431 Quiz 2 with Answer Sketch

Thomas E. Love

due 2018-12-11 at 5 PM. Version 2018-12-11 23:52:17

General Instructions

The deadline for completing this quiz is 5 PM on Tuesday 2018-12-11, and this is a firm deadline.

If you wish to work on some of the quiz and then return to it later, you can do this by [1] scrolling down to the final question which asks you to type in your full name, affirming that you have done the work for the quiz alone, 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 without losing your progress.

There are 40 questions, labeled Q01 through Q40. The maximum score available is 110 points.

- Each question is worth between 2 and 6 points. Partial credit is available on some questions.
- Note that Q01 is probably the question which will take the longest amount of time for you to answer, so don't worry that the other 39 questions are as lengthy as that one.
- The order of the questions is arbitrary. Some are meant to be easy, some are not.
- You should attempt to answer every question. There is no advantage to failing to give a response, as an incorrect response is always at least as valuable as a missing one.
- Please select or type in your best response (or responses, as indicated) for each question.

Data sets are available for several of the questions. You will find those data sets at https://github.com/THOMASELOVE/431-2018/tree/master/quizzes/quiz02

- The `oscar_A` and `oscar_B` data sets apply to Q01 and Q02.
- The `swordfish` data set applies to Q03 and Q04.
- The `limestone` data set applies to Q12.
- The `wc_code.R` bit of R code applies to Q22.
- The `hospsim` data set applies to Q26 - Q32.
- The `wcgs` data set was used to develop Q19, but you will not actually use it in responding to the quiz.
- Our usual `Love-boost.R` script will also be used in developing the answer sketch.

For any question where we do not specify something different, you should assume the following:

- that you are looking to do two-sided statistical inference with a 95% confidence level.
- that you should round all numerical responses to two decimal places.
- that two-by-two tables should be built without Bayesian adjustments

Please use `set.seed(2018)` whenever you need to do work that requires random sampling.

The packages we loaded in R to complete the Answer Sketch for this Quiz were:

- `Hmisc`, `fivethirtyeight`, `car`, `broom`, `Epi`, `magrittr`, and the `tidyverse`.
- We also made use of functions from `knitr`, `mosaic` and `gridExtra` and those were installed, but not loaded with `library()`.
- You may also need to use a function from the `vcd` package, which should also be installed on your machine.
- As mentioned above, we also sourced in the `Love-boost.R` script.

You are welcome to consult the materials provided on the course website, but you are not allowed to discuss the questions on this quiz with anyone other than Professor Love and the teaching assistants at `431-help at case dot edu`. Please submit any questions you have about the Quiz to `431-help` through email.

Good Luck!

1 Q01 (6 points)

1.1 The Question

The `oscar_A.csv` and `oscar_B.csv` data files are available to support your work in Q01 and Q02.

Every year, the Academy Awards (also called the Oscars) are presented for artistic and technical merit in the American film industry. The `oscar_A.csv` and `oscar_B.csv` data files that are available to you each contain the names and ages of 51 winners of the Best Actor in a Leading Role and the Best Actress in a Leading Role awards since 1967. The two files simply arrange the same data in two different formats. Use whichever format works better for you.

The key question you will address in Q01 is "At the 5% significance level, which group (Best Actors or Best Actresses) tends to have older Oscar winners, and by how many years on average?"

In your response to Q01, we expect you to

a. specify the test or confidence interval procedure you used (this includes specifying whether the samples are paired or independent),
b. justify clearly why you chose that procedure,
c. state clearly what the calculated interval estimate you developed is (rounded to two decimal places), and
d. state clearly what your conclusion is regarding the comparison of Best Actor ages to Best Actress ages based on this sample, using complete English sentences, and using a maximum of 1500 characters.

The form will limit you to a maximum of 1,500 characters in your response. Dr. Love's response in the answer sketch is approximately 600 characters. Use complete English sentences, and address all four issues (a, b, c, and d) specified in the directions, and anything else you think is important in addressing the key question "At the 5% significance level, which group (Best Actors or Best Actresses) tends to have older Oscar winners, and by how many years on average?".

1.2 The Answer is below.

> The outcome here is age, and the exposure of interest is "Actor vs. Actress". We are comparing two samples, the data are quantitative and paired, by year. The paired men - women differences are well-approximated by the Normal distribution. So a paired t test is reasonable, as is a paired comparison using the bootstrap.
>
> Using either approach, at the 5% significance level, the Actors (men) are significantly older, by 7 $\pm$ 4 years. The 95% confidence interval for the mean (Actor - Actress) difference is (2.93, 11.15) from a paired t test, and it is (2.94, 10.98) from the bootstrap procedure.

Here are the details on how I got to that answer. I used the `oscar_B` data set, since it is in the wide format, anticipating paired samples.

```
oscar <- read_csv("data/oscar_B.csv") %>%
  mutate(agediff = age_lead_male - age_lead_female)
```

```
Parsed with column specification:
cols(
  year = col_double(),
  lead_male = col_character(),
  age_lead_male = col_double(),
  lead_female = col_character(),
  age_lead_female = col_double()
)
```

```r
oscar
```

```
# A tibble: 51 x 6
    year lead_male    age_lead_male lead_female    age_lead_female agediff
   <dbl> <chr>                <dbl> <chr>                    <dbl>   <dbl>
 1  1967 Paul Scofie~            45 Elizabeth Tay~              35      10
 2  1968 Rod Steiger             42 Katharine Hep~              60     -18
 3  1969 Cliff Rober~            45 Katharine Hep~              61     -16
 4  1970 John Wayne              62 Maggie Smith                35      27
 5  1971 George C. S~            43 Glenda Jackson              34       9
 6  1972 Gene Jackman            42 Jane Fonda                  34       8
 7  1973 Marlon Bran~            48 Liza Minnelli               27      21
 8  1974 Jack Lemmon             49 Glenda Jackson              37      12
 9  1975 Art Carney              56 Ellen Burstyn               42      14
10  1976 Jack Nichol~            38 Louise Fletch~              41      -3
# ... with 41 more rows
```

Pairing actually doesn't help to reduce variation here very much, as the data are paired by year, but the correlation of lead actor and lead actress ages is only slightly positive (r = 0.08).

```r
cor(oscar$age_lead_male, oscar$age_lead_female)
```

```
[1] 0.08498332
```

Next, we'll check to see if those paired male - female differences are well approximated by the Normal distribution. Here are some numerical summaries of the paired differences:

```r
mosaic::favstats(~ agediff, data = oscar)
```

```
 min   Q1 median Q3 max     mean       sd  n missing
 -25 -0.5      8 15  41 7.039216 14.61911 51       0
```

While we're at it, we might as well check our usual non-parametric skewness measure.

```r
skew1(oscar$agediff)
```

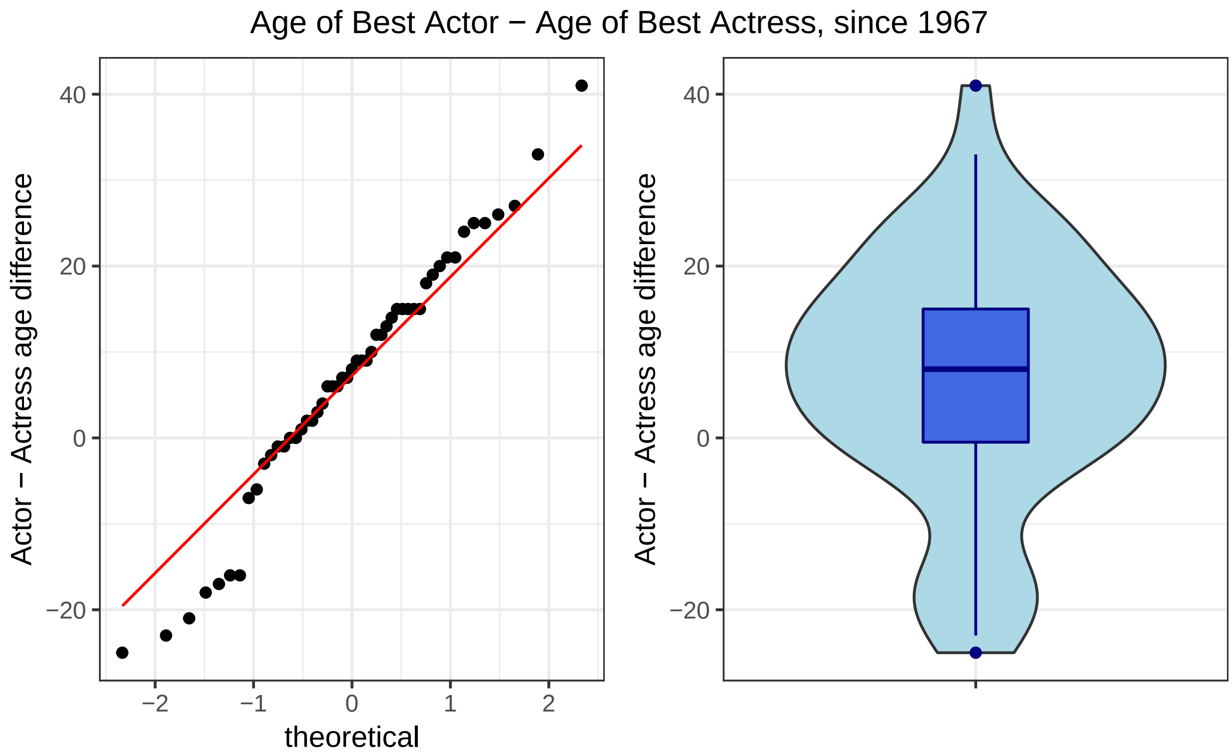
```
[1] -0.06572111
```

Here, the mean and median are pretty close to one another, and separated by less than 7% of a standard deviation.

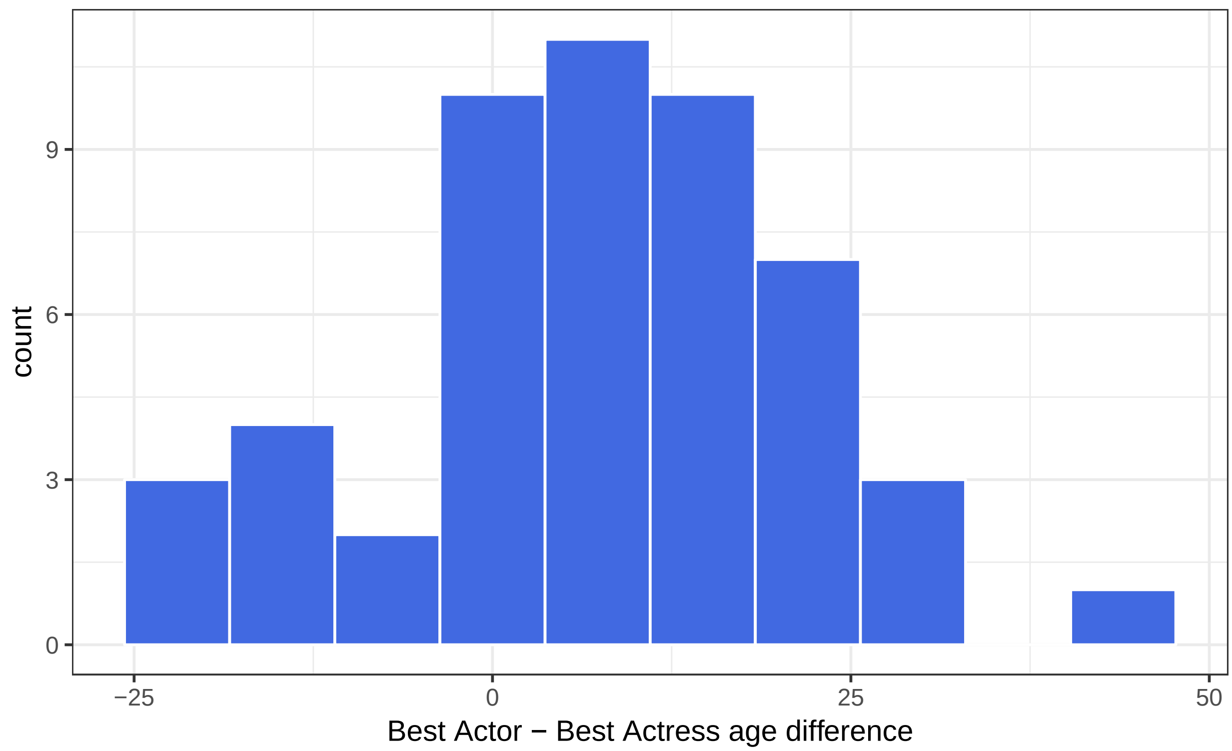
```r
# Normal Q-Q plot
p1 <- ggplot(oscar, aes(sample = agediff)) +
  geom_qq() + geom_qq_line(col = "red") +
  theme_bw() +
  labs(y = "Actor - Actress age difference")

# Violin + Boxplot
p2 <- ggplot(oscar, aes(x = "", y = agediff)) +
  geom_violin(fill = "lightblue") +
  geom_boxplot(fill = "royalblue", col = "navy",
               width = 0.25) +
  theme_bw() +
  labs(x = "", y = "Actor - Actress age difference")

gridExtra::grid.arrange(p1, p2, nrow = 1,
                        top = "Age of Best Actor - Age of Best Actress, since 1967")
```




```r
ggplot(oscar, aes(x = agediff)) +
  geom_histogram(bins = 10, fill = "royalblue", col = "white") +
  theme_bw() +
  labs(x = "Best Actor - Best Actress age difference")
```



Assesing the Normality assumption, there's nothing too problematic with the assumption of a Normal model for these data, in my view. It certainly would have been OK to select a paired t test. You could, I suppose, make a modest case for a symmetric with outliers conclusion, in which case a bootstrap might also be an appealing choice. So we'll look at each of those, but I don't see any reason to use a Wilcoxon signed rank test in this setting, so I'll skip that.

```
t.test(oscar$agediff)
```

```

	One Sample t-test

data:  oscar$agediff
t = 3.4387, df = 50, p-value = 0.001187
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
  2.927524 11.150907
sample estimates:
mean of x 
 7.039216 
```

```
set.seed(2018)
smean.cl.boot(oscar$agediff)
```

```
     Mean     Lower     Upper 
 7.039216  2.940196 10.981373 
```

If you'd instead used the `oscar_A` presentation of the data, and assumed independent samples, you'd obtain similar answers but ignored the pairing, so that's much less appropriate.

1.3 Grading Q01

Q01 is worth a maximum of 6 points. You needed to do all five of the following things to receive the full 6 points.

a. Assert that the data came from paired samples, or actively assert that sample oscar_B was the right choice. (10 of 48 students got this right.)
b. Verify that the paired differences were well described by a Normal model. (To get this right, you'd have had to recognize these as paired samples, of course. So only 3 of 48 students got this right.)
c. Specify that you would use a t test or bootstrap approach. (We gave credit here more leniently, regardless of whether you were using paired or independent samples, but we probably shouldn't have, so 42 of 48 got this right.)
d. Specify a correct point estimate of 7 years difference in age, with males older. (43 of 48 got this right.)
e. Specify the correct confidence interval of either (2.93, 11.15) from a paired t test, or (2.94, 10.98) from the bootstrap procedure. (9 of 48 got this right.)

The grading roster shows my Y/N evaluation for you on each of these in the Q01a - Q01e columns. There was partial credit involved.

If you got	Points	# of students
all five of these	6	3
4 of 5 (which was `a`, `c`, `d` and `e`)	5	4
three of these	3	4
any two of these	2	29
any one of these	1	6
none	0	2

2 Q02 (3 points)

2.1 The Question

If you look closely at the data in the `oscar_A` and `oscar_B` files discussed in Q01, you'll see they are missing the 1990 data. In that year, Daniel Day-Lewis won his first Oscar (for "My Left Foot") at the age of 32, while Jessica Tandy also won her first Oscar (for "Driving Miss Daisy") at the age of 80.

Which of the following statements are true about what would happen, if the 1990 data were included along with the 51 observations we've already used in Q01? (Note that no calculations are required here.)

[*Set up as TRUE or FALSE in each case.*]

- a. Whether the data were best analyzed as paired or independent samples would change.
- b. The point estimate of the population age difference incorporating the new data would shift closer to 0.
- c. A t-based confidence interval estimate incorporating the new data would be wider.

2.2 The Answer is that statement `a` is FALSE, statements `b` and `c` are TRUE.

- Statement `a` is FALSE.
  - The data remain paired by year with the addition of the 1990 data
- Statement `b` is TRUE.
  - We're adding in an outlier difference (32 - 80 = -48 is much smaller than any value we saw in the initial data) on the low end of the distribution. That will drag the sample mean of the paired differences down closer to zero, since that mean is currently positive.
- Statement `c` is also TRUE.
  - The standard error of our confidence interval would increase as a result of adding this outlier to the data, so that interval would in fact become wider.

2.3 Grading Q02

Q02 is worth 3 points, with 1 point awarded per correct response.

- Nearly all students got statement `a` right.
- 42/48 got statement `b` right.
- 33/48 got statement `c` right.

3 Q03 (4 points)

3.1 The Question

The `swordfish.csv` data file is available to support your work in Q03 and Q04.

Swordfish absorb mercury in their bodies, and it is thought that a mercury concentration of more than 1.00 ppm (parts per million) is not good for human consumption.

In a random sample of 28 swordfish, the concentrations listed below in the Output for Q03 and Q04 (and also in the `swordfish.csv` data file) were found.

```
0.07 0.24 0.39 0.54 0.61 0.72 0.81
0.82 0.84 0.90 0.95 0.98 1.02 1.08
1.14 1.20 1.20 1.26 1.29 1.30 1.37
1.40 1.44 1.58 1.62 1.68 1.85 2.10
```

The summary statistics from (`mosaic::favstats`) are:

```
  min     Q1 median     Q3 max     mean        sd  n missing
 0.07 0.8175   1.11 1.3775 2.1 1.085714 0.4757951 28       0
```

Suppose you want to know whether there is evidence at the 5% significance level that the mean concentration in the population of swordfish is different than 1.00 ppm, and you plan to use a two-sided bootstrap confidence interval with 1,000 bootstrap replications (which is the default choice) to make this decision.

In your response to Q03, (a) specify an appropriate confidence interval, and (b) clearly state your decision as to whether or not there is statistically significant evidence to support this claim. Do this in two or more complete sentences.

As is our general rule, round your interval to two decimal places. The form will limit you to a maximum of 500 characters in your response. Dr. Love's response in the Answer Sketch is approximately 160 characters.

3.2 The Answer is below.

> The bootstrap confidence interval is (0.92, 1.26) ppm. So no, there is no significant evidence of a difference from 1.00 ppm at the 5% significance level.

```r
swordfish <- read_csv("data/swordfish.csv")
```

```
Parsed with column specification:
cols(
  concentration = col_double()
)
```

```r
set.seed(2018)
smean.cl.boot(swordfish$concentration, B = 1000)
```

```
     Mean     Lower     Upper
1.0857143 0.9181429 1.2561250
```

3.3 Grading Q03

Q03 is worth 4 points.

- 4 points for a correct response and interpretation, so long as your bootstrap result has a lower bound of 0.91 or 0.92 and an upper bound of 1.25 or 1.26.

- 3 points for a slightly incorrect CI (close to but outside the choices above) but correct conclusion, or for an essentially correct statement but with a serious error in phrasing, like "confidential interval" or a description of a "difference in means" when there's only one mean here, or using a non-existent term like "statistically different" or for problems rounding.
- at most 2 points for a correct CI but comparing the CI to 0 rather than 1, or in some other way giving a poor explanation of the reason for your conclusion, or coming to no clear conclusion at all.
- at most 2 points for a correct (or mostly correct) CI but using a p value incorrectly to determine significance.
- at most 1 point for a seriously flawed CI, or for a correct CI but a terrible conclusion like "There is statistically significant evidence to support that the mean concentration is 1.00ppm" which is very, very wrong, as is "There is not statistically significant evidence to support this claim, in fact, there is statistically significant evidence to conclude the opposite." The first half of that was OK, but the second half is completely untrue.

Scores were:

Points	0	1	2	3	4
Students	2	5	11	12	18

4 Q04 (3 points)

4.1 The Question

Estimate a 95% confidence interval for the probability that a randomly selected swordfish will have a concentration of mercury above 1.00 ppm, based on the data in the `swordfish.csv` file, using the SAIFS approach. Express this probability as a proportion (between 0 and 1) rather than as a percentage (between 0 and 100). As is our general rule, round the endpoints of your confidence interval to two decimal places.

4.2 The Answer is (0.36, 0.78)

16 of the 28 swordfish in the sample have a concentration higher than 1.00 bpm.

```
swordfish %>% count(concentration > 1.00)
```

```
# A tibble: 2 x 2
  `concentration > 1`     n
  <lgl>               <int>
1 FALSE                  12
2 TRUE                   16
```

So the SAIFS interval is…

```
saifs.ci(x = 16, n = 28)
```

```
Sample Proportion             0.025             0.975
            0.571             0.359             0.777
```

4.3 Grading Q04

- 3 points for (0.36, 0.78)
- at most 2 points for a rounding error or for not rounding to two decimal places.
- at most 1 point for misspecifying the x or n value
- no other partial credit

Scores were:

Points	0	1	2	3
Students	11	3	1	33

5 Q05 (3 points)

5.1 The Question

In *The Signal and The Noise*, Nate Silver writes repeatedly about a Bayesian way of thinking about uncertainty, for instance in Chapters 8 and 13. Which of the following statistical methods is **NOT** consistent with a Bayesian approach to thinking about variation and uncertainty?

- a. Gambling using a strategy derived from a probability model.
- b. Combining information from multiple sources to build a model.
- c. Establishing a researchable hypothesis prior to data collection.
- d. Significance testing of a null hypothesis, using, say, Fisher's exact test.
- e. Updating our forecasts as new information appears.

5.2 The answer is `d`

See, for instance, this quote from Silver in the "Bob the Bayesian" section of Chapter 8.

> The problem with Fisher's notion of hypothesis testing is not with having hypotheses but with the way Fisher recommends that we test them.

Each of the other strategies mentioned (besides `d`) is clearly part of the Bayesian approach, and is explicitly described as such in the book.

5.3 Grading Q05

Q05 is worth 3 points. No partial credit is available. 38/48 got it right.

Background Information for Q06 - Q10

Suppose you fit four candidate regression models to predict the natural logarithm of a measure of predatory behavior in leopards.

The four models are nested, in that D is a proper subset of C, which is a proper subset of B, which is a proper subset of A.

Specifically, Model A contains seven predictors, Model B contains five of those seven predictors, and Model C contains three of the five Model B predictors, while Model D is a simple regression, using one of the predictors in Model C.

You obtain the results shown below in the Output for Q06 - Q10.

```r
modelA <- lm(log(predatory.behavior) ~
               x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8, data = leopard)
modelB <- lm(log(predatory.behavior) ~
               x1 + x2 + x3 + x4 + x5, data = leopard)
modelC <- lm(log(predatory.behavior) ~
               x1 + x2 + x3, data = leopard)
modelD <- lm(log(predatory.behavior) ~
               x1 , data = leopard)

BIC(modelA, modelB, modelC, modelD)
```

```
       df      BIC
modelA 10 4486.338
modelB  7 4470.056
modelC  5 5125.469
modelD  3 5304.291
```

6 Q06 (2 points)

6.1 The Question

Which of these models does the output above suggest will be the best choice to predict the natural logarithm of the predatory behavior measure?

- a. Model A
- b. Model B
- c. Model C
- d. Model D
- e. The output doesn't suggest a "best" choice.

6.2 The Answer is `b`

The model with the smallest BIC is model B. That's the winning model, based on this output.

6.3 Grading Q06

Q06 is worth 2 points. No partial credit is available. Nearly all students got it right.

7 Q07 (2 points)

7.1 The Question

The following output relates to `modelA` described in the previous question.

```
round(car::vif(modelA),3)
```

```
   x1    x2    x3    x4    x5    x6    x7    x8 
1.018 1.014 1.074 1.007 1.014 2.665 2.554 1.013 
```

Which of the following statements is the best conclusion from the output for Q07 shown above?

- a. Model A's residuals will show no problem with independence.
- b. Model A's residuals will show a serious problem with independence.
- c. Model A has no sign of meaningful collinearity.
- d. Model A has a serious problem with collinearity.
- e. Model A's residual variance will be larger than the residual variance of Model B, which is the model that includes predictors `x1, x2, x3, x4` and `x5`, only.

7.2 The Answer is `c`

The `vif`, or variance inflation factor, helps us measure collinearity. Since none of the VIF values exceed (or even approach) 5, this indicates that there is no serious collinearity in Model A. Statements `a`, `b`, and `e` are unrelated to the VIF, and statement `d` is incorrect.

7.3 Grading Q07

Q07 is worth 2 points. No partial credit is available. 42/48 got it right.

8 Q08 (2 points)

8.1 The Question

Which of the following R commands would provide fitted values of `log(predatory.behavior)` using the equation in Model A (from the previous two questions), for a new set of data contained in the `newleopard` tibble?

- a. `tidy(modelA, newdata = newleopard)`
- b. `glance(modelA, newdata = newleopard)`
- c. `augment(modelA, newdata = newleopard)`
- d. `split(modelA, newdata = newleopard)`
- e. None of these.

8.2 The Answer is `c`

The `augment` function in choice `c` does this job. The others do not.

8.3 Grading Q08

Q08 is worth 2 points. No partial credit is available. 35/48 got it right.

9 Q09 (2 points)

9.1 The Question

Suppose the first predicted subject in the `newleopard` tibble yields a prediction of `log(predatory.behavior)` of 3.5, with a 95% uncertainty interval of (3, 4). To convert that uncertainty interval back to the original scale on which the predatory behavior measurements were obtained, we would obtain which of the following results?

- a. 3 to 4
- b. `log(3)` to `log(4)`
- c. `10*3` to `10*4`
- d. `exp(3)` to `exp(4)`
- e. None of these would work.

9.2 The Answer is `d`

To back out of the natural logarithm (`log`) and thus untransform back to our original scale, we'd use the `exp` function. `d` is correct.

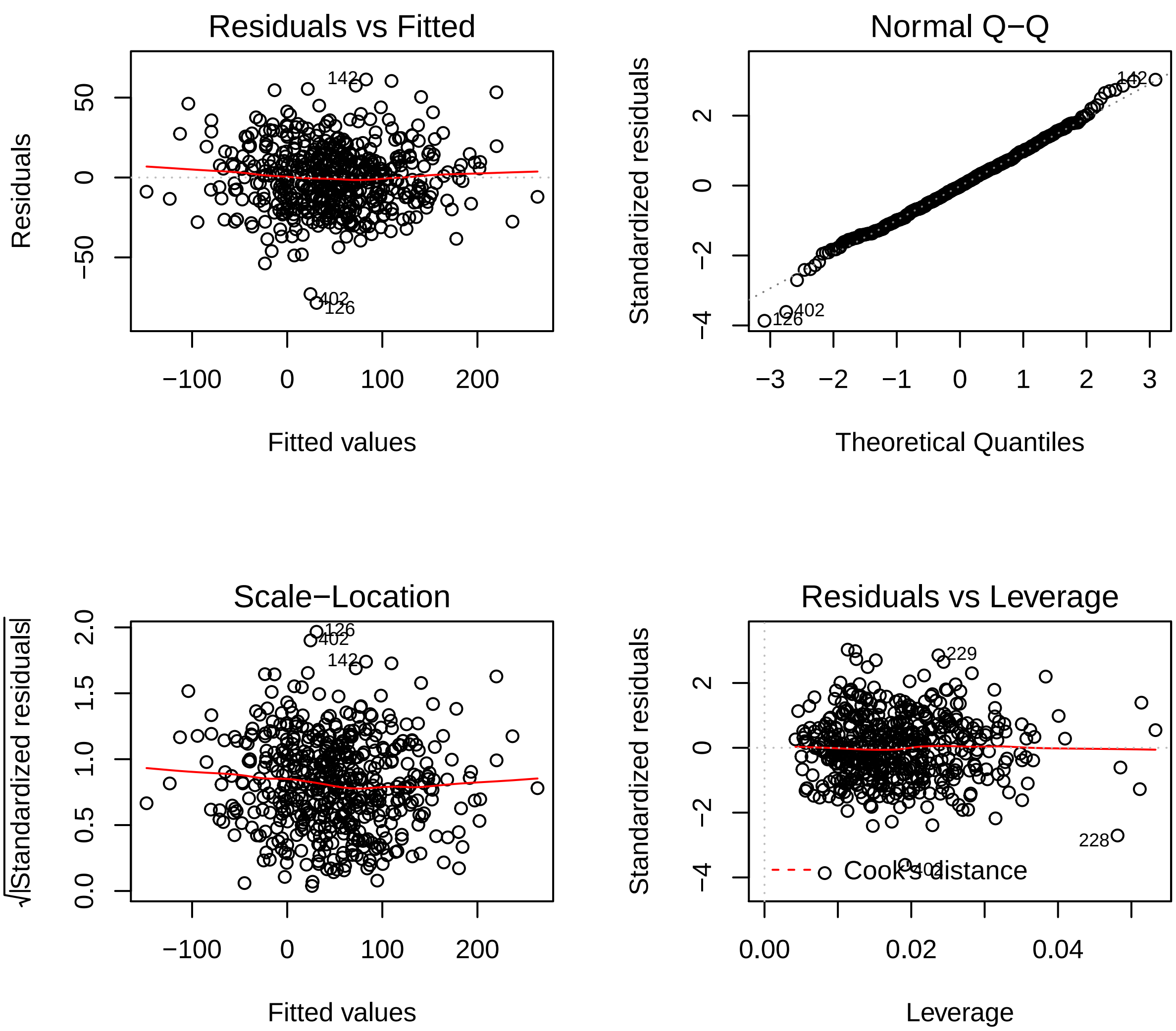
For those of you who chose `c`, remember that `log10` in R gives the base-10 logarithm, and not `log`.

9.3 Grading Q09

Q09 is worth 2 points. There is no partial credit available. 43/48 got it right.

10 Q10 (3 points)

Behold the residual plots for Model A that we have been discussing since Q06.



In the output for Q10 provided above, you see the residual plots for the Model A that we have been discussing since Q06. Which of the following conclusions is most appropriate?

- a. There is a serious problem with the assumption of linearity.
- b. There is a serious problem with the assumption of Normality.
- c. There is a serious problem with the assumption of constant variance
- d. There are no serious problems evident in these residual plots.
- e. None of these conclusions are appropriate.

10.1 The Answer is `d`.

I see no problems. Choice `d` is correct.

- I see no serious curve in the plot of residuals vs. fitted values, so there's no clear problem with *linearity*.
- The Normal Q-Q plot shows no sign of substantial problems with the assumption of Normality.
- Nor are there any clearly influential points.
- I see no serious fan shape in the residuals vs. fitted values, and no clear rise or fall in the scale-location plot, so there's no clear problem with the assumption of constant variance.

10.2 Grading Q10

Q10 is worth 3 points. No partial credit is available. 45/48 got it right.

11 Q11 (2 points)

11.1 The Question

Once a confidence interval is calculated, several design changes may be used by a researcher to make a confidence interval wider or narrower. For each of the changes listed below, indicate the impact on the width of the confidence interval.

- Rows
  - a. Increase the level of confidence.
  - b. Increase the sample size.
  - c. Increase the standard error of the estimate.
  - d. Use a bootstrap approach to estimate the CI.
- Columns
  - 1. CI will become wider
  - 2. CI will become narrower
  - 3. CI width will not change
  - 4. It is impossible to tell

11.2 The Answer is `a` = 1, `b` = 2, `c` = 1, and `d` = 4

Increasing the level of confidence, or increasing the standard error, will automatically increase the width of the confidence interval. Increasing the sample size will decrease the standard error, and thus increase the confidence interval. If we use a bootstrap, our interval might get wider and it might get narrower, so we cannot tell.

11.3 Grading Q11

Q11 is worth 2 points, with 0.5 point awarded per correct response.

- 41/48 students got part `a` right.
- 40/48 got part `b` right.
- 43/48 got part `c` right.
- 26/48 got part `d` right.

12 Q12 (3 points)

12.1 The Question

The `limestone.csv` data file is available to support your work in Q12.

A geologist collects 130 hand-specimen sized pieces of limestone from a particular area. A qualitative assessment of both texture (either Fine, Medium or Coarse) and color (either Red, Orange or Brown) is made with the results shown in the Output for Q12.

—	Red	Orange	Brown
Fine	9	21	9
Medium	8	17	18
Coarse	19	22	7

Suppose you want to know if there is evidence of an association (at the 99% confidence level) between color and texture for these limestones. Which of the following conclusions is most appropriate?

- a. There is evidence of a significant color-texture association, since the p value is greater than 0.01
- b. There is evidence of a significant color-texture association, since the p value is less than 0.01
- c. There is no evidence of a significant color-texture association, since the p value is greater than 0.01
- d. There is no evidence of a significant color-texture association, since the p value is less than 0.01
- e. It is impossible to tell from the information provided.

12.2 The Answer is `c.`

To answer the question, we need to build the table in R, and then run the appropriate chi-squared test, yielding a p value of about 0.02, which is greater than 0.01, and thus indicates no significant color-texture association at the $\alpha = 0.01$ level.

```r
limestone <- read.csv("data/limestone.csv") %>% tbl_df

lime_tab <- limestone %>%
  mutate(color = fct_relevel(color, "Red", "Orange", "Brown"),
         texture = fct_relevel(texture, "Fine", "Medium", "Coarse")) %>%
  select(texture, color) %>%
  table

knitr::kable(addmargins(lime_tab))
```

	Red	Orange	Brown	Sum
Fine	9	21	9	39
Medium	8	17	18	43
Coarse	19	22	7	48
Sum	36	60	34	130

```r
chisq.test(lime_tab)
```

```
    Pearson's Chi-squared test

data:  lime_tab
X-squared = 11.597, df = 4, p-value = 0.02062
```

12.3 Grading Q12

Q12 is worth 3 points. No partial credit is available. Nearly all students got it right.

13 Q13 (3 points)

13.1 The Question

Suppose you have a data frame named `dat` containing a variable called `height`, which shows the subject's height in centimeters. Which of the following lines of code will create a new variable `tall` in the `dat` data frame which takes the value **TRUE** when a subject is more than 175 cm tall, and **FALSE** when a subject's height is at most 175 cm?

- a. `dat %>% mutate(tall = height > 175)`
- b. `dat %>% tall <- height > 175`
- c. `dat$tall <- ifelse(dat$height > 175, "YES", "NO")`
- d. `tall <- dat %>% filter(height > 175)`
- e. None of these will do the job

13.2 The Answer I had in mind was `a` but the correct answer is actually `e`

- Approach `a` does what we're looking for, except it doesn't actually store `tall` in the `dat` frame. To do that you would need: `dat <- dat %>% mutate(tall = height > 175)`.
- Approach `b` will throw an error message.
- Approach `c` will create a YES/NO, rather than a TRUE/FALSE variable. I expect this to be the most common incorrect response.
- Approach `d` will pull the tall people into a data frame called tall.
- So `e` is actually correct.

13.3 Grading Q13

Q13 is worth 3 points. I gave full credit to all students who answered `a` or `e`. 40/48 students did.

14 Q14 (3 points)

14.1 The Question

The lab component of a core course in biology is taught at the Watchmaker's Technical Institute by a set of five teaching assistants, whose names, conveniently, are Amy, Beth, Carmen, Donna and Elena. On the second examination of the semester (each section takes the same set of exams) an administrator at WTI wants to compare the mean scores across lab sections. She produces the following output in R.

```
Analysis of Variance Table

Response: exam2
           Df  Sum Sq Mean Sq F value  Pr(>F)
ta          4  1199.8 299.950  3.3355 0.01174
Residuals 165 14837.8  89.926
```

Emboldened by this result, the administrator decides to compare mean `exam2` scores for each possible pair of TAs, using a Bonferroni correction. If she wants to maintain an overall α level of 0.05 for the resulting suite of pairwise comparisons, and plans to do each of them separately with a two-sample t test, then what significance level should she use for each of the individual two-sample t tests?

- a. She should use a significance level of 0.20 on each test.
- b. She should use 0.005 on each test.
- c. She should use 0.0125 on each test.
- d. She should use 0.05 on each test.
- e. None of these answers are correct.

14.2 The Answer is `b`.

In total, there are **ten** pairwise comparisons to be made at Watchmaker's Technical Institute: A[my] vs. B[eth], A vs. C[armen], A vs. D[onna], A vs. E[lena], B vs. C, B vs. D, B vs. E, C vs. D, C vs. E and D vs. E.

- So if we want to retain a 5% significance level with a Bonferroni correction, we'd have to run the two-sample t tests at a significance level 1/10 that size, or 0.005.
- Only that rate (contained in answer `b`) will ensure that our overall error rate across all 10 comparisons will be no more than 0.05.
- So you'll need to make sure that's the approach taken the administrator at Tick-Tock-Tech.

14.3 Grading Q14

Q14 is worth 3 points. There is no partial credit available. 33/48 students got this right.

15 Q15 (4 points)

15.1 The Question

Suppose you have completed a pilot study of average birth weight for full term infants whose gestational age is 40 weeks. In that sample, the infants whose mothers smoked during pregnancy had a mean birth weight that was 300 grams lighter than the mean birth weight of the infants whose mothers did not smoke during pregnancy, while the standard deviation of the birth weights was about 430 grams in each group.

Assume that you are planning to conduct a new study, with a balanced design where you will initially enroll in the study a total of 400 infants (again, who were at 40 weeks gestation) in a balanced design between the two exposure groups (by the mother's smoking status.) In this new study, you want to be able to detect a difference in means that is at least half as large as what you observed in the pilot study.

If 10% of your initially enrolled subjects (in each smoking group) cannot be used in the final analysis for some reason, then with what power will you be able to detect the desired effect? Please present your power estimate as a percentage, rather than as a proportion, and round it to the nearest integer.

15.2 The Answer is 90%.

- Our initial enrollment was 200 infants in each sample (smoking mom and non-smoking mom) but we lost 10% of that, so we wind up with 180 infants in each sample.
- The desired "minimum important" clinical effect size to detect is a `delta` of 150 g (half of the 300 g we observed initially.)
- We assume the initial standard deviation of 430 g from the pilot study still holds.
- We will use a 5% significance level, and a two-sided test, as the instructions of the Quiz suggest are appropriate when not otherwise specified.
- This is an independent samples comparison. The infants are not paired or matched.

So we calculate:

```r
power.t.test(n = 180, delta = 150, sd = 435, sig.level = 0.05, type = "two.sample")
```

```

     Two-sample t test power calculation

              n = 180
          delta = 150
             sd = 435
      sig.level = 0.05
          power = 0.9036391
    alternative = two.sided

NOTE: n is number in *each* group
```

and presenting this as a percentage, we get power = 90.36%, which rounded to the nearest integer is 90.

15.3 Grading Q15

Q15 is worth 4 points.

- An unrounded response, like 90.3% will receive 3 points.
- Rounding 90.36% up to 91% is the opposite of what you want to do here. This isn't calculating a sample size, it's calculating a power. So that also receives 3 points, and that was by far the most common response, which makes me think that there might have been some other way to get 91%.

- Presenting the correct power as a proportion like 0.90 or 0.904 will also yield 3 points.
- No other partial credit will be awarded.

Points	0	3	4
Students	15	32	1

16 Q16 (3 points)

16.1 The Question

Suppose that 80 of 100 male applicants to a graduate school are accepted, while 60 of 100 female applicants are accepted. Estimate a two-sided 95% confidence interval for the relative risk of acceptance for a male applicant as compared to a female one. Round all parts of your response to one decimal place.

16.2 The Answer is (1.1, 1.6)

We can calculate the resulting confidence interval using the `twobytwo` function in `Love-boost.R`.

```r
twobytwo(80, 20, 60, 40, "Male", "Female", "Accepted", "Not Accepted")
```

```
2 by 2 table analysis:
------------------------------------------------------
Outcome   : Accepted
Comparing : Male vs. Female

       Accepted Not Accepted    P(Accepted) 95% conf. interval
Male         80           20            0.8    0.7102   0.8672
Female       60           40            0.6    0.5013   0.6912

                                   95% conf. interval
             Relative Risk: 1.3333    1.1052   1.6086
         Sample Odds Ratio: 2.6667    1.4166   5.0199
Conditional MLE Odds Ratio: 2.6534    1.3577   5.3125
    Probability difference: 0.2000    0.0731   0.3185

             Exact P-value: 0.0032
        Asymptotic P-value: 0.0024
------------------------------------------------------
```

16.3 Grading Q16

Q16 is worth 3 points.

- 3 points for a correct response of (1.1, 1.6)
- 2 points for an unrounded but otherwise correct response, like (1.11, 1.61)
- 1.5 points for the relative risk without the CI (1.3)

Points	0	1.5	2	3
Students	7	1	3	37

17 Q17 (3 points)

17.1 The Question

Breaking down the applications described in Q16 into the school's two separate programs, we find that program A accepted 72 of its 80 male applicants, and program B accepted 41 of its 80 female applicants. Each of the 200 students described in Q16 applied to either program A or program B, and not to both. Which type of student (males or females) had lower odds of being accepted by the school ...

Rows:

- a. into Program A?
- b. into Program B?
- c. into the school overall?

Columns:

- Females had lower odds.
- Males had lower odds.

17.2 The Answer is that a is MALE, b is MALE and c is FEMALE

This is a classic example of Simpson's Paradox. The relevant 2x2 tables are:

- for the school, we know that 80 of 100 male and 60 of 100 female applicants were admitted.

```
twobytwo(80, 20, 60, 40, "Male", "Female", "Accepted by School", "Not Accepted by School")
```

```
2 by 2 table analysis:
------------------------------------------------------
Outcome   : Accepted by School
Comparing : Male vs. Female

       Accepted by School Not Accepted by School    P(Accepted by School)
Male                   80                     20                      0.8
Female                 60                     40                      0.6
       95% conf. interval
Male      0.7102   0.8672
Female    0.5013   0.6912

                                  95% conf. interval
             Relative Risk: 1.3333    1.1052   1.6086
         Sample Odds Ratio: 2.6667    1.4166   5.0199
Conditional MLE Odds Ratio: 2.6534    1.3577   5.3125
    Probability difference: 0.2000    0.0731   0.3185

             Exact P-value: 0.0032
        Asymptotic P-value: 0.0024
------------------------------------------------------
```

- and so the odds ratio is 2.67, which is, of course, greater than 1.
- for Program A, we know that 72 of 80 male applicants were admitted, which means that the other 20 male applicants applied to program B and 8 of those 20 were admitted, in order to make the totals for Males add up properly across the school's two programs.

- for Program B, we know that 41 of 80 female applicants were admitted, which means that the other 20 female applicants applied to program A, and that 19 of those 20 were admitted, in order to make the totals for Females add up properly across the school's two programs.

So, for Program A, we have an odds ratio of 0.47.

```r
twobytwo(72, 8, 19, 1, "Male", "Female", "Accepted by A", "Not Accepted by A")
```

```
2 by 2 table analysis:
------------------------------------------------------
Outcome   : Accepted by A
Comparing : Male vs. Female

       Accepted by A Not Accepted by A    P(Accepted by A) 95% conf.
Male              72                 8                0.90    0.8126
Female            19                 1                0.95    0.7178
       interval
Male     0.9492
Female   0.9930

                                    95% conf. interval
             Relative Risk:  0.9474    0.8367   1.0727
         Sample Odds Ratio:  0.4737    0.0558   4.0238
Conditional MLE Odds Ratio:  0.4766    0.0102   3.9438
    Probability difference: -0.0500   -0.1445   0.1423

             Exact P-value: 0.6827
        Asymptotic P-value: 0.4936
------------------------------------------------------
```

And, for Program B, we have an odds ratio of 0.63.

```r
twobytwo(8, 12, 41, 39, "Male", "Female", "Accepted by B", "Not Accepted by B")
```

```
2 by 2 table analysis:
------------------------------------------------------
Outcome   : Accepted by B
Comparing : Male vs. Female

       Accepted by B Not Accepted by B    P(Accepted by B) 95% conf.
Male               8                12              0.4000    0.2142
Female            41                39              0.5125    0.4041
       interval
Male     0.6199
Female   0.6197

                                    95% conf. interval
             Relative Risk:  0.7805    0.4380   1.3908
         Sample Odds Ratio:  0.6341    0.2342   1.7173
Conditional MLE Odds Ratio:  0.6370    0.2023   1.9079
    Probability difference: -0.1125   -0.3226   0.1265

             Exact P-value: 0.4562
        Asymptotic P-value: 0.3702
------------------------------------------------------
```

Males were more likely to be accepted than Females overall, but were less likely to be accepted to each of the two programs. This happens because it was substantially easier to be accepted into Program A, and Males were also far more likely to apply to Program A.

17.3 Grading Q17

Q17 is worth 3 points, with 1 point awarded per correct response.

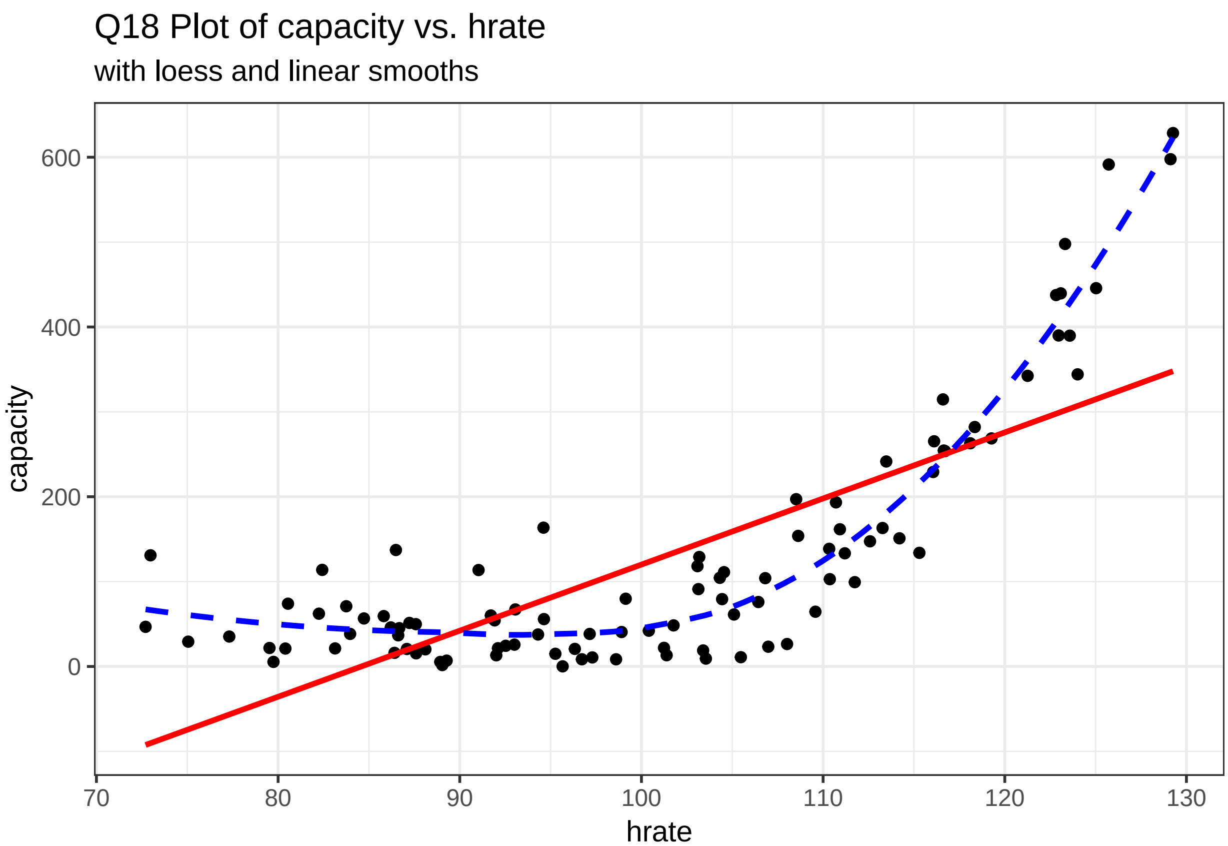
Only 31/48 students got `a` right. Nearly all students got `b` and `c` right.

18 Q18 (2 points)

18.1 The Question

Suppose we plotted the relationship between an outcome related to blood flow capacity, labeled `capacity`, and a predictor called `hrate`, which is a measure of peak comfortable heart rate. Each is measured for a cross-section of 100 subjects.

We then used the `geom_smooth` function in `ggplot2` to fit both a linear smooth and a loess smooth, producing the plot shown below in the output for Q18. One smooth is shown with blue dashes and the other is shown as a red solid line.



Which of the following statements is true?

- a. The linear fit is shown as a blue dashed line in this plot.
- b. The linear model describing `capacity` using `hrate` has a problem with independence.
- c. The Pearson correlation of `hrate` and `capacity` is negative.
- d. The linear model provides a better fit to the data than does the loess smooth.
- e. None of these statements are true.

18.2 The answer is `e`

`e` is correct, because the other four statements are false.

- `a` is false. The linear fit is clearly shown in red, not blue, and is a solid, not dashed line.

- `b` is false. The data are taken from a cross-section of subjects. There's no time ordering here, so there's no way to have a problem with the independence assumption in a regression model.
- `c` is false. The Pearson correlation will have the same sign as the slope of the regression line, which is positive.
- `d` is false. The loess smooth is a much better fit to this curved association.

18.3 Grading Q18

Q18 is worth 2 points. No partial credit is available. 41/48 got this right.

19 Q19 (3 points)

19.1 The Question

Data describing a sample of subjects participating in the Western Collaborative Group Study (discussed in our Course Notes in several places) were used here to fit a model to predict the natural logarithm of systolic blood pressure (`sbp`) using the subject's age, height, smoking status (yes/no) and the natural logarithm of their weight. The `wcgs` file is on our website, but this question uses a sample from that data that is unknown to you, so you will not be able to duplicate the output that follows.

```
m19 <- lm(log(sbp) ~ age + log(weight) + height + smoke, data = wcgs19)
tidy(m19, conf.int = TRUE) %>% knitr::kable(digits = 3)
```

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	4.528	0.414	10.941	0.000	3.710	5.346
age	0.008	0.002	3.977	0.000	0.004	0.011
log(weight)	0.163	0.084	1.934	0.055	-0.004	0.330
height	-0.012	0.004	-2.682	0.008	-0.021	-0.003
smokeYes	-0.023	0.021	-1.122	0.264	-0.064	0.018

```
glance(m19) %>%
  select(r.squared, adj.r.squared, sigma, statistic, df,
         p.value, AIC, BIC) %>%
  knitr::kable(digits = 3)
```

	r.squared	adj.r.squared	sigma	statistic	df	p.value	AIC	BIC
value	0.171	0.148	0.122	7.491	5	0	-199.285	-181.221

What conclusions can you draw from this output, using a 5% significance level?

[*SET UP as SEPARATE TRUE-FALSE items*]

- a. Smokers have significantly lower systolic blood pressures than non-smokers, after we account for age and size (height and weight).
- b. Larger height is associated with significantly lower blood pressure, even after we've accounted for age, weight and smoking status.
- c. This model accounts for more than 15% of the variation in the log of systolic blood pressure.

19.2 Answer 19 is `a` = FALSE, `b` = TRUE, `c` = TRUE

- `a` is False, because the estimated coefficient for smokeYes is -0.023, barely larger than its associated standard error (0.021), and the 95% uncertainty interval includes 0. We can also see this from the p-value, which is not significant for `smoke` as the last predictor into the model after accounting for the `age` and size variables.
- `b` is True. `height` is negatively associated with log(`sbp`) and the 95% confidence interval for the slope of height is (-0.021, -0.003), so additional height is associated with lower BP after accounting for the other variables in the model.
- `c` is True. The R^2 here is 0.171, which is larger than 15%. It's important to use the R^2, rather than the adjusted R^2, which is an index, but isn't a proportion of anything.

19.3 Grading Q19

Q19 is worth 3 points. 1 point for each correct answer, to each of the three parts. No partial credit.

- 43/48 got `a` right, while 39/48 got `b` and 39/48 also got `c` right.

20 Q20 (2 points)

20.1 The Question

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 can be 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`

20.2 The Answer is `b`.

`fct_relevel` lets you specify a new order for factor levels "by hand" which is what you'd need to do here, and `facet_wrap` is the easier approach to getting the individual subsetted histograms in this context.

Choice `a` is probably going to be popular, too, but `fct_reorder` lets you reorder the factor levels by sorting along another variable's values, and that's not what we're doing here.

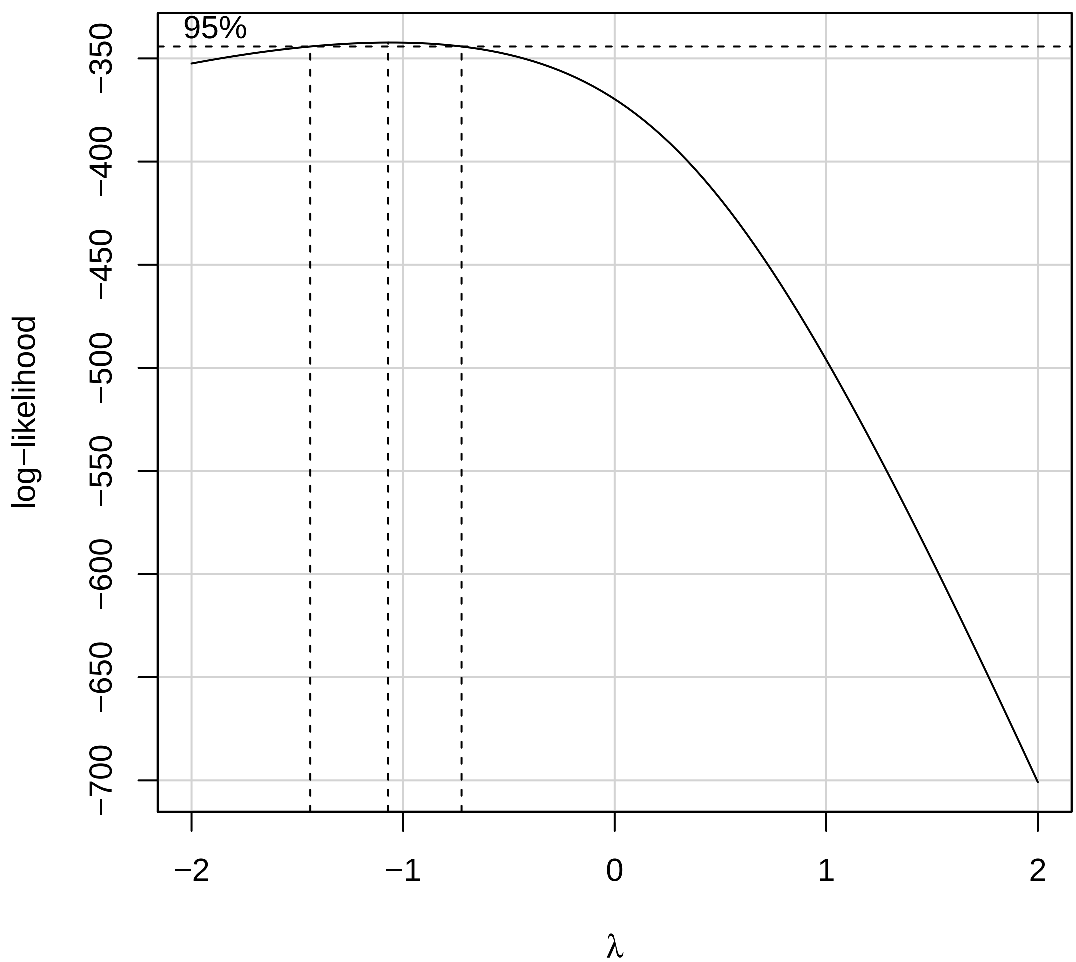
20.3 Grading Q20

Q20 is worth 2 points. No partial credit is available. 42/48 got this right.

21 Q21 (2 points)

21.1 The Question

Consider the Box-Cox plot below, which addresses a model built to predict an outcome called `score` using four predictors.



What transformation of our response does this plot suggest?

- a. The square of our outcome, $score^2$.
- b. The square root of our outcome, $\sqrt{score}$.
- c. The logarithm of our outcome, $log(score)$.
- d. The inverse of our outcome, $1/score$.
- e. The original, untransformed outcome, $score$.

21.2 The Answer is `d`

The suggested power is clearly near -1, which indicates the inverse of our outcome.

21.3 Grading Q21

Q21 is worth 2 points. Everyone got it right. Hooray!

22 Q22 (2 points)

22.1 The Question

The code snippet `wc_code.R` is available to support your work in Q22 - Q24.

Consider the `weather_check` data frame within the `fivethirtyeight` package. We will use these data for Q22-Q24.

Suppose you want to build a table containing information from the `female`, `ck_weather` and `age` variables in that data frame. I suggest you use the following approach to place the data in the `wc` tibble, and adjust some of the coding.

Note I have provided this code snippet to you in a file called `wc_code.R`.

```
wc <- fivethirtyeight::weather_check %>%
  select(female, ck_weather, age) %>%
  mutate(female = fct_recode(factor(female),
                             "Female" = "TRUE",
                             "Male" = "FALSE"),
         ck_weather = fct_recode(factor(ck_weather),
                            "Check" = "TRUE",
                            "No Check" = "FALSE")) %>%
  mutate(female = fct_relevel(female, "Female"),
         ck_weather = fct_relevel(ck_weather, "Check"))
```

Build the specified table using your `wc` tibble. Which age group has exactly 105 female respondents who indicated that they typically check a daily weather report?

- a. Ages 18-29
- b. Ages 30-44
- c. Ages 45-59
- d. Ages 60+
- e. None of these.

22.2 The Answer is `b`

Here's the code and the resulting (flat) table.

```
wc <- fivethirtyeight::weather_check %>%
  select(female, ck_weather, age) %>%
  mutate(female = fct_recode(factor(female),
                             "Female" = "TRUE",
                             "Male" = "FALSE"),
         ck_weather = fct_recode(factor(ck_weather),
                            "Check" = "TRUE",
                            "No Check" = "FALSE")) %>%
  mutate(female = fct_relevel(female, "Female"),
         ck_weather = fct_relevel(ck_weather, "Check"))

wc_table <- table(wc$female, wc$ck_weather, wc$age)

ftable(wc_table)
```

		18 - 29	30 - 44	45 - 59	60+
Female	Check	79	105	115	121
	No Check	30	28	29	20
Male	Check	41	56	119	103
	No Check	26	15	15	14

22.3 Grading Q22

Q22 is worth 2 points. There is no partial credit available. 43/48 got it right.

23 Q23 (3 points)

23.1 The Question

Perform an appropriate test to see if the odds ratio for a Yes (TRUE) response to "Do you typically check a daily weather report?" comparing Female to Male respondents is essentially consistent across age categories. What is the name of the test that you ran, and what is the conclusion? As usual, use a 5% significance level here.

- a. I ran a chi-square test on a 2x2 table using the `Epi` package's `twoby2` function, and the conclusion is that there is a significant association.
- b. I ran a chi-square test on a 2x2 table using the `Epi` package's `twoby2` function, and the conclusion is that there is not a significant association.
- c. I ran Woolf's test (`woolf_test`) to assess the homogeneity of odds ratios from the `vcd` package, and I conclude that the odds ratio is sufficiently consistent across age categories to allow me to collapse on age.
- d. I ran Woolf's test (`woolf_test`) to assess the homogeneity of odds ratios from the `vcd` package, and I conclude that the odds ratio is NOT sufficiently consistent across age categories to allow me to collapse on age.
- e. None of these statements describe an appropriate test.

23.2 The Answer is `c`.

We need to run a Woolf test on the table we just built. Our conclusion matches `c`. There is no significant effect at the 5% level.

```
vcd::woolf_test(wc_table)
```

```

    Woolf-test on Homogeneity of Odds Ratios (no 3-Way assoc.)

data:  wc_table
X-squared = 6.5613, df = 3, p-value = 0.08728
```

People might have trouble fitting the `woolf_test`, because it works on a table, and not a data frame, but I hope not.

23.3 Grading Q23

Q23 is worth 3 points. No partial credit is available. 29/48 got it right.

24 Q24 (3 points)

24.1 The Question

Use the data we have been working with in the previous two questions, regardless of how you answered those questions. Suppose we want to use the Cochran-Mantel-Haenszel approach to estimate the common odds ratio across all age categories comparing Females to Males as to whether they check the weather daily. Which of the following statements is true?

- a. Females have higher odds of checking the weather, and a 95% confidence interval includes 1.
- b. Females have higher odds of checking the weather, and a 95% confidence interval does not include 1.
- c. Females have lower odds of checking the weather, and a 95% confidence interval includes 1.
- d. Females have lower odds of checking the weather, and a 95% confidence interval does not include 1.
- e. None of these statements are true.

24.2 The Answer is `c`.

Now we will run the Cochran-Mantel-Haenszel test.

```
mantelhaen.test(wc_table)
```

```

	Mantel-Haenszel chi-squared test with continuity correction

data:  wc_table
Mantel-Haenszel X-squared = 0.17397, df = 1, p-value = 0.6766
alternative hypothesis: true common odds ratio is not equal to 1
95 percent confidence interval:
 0.6533502 1.2870384
sample estimates:
common odds ratio 
        0.9169988 
```

The odds ratio is 0.92 for Females vs. Males on checking the weather, so Females checked less frequently in the sample (because the sample odds ratio is less than 1.) The 95% confidence interval is (0.65, 1.29) and that certainly includes 1.

24.3 Grading Q24

Q24 is worth 3 points. No partial credit is available. 30/48 got it right.

25 Q25 (3 points)

25.1 The Question

Which of the following statements is **NOT** part of what Silver is trying to tell us in *The Signal and The Noise*? (You may wish to focus on Chapter 13, which summarizes the preceding arguments nicely.)

- a. Our bias is to think we are better at prediction than we really are.
- b. Make a lot of forecasts. It's the only way to get better.
- c. State, explicitly, how likely we believe an event is to occur before we begin to weigh the evidence.
- d. Revise and improve your estimates as you encounter new information.
- e. Nature's laws change quickly, and do so all the time.

25.2 The Answer is `e`

Each of the other statements is either an exact quote or a very close paraphrasing of actual text in Chapter 13 of Silver. But statement `e` is the opposite of what Nate writes, which is "Nature's laws do not change very much."

25.3 Grading Q25

Q25 is worth 3 points. No partial credit is available. 35/48 got it right.

Setup for Q26-32

For Q26 - Q32, consider the data I have provided in the `hospsim.csv` file. The data describe 750 patients seen for care in the past year at a metropolitan hospital system. They are simulated. Available are:

- `subject.id` = Subject Identification Number (not a meaningful code)
- `age` = the patient's age, in years (all subjects are between 21 and 75)
- `ehr_time` = Continuing or New, where Continuing means their electronic health record indicates they were also seen for care in this hospital system last year. New means they are "new" to the system this year.
- `a1c` = the patient's hemoglobin A1c level (in %)
- `ldl` = the patient's LDL cholesterol level (in mg/dl)
- `sbp` = the patient's systolic blood pressure (in mm Hg)
- `bmi` = the patient's body mass index (in kg/square meter)
- `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)
- `clinic.type` = whether the patient goes to a newly built clinic or an old clinic

26 Q26 (3 points)

26.1 The Question

Using the `hospsim` data, what is the 95% confidence interval for the odds ratio which compares the odds of receiving a statin if you were seen last year as well as this year divided by the odds of receiving a statin if you were a new patient this year. Do **NOT** use a Bayesian augmentation.

- a. Odds Ratio is 0.51, CI is (0.36, 0.71)
- b. Odds Ratio is 0.72, CI is (0.60, 0.86)
- c. Odds Ratio is 1.18, CI is (1.08, 1.28)
- d. Odds Ratio is 1.98, CI is (1.41, 2.78)
- e. None of these answers are correct

26.2 The Answer is `d.`

The correct odds ratio is 1.98, with 95% CI (1.41, 2.78).

```
hospsim <- read.csv("data/hospsim.csv") %>% tbl_df

table(hospsim$ehr_time, hospsim$statin)
```

```

              No Yes
  Continuing  76 339
  New        103 232
```

We will need to reorder the `statin` variable to get the table we want, with Continuing and Yes in the top left corner.

```r
hospsim <- hospsim %>%
  mutate(statin = fct_relevel(statin, "Yes", "No")
  )

table(hospsim$ehr_time, hospsim$statin)
```

```
             Yes  No
  Continuing 339  76
  New        232 103
```

```r
twoby2(table(hospsim$ehr_time, hospsim$statin))
```

```
2 by 2 table analysis:
------------------------------------------------------
Outcome   : Yes
Comparing : Continuing vs. New

           Yes  No    P(Yes) 95% conf. interval
Continuing 339  76    0.8169    0.7767   0.8512
New        232 103    0.6925    0.6411   0.7396

                                   95% conf. interval
             Relative Risk: 1.1795    1.0838   1.2837
         Sample Odds Ratio: 1.9803    1.4093   2.7828
Conditional MLE Odds Ratio: 1.9785    1.3897   2.8261
    Probability difference: 0.1243    0.0626   0.1861

             Exact P-value: 1e-04
        Asymptotic P-value: 1e-04
------------------------------------------------------
```

26.3 Grading Q26

Q26 is worth 3 points. There is no partial credit available. 29/48 got it right.

27 Q27 (3 points)

27.1 The Question

Perform an appropriate analysis to determine whether insurance type is associated with the education (`hsgrads`) variable, ignoring all other information in the `hospsim` data. Which of the following conclusions is most appropriate based on your significance tests?

- a. The ANOVA F test is not significant, so it doesn't make sense to compare insurance types pairwise.
- b. The ANOVA F test is significant, and a Tukey HSD comparison reveals that Medicare shows significantly higher education levels than Uninsured.
- c. The ANOVA F test is significant, and a Tukey HSD comparison reveals that Medicaid's education level is significantly lower than either Medicare or Commercial.
- d. The ANOVA F test is significant, and a Tukey HSD comparison reveals that Uninsured's education level is significantly lower than Commercial or Medicare.
- e. None of these conclusions is appropriate.

27.2 The Answer is `c`

There is clearly a highly significant ANOVA test in this case.

```r
m27 <- lm(hsgrads ~ insurance, data = hospsim)

anova(m27)
```

```
Analysis of Variance Table

Response: hsgrads
           Df Sum Sq Mean Sq F value   Pr(>F)
insurance   3   3511 1170.19  10.575 8.14e-07 ***
Residuals 746  82550  110.66
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Now, we'll run the Tukey HSD comparison of the pairs of `insurance` means.

```r
tidy(TukeyHSD(aov(m27), ordered = TRUE)) %>%
  knitr::kable(digits = 2)
```

term	comparison	estimate	conf.low	conf.high	adj.p.value
insurance	UNINSURED-MEDICAID	3.07	-2.61	8.74	0.51
insurance	MEDICARE-MEDICAID	5.63	2.56	8.71	0.00
insurance	COMMERCIAL-MEDICAID	6.41	3.37	9.44	0.00
insurance	MEDICARE-UNINSURED	2.57	-2.71	7.84	0.59
insurance	COMMERCIAL-UNINSURED	3.34	-1.91	8.59	0.36
insurance	COMMERCIAL-MEDICARE	0.77	-1.41	2.96	0.80

There are two pairwise differences that pop up here as significant. Since both the Medicare-Medicaid difference and the Commercial-Medicaid difference are significantly below 0, we can conclude that the Medicaid mean is significantly lower than either the Commercial or Medicare means. That's option `c`.

27.3 Grading Q27

Q27 is worth 3 points. There is no partial credit available. 44/48 got it right.

28 Q28 (3 points)

28.1 The Question

Build a model to predict LDL cholesterol using all of the other available variables except subject ID. After adjusting for all of the other variables, which of the following statements appears true? Do not transform your outcome.

- a. Whether you were in an old or new clinic type doesn't seem to matter significantly for LDL.
- b. Older clinics had significantly higher LDL levels, holding everything else constant, and the model accounts for less than 20% of the variation in LDL.
- c. Older clinics had significantly lower LDL levels, holding everything else constant, and the model accounts for less than 20% of the variation in LDL.
- d. Older clinics had significantly higher LDL levels, holding everything else constant, and the model accounts for 20% or more of the variation in LDL.
- e. Older clinics had significantly lower LDL levels, holding everything else constant, and the model accounts for 20% or more of the variation in LDL.

28.2 The answer is `b`

```r
model_q28 <- lm(ldl ~ clinic.type + age + ehr_time +
                  insurance + hsgrads + a1c + bmi +
                  sbp + statin, data = hospsim)

summary(model_q28)
```

```

Call:
lm(formula = ldl ~ clinic.type + age + ehr_time + insurance +
    hsgrads + a1c + bmi + sbp + statin, data = hospsim)

Residuals:
   Min     1Q Median     3Q    Max
-81.32 -25.92  -6.19  19.15 142.50

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        70.71253   21.45571   3.296 0.001029 **
clinic.typeOLD      9.39474    3.13053   3.001 0.002782 **
age                -0.31976    0.17065  -1.874 0.061357 .
ehr_timeNew        -9.12444    2.76501  -3.300 0.001013 **
insuranceMEDICAID  -2.17854    4.58346  -0.475 0.634711
insuranceMEDICARE  -3.51159    3.50029  -1.003 0.316080
insuranceUNINSURED 11.16565    7.24439   1.541 0.123677
hsgrads             0.08475    0.12994   0.652 0.514466
a1c                 2.38606    0.69627   3.427 0.000644 ***
bmi                -0.35244    0.18708  -1.884 0.059975 .
sbp                 0.29011    0.09221   3.146 0.001720 **
statinNo            0.93219    3.18013   0.293 0.769506
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 36.61 on 738 degrees of freedom
Multiple R-squared:  0.08075,   Adjusted R-squared:  0.06705
F-statistic: 5.894 on 11 and 738 DF,  p-value: 3.149e-09
```

According to our model, older Clinics have significantly higher LDL cholesterol than newer ones after accounting for all of the other variables, as we can see by the significant p value from the t-test for `clinic.type`.

The overall R-square is 8.1%, well less than 20%. So the correct answer is `b`.

28.3 Grading Q28

Q28 is worth 3 points. No partial credit is available. 41/48 got it right.

29 Q29 (3 points)

29.1 The Question

Run a backwards elimination stepwise procedure. After doing so, how many of the original nine regression inputs (`clinic.type`, `age`, `ehr_time`, `insurance`, `hsgrads`, `a1c`, `bmi`, `sbp` and `statin`) remain in the model?

- a. 1, 2, or 3
- b. 4
- c. 5
- d. 6
- e. 7 or 8

29.2 The Answer is `d`

Six inputs remain. The three that drop out are `insurance`, `hsgrads` and `statin`.

```r
tidy(step(model_q28))
```

```
Start:  AIC=5412.2
ldl ~ clinic.type + age + ehr_time + insurance + hsgrads + a1c + 
    bmi + sbp + statin

              Df Sum of Sq     RSS    AIC
- statin       1     115.1  989004 5410.3
- insurance    3    5822.4  994712 5410.6
- hsgrads      1     570.0  989459 5410.6
<none>                      988889 5412.2
- age          1    4704.6  993594 5413.8
- bmi          1    4755.5  993645 5413.8
- clinic.type  1   12067.7 1000957 5419.3
- sbp          1   13264.9 1002154 5420.2
- ehr_time     1   14591.9 1003481 5421.2
- a1c          1   15736.2 1004625 5422.0

Step:  AIC=5410.29
ldl ~ clinic.type + age + ehr_time + insurance + hsgrads + a1c + 
    bmi + sbp

              Df Sum of Sq     RSS    AIC
- insurance    3    5885.5  994890 5408.7
- hsgrads      1     598.5  989603 5408.7
<none>                      989004 5410.3
- age          1    4716.4  993721 5411.9
- bmi          1    4759.6  993764 5411.9
- clinic.type  1   12111.3 1001116 5417.4
- sbp          1   13331.8 1002336 5418.3
- ehr_time     1   14507.6 1003512 5419.2
- a1c          1   15733.0 1004737 5420.1

Step:  AIC=5408.74
ldl ~ clinic.type + age + ehr_time + hsgrads + a1c + bmi + sbp
```

```
              Df Sum of Sq     RSS    AIC
- hsgrads      1       733.8  995624 5407.3
<none>                        994890 5408.7
- bmi          1     5655.8 1000546 5411.0
- age          1    12088.6 1006979 5415.8
- clinic.type  1    12995.9 1007886 5416.5
- sbp          1    13357.5 1008247 5416.7
- ehr_time     1    13908.9 1008799 5417.1
- a1c          1    16887.2 1011777 5419.4

Step:  AIC=5407.29
ldl ~ clinic.type + age + ehr_time + a1c + bmi + sbp

              Df Sum of Sq     RSS    AIC
<none>                       995624 5407.3
- bmi          1     5553.3 1001177 5409.5
- age          1    11761.8 1007386 5414.1
- clinic.type  1    12327.6 1007951 5414.5
- sbp          1    12925.8 1008549 5415.0
- ehr_time     1    13367.9 1008992 5415.3
- a1c          1    16611.9 1012236 5417.7
```

```
# A tibble: 7 x 5
  term           estimate std.error statistic    p.value
  <chr>             <dbl>     <dbl>     <dbl>      <dbl>
1 (Intercept)      82.8     17.1         4.83 0.00000166
2 clinic.typeOLD    8.97     2.96        3.03 0.00250
3 age              -0.402    0.136      -2.96 0.00315
4 ehr_timeNew      -8.59     2.72       -3.16 0.00165
5 a1c               2.44     0.694       3.52 0.000456
6 bmi              -0.379    0.186      -2.04 0.0421
7 sbp               0.285    0.0917      3.11 0.00197
```

29.3 Grading Q29

Q29 is worth 3 points. No partial credit is available. 44/48 got it right.

30 Q30 (3 points)

30.1 The Question

Compare your initial "kitchen sink" model with all 9 inputs to the model generated by the stepwise approach in Q29 using adjusted R-squared, AIC and BIC. For each summary approach, which of the two models you are comparing gives BETTER results?

Rows:

- a. AIC
- b. BIC
- c. Adjusted R^2

Columns:

1. [SMALLER]. The smaller model (stepwise result from Q29)
2. [KITCHEN SINK]. The kitchen sink model with all 9 predictors

30.2 The Answer is `a` = SMALLER, `b` = SMALLER, `c` = KITCHEN SINK

```
model_q28 %>% glance() %>% select(adj.r.squared, AIC, BIC)
```

```
# A tibble: 1 x 3
  adj.r.squared   AIC   BIC
*         <dbl> <dbl> <dbl>
1        0.0671 7543. 7603.
```

```
model_q29 <- lm(ldl ~ clinic.type + age + ehr_time + a1c +
                  bmi + sbp, data = hospsim)

model_q29 %>% glance() %>% select(adj.r.squared, AIC, BIC)
```

```
# A tibble: 1 x 3
  adj.r.squared   AIC   BIC
*         <dbl> <dbl> <dbl>
1        0.0670 7538. 7575.
```

- The AIC and BIC values are lower (better) in the 6-predictor model (`model_q29`).
- The Adjusted R^2 is a little better (higher in this case) in the original 9-predictor model (`model_q28`).

30.3 Grading Q30

Q30 is worth 3 points, with 1 point awarded per correct response.

Nearly everyone got `a` and `b` right. 34/48 got `c` right.

31 Q31 (3 points)

31.1 The Question

Now build a model using ehr_time and insurance type to predict a different outcome, hemoglobin A1c. Which of the following statements best describes the result?

- a. The model R^2 is below 10%, and both ehr_time and insurance type have a significant impact on hemoglobin A1c given the other predictor.
- b. The model R^2 is above 10%, and both ehr_time and insurance type have a significant impact on hemoglobin A1c given the other predictor.
- c. The model R^2 is below 10%, and neither ehr_time nor insurance type have a significant impact on hemoglobin A1c given the other predictor.
- d. The model R^2 is above 10%, although neither ehr_time nor insurance type have a significant impact on hemoglobin A1c given the other predictor.
- e. None of these statements are true.

31.2 The Answer is a

Both `ehr_time` and `insurance` add significant value, but the R^2 is quite low.

```r
model_q31 <- lm(a1c ~ ehr_time + insurance, data = hospsim)

summary(model_q31)
```

```

Call:
lm(formula = a1c ~ ehr_time + insurance, data = hospsim)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.5625 -1.3500 -0.4981  0.9005  8.3159 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)          7.7995     0.1295  60.213  < 2e-16 ***
ehr_timeNew          0.3125     0.1443   2.165  0.03069 *  
insuranceMEDICAID    0.2505     0.2191   1.143  0.25332    
insuranceMEDICARE   -0.5154     0.1586  -3.250  0.00121 ** 
insuranceUNINSURED   0.6354     0.3794   1.675  0.09440 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.957 on 745 degrees of freedom
Multiple R-squared:  0.03551,	Adjusted R-squared:  0.03033 
F-statistic: 6.857 on 4 and 745 DF,  p-value: 2.013e-05
```

```r
anova(model_q31)
```

```
Analysis of Variance Table

Response: a1c
```

```
           Df  Sum Sq Mean Sq F value    Pr(>F)
ehr_time    1   24.46 24.4604  6.3863 0.0117064 *
insurance   3   80.60 26.8665  7.0145 0.0001178 ***
Residuals 745 2853.46  3.8301
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

31.3 Grading Q31

Q31 is worth 3 points. No partial credit is available. 39/48 got it right.

32 Q32 (3 points)

32.1 The Question

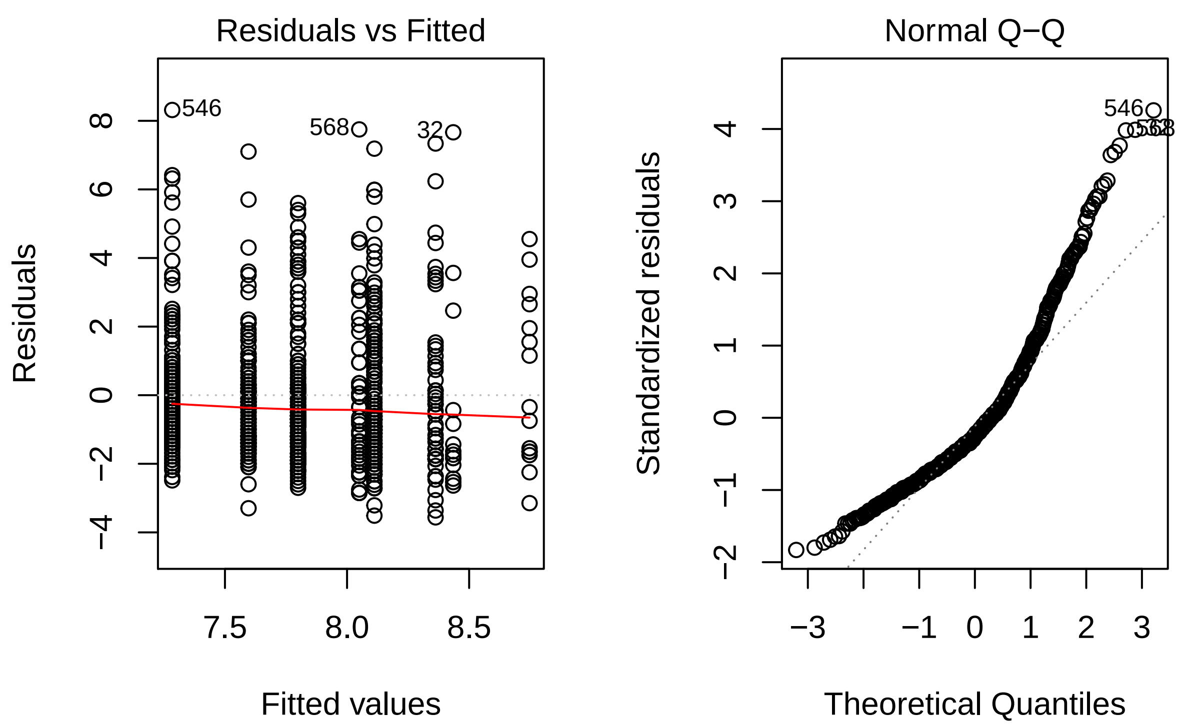
In your model for Q31, identify the subject with the largest residual. Which of the following set of features best describe this subject?

- a. This is a continuing Medicare patient who is less than 65 years of age.
- b. This is a continuing Medicare patient who is 65 years old, or older.
- c. This is a new Medicare patient who is less than 65 years of age.
- d. This is a new Medicare patient who is 65 years old, or older.
- e. None of these accurately describe the subject in question.

32.2 The Answer is a

One approach is to plot the residuals, and observe that the subject in row 546 has the largest residual.

```
par(mfrow=c(1,2))
plot(model_q31, which = 1:2)
```




```
par(mfrow=c(1,1))
```

Let's look at the details for the subject in row 546.

```r
hospsim %>% slice(546)
```

```
# A tibble: 1 x 11
  subject.id clinic.type   age ehr_time insurance hsgrads   a1c   ldl   bmi
  <fct>      <fct>       <int> <fct>    <fct>       <dbl> <dbl> <int> <dbl>
1 X1546      OLD            28 Continu~ MEDICARE     89.9  15.6   179  25.1
# ... with 2 more variables: sbp <int>, statin <fct>
```

32.3 Grading Q32

Q32 is worth 3 points. No partial credit is available. 41/48 got it right.

33 Q33 (2 points)

33.1 The Question

For each listed research question, decide what statistical procedure (of those listed) would be **most** useful in answering the question posed. Assume all assumptions have been met for using the procedure.

Rows:

- a. Do college grade point averages differ for male athletes in major sports (e.g., football), minor sports (e.g., swimming), and in intramural sports?
- b. Does intelligence as measured by IQ score differ between college students on academic probation and those not on probation?
- c. Does support for a school bond issue (For or Against) differ by neighborhood in the city?
- d. In twins of opposite sex, does the boy score higher or lower on a test of reading achievement?

Columns:

- 1. Independent Samples t test / CI.
- 2. Paired Samples t test / CI.
- 3. One-Way ANOVA comparing more than two means.
- 4. Chi-Square test of Association.

33.2 The Answer is a = 3, b = 1, c = 4, d = 2

These are meant to be straightforward.

33.3 Grading Q33

Q33 is worth 2 points. 0.5 point for each correct response.

- Everyone got `a` right.
- Nearly everyone got `b` and `c` right.
- 42/48 got `d` right.

34 Q34 (2 points)

Suppose we have several potential models for a particular outcome, and we obtain the following output.

Model	Multiple R-squared	Adjusted R-squared
A	0.41	0.40
B	0.49	0.41
C	0.53	0.43
D	0.55	0.47

34.1 The Question

Which of these models is most likely to retain its nominal R-square value in predicting new data?

- a. Model A
- b. Model B
- c. Model C
- d. Model D
- e. It is impossible to tell from the information provided.

34.2 The Answer is `a`.

Model A has a much smaller gap between its Multiple R^2 and its Adjusted R^2 than do its comparison models.

Choice `d` is likely to be a popular error. Model D has a nominal R^2 value of 0.55, so while its Adjusted R^2 is larger than, say, Model A, it is still a good distance away from the nominal value.

34.3 Grading Q34

Q34 is worth 2 points. No partial credit is available. 44/48 got it right.

35 Q35 (3 points)

Anne is a researcher on the West Coast of the U.S. who wants to estimate the amount of a newly discovered antibody in human blood. Anne's research funds will only let her obtain blood samples from 41 people, so she decides to construct a two-sided 90% confidence interval thinking it will give her a more precise estimate of the mean antibody level in the population. Bill is a researcher on the East Coast of the United States who is researching the same antibody, but he has more research funding and can afford to obtain blood samples from 121 people. Bill decides to construct a two-sided 99% confidence interval. Suppose that the sample standard deviation will be about 10 in each of the samples, and that each researcher plans to use a t-based interval.

35.1 The Question

Which estimate, Anne's or Bill's, will produce the more precise estimate, if more precise is taken to mean the interval estimate with the smaller width?

a. Anne's estimate will be more precise.
b. Bill's estimate will be more precise.
c. The estimates will be equally precise.
d. It is impossible to tell from the information provided.

35.2 The Answer is `b`

The width of a two-sided confidence interval for a mean based on a t distribution is

$$2 \times t_{\alpha/2,n-1} \frac{s}{\sqrt{n}}$$

.

- In Anne's case, we have 90% confidence, so $\alpha = 0.10$ and a sample of $n = 41$ observations, so the width of the interval will be
$$2 \times t_{.05,40} \frac{s}{\sqrt{40}}$$
. Note that $t_{0.05,40} = 1.68$ (use `qt(p = 0.05, df = 40, lower.tail = FALSE)`), and s = 10, so the width of the interval will be
$$2 \times 1.68 \frac{10}{6.32} = 5.32$$
- In Bill's case, we have 99% confidence, so $\alpha = 0.01$ and a sample of $n = 121$ observations, so the width of the interval will be
$$2 \times t_{.005,120} \frac{s}{\sqrt{120}}$$
. Note that $t_{0.005,120} = 2.62$ (use `qt(p = 0.005, df = 120, lower.tail = FALSE)`) and s = 10, so the width of the interval will be
$$2 \times 2.62 \frac{10}{10.95} = 4.79$$
.

So Bill's estimate will be more precise, because $4.79 < 5.32$.

35.3 Grading Q35

Q35 is worth 3 points. No partial credit is available. I expected this to be a difficult question, and only 20/48 got it right.

36 Q36 (2 points)

36.1 The Question

You have a tibble called `mydat` that contains 500 observations on 1 outcome and 5 predictors. Which of the following codes would most appropriately split the data into a test sample (called `mydat.test`) containing 20% of the observations, and a training sample containing the rest?

- a. `mydat.test <- sample_n(mydat, 100)` and `mydat.train = anti_join(mydat, mydat.test)`
- b. `mydat.test <- partition(mydat, 400:100)` and `mydat.train = anti_join(mydat, mydat.test)`
- c. `mydat.test <- slice(mydat, 100)` and `mydat.train = anti_join(mydat, mydat.test)`
- d. `mydat.test <- sample_frac(mydat, 0.80)` and `mydat.train = anti_join(mydat, mydat.test)`
- e. None of these approaches would work.

36.2 The Answer is `a`.

Approach `a` accomplishes the specified task. The others do not.

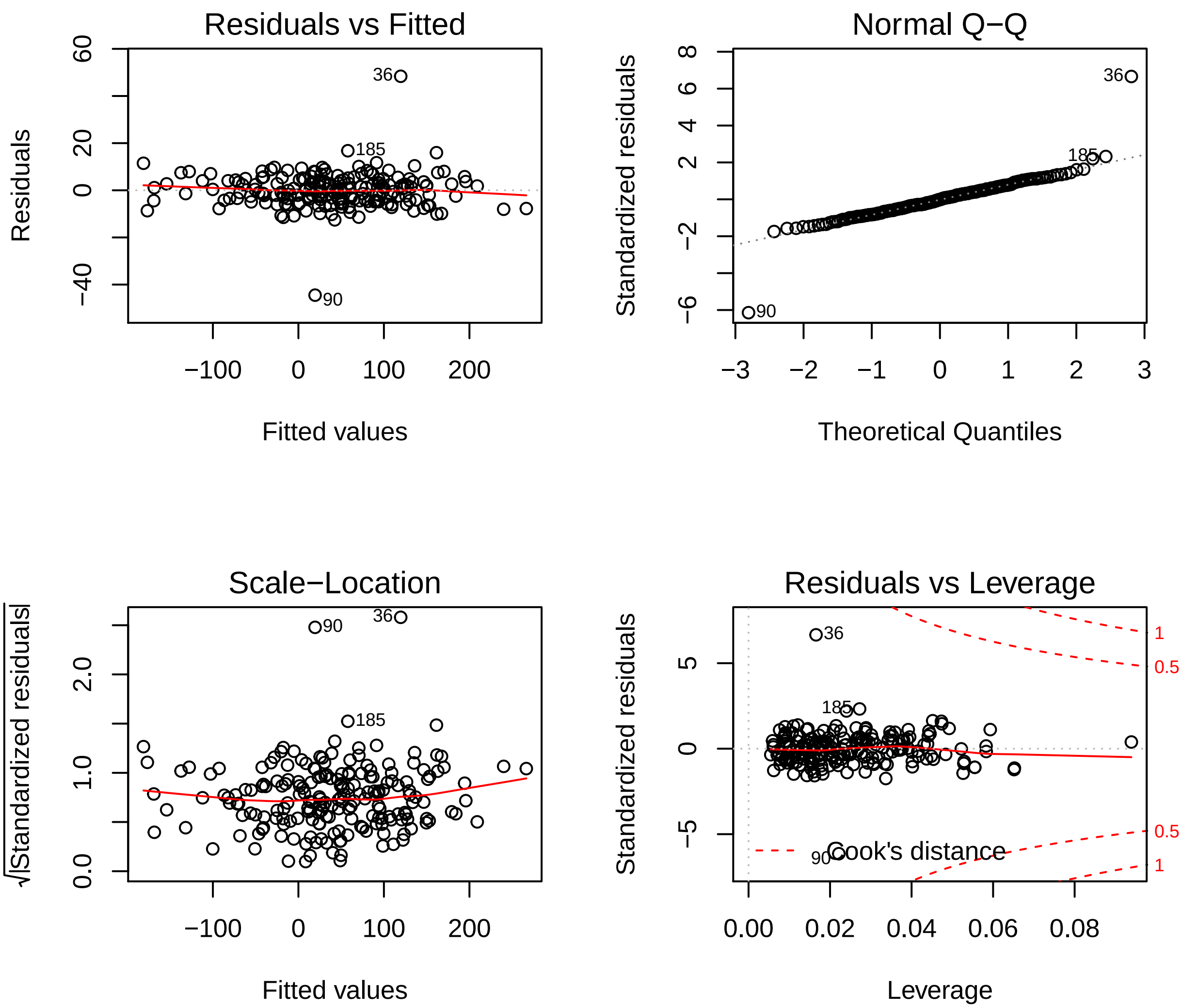
Option `d` might be popular. That would create a test sample with 80% of the observations and a training sample with the remaining 20%, which is the opposite of what we're trying to do.

36.3 Grading Q36

Q36 is worth 2 points. No partial credit is available. 31/48 got it right.

37 Q37 (2 points)

A regression model was developed to predict an outcome, `y`, based on a linear model using the four predictors `x1`, `x2`, `x3` and `x4`, in a sample of 200 subjects. The following residual plots emerged from R.



Which of the following conclusions best describes this situation, based on the output?

- a. Our main problem is with collinearity.
- b. Our main problem is with the assumption of linearity.
- c. Our main problem is with the assumption of constant variance.
- d. Our main problem is with the assumption of normality.
- e. We have no apparent problems with regression assumptions.

37.1 The Answer is `d`.

We have two especially poorly fitted points in the data, and these appear to be rows 36 and 90. These two outliers don't have especially large leverage or influence, but they are very poorly fit, with standardized residuals around 6-7 (one positive, one negative). This is primarily a problem with the assumption of Normality. If those two points were removed, there's no suggestion in any of the other plots of a problem.

Residuals of this size are definitely a violation of the Normality assumption. It is certainly true that in this case, the solution appears to be straightforward: removing these two outliers and refitting the model probably won't change our conclusions appreciably, based on the current leverage and influence profile.

37.2 Grading Q37

Q37 is worth 2 points. No partial credit is available. This should have gone better, but only 13/48 got it right. The most popular wrong answer was `e`.0 0

38 Q38 (3 points)

38.1 The Question

Suppose now that we want to build a study of the efficacy of a new drug formulation, as compared to the old formulation. We have decided that about 45% of people respond to the old formulation, and we want to declare a statistically significant effect of the new drug if we complete a two-sided test using a 5% significance level, if at least 55% of those receiving the new drug respond. If we want at least 90% power, and plan a balanced design, how many subjects will we need to enroll, in total, across the two formulation groups? Assume that no enrolled subjects will drop out of the study.

38.2 The Answer is 462 subjects.

Let's run `power.prop.test` to obtain an estimate.

```r
power.prop.test(p1 = .45, p2 = .55, power = 0.90)
```

```
     Two-sample comparison of proportions power calculation

              n = 523.2909
             p1 = 0.45
             p2 = 0.55
      sig.level = 0.05
          power = 0.9
    alternative = two.sided

NOTE: n is number in *each* group
```

We will need a minimum of 524 subjects providing data in each group, for a total of 1048 enrolled subjects across the two arms of the study in a balanced design.

38.3 Grading Q38

Q38 is worth 3 points.

- 3 points for the correct response (1048)
- 2 points for poor rounding (1046 or 1047) or forgetting to double the R estimate (524) or saying 524 in both groups, which is misleading.
- 1 point for both poor rounding and forgetting to double the R estimate (523)
- No other partial credit is available.

Points	0	1	2	3
Students	10	1	6	31

39 Q39 (2 points)

39.1 The Question

According to Jeff Leek in *The Elements of Data Analytic Style*, which of the following is **NOT** a good reason to create graphs for data exploration?

- a. To understand properties of the data.
- b. To inspect qualitative features of the data rather than a huge table of raw data.
- c. To discover new patterns or associations.
- d. To consider whether transformations may be of use.
- e. To look for statistical significance without first exploring the data.

39.2 The Answer is `e`.

If this isn't clear, take a look at Chapter 5 of Jeff's book.

39.3 Grading Q39

Q39 is worth 2 points. No partial credit is available. Everyone got this right. Hooray!

40 Q40 (2 points)

A special method using regression strategies uses a sample of data to estimate a parameter as 2.35, with a standard error of 0.5. Which of the following statements best describes a 95% uncertainty interval (confidence interval) for that parameter, based on this sample?

- a. (2.35 - 0.5, 2.35 + 0.5)
- b. (2.35 - 0.1, 2.35 + 0.1)
- c. (2 - 2.35, 2 + 2.35)
- d. (2.35 - 0.35, 2.35 + 0.35)
- e. None of these.

40.1 The Answer is `e`.

A good estimate would be to take the estimate and add and subtract 2 times the standard error. That would be `(2.35 - 2*0.5, 2.35 + 2*0.5)`, or (1.35, 3.35). None of our available responses match this result, so we must choose `e`.

40.2 Grading Q40

Q40 is worth 2 points. No partial credit is available. 35/48 got this right.

41 Answer Key

Question	Points	Correct Response
Q01	6	Essay
Q02a	1	FALSE
Q02b	1	TRUE
Q02c	1	TRUE
Q03	4	Essay
Q04	3	(0.36, 0.78)
Q05	3	d
Q06	2	b
Q07	2	c
Q08	2	c
Q09	2	d
Q10	3	d

Question	Points	Correct Response
Q11a	0.5	option 1
Q11b	0.5	option 2
Q11c	0.5	option 1
Q11d	0.5	option 4
Q12	3	c
Q13	3	a
Q14	3	b
Q15	4	90, or 90%.
Q16	3	(1.1, 1.6)
Q17a	1	Males had lower odds.
Q17b	1	Males had lower odds.
Q17c	1	Females had lower odds.
Q18	2	e
Q19a	1	FALSE
Q19b	1	TRUE
Q19c	1	TRUE
Q20	2	b

Question	Points	Correct Response
Q21	2	d
Q22	2	b
Q23	3	c
Q24	3	c
Q25	3	e
Q26	3	d
Q27	3	c
Q28	3	b
Q29	3	d
Q30a	1	SMALLER
Q30b	1	SMALLER
Q30c	1	KITCHEN SINK

Question	Points	Correct Response
Q31	3	`a`
Q32	3	`a`
Q33a	0.5	option `3` (one-way ANOVA)
Q33b	0.5	option `1` (indep. samples t)
Q33c	0.5	option `4` (chi-square)
Q33d	0.5	option `2` (paired t)
Q34	2	`a`
Q35	3	`b`
Q36	2	`a`
Q37	2	`d`
Q38	3	462 subjects
Q39	2	`e`
Q40	2	`e`

42 Grades

Sum your points across the 40 items. That's out of 110 points.

- I decided to treat this as if it was out of 100 points, without any further adjustment.

As a result, the spread of grades was:

- 4 students (8%) scored above 100, which is certainly an A.
- 17 students (35%) scored between 90 and 100, which I would consider an A.
- 4 students (8%) scored between 85 and 89.5, which I would consider as an A- or B+.
- 10 students (21%) scored between 75 and 84.5, which I would consider a B.
- 4 students (8%) scored between 70 and 74.5, which I would consider as a low B.
- with the remaining students scoring below 70.

TEL