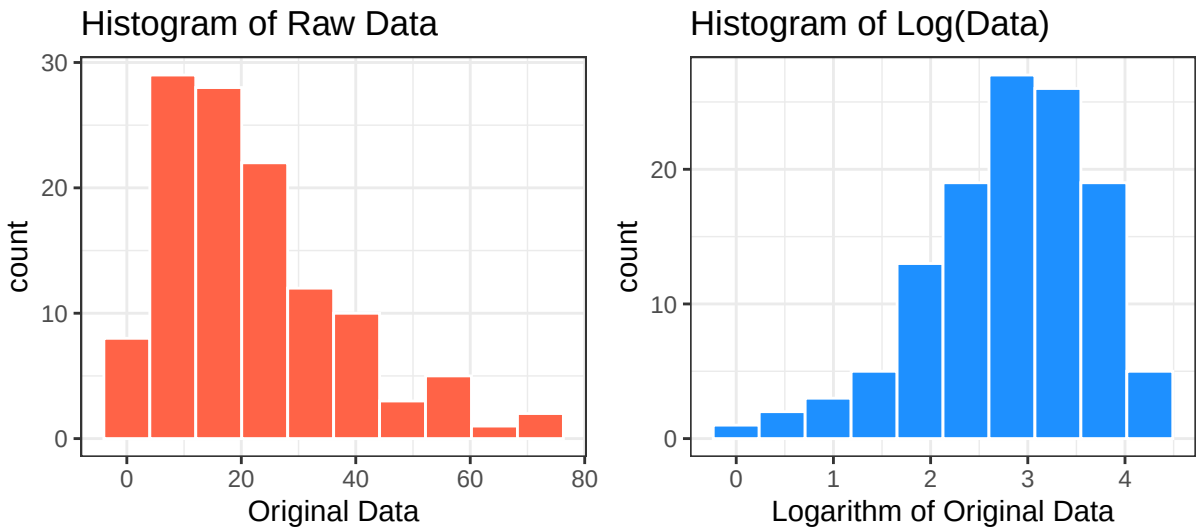
Specify your point estimate, and then the lower and upper bound, each rounded to a single decimal place, and be sure to specify the units of measurement.

12 Question 12 (4 points)

Consider the two histograms shown in the Figure for Question 12. On the left, we show the original data set, in a red color. On the right, we show the natural logarithm of the data, in a blue color. Assuming you are unsatisfied with assuming a Normal distribution for each of these expressions of the data, what transformation would the ladder of power transformations recommend next, in an effort to re-express the data in a form that could be modeled effectively using a Normal distribution?

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

Figure for Question 12



13 Question 13 (4 points)

A series of 96 models were built by a team of researchers interested in systems biology. 45 of the models showed promising results in an attempt to validate them out of sample. Define the hit rate as the percentage of models built that show these promising results. Consider the following approaches to form a confidence interval that describes the uncertainty we have around a hit rate estimate in this setting, while permitting a 10% rate of Type I error. Some of the methods below are appropriate in this setting.

Which of the methods below are appropriate in this setting? (CHECK ALL THAT APPLY.)

a. An Agresti-Coull procedure.
b. A linear regression procedure.
c. A Clopper-Pearson procedure.
d. A "plus4" procedure.
e. A rank-based procedure.
f. A Score procedure without continuity correction.

14 Question 14 (4 points)

Which one (of those appropriate methods you identified in Question 13) provides the narrowest confidence interval for the setting described in Question 13? Pick the one best response.

a. An Agresti-Coull procedure.
b. A linear regression procedure.
c. A Clopper-Pearson procedure.
d. A "plus4" procedure.
e. A rank-based procedure.
f. A Score procedure without continuity correction.

15 Question 15 (6 points)

An investigator plans to replicate part of a study of the gut hormone fragment peptide YY_{3-36} (PYY) which reduces appetite and food intake when infused into subjects of normal weight. The original study is found at https://www.nejm.org/doi/full/10.1056/nejmoa030204, if you are curious.

> In common with the adipocyte hormone leptin, PYY reduces food intake by modulating appetite circuits in the hypothalamus. However, in obesity there is a marked resistance to the action of leptin, which greatly limits its therapeutic effectiveness.

The investigator wants to know whether obese subjects are also resistant to the anorectic effects of PYY. She intends to perform a randomized, placebo-controlled, double-blind crossover study on healthy obese subjects (including **an equal number of male subjects and female subjects**), with each subject studied on two occasions one week apart.

The subjects will be screened by a dietitian who will assess their eating behavior with (several established scales). The protocol specifies that:

> (Study participants will) complete a three-day diet diary to permit us to assess their usual eating habits. Food preferences were assessed at screening (to) ensure that the food offered at the buffet lunch is acceptable. ... The subjects' food intake for the 48 hours before each study day (will be) standardized, and during this period they (will complete) food diaries to confirm compliance. ... (On each study day, cannulas will be) inserted into veins in both forearms, one for the collection of blood and the other for the infusion of PYY or saline. ... Two hours after the infusion, the subjects (will be) offered a buffet lunch with food in such excess that all appetites (will be) satisfied. The amounts of food and water (will be) quantified preprandially and postprandially, and the caloric intake calculated.

On two consecutive Thursdays, we will measure caloric intake during a buffet lunch offered two hours after the infusion of that week's exposure. In one of the weeks, the subject will receive an infusion of PYY, and in the other week (with the order of the weeks determined at random) the subject will receive a placebo. The number of calories consumed at each lunch is measured and then converted to an appetite rating. Our primary outcome is the difference between the appetite rating after PYY and the appetite rating after placebo.

A clinically meaningful difference, the investigator tells you, would be one in which these comparisons would differ by 30 or more points on the appetite rating scale comparing the two infusions, which is 60% of the anticipated standard deviation of these results. The investigator then asks what the *smallest possible* number of patients is that she will have to enroll and gather data from in order to have at least 90% power to detect an effect of this size using a 5% two-tailed significance level, and to meet all other requirements described above.

This question is in two parts.

In part a, specify the number that the investigator asked for.

In part b, specify a single line of R code (you may use up to two pipes) that yields the information necessary to complete part a.

16 Question 16 (4 points)

I fit four models using linear regression to predict left ventricular ejection fraction (LVEF) for a sample of 120 adults who are suspected of having heart failure. A summary of some key results follows:

Model	Predictors used	R^2	Adjusted R^2	AIC	Residual SD
A	use of hydralazine	.135	.134	503	5.24
B	age, use of hydralazine	.365	.357	422	3.12
C	age, sex, use of hydralazine	.404	.343	458	2.92
D	age, use of hydralazine, angina, sex	.406	.354	456	2.91

Identify the model that ...

Rows:

1. explains the largest fraction of the outcome's variation for these 120 adults
2. shows the smallest violations of linear regression modeling assumptions
3. displays the best performance using the Akaike information criterion
4. would be most improved by using a transformation of LVEF as the outcome

Columns:

a. Model A
b. Model B
c. Model C
d. Model D
e. It is impossible to tell.

17 Question 17 (4 points)

Based on data about which you know very little, a researcher compares populations using means and obtains a point estimate of 0.81, and a 90% confidence interval with endpoints (0.51, 1.02), then comes to you for help.

Which of the following procedures could have created such an interval? (CHECK ALL THAT APPLY.)

a. A two-sample pooled t procedure.
b. A paired-samples t procedure.
c. A bootstrap procedure applied to paired samples.
d. A Wilcoxon rank sum test procedure.
e. None of the above.

18 Question 18 (4 points)

Suppose you have a data frame named `mydata` containing a variable called `sbp`, which shows the participant's systolic blood pressure in millimeters of mercury. Which of the following lines of code will create a new variable `badbp` within the `mydata` data frame which takes the value **TRUE** when a subject has a systolic blood pressure that is at least 120 mm Hg, and **FALSE** when a subject's systolic is less than 120 mm Hg.

a. `mydata %>% badbp <- sbp >= 120`
b. `mydata$badbp <- ifelse(mydata$sbp >= 120, "YES", "NO")`
c. `badbp <- mydata %>% filter(sbp >= 120)`
d. `mydata %>% mutate(badbp = sbp < 120)`
e. None of these will do the job.

19 Question 19 (4 points)

If the characteristics of a sample approximate the characteristics of its population in every respect, then which of the statements below is true? (**CHECK ALL THAT APPLY.**)

a. The sample is random
b. The sample is accidental
c. The sample is stratified
d. The sample is systematic
e. The sample is representative
f. None of the above

20 Question 20 (4 points)

Suppose we wanted to do a new study of people guessing my age, at the 5% significance level. We believe we can enroll 225 people in total, 100 of whom are male, who can then be asked to guess my age.

Find the power for a two-sided t test to compare age guesses, if we assume a minimum meaningful difference between male and female observers in terms of age guesses is 2.5 years, and assuming that the standard deviation of age guesses is 5 years in both males and females. Which of the following options best describes the power we obtain in this case?

a. Less than 80%.
b. Between 80% and 84.9%.
c. Between 85% and 89.9%.
d. Between 90% and 94.9%.
e. 95% or higher.
f. It is impossible to estimate the power in this setting.

21 Question 21 (5 points)

The `quiz_hosp` data I have provided describe five characteristics of 680 simulated patients at seen for out-patient care at a local hospital. Available are:

- `person` = Subject Identification Number (not a meaningful code)
- `sex` = the patient's sex (FEMALE or MALE)
- `statin` = does the patient have a prescription for a statin medication (YES or NO)
- `insurance` = the patient's insurance type (MEDICARE, COMMERCIAL, MEDICAID, UNINSURED)
- `hsgrads` = the percentage of adults in the patient's home neighborhood who have at least a high school diploma (this measure of educational attainment is used as an indicator of the socio-economic place in which the patient lives)

Using the `quiz_hosp` data, what is the 95% confidence interval for the odds ratio which compares the odds of receiving a statin if you are MALE divided by the odds of receiving a statin if you are FEMALE. Show the point and interval estimates, rounded to two decimal places. Do **NOT** use a Bayesian augmentation in responding to this question.

22 Question 22 (6 points)

Again using the `quiz_hosp` data we discussed in the previous question, begin by identifying the two largest insurance groups (in terms of numbers of people in the data with that insurance type.)

Your task for this Question is to compare the population average of "percentage of adults in the patient's home neighborhood who have at least a high school diploma" across those two insurance groups. Use a 90% confidence interval and an appropriate bootstrap procedure, with the default number of bootstrap replications. Be sure to set your seed to `4312021` immediately before running your code.

Your response should be a single complete sentence in English which specifies clearly the point estimate and the end points of the confidence interval.

- All numbers in your sentence should be rounded to two decimal places.
- When we read your sentence, we must learn which insurance groups you are comparing, and we must understand the direction of the effect you observed (at least in your point estimate.)
- Do not overstate your conclusions.

Session Information (Dr. Love's Setup)

```
sessionInfo()
```

```
R version 4.1.1 (2021-08-10)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 19043)

Matrix products: default

locale:
[1] LC_COLLATE=English_United States.1252
[2] LC_CTYPE=English_United States.1252
[3] LC_MONETARY=English_United States.1252
[4] LC_NUMERIC=C
[5] LC_TIME=English_United States.1252
```

```
attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
 [1] forcats_0.5.1      stringr_1.4.0      purrr_0.3.4        readr_2.0.2
 [5] tidyr_1.1.4        tibble_3.1.5       tidyverse_1.3.1    pwr_1.3-0
 [9] patchwork_1.1.1    naniar_0.6.1       mosaic_1.8.3       ggridges_0.5.3
[13] mosaicData_0.20.2  ggformula_0.10.1   ggstance_0.3.5     dplyr_1.0.7
[17] Matrix_1.3-4       magrittr_2.0.1     knitr_1.36         janitor_2.1.0
[21] Hmisc_4.6-0        Formula_1.2-4      survival_3.2-13    lattice_0.20-45
[25] ggrepel_0.9.1      ggplot2_3.3.5      glue_1.4.2         equatiomatic_0.3.0
[29] Epi_2.44           car_3.0-11         carData_3.0-4      broom_0.7.9

loaded via a namespace (and not attached):
 [1] colorspace_2.0-2    ellipsis_0.3.2      rio_0.5.27
 [4] visdat_0.5.3        leaflet_2.0.4.1     snakecase_0.11.0
 [7] htmlTable_2.3.0     fs_1.5.0            base64enc_0.1-3
[10] ggdendro_0.1.22     rstudioapi_0.13     farver_2.1.0
[13] bit64_4.0.5         fansi_0.5.0         lubridate_1.8.0
[16] xml2_1.3.2          splines_4.1.1       polyclip_1.10-0
[19] jsonlite_1.7.2      dbplyr_2.1.1        cluster_2.1.2
[22] png_0.1-7           ggforce_0.3.3       shiny_1.7.1
[25] httr_1.4.2          compiler_4.1.1      backports_1.2.1
[28] assertthat_0.2.1    fastmap_1.1.0       cli_3.0.1
[31] later_1.3.0         tweenr_1.0.2        htmltools_0.5.2
[34] tools_4.1.1         gtable_0.3.0        Rcpp_1.0.7
[37] cellranger_1.1.0    vctrs_0.3.8         nlme_3.1-152
[40] crosstalk_1.1.1     xfun_0.27           rvest_1.0.2
[43] openxlsx_4.2.4      mime_0.12           lifecycle_1.0.1
[46] mosaicCore_0.9.0    MASS_7.3-54         zoo_1.8-9
[49] scales_1.1.1        vroom_1.5.5         hms_1.1.1
[52] promises_1.2.0.1    parallel_4.1.1      RColorBrewer_1.1-2
[55] yaml_2.2.1          curl_4.3.2          gridExtra_2.3
[58] labelled_2.8.0      rpart_4.1-15        latticeExtra_0.6-29
[61] stringi_1.7.5       highr_0.9           checkmate_2.0.0
[64] zip_2.2.0           rlang_0.4.12        pkgconfig_2.0.3
[67] evaluate_0.14       labeling_0.4.2      htmlwidgets_1.5.4
[70] cmprsk_2.2-10       bit_4.0.4           tidyselect_1.1.1
[73] plyr_1.8.6          R6_2.5.1            generics_0.1.0
[76] DBI_1.1.1           pillar_1.6.4        haven_2.4.3
[79] foreign_0.8-81      withr_2.4.2         mgcv_1.8-36
[82] abind_1.4-5         nnet_7.3-16         etm_1.1.1
[85] modelr_0.1.8        crayon_1.4.1        utf8_1.2.2
[88] tzdb_0.1.2          rmarkdown_2.11      jpeg_0.1-9
[91] grid_4.1.1          readxl_1.3.1        data.table_1.14.2
[94] reprex_2.0.1        digest_0.6.28       xtable_1.8-4
[97] httpuv_1.6.3        numDeriv_2016.8-1.1 munsell_0.5.0
```

THIS IS THE END OF THE QUIZ.