

Answer Sketch: 431 Quiz 2: Fall 2019

Thomas E. Love

due 2019-11-18 at Noon, version 2019-11-21

Contents

Instructions	3
The Materials	3
Key Things To Remember	4
1 Question 01 (4 points)	4
1.1 Answer 01 is 4.1 with 90% CI (-4.6, 12.8) kg/m^2	5
1.2 Results 01	6
2 Question 02 (4 points)	6
2.1 Answer 02 is 80.5% with 90% CI (67.8%, 91.3%).	7
2.2 Results 02	7
3 Question 03	7
3.1 Answer 03 is $a = 3, b = 1, c = 4, d = 2$	8
3.2 Results 03	8
4 Question 04	9
4.1 Answer 04 is b.	9
4.2 Results 04	9
5 Question 05 (4 points)	9
5.1 Analysis D	10
5.2 Analysis E	11
5.3 Analysis F	11
5.4 Analysis G	11
5.5 Answer 05 is d	11
5.6 Results 05	12
6 Question 06 (4 points)	12
6.1 Analysis J	12
6.2 Analysis K	13
6.3 Analysis L	14
6.4 Analysis M	14
6.5 Answer 06 was meant to be Analysis M, which produces the right CI for the odds ratio, but it really could have been N, too.	15
6.6 Results 06	16
7 Question 07	16
7.1 Answer 07 is 0.40	16
7.2 Results 07	16
8 Question 08 (4 points)	17
8.1 Answer 08 is a line of R Code	17
8.2 Results 08	17

9 Question 09	18
9.1 Answer 09 is a.	18
9.2 Results 09	19
10 Question 10	19
10.1 Answer 10a is 4 (Ca) and 10b is 5 (Na)..	21
10.2 Results 10	21
11 Question 11	21
11.1 Answer 11 should have been e but I accepted d or e.	22
11.2 Results 11	22
12 Question 12	22
12.1 Answer 12 is e.	22
12.2 Results 12	22
13 Question 13	23
13.1 Answer 13 is e, and only e.	23
13.2 Results 13	23
13.3 Setup for Questions 14-15	23
14 Question 14 (4 points)	24
14.1 Answer 14 odds ratio is 1.96, with 95% CI (1.33, 2.88)..	24
14.2 Results 14	24
15 Question 15	25
15.1 Answer 15 is c.	25
15.2 Results 15	26
16 Question 16	26
16.1 Answer 16: a = 1, b = 2, c = 1, d = 4	26
16.2 Results 16	26
17 Question 17 (4 points)	27
17.1 Answer 17 was intended to be that b, d and f get the direction right and are reasonable, but the rest are not. But there were problems. So I also accepted b alone or with d only or with f only.	27
17.2 Problems with Question 17	29
17.3 Results 17	30
18 Question 18	30
18.1 Answer 18 is a = 2, b = 1.	31
18.2 Results 18	32
19 Question 19	32
19.1 Answer 19 is d.	32
19.2 Results 19	33
20 Question 20	33
20.1 Answer 20 is d.	33
20.2 Results 20	34
21 Question 21	34
21.1 Answer 21 is c.	34
21.2 Results 21	34

22 Question 22 (4 points)	35
22.1 Answer 22 is b	35
22.2 Results 22	35
23 Question 23	35
23.1 Answer 23 is c	36
23.2 Results 23	36
24 Question 24	36
24.1 Answer 24 is -0.0075 as a proportion or -0.75 percentage points.	36
24.2 Results 24	38
25 Question 25	38
25.1 Answer 25 is c and c only.	38
25.2 Results 25	38
26 Question 26	39
26.1 Answer 26 is c, only.	39
26.2 Results 26	39
27 Question 27 (4 points)	39
27.1 Answer 27: $RR = 1.15$ with 95% CI of (1.03, 1.30)	40
27.2 Results 27	40
28 Question 28	40
28.1 Answer 28 is a, b and c.	41
28.2 Results 28	43
29 Question 29	43
29.1 Answer 29 is e	44
29.2 Results 29	44
30 Question 30 (4 points)	44
30.1 Analysis W	45
30.2 Analysis X	46
30.3 Analysis Y	46
30.4 Analysis Z	47
30.5 Answer 30 is d	48
30.6 Results 30	48
31 Answer Key	49
32 Grading	49

Instructions

All of the links for Quiz 2 Materials are at

<https://github.com/THOMASELOVE/2019-431/tree/master/QUIZZES/QUIZ2>

The Materials

To complete the Quiz you'll need three things, all of which are linked at the URL above.

1. The `2019-431-quiz02-questions.PDF` file. This contains all of the instructions, questions and potential responses. Be sure that you see all 30 questions, and all 27 pages.
2. Five data files, named `quiz_data_states.csv`, `quiz_hosp.csv`, `quiz_ra.csv`, `quiz_sim_nejm.csv` and `quiz_statin.csv`, which may be useful to you.
3. The **Quiz 2 Answer Sheet** which is a Google Form.

Use the PDF file to read the quiz and craft your responses (occasionally making use of the provided data sets), and then place those responses into the Answer Sheet Google Form. When using the Answer Sheet, please select or type in your best response (or responses, as indicated) for each question. All of your responses must be in the Answer Sheet by the deadline.

Key Things To Remember

The **deadline** for completing the Answer Sheet is Noon on Monday 2019-11-18, and this is a firm deadline, without the grace period we allow for in turning in Homework.

The questions are not arranged in any particular order, and your score is based on the number of correct responses, so you should answer all questions. There are **30** questions, and each is worth either 3 or 4 points. The maximum possible score on the quiz is **100** points. Questions 01, 02, 05, 06, 08, 14, 17, 22, 27 and 30 are worth 4 points each. They are marked to indicate this.

If you wish to work on some of the quiz on the Answer Sheet and then return later, you can do this by [1] completing the final question which asks you to type in your full name, and then [2] submitting the Answer Sheet. You will then receive a link which allows you to return to the Answer Sheet without losing your progress.

Occasionally, I ask you to provide a single line of code. In all cases, a single line of code can include at most one pipe for these purposes, although you may or may not need the pipe in any particular setting. Moreover, you need not include the library command at any time for any of your code. Assume in all questions that all relevant packages have been loaded in R. Any reference to a logarithm refers to a natural logarithm. If you need to set a seed, use `set.seed(2019)` throughout this Quiz.

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. Thank you, and good luck.

1 Question 01 (4 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. Don't assume 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:

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

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.

1.1 Answer 01 is 4.1 with 90% CI (-4.6, 12.8) kg/m^2 .

```
sw <- starwars %>%
  mutate(bmi = 10000 * mass / (height^2)) %>%
  filter(species == "Human") %>%
  filter(complete.cases(height, mass))

mosaic::favstats(bmi ~ gender, data = sw)

  gender      min      Q1  median      Q3      max      mean      sd  n
1 female 16.52893 19.15335 21.77778 24.66299 27.54821 21.95164 5.511699  3
2 male  21.47709 23.67812 24.83565 25.98957 37.87401 26.04427 4.291291 19
missing
1      0
2      0

tidy(t.test(bmi ~ gender, data = sw, conf = 0.9)) %>%
  select(estimate, estimate1, estimate2, conf.low, conf.high)

# A tibble: 1 x 5
  estimate estimate1 estimate2 conf.low conf.high
  <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
1   -4.09      22.0      26.0     -12.8       4.62

sw %>% t.test(bmi ~ gender)
```

Welch Two Sample t-test

```
data:  bmi by gender
t = -1.2287, df = 2.3987, p-value = 0.326
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -16.36724   8.18197
sample estimates:
mean in group female    mean in group male
      21.95164           26.04427
```

That output provides a confidence interval for the female - male difference. We reverse the sign of the point estimate and confidence interval bounds to get the male - female difference.

Or, we can force R to estimate the male - female difference directly, by reordering the levels of the `gender` factor, as shown below.

```
sw <- sw %>% mutate(gen2 = fct_relevel(gender, "male"))

tidy(t.test(bmi ~ gen2, data = sw, conf = 0.9)) %>%
  select(estimate, estimate1, estimate2, conf.low, conf.high)

# A tibble: 1 x 5
  estimate estimate1 estimate2 conf.low conf.high
  <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
1    4.09      26.0      22.0     -4.62      12.8
```

Suppose that instead of doing a Welch test, you opted for a t test assuming equal variances, then you'd have:

```
sw <- sw %>% mutate(gen2 = fct_relevel(gender, "male"))
```

```
tidy(t.test(bmi ~ gen2, data = sw, var.equal = TRUE, conf = 0.9)) %>%
  select(estimate1, estimate2, conf.low, conf.high)
```

```
# A tibble: 1 x 4
  estimate1 estimate2 conf.low conf.high
    <dbl>     <dbl>   <dbl>   <dbl>
1    26.0     22.0  -0.652    8.84
```

but of course, I'd asked you not to assume equal population variances, so that's inappropriate.

1.2 Results 01

Grades on Question 01	4 points
% receiving full credit	26.7
% of available points awarded	59.6

- Point Estimate (44/60 were correct)
 - 1 point for 4.1
 - 0.5 point for -4.1
 - 0 points for anything else
- Units (38/60 were correct)
 - 1 point for the correct units (kg/m^2)
 - 0 points for incorrect or missing units
- Confidence Interval (24/60 were correct)
 - 2 points for (-4.6, 12.8)
 - 0.5 point for (-12.8, 4.6)
 - 0.5 point for (-0.7, 8.8)
 - 0 points for anything else

2 Question 02 (4 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 — 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.

2.1 Answer 02 is 80.5% with 90% CI (67.8%, 91.3%).

```
saifs.ci(33, 41, conf.level = 0.90)
```

Sample Proportion	0.05	0.95
	0.805	0.913

Remember that the SAIFS procedure automatically includes a Bayesian adjustment, so `saifs.ci(34, 42, conf.level = 0.90)` would *not* be appropriate.

2.2 Results 02

Grades on Question 02		4 points
% receiving full credit		66.7
% of available points awarded		79.8

- Point Estimate (50/60 were correct)
 - 1 point for 80.5
 - 0 points for anything else, including 80, 81 or no estimate
- Units (more than 55/60 were correct)
 - 1 point for the correct units (%)
 - 0 points for incorrect or missing units
- Confidence Interval (41/60 were correct)
 - 2 points for (67.8, 91.3)
 - 0.5 point for (65.6, 93.3), which is the 95% CI
 - 0 points for anything else

3 Question 03

Alex, Beth, Cara and Dave independently select random samples from the same population. The sample sizes are 200 for Alex, 400 for Beth, 125 for Cara, and 300 for Dave. Each researcher constructs a 95% confidence interval from their data using the same statistical method. The half-widths (margins of error) for those confidence intervals are 1.45, 1.74, 1.96 and 2.43. Match each interval's margin of error with its researcher.

Rows:

- Alex, who took a sample of $n = 200$ people.
- Beth, who took a sample of $n = 400$ people.
- Cara, who took a sample of $n = 125$ people.
- Dave, who took a sample of $n = 300$ people.

Columns:

- 1.45
- 1.74
- 1.96
- 2.43

3.1 Answer 03 is a = 3, b = 1, c = 4, d = 2

The smaller the sample size, holding everything else equal, the larger the half-width. Here's a demonstration:

```
set.seed(431123)
alex <- rnorm(200, mean = 0, sd = 15)
beth <- rnorm(400, mean = 0, sd = 15)
cara <- rnorm(150, mean = 0, sd = 15)
dave <- rnorm(300, mean = 0, sd = 15)

tidy(t.test(alex)) %>%
  mutate(diff = estimate - conf.low)

# A tibble: 1 x 9
  estimate statistic p.value parameter conf.low conf.high method
  <dbl>      <dbl>   <dbl>      <dbl>   <dbl>   <dbl> <chr>
1     1.06      1.06  0.290        199   -0.905    3.02 One S~
# ... with 2 more variables: alternative <chr>, diff <dbl>

tidy(t.test(beth)) %>%
  mutate(diff = estimate - conf.low)

# A tibble: 1 x 9
  estimate statistic p.value parameter conf.low conf.high method
  <dbl>      <dbl>   <dbl>      <dbl>   <dbl>   <dbl> <chr>
1   -0.528   -0.714  0.476        399   -1.98    0.925 One S~
# ... with 2 more variables: alternative <chr>, diff <dbl>

tidy(t.test(cara)) %>%
  mutate(diff = estimate - conf.low)

# A tibble: 1 x 9
  estimate statistic p.value parameter conf.low conf.high method
  <dbl>      <dbl>   <dbl>      <dbl>   <dbl>   <dbl> <chr>
1     1.90     1.54  0.125        149   -0.531    4.34 One S~
# ... with 2 more variables: alternative <chr>, diff <dbl>

tidy(t.test(dave)) %>%
  mutate(diff = estimate - conf.low)

# A tibble: 1 x 9
  estimate statistic p.value parameter conf.low conf.high method
  <dbl>      <dbl>   <dbl>      <dbl>   <dbl>   <dbl> <chr>
1    -1.08    -1.22  0.223        299   -2.82    0.660 One S~
# ... with 2 more variables: alternative <chr>, diff <dbl>
```

3.2 Results 03

Grades on Question 03		3 points
% receiving full credit		88.3
% of available points awarded		90.6

- 3 points for a correct response (a-3, b-1, c-4, d-2)
- 1 point for getting two of the four parts right
- 0 points otherwise

4 Question 04

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`

4.1 Answer 04 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.

4.2 Results 04

Grades on Question 04	3 points
% receiving full credit	88.3
% of available points awarded	88.3

- Partial credit was not available on this question.
- The most common incorrect response was a, 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.

5 Question 05 (4 points)

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

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

A student completed four analyses, shown below. Which of the following 90% confidence intervals for the population mean change in DAS28 at 48 weeks most appropriately compares the pricier therapy to the cheaper one?

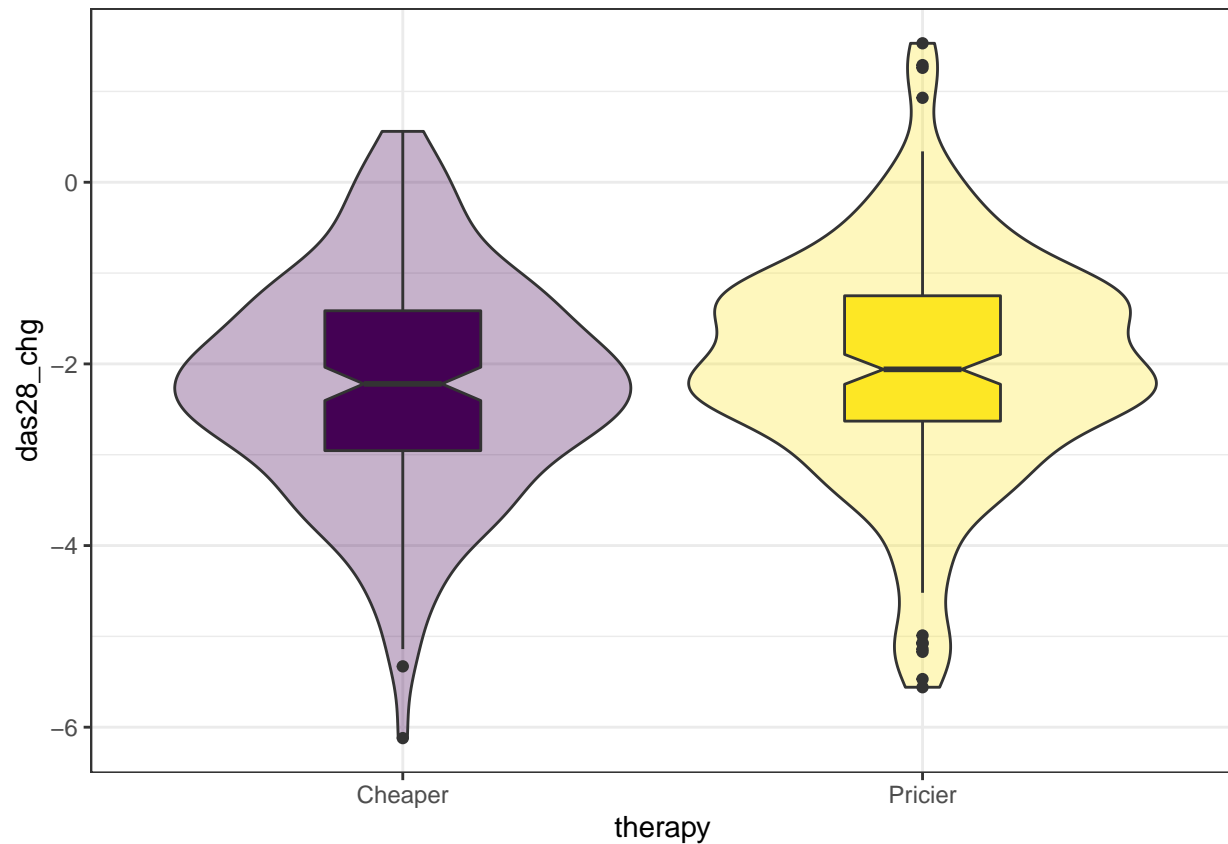
- d. Analysis D
- e. Analysis E
- f. Analysis F
- g. Analysis G

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

```
mosaic::favstats(das28_chg ~ therapy, data = ra)
```

	therapy	min	Q1	median	Q3	max	mean	sd	n	missing
1	Cheaper	-6.12	-2.955	-2.22	-1.415	0.56	-2.250857	1.208183	175	0
2	Pricier	-5.56	-2.630	-2.06	-1.250	1.53	-2.027486	1.260694	175	0

```
ggplot(data = ra, aes(x = therapy, y = das28_chg, fill = therapy)) +  
  geom_violin(alpha = 0.3) + geom_boxplot(width = 0.3, notch = TRUE) +  
  theme_bw() + guides(fill = FALSE) + scale_fill_viridis_d()
```



5.1 Analysis D

```
ra %>% t.test(das28_chg ~ therapy, var.equal = TRUE, conf.level = 0.90) %>%  
  tidy(conf.int = TRUE, conf.level = 0.90) %>%  
  mutate(estimate = estimate1 - estimate2) %>%  
  select(estimate, conf.low, conf.high, method)
```

```
# A tibble: 1 x 4  
  estimate conf.low conf.high method  
    <dbl>    <dbl>    <dbl> <chr>  
1  -0.223  -0.441  -0.00568 " Two Sample t-test"
```

5.2 Analysis E

```
ra %$% t.test(das28_chg ~ therapy, paired = TRUE) %>%
  tidy(conf.int = TRUE, conf.level = 0.90) %>%
  select(estimate, conf.low, conf.high, method)
```

```
# A tibble: 1 x 4
  estimate conf.low conf.high method
  <dbl>    <dbl>    <dbl> <chr>
1  -0.223  -0.250  -0.197 Paired t-test
```

5.3 Analysis F

```
ra %$% wilcox.test(das28_chg ~ therapy, paired = TRUE,
  conf.int = TRUE, conf.level = 0.90) %>%
  tidy() %>%
  select(estimate, conf.low, conf.high, method)
```

```
# A tibble: 1 x 4
  estimate conf.low conf.high method
  <dbl>    <dbl>    <dbl> <chr>
1  -0.230  -0.245  -0.215 Wilcoxon signed rank test with continuity co~
```

5.4 Analysis G

```
ra %$% wilcox.test(das28_chg ~ therapy, conf.int = TRUE, conf.level = 0.90) %>%
  tidy() %>%
  select(estimate, conf.low, conf.high, method)
```

```
# A tibble: 1 x 4
  estimate conf.low conf.high method
  <dbl>    <dbl>    <dbl> <chr>
1  -0.240  -0.450  -0.0300 Wilcoxon rank sum test with continuity corre~
```

5.5 Answer 05 is d

It's Analysis D, because these are independent samples, and the t test compares the means.

The Wilcoxon rank sum test (Analysis G) isn't appropriate as the data are reasonably close to a Normal distribution in each sample and we want to compare means, but I did award 1 point for this response for at least recognizing that independent samples are involved.

The data aren't paired, so E and F are clearly incorrect.

BUT there's a problem in that D doesn't actually give a 90% interval, to get that I needed the following... (of course, this doesn't change the answer though.)

```
ra %$% t.test(das28_chg ~ therapy, var.equal = TRUE, conf.level = 0.90) %>%
  tidy() %>%
  mutate(estimate = estimate1 - estimate2) %>%
  select(estimate, conf.low, conf.high, method)
```

```
# A tibble: 1 x 4
  estimate conf.low conf.high method
  <dbl>     <dbl>     <dbl> <chr>
1  -0.223   -0.441   -0.00568 " Two Sample t-test"
```

5.6 Results 05

Grades on Question 05		4 points
% receiving full credit		58.3
% of available points awarded		65.8

- 4 points for the correct response, d.
- 1 point for g, which was also the most common incorrect response.
- 0 points for any other response.

6 Question 06 (4 points)

Referring again to the study initially described in Question 05, which of the following analyses provides an appropriate 90% confidence interval for the difference (cheaper - pricier) in the proportion of participants who had a clinically meaningful improvement (DAS28 change of -1.2 or below) at 48 weeks?

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

6.1 Analysis J

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

ra <- ra %>%
  mutate(improved = das28_chg < -1.2) %>%
  mutate(improved = fct_relevel(factor(improved), "FALSE"))

ra %>% tabyl(improved, therapy)
```

```
improved Cheaper Pricier
FALSE      31      41
TRUE      144     134
```

```
twobytwo(31, 41, 144, 134,
  "improved", "didn't improve",
  "cheaper", "pricier")
```

2 by 2 table analysis:

```
-----
Outcome      : cheaper
Comparing    : improved vs. didn't improve
```

	cheaper	pricier	P(cheaper)	95% conf. interval
improved	31	41	0.4306	0.3217 0.5466
didn't improve	144	134	0.5180	0.4593 0.5762

		95% conf. interval
Relative Risk:	0.8312	0.6227 1.1096
Sample Odds Ratio:	0.7036	0.4173 1.1864
Conditional MLE Odds Ratio:	0.7043	0.4019 1.2246
Probability difference:	-0.0874	-0.2100 0.0416

Exact P-value: 0.2339
Asymptotic P-value: 0.1872

6.2 Analysis K

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

ra <- ra %>%
  mutate(improved = das28_chg <= -1.2) %>%
  mutate(improved = fct_relevel(factor(improved), "TRUE"))

ra %>% tabyl(improved, therapy)
```

```
improved Cheaper Pricier
TRUE      144      134
FALSE      31      41

twobytwo(144, 134, 31, 41,
  "improved", "didn't improve",
  "cheaper", "pricier")
```

2 by 2 table analysis:

Outcome : cheaper
Comparing : improved vs. didn't improve

	cheaper	pricier	P(cheaper)	95% conf. interval
improved	144	134	0.5180	0.4593 0.5762
didn't improve	31	41	0.4306	0.3217 0.5466

		95% conf. interval
Relative Risk:	1.2031	0.9013 1.6059
Sample Odds Ratio:	1.4213	0.8429 2.3965
Conditional MLE Odds Ratio:	1.4198	0.8166 2.4880
Probability difference:	0.0874	-0.0416 0.2100

Exact P-value: 0.2339
Asymptotic P-value: 0.1872

6.3 Analysis L

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

ra <- ra %>%
  mutate(improved = das28_chg < -1.2) %>%
  mutate(improved = fct_relevel(factor(improved), "FALSE"))

ra %>% tabyl(improved, therapy)
```

improved	Cheaper	Pricier
FALSE	31	41
TRUE	144	134

```
twobytwo(31, 41, 144, 134, conf.level = 0.90,
  "improved", "didn't improve",
  "cheaper", "pricier")
```

2 by 2 table analysis:

Outcome : cheaper
Comparing : improved vs. didn't improve

	cheaper	pricier	P(cheaper)	90% conf. interval
improved	31	41	0.4306	0.3383 0.5279
didn't improve	144	134	0.5180	0.4687 0.5669

		90% conf. interval
Relative Risk:	0.8312	0.6523 1.0592
Sample Odds Ratio:	0.7036	0.4538 1.0908
Conditional MLE Odds Ratio:	0.7043	0.4379 1.1271
Probability difference:	-0.0874	-0.1914 0.0212

Exact P-value: 0.2339
Asymptotic P-value: 0.1872

6.4 Analysis M

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

ra <- ra %>%
  mutate(improved = das28_chg <= -1.2) %>%
  mutate(improved = fct_relevel(factor(improved), "TRUE"))

ra %>% tabyl(improved, therapy)
```

improved	Cheaper	Pricier
TRUE	144	134
FALSE	31	41

```
twobytwo(144, 134, 31, 41, conf.level = 0.90,
  "improved", "didn't improve",
  "cheaper", "pricier")
```

2 by 2 table analysis:

Outcome : cheaper

Comparing : improved vs. didn't improve

	cheaper	pricier	P(cheaper)	90% conf. interval
improved	144	134	0.5180	0.4687 0.5669
didn't improve	31	41	0.4306	0.3383 0.5279

	90% conf. interval
Relative Risk: 1.2031	0.9441 1.5331
Sample Odds Ratio: 1.4213	0.9168 2.2034
Conditional MLE Odds Ratio: 1.4198	0.8872 2.2838
Probability difference: 0.0874	-0.0212 0.1914

Exact P-value: 0.2339

Asymptotic P-value: 0.1872

6.5 Answer 06 was meant to be Analysis M, which produces the right CI for the odds ratio, but it really could have been N, too.

- It's M, because that's the only one that gets the two-by-two table right and also uses 90% confidence intervals, rather than 95%.
- But in fact, we probably want to invert the rows and columns and instead run:

```
twobytwo(144, 31, 134, 41, conf.level = 0.90,  
          "cheaper", "pricier", "improved", "not improved")
```

2 by 2 table analysis:

Outcome : improved

Comparing : cheaper vs. pricier

	improved	not improved	P(improved)	90% conf. interval
cheaper	144	31	0.8229	0.7703 0.8655
pricier	134	41	0.7657	0.7090 0.8142

	90% conf. interval
Relative Risk: 1.0746	0.9824 1.1756
Sample Odds Ratio: 1.4213	0.9168 2.2034
Conditional MLE Odds Ratio: 1.4198	0.8872 2.2838
Probability difference: 0.0571	-0.0141 0.1278

Exact P-value: 0.2339

Asymptotic P-value: 0.1872

As you can see, the odds ratio is the same, but this is better. So I gave full credit to both M and N as responses.

6.6 Results 06

Grades on Question 06	4 points
% receiving full credit	95.0
% of available points awarded	95.0

- m and n received 4 points, while all other responses received 0.

7 Question 07

In response to unexpectedly low enrollment, the protocol was amended part-way through the trial described in Question 05 to change the primary outcome from a binary outcome to a continuous outcome in order to increase the power of the study.

Originally, the proposed primary outcome was the difference in the proportion of participants who had a DAS28 of 3.2 or less at week 48. The original power analysis established a sample size target of 225 completed enrollments in each therapy group, based on a two-sided 10% significance level, and a desire for 90% power. In that initial power analysis, the proportion of participants with a DAS28 of 3.2 or less at week 48 was assumed to be 0.27 under the less effective of the two therapies.

What value was used in the power calculation for the proportion of participants with DAS28 of 3.2 or less at week 48 for the more effective therapy? State your answer rounded to two decimal places.

7.1 Answer 07 is 0.40

```
power.prop.test(n = 225, p1 = 0.27, sig.level = 0.1, power = .9)
```

Two-sample comparison of proportions power calculation

```
n = 225
p1 = 0.27
p2 = 0.3996866
sig.level = 0.1
power = 0.9
alternative = two.sided
```

NOTE: n is number in *each* group

7.2 Results 07

Grades on Question 07	3 points
% receiving full credit	63.3
% of available points awarded	63.9

- Partial credit was not available on this question, except that I gave 1 point for 0.39.
- Incorrect responses ranged from 0.075 to values above 1.

8 Question 08 (4 points)

In the trial described in Question 05, 21 of the 222 subjects originally assigned to receive the cheaper therapy and 35 of the 219 subjects originally assigned to receive the pricier therapy experienced a serious adverse event (which included infections, gastrointestinal, renal, urinary, cardiac or vascular disorders, as well as surgical or medical procedures.)

Suppose you wanted to determine whether or not there was a statistically detectable difference in the rates of serious adverse events in the two therapy groups at the 5% significance level? How could you do this in R, in a single line of code?

8.1 Answer 08 is a line of R Code

```
twobytwo(21, 201, 35, 184, "Cheaper", "Pricier", "SAE", "No SAE")
```

2 by 2 table analysis:

Outcome : SAE

Comparing : Cheaper vs. Pricier

	SAE	No SAE	P(SAE)	95% conf. interval	
Cheaper	21	201	0.0946	0.0625	0.1407
Pricier	35	184	0.1598	0.1170	0.2145

	95% conf. interval		
Relative Risk:	0.5919	0.3562	0.9835
Sample Odds Ratio:	0.5493	0.3085	0.9778
Conditional MLE Odds Ratio:	0.5500	0.2928	1.0118
Probability difference:	-0.0652	-0.1282	-0.0028

Exact P-value: 0.0454

Asymptotic P-value: 0.0417

8.2 Results 08

Grades on Question 08	4 points
% receiving full credit	68.3
% of available points awarded	72.5

- A `twobytwo` approach, properly labeled, was OK. You didn't need to specify the confidence level - it was 95% by default.
 - I'd have taken `twobytwo(22, 202, 36, 185)` as well (using a Bayesian augmentation)
 - Some people used `twobytwo(21, 222-21, 35, 219-35)` and that's OK, too.
 - Another acceptable response would have been to flip the columns, with `twobytwo(201, 21, 184, 35, "Cheaper", "Pricier", "no SAE", "SAE")`
 - If you flipped the columns but not the labels, like `twobytwo(201, 21, 184, 35, "Cheaper", "Pricier", "SAE", "no SAE")` then that got you 1 point.
 - If you mistyped one of the four values, like `twobytwo(21, 201, 25, 184, "Cheaper", "Pricier", "SAE", "No SAE")` that got you 1 point.
 - If you didn't subtract, and just did `twobytwo(21, 222, 35, 219)` you got 1 point.

- If you switched the rows and columns, like `twobytwo(21, 35, 201, 184, "SAE", "no SAE", "Cheaper", "Pricier")` then that was a major problem because the comparison you're doing inverts the outcome and the exposure, so that got you 1 point.
- Using `twoby2` from `Epi` was less appropriate, but I gave full credit if (and only if) you built a matrix of appropriate values, as in `Epi::twoby2(matrix(c(21, 35, 222-21, 219-35), nrow = 2))`.
- This requires the comparison of two **proportions** using **independent** samples, so `prop.test` and `binom.test` and `anova` and `t.test` are all inappropriate.
- These aren't paired samples, so no use of `diffpropci.Wald.mp` could work, either.

9 Question 09

The `Pottery` data are part of the `carData` package in R. Included are data describing the chemical composition of ancient pottery found at four sites in Great Britain. This data set will also be used in Question 10. In this question, we will focus on the Na (Sodium) levels, and our goal is to compare the mean Na levels across the four sites.

```
anova(lm(Na ~ Site, data = carData::Pottery))
```

Analysis of Variance Table

Response: Na

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Site	3	0.25825	0.086082	9.5026	0.0003209 ***
Residuals	22	0.19929	0.009059		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Which of the following conclusions is most appropriate, based on the output above?

- The F test allows us to conclude that the population mean Na level in at least one of the four sites is detectably different than the others, at a 1% significance level.
- The F test allows us to conclude that the population mean Na level in each of the four sites is detectably different than each of the others, at a 1% significance level.
- The F test allows us to conclude that the population mean Na level is the same in all four sites, at a 1% significance level.
- The F test allows us to conclude that the population mean Na level may not be the same in all sites, but is not detectably different at the 1% level.
- None of these conclusions are appropriate.

9.1 Answer 09 is a.

The ANOVA F test gives us a p value well below 0.01, so we can certainly conclude that the population mean Na level in at least one of the four sites is different than the others, at a 1% significance level. That is, in fact, what **a** says. The other statements (**b**, **c** and **d**) show various mistaken versions of what the ANOVA F test might do.

AFTER THE QUIZ some people have suggested to me that the data don't follow a Normal distribution very well, as part of an argument for choosing **e** instead. Three thoughts:

- The samples are very small (with 5, 2, 5 and 14 observations), so we really cannot judge Normality very well.

```
carData::Pottery %>% mosaic::favstats(Na ~ Site)
```

Site	min	Q1	median	Q3	max	mean	sd	n
------	-----	----	--------	----	-----	------	----	---

```

1 AshleyRails 0.03 0.050 0.05 0.0500 0.06 0.0480000 0.01095445 5
2 Caldicot 0.04 0.045 0.05 0.0550 0.06 0.0500000 0.01414214 2
3 IsleThorns 0.03 0.040 0.04 0.0600 0.10 0.0540000 0.02792848 5
4 Llanedyrn 0.14 0.185 0.21 0.2375 0.54 0.2507143 0.12262916 14
missing
1 0
2 0
3 0
4 0

```

2. ANOVA is very robust to problems with Normality, so I don't really agree. One way to see this is that the ANOVA results are almost identical to the Kruskal-Wallis alternative here.

```
kruskal.test(Na ~ Site, data = carData::Pottery)
```

Kruskal-Wallis rank sum test

data: Na by Site

Kruskal-Wallis chi-squared = 18.794, df = 3, p-value = 0.0003015

3. Statement **a** is factually true - that is what the F test concludes, regardless of whether a Kruskal-Wallis approach might be marginally better or not.

So, sorry, but I'm only accepting **a** as a response. I'll write more carefully about this in future Quizzes if it comes up.

9.2 Results 09

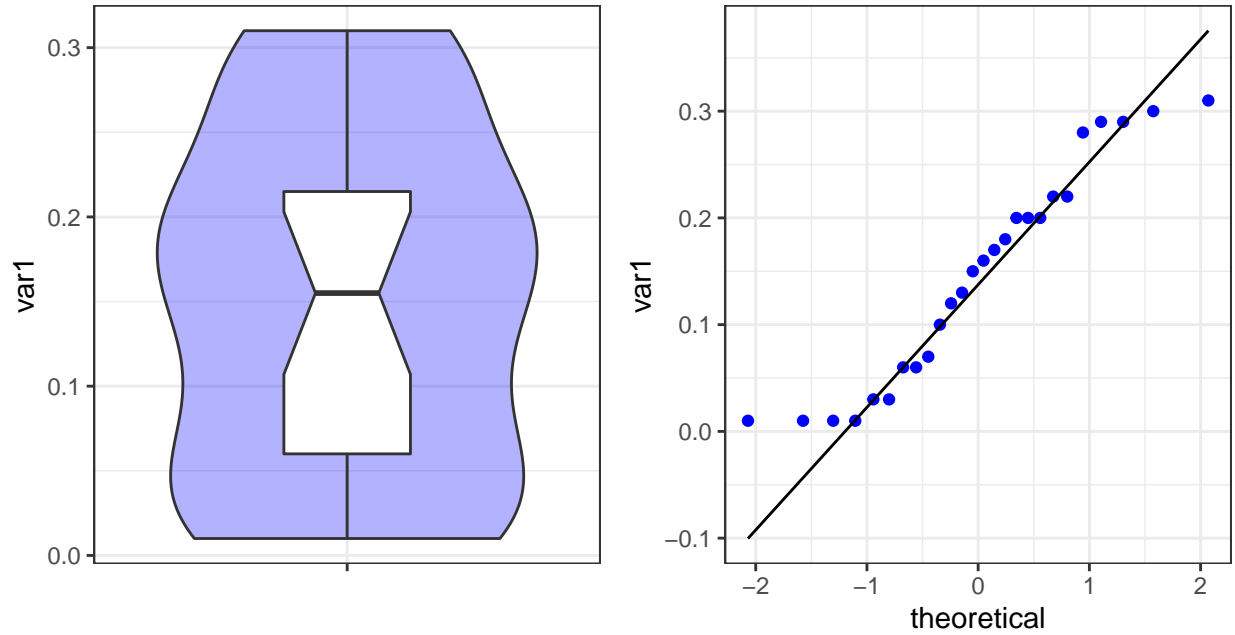
Grades on Question 09	3 points
% receiving full credit	78.3
% of available points awarded	78.3

- Partial credit was not available on this question.
- The most common incorrect responses were **e** and then **b**.

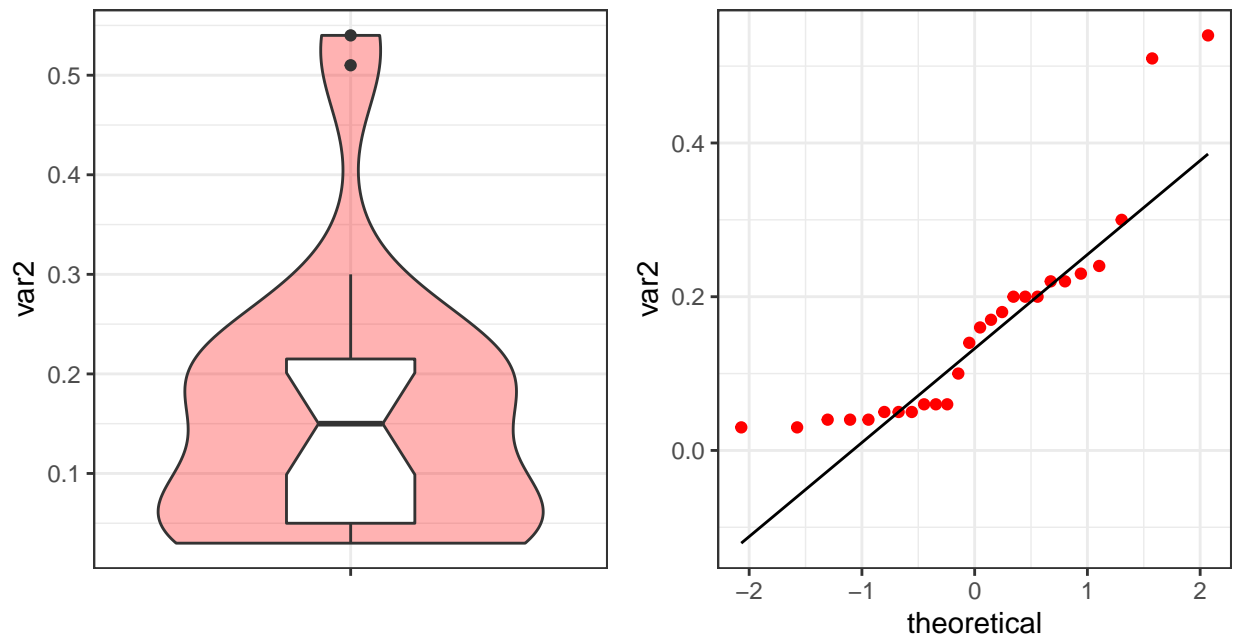
10 Question 10

Consider these two sets of plots, generated to describe variables from the Pottery data set within the `carData` package.

Plot 1 for Question 10



Plot 2 for Question 10



And now, here are summary statistics from the `mosaic::inspect` function describing the variables contained in the Pottery data set.

```
categorical variables:
  name  class levels  n missing
1 Site factor      4 26      0

distribution
```

```
1 Llanedyrn (53.8%), AshleyRails (19.2%) ...
```

```
quantitative variables:
```

```
  name  class  min    Q1 median    Q3   max    mean    sd  n
1   Al numeric 10.10 11.95 13.800 17.4500 20.80 14.4923077 2.9926474 26
2   Fe numeric  0.92  1.70  5.465  6.5900  7.09  4.4676923 2.4097507 26
3   Mg numeric  0.53  0.67  3.825  4.5025  7.23  3.1415385 2.1797260 26
4   Ca numeric  0.01  0.06  0.155  0.2150  0.31  0.1465385 0.1012301 26
5   Na numeric  0.03  0.05  0.150  0.2150  0.54  0.1584615 0.1352832 26
missing
1      0
2      0
3      0
4      0
5      0
```

Based on this output, and whatever other work you need to do, which of the statements below is true, about Variable 1 (as shown in Plot 1) and Variable 2 (shown in Plot 2)?

- a. Var1 is ...
- b. Var2 is ...

Choices are:

- 1. Al
- 2. Fe
- 3. Mg
- 4. Ca
- 5. Na

10.1 Answer 10a is 4 (Ca) and 10b is 5 (Na)..

Most people got this right. The explanation is pretty straightforward, I think.

- For **var1** the plots place the minimum near 0 and the maximum around 0.3, with a median at about 0.15.
- For **var2** the plots place the minimum near 0 and the maximum around 0.6, with a median also around 0.15.

10.2 Results 10

Grades on Question 10	3 points
% receiving full credit	over 95
% of available points awarded	over 95

- Partial credit was available on this question, but no one got exactly one of **a** and **b** right, so I didn't award any.

11 Question 11

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` in 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.

11.1 Answer 11 should have been e but I accepted d or e.

- Approach a will throw an error message.
- Approach b will create a YES/NO, rather than a TRUE/FALSE variable.
- Approach c will pull the people with bad blood pressures into a data frame called `badbp`.
- Approach d does what we're looking for, sort of, but doesn't permanently change the data frame.

So the best response would be e, because what we really wanted was something like:

```
mydata <- mydata %>% mutate(badbp = sbp >= 120)
```

11.2 Results 11

Grades on Question 11	3 points
% receiving full credit	90.0
% of available points awarded	90.0

- I gave full credit for d or e. Otherwise, partial credit was not available.
- The most common incorrect responses were b and c.

12 Question 12

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 more effectively than a huge table of raw data would allow.
- 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.

12.1 Answer 12 is e.

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

12.2 Results 12

Grades on Question 12	3 points
% receiving full credit	over 95
% of available points awarded	over 95

- Partial credit was not available on this question.

13 Question 13

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.**)

- The sample is random
- The sample is accidental
- The sample is stratified
- The sample is systematic
- The sample is representative
- None of the above

13.1 Answer 13 is e, and only e.

This is the definition of a representative sample. None of the other descriptions need necessarily hold in this case. In particular, a random sample is representative in expectation, but there are no guarantees that if the sample is representative it will also be random.

13.2 Results 13

Grades on Question 13	3 points
% receiving full credit	53.3
% of available points awarded	78.9

- Everyone selected **e**, but many people also chose other responses, none of which are correct.
- In grading, you lost 1 point for each additional response (besides **e**) that you selected, with a minimum final score on the item of 0.5.

Responses Selected	Grade
e	3
a, b, c, d, e	0.5
a, c, e	1
a, d, e	1
a, e	2
c, e	2
d, e	2

13.3 Setup for Questions 14-15

For Questions 14 and 15, consider the data I have provided in the `quiz_hosp.csv` file. The data describe 700 simulated patients at a metropolitan hospital. Available are:

- `subject.id` = 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)

14 Question 14 (4 points)

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 here.

14.1 Answer 14 odds ratio is 1.96, with 95% CI (1.33, 2.88).

```
quiz_hosp <- read.csv("data/quiz_hosp.csv") %>% tbl_df
table(quiz_hosp$sex, quiz_hosp$statin)
```

```
      NO YES
FEMALE 100 288
MALE    47 265
```

```
twobytwo(265, 47, 288, 100, "Male", "Female", "Yes", "No")
```

2 by 2 table analysis:

Outcome : Yes

Comparing : Male vs. Female

	Yes	No	P(Yes)	95% conf. interval
Male	265	47	0.8494	0.8052 0.8849
Female	288	100	0.7423	0.6964 0.7833

	95% conf. interval
Relative Risk: 1.1443	1.0616 1.2334
Sample Odds Ratio: 1.9577	1.3326 2.8762
Conditional MLE Odds Ratio: 1.9559	1.3131 2.9437
Probability difference: 0.1071	0.0470 0.1650

Exact P-value: 0.0005

Asymptotic P-value: 0.0006

14.2 Results 14

Grades on Question 14	4 points
% receiving full credit	71.7
% of available points awarded	81.3

- Point Estimate (46/60 were correct)
 - 2 points for 1.96

- 1.5 points for 2.0 or 1.95 or 1.94
- 0 points for anything else
- Confidence Interval (44/60 were correct)
 - 2 points for (1.33, 2.88)
 - 1.5 points for (1.33, 2.87) or (1.32, 2.84) or (1.3, 2.9)
 - 0 points for anything else

15 Question 15

Perform an appropriate analysis to determine whether insurance type is associated with the education (`hsgrads`) variable, ignoring all other information in the `quiz_hosp` data. Which of the following conclusions is most appropriate based on your analyses, using a 5% significance level?

- The ANOVA F test shows no detectable effect of insurance on `hsgrads`, so it doesn't make sense to compare pairs of insurance types.
- The ANOVA F test shows a detectable effect of insurance on `hsgrads`, and a Tukey HSD comparison reveals that Medicare shows detectably higher education levels than Uninsured.
- The ANOVA F test shows a detectable effect of insurance on `hsgrads`, and a Tukey HSD comparison reveals that Medicaid's education level is detectably lower than either Medicare or Commercial.
- The ANOVA F test shows a detectable effect of insurance on `hsgrads`, and a Tukey HSD comparison reveals that Uninsured's education level is detectably lower than Commercial or Medicare.
- None of these conclusions is appropriate.

15.1 Answer 15 is c.

```
anova(lm(hsgrads ~ insurance, data = quiz_hosp))
```

Analysis of Variance Table

Response: hsgrads

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
insurance	3	3617	1205.62	11.082	4.105e-07 ***
Residuals	696	75717	108.79		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
TukeyHSD(aov(hsgrads ~ insurance, data = quiz_hosp))
```

Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = hsgrads ~ insurance, data = quiz_hosp)

\$insurance		diff	lwr	upr	p adj
MEDICAID-COMMERCIAL		-6.7960790	-9.914976	-3.677182	0.0000002
MEDICARE-COMMERCIAL		-0.8069457	-3.049217	1.435325	0.7905104
UNINSURED-COMMERCIAL		-3.0478058	-8.535487	2.439875	0.4808110
MEDICARE-MEDICAID		5.9891332	2.823016	9.155250	0.0000082
UNINSURED-MEDICAID		3.7482732	-2.177194	9.673740	0.3629021
UNINSURED-MEDICARE		-2.2408601	-7.755515	3.273795	0.7221248

Medicaid is detectably lower than either Commercial or Medicare at a 5% level using the Tukey HSD comparisons. That's option **c**.

15.2 Results 15

Grades on Question 15	3 points
% receiving full credit	93.3
% of available points awarded	93.3

- Partial credit was not available on this question.
- The most common incorrect response was **e**.

16 Question 16

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 are

- Increase the level of confidence.
- Increase the sample size.
- Increase the standard error of the estimate.
- Use a bootstrap approach to estimate the CI.

Columns are

- CI will become wider
- CI will become narrower
- CI width will not change
- It is impossible to tell

16.1 Answer 16: a = 1, b = 2, c = 1, 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.

16.2 Results 16

Grades on Question 16	3 points
% receiving full credit	56.7
% of available points awarded	84.4

- Subpart Results
 - 93% of students got **a** right
 - Over 95% got **b** right
 - 93% got **c** right

- 70% got d right
- Grading
 - 4 correct responses got you 3 points
 - 3 correct got you 2 points
 - 2 correct got you 1 point

17 Question 17 (4 points)

The data in the `quiz_statin.csv` file provided to you describe the results of a study of 180 patients who have a history of high cholesterol. Patients in the study were randomly assigned to the use of a new statin medication, or to retain their current regimen. The columns in the data set show a patient identification code, whether or not the patient was assigned to the new statin (Yes or No) and their LDL cholesterol value (in mg/dl) at the end of the study. Produce a 95% confidence interval comparing the LDL levels across the two statin groups (including both a point estimate and appropriate confidence interval), and then describe your result in context in a single English sentence.

- LDL levels using the new statin were 4.95 mg/dl higher with 95% CI (0.65, 9.24) mg/dl, based on an indicator variable regression model, which replicates a two-sample t test assuming equal variances.
- LDL levels using the new statin were 4.95 mg/dl lower with 95% CI (0.65, 9.24) mg/dl, based on an indicator variable regression model, which replicates a two-sample t test assuming equal variances.
- LDL levels using the new statin were 4.95 mg/dl higher with 95% CI (0.56, 9.33) mg/dl, based on a Welch two-sample t test not assuming equal variances.
- LDL levels using the new statin were 4.95 mg/dl lower with 95% CI (0.56, 9.33) mg/dl, based on a Welch two-sample t test not assuming equal variances.
- LDL levels using the new statin were 4.95 mg/dl higher with 95% CI (0.94, 9.21) mg/dl, based on a bootstrap comparison of the population means and using the seed 2019.
- LDL levels using the new statin were 4.95 mg/dl lower with 95% CI (0.94, 9.21) mg/dl, based on a bootstrap comparison of the population means and using the seed 2019.
- None of the above are appropriate, since we should be using a paired samples analysis with these data.

17.1 Answer 17 was intended to be that b, d and f get the direction right and are reasonable, but the rest are not. But there were problems. So I also accepted b alone or with d only or with f only.

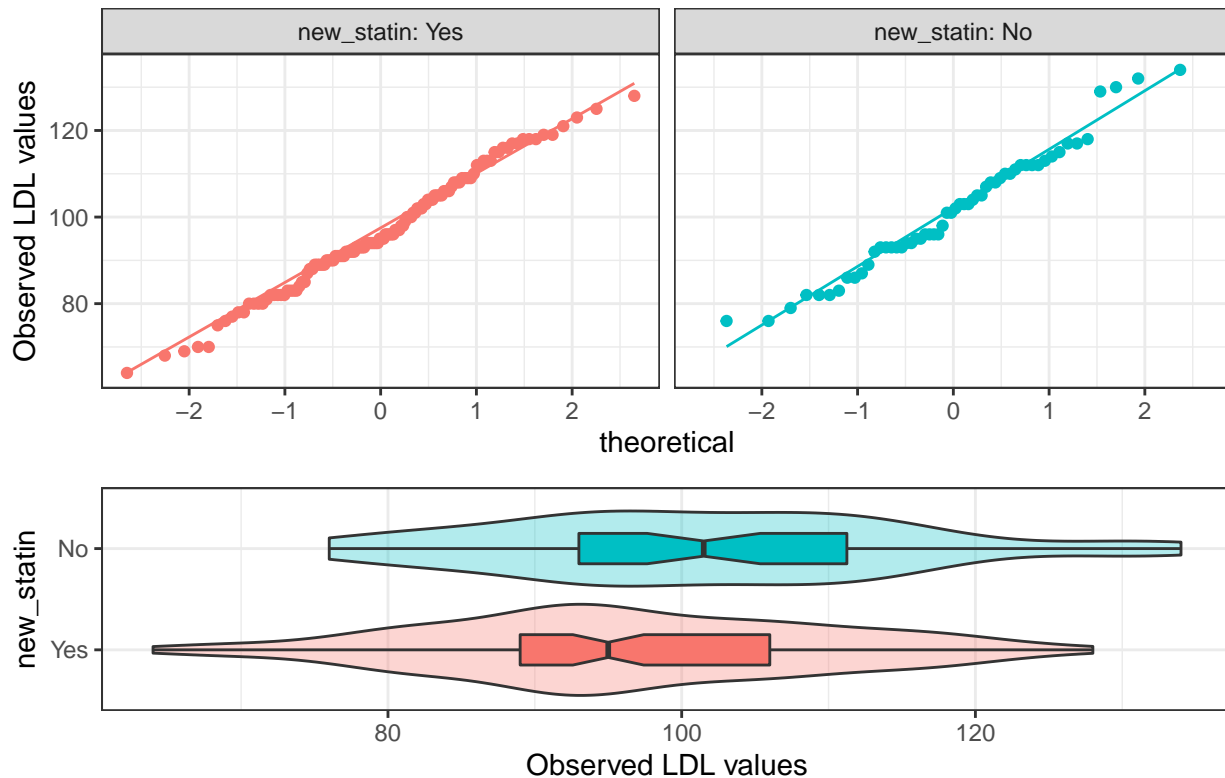
```
qs <- read.csv("data/quiz_statin.csv") %>% tbl_df() %>%
  mutate(new_statin = fct_relevel(new_statin, "Yes"))

p1 <- ggplot(qs, aes(sample = ldl, color = new_statin)) +
  geom_qq() + geom_qq_line() +
  facet_wrap(~ new_statin, labeller = "label_both") +
  guides(color = FALSE) +
  labs(y = "Observed LDL values") +
  theme_bw()

p2 <- ggplot(qs, aes(x = new_statin, y = ldl, fill = new_statin)) +
  geom_violin(alpha = 0.3) +
  geom_boxplot(width = 0.3, notch = TRUE) +
  guides(fill = FALSE) +
  coord_flip() +
  theme_bw() +
  labs(y = "Observed LDL values")
```

```
p1 + p2 + plot_layout(ncol = 1, height = c(3, 2)) +
  plot_annotation(title =
    "LDL distributions approximated well by the Normal model")
```

LDL distributions approximated well by the Normal model



Neither group shows meaningful skew or outliers, according to the plot, so I'm willing to consider a confidence interval based on the t distribution. Since the sample sizes are not balanced, we should look at the sample standard deviations to help make a decision between a Welch test and a pooled t test.

```
mosaic::favstats(ldl ~ new_statin, data = qs)
```

	new_statin	min	Q1	median	Q3	max	mean	sd	n	missing
1	Yes	64	89	95.0	106.00	128	96.51613	13.35962	124	0
2	No	76	93	101.5	111.25	134	101.46429	13.87079	56	0

Given the modest difference in sample variances, and the similar spreads of the data by IQR or range, I'm inclined to use a pooled t approach, as we get from indicator variable regression.

```
qs %>% lm(ldl ~ new_statin) %>% tidy(., conf.int = TRUE) %>%
  select(term, estimate, conf.low, conf.high)
```

```
# A tibble: 2 x 4
  term          estimate conf.low conf.high
  <chr>         <dbl>    <dbl>    <dbl>
1 (Intercept)    96.5     94.1     98.9
2 new_statinNo    4.95     0.653    9.24
```

You could also have used a Welch test, if you wanted to assume equal population variances.

```
qs %$% t.test(ldl ~ new_statin) %>% tidy(., conf.int = TRUE) %>%
  select(estimate, estimate1, estimate2, conf.low, conf.high)
```

```
# A tibble: 1 x 5
  estimate estimate1 estimate2 conf.low conf.high
    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
1   -4.95     96.5     101.    -9.33    -0.569
```

The other potentially reasonable response was a two-sample bootstrap comparison, although I don't have any problem with the assumption of Normality here, so this isn't quite as good a choice:

```
set.seed(2019)
qs %$% bootdif(ldl, new_statin)
```

```
Mean Difference      0.025      0.975
      4.9481567      0.9409562      9.2080213
```

What you couldn't have done was a paired samples analysis, not just because you couldn't arrange the data to get R to display a result, but because these samples aren't paired - among other things, the sample sizes in the two groups are not the same.

The main points are:

- these are independent, not paired samples
- each of the methods specified for independent samples was reasonable in this setting
- so long as you get the direction right

So, choices (b), (d) and (f) are all fairly reasonable, but (a), (c) and (e) are not.

17.2 Problems with Question 17

There were, I believe at least two problems with Question 17, and maybe three. All of the main points above are correct, but ...

1. I made a rounding error in choices **d** and **e** (lower bound should have been 0.57.)
2. If you did the bootstrap comparing the statins in the direction I expected you to, you got my answer shown in **f** or **g**. But if you did the bootstrap comparing the statins in the opposite direction, then even if you used the seed of 2019 you wouldn't have gotten my response.

Let me demonstrate:

- In the sketch, I re-ordered the **new_statin** factor, to produce tables (or tabyls) with Yes first:

```
qs %>% tabyl(new_statin)
```

```
new_statin  n  percent
      Yes 124 0.6888889
      No  56 0.3111111
```

So if you produce a comparison for Yes - No this way using bootdif, you get:

```
set.seed(2019)
qs %$% bootdif(ldl, new_statin)
```

```
Mean Difference      0.025      0.975
      4.9481567      0.9409562      9.2080213
```

- But what if you didn't re-order the levels of the **new_statin** factor.

```
qs2 <- read.csv("data/quiz_statin.csv") %>% tbl_df()
qs2 %>% tabyl(new_statin)
```

```
new_statin  n    percent
      No   56 0.3111111
      Yes 124 0.6888889
```

Now, “No” is first. Let’s try bootdif...

```
set.seed(2019)
qs2 %$% bootdif(1dl, new_statin)
```

```
Mean Difference      0.025      0.975
      -4.9481567      -9.3010945      -0.9099654
```

Aha! So the result would then be 4.95 with 95% CI (0.91, 9.30). This makes sense because in the first case the machine is pulling a random sample on Yes first then No, and in the second case the machine pulls No first then Yes, so that even though the sample is pulling the same way, the values are re-ordered.

17.2.1 So where does that leave us?

I had intended to make the answer **b**, **d** and **f**, but you could have correctly done the work and decided that **d** was wrong because of my rounding error and that **f** was wrong because you didn’t relevel the **new_statin** variable. So I’ve now decided to award full credit to anyone with the answer **b** either alone, or with **d** or **f** or both.

So as to not have to remove points from people, I am retaining the scoring originally planned for other options.

17.3 Results 17

Grades on Question 17		4 points
% receiving full credit		75.0
% of available points awarded		82.9

- In revising the grading, I used the following table to award partial credit.

		Responses Selected	Grade
		b, d, and f	4
		b and d	4
		b and f	4
balone	4dandf 3b,candf 1 ``d alone	d alone	1
		f alone	1
		g	0

18 Question 18

A hospital system has about 1 million records in its electronic health record database who meet our study’s qualifying requirements for inclusion and exclusion. We believe that about 20% of the subjects who qualify

by these criteria will need a particular blood test.

- a. Which will provide a confidence interval with smaller width for the proportion needing the blood test, using a Wald approach?
 - b. Which will provide a better confidence interval estimate for the sample proportion of eligible subjects who need the blood test?
1. A random sample of 85 subjects who meet the qualifying requirements.
 2. A non-random sample of 850,000 of the subjects who met the qualifying requirements in the past year.

18.1 Answer 18 is a = 2, b = 1.

As a demonstration of each result, consider the following situation. It's easy for the big data sample of 850,000 to be substantially biased, but a random sample avoids this problem (at least in expectation.) Here I created one million records, where the first 800,000 are 1 and the remaining 200,000 are zero. My random sample is just that. My non-random sample includes all of the 1s and only 50,000 of the zeros, so it's heavily biased. The width of the confidence interval follows the sample size, though, even though the non-random sample CI is substantially biased.

```
whole_pop <- tibble(status = c(rep(1, 800000), rep(0, 200000)))
```

```
set.seed(201943123)
```

```
sample_a <- sample_n(whole_pop, size = 85, replace = FALSE)
```

```
sample_b <- slice(whole_pop, 1:850000)
```

```
mosaic::favstats(~ status, data = whole_pop)
```

min	Q1	median	Q3	max	mean	sd	n	missing
0	1	1	1	1	0.8	0.4000002	1000000	0

```
mosaic::favstats(~ status, data = sample_a)
```

min	Q1	median	Q3	max	mean	sd	n	missing
0	1	1	1	1	0.7764706	0.4190826	85	0

```
mosaic::favstats(~ status, data = sample_b)
```

min	Q1	median	Q3	max	mean	sd	n	missing
0	1	1	1	1	0.9411765	0.2352943	850000	0

```
x_a <- sum(sample_a$status); n_a <- nrow(sample_a); prop.test(x_a, n_a)
```

1-sample proportions test with continuity correction

```
data: x_a out of n_a, null probability 0.5
X-squared = 24.894, df = 1, p-value = 6.057e-07
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.6706790 0.8568144
sample estimates:
```

```
      p
0.7764706
```

```
x_b <- sum(sample_b$status); n_b <- nrow(sample_b); prop.test(x_b, n_b)
```

1-sample proportions test with continuity correction

```
data:  x_b out of n_b, null probability 0.5
X-squared = 661763, df = 1, p-value < 2.2e-16
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.9406737 0.9416753
sample estimates:
      p
0.9411765
```

18.2 Results 18

Grades on Question 18	3 points
% receiving full credit	73.3
% of available points awarded	80.0

- I awarded 1.5 point for a correct response to each part (a and b.)
 - 81.7% of students got a right.
 - 78.3% of students got b right.

19 Question 19

A series of 88 models were built by a team of researchers interested in systems biology. 36 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. Which of the following intervals appropriately describes the uncertainty we have around a hit rate estimate in this setting, using a Wald confidence interval approach with a Bayesian augmentation and permitting a 10% rate of Type I error?

- (31.8%, 50.3%)
- (32.2%, 50.2%)
- 0.411 plus or minus 9 percentage points
- (32.4%, 50.3%)
- None of these intervals.

19.1 Answer 19 is d.

This is just a fancy way of asking for `prop.test` with $x = 36$ and $n = 88$, but with an added success and an added failure, so we have $x = 37$ and $n = 90$ entering our output.

```
prop.test(x = 37, n = 90, conf.level = 0.90)
```

1-sample proportions test with continuity correction

```
data:  37 out of 90, null probability 0.5
X-squared = 2.5, df = 1, p-value = 0.1138
alternative hypothesis: true p is not equal to 0.5
90 percent confidence interval:
 0.3243872 0.5033561
sample estimates:
```


p
0.4111111

So the answer (in terms of a percentage rather than a proportion) is (32.4%, 50.3%), or choice **d**.

19.2 Results 19

Grades on Question 19	3 points
% receiving full credit	53.3
% of available points awarded	53.3

- Partial credit was not available on this question.
- The most common incorrect response was **e** and then **b**.

20 Question 20

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 quiz of the semester (each section takes the same set of quizzes) 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   971.5  242.868   2.7716 0.02898
Residuals 165 14458.4   87.627
```

Emboldened by this result, the administrator decides to compare mean **exam2** scores for each possible pair of TAs, using a Bonferroni correction. Suppose she's not heard of **pairwise.t.test()** and therefore plans to make each comparison separately with two-sample t tests. If she wants to maintain an overall α level of 0.10 for the resulting suite of pairwise comparisons using the Bonferroni correction, then what significance level should she use for each of the individual two-sample t tests?

- a. She should use a significance level of 0.10 on each test.
- b. She should use 0.05 on each test.
- c. She should use 0.025 on each test.
- d. She should use 0.01 on each test.
- e. She should use 0.001 on each test.
- f. None of these answers are correct.

20.1 Answer 20 is d.

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 10% 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 0.01.
- Only that rate (contained in answer **c**) will ensure that our overall error rate across all 10 comparisons will be no more than 0.10.
- So you'll need to make sure that's the approach taken the administrator at Tick-Tock-Tech.

20.2 Results 20

Grades on Question 20	3 points
% receiving full credit	71.7
% of available points awarded	71.7

- Partial credit was not available on this question.
- The most common incorrect response was **f** and then **c**.

21 Question 21

If the administrator at the Watchmaker's Technical Institute that we mentioned in Question 20 instead used a Tukey HSD approach to make her comparisons, she might have obtained the following output.

Tukey multiple comparisons of exam2 means, 90% family-wise confidence level

	diff	lwr	upr			diff	lwr	upr
	-----	-----	-----			-----	-----	-----
Beth-Amy	1.21	-4.43	6.83		Donna-Beth	-6.53	-12.16	-0.90
Carmen-Amy	-1.41	-7.04	4.22		Elena-Beth	-0.24	-5.87	5.40
Donna-Amy	-5.32	-10.96	0.31		Donna-Carmen	-3.91	-9.54	1.72
Elena-Amy	0.97	-4.66	6.60		Elena-Carmen	2.38	-3.25	8.01
Carmen-Beth	-2.62	-8.25	3.01		Elena-Donna	6.29	0.66	11.93

Note that when we refer in the responses below to Beth's scores, we mean the scores of students who were in Beth's lab section. Which conclusion of those presented below would be most appropriate?

- Amy's scores are significantly higher than Carmen's or Elena's.
- Beth's scores were significantly higher than Amy's.
- Donna's scores are significantly lower than Beth's or Elena's.
- Elena's scores are significantly lower than Donna's.
- None of these answers are correct.

21.1 Answer 21 is c.

There are ten comparisons being made, but only two are statistically significant, in that the Tukey HSD interval does not include a mean difference of zero. Those are the comparisons of Donna to Beth and Donna to Elena. In each case, Donna has lower scores, since the D-B difference is negative and the E-D difference is positive.

21.2 Results 21

Grades on Question 21	3 points
% receiving full credit	80.0
% of available points awarded	80.0

- Partial credit was not available on this question.
- The most common incorrect response was **e** and then **d**.

22 Question 22 (4 points)

The `quiz_data_states` data set contains information on several variables related to the 50 United States plus the District of Columbia. The available data include 102 rows of information on six columns, and those columns are:

- `code`: the two-letter abbreviation for the “state” (DC = Washington DC, etc.)
- `state`: the “state” name
- `year`: 2019 or 2010, the year for which the remaining variables were obtained
- `population`: number of people living in the “state”
- `poverty_people`: number of people in the “state” living below the poverty line
- `poverty_rate`: % of people living in the “state” who are below the poverty line

Our eventual goal is to use the `quiz_data_states` data to produce an appropriate 90% confidence interval for the change from 2010 to 2019 in poverty rate, based on an analysis of the data at the level of the 51 “states”.

Which of the following statements is most true?

- This should be done using a paired samples analysis, and the `quiz_data_states` data require us to calculate the paired differences, but are otherwise ready to plot now.
- This should be done using a paired samples analysis, and the `quiz_data_states` data require us to pivot the data to make them wider, and then calculate the paired differences and plot them.
- This should be done using a paired samples analysis, and the `quiz_data_states` data require us to pivot the data to make them longer, and then calculate the paired differences and plot them.
- This should be done using an independent samples analysis, and the `quiz_data_states` data are ready to be plotted appropriately now.
- This should be done using an independent samples analysis, and the `quiz_data_states` data require us to pivot the data to make them wider, and then plot the distributions of the two samples.
- This should be done using an independent samples analysis, and the `quiz_data_states` data require us to pivot the data to make them longer, and then plot the distributions of the two samples.

22.1 Answer 22 is b

The data are paired by state. We need to widen the data with `pivot_wider` so that we get one row per state, showing both 2019 and 2010 rates, then subtract to get the paired differences, before plotting. So that’s b.

22.2 Results 22

Grades on Question 22		4 points
% receiving full credit		85.0
% of available points awarded		92.1

- I gave 4 points for **b** because it’s correct.
- I gave 3 points for **a** because it’s only a partial response, but it’s not wrong.
- I gave 1 point for **c** or **d** which was a mistake in grading (for **d**) but I won’t fix it now.

23 Question 23

Which of the following is the most appropriate way to complete the development of the confidence interval proposed in Question 22?

- a. Tukey HSD comparisons following an Analysis of Variance
- b. Applying `tidy()` to an Indicator Variable Regression
- c. Applying `tidy()` to an Intercept-only Regression
- d. A Wilcoxon-Mann-Whitney Rank Sum Confidence Interval
- e. A bootstrap on the `poverty_people` values across the states

23.1 Answer 23 is c

You want to find the paired poverty rate differences, then get a paired t test to compare them, and that's an intercept-only regression model.

- If these were independent samples, then either an indicator variable regression (b) or even (a) might be appropriate. But these are paired samples.
- Similarly, a Wilcoxon-Mann-Whitney Rank Sum would be appropriate only for independent samples (it's the Wilcoxon signed rank test we might use in paired samples.)
- A bootstrap would only be appropriate for the correct response, which `poverty_people` certainly is not in this context, since we want to compare differences of poverty rates.

23.2 Results 23

Grades on Question 23	3 points
% receiving full credit	61.7
% of available points awarded	61.7

- Partial credit was not available on this question.
- The most common incorrect response was d but each of the wrong responses were selected a few times.

24 Question 24

Use the data you have been provided in `quiz_data_states.csv` to provide a point estimate of the change from 2010 to 2019 in the poverty rate in the United States as a whole. Provide your response as a proportion with four decimal places. Note carefully what I am asking for (and not asking for) here.

24.1 Answer 24 is -0.0075 as a proportion or -0.75 percentage points.

I was looking for you to realize that the average of the 51 “states” is not correct here. You need a weighted average, based on population. Instead, I just wanted you to add up the `poverty_people` counts and `population` counts for each year, and then calculate the rate in each year, then take the difference. So in 2010, we have:

```
quiz_data_states <- read_csv("data/quiz_data_states.csv")
```

Parsed with column specification:

```
cols(
  code = col_character(),
  state = col_character(),
  year = col_double(),
  population = col_double(),
  poverty_people = col_double(),
  poverty_rate = col_double()
```

```
)
quiz_data_states %>%
  group_by(year) %>%
  summarize(pop = sum(population), pov_n = sum(poverty_people), rate = pov_n/pop)
```

```
# A tibble: 2 x 4
  year      pop    pov_n rate
  <dbl>   <dbl>   <dbl> <dbl>
1  2010 308745538 46215956 0.150
2  2019 321004407 45650345 0.142
```

and so we have the point estimate for the (2019 - 2010) difference as $0.1422 - 0.1497 = -0.0075$. The poverty rate fell 0.75 percentage points in that nine-year period.

If you mistakenly took the (unweighted by population) average poverty rate across the 51 “states”, you would have gotten:

```
quiz_data_states <- read_csv("data/quiz_data_states.csv")

quiz_data_states_wider <- quiz_data_states %>%
  pivot_wider(id = c(code, state),
              names_prefix = "in", names_from = year,
              values_from = poverty_rate) %>%
  mutate(diff = in2019 - in2010)

quiz_data_states_wider
```

```
# A tibble: 51 x 5
  code state      in2019 in2010      diff
  <chr> <chr>      <dbl> <dbl>      <dbl>
1 AL   Alabama      0.175  0.186    -0.0106
2 AK   Alaska       0.0994 0.0975    0.00190
3 AZ   Arizona       0.166  0.171   -0.00560
4 AR   Arkansas      0.176  0.183   -0.0074
5 CA   California    0.148  0.155   -0.00710
6 CO   Colorado      0.113  0.131   -0.0186
7 CT   Connecticut   0.0976 0.098   -0.000400
8 DE   Delaware      0.118  0.115    0.0027
9 DC   District of Columbia 0.165  0.182   -0.0167
10 FL   Florida       0.151  0.162   -0.0107
# ... with 41 more rows
```

And then, if we just take the mean of those paired differences, we get -0.0070, instead of -0.0075, because not every “state” should carry the same weight, because they don’t all have the same population.

```
mosaic::favstats(~ diff, data = quiz_data_states_wider)

      min      Q1  median      Q3     max      mean      sd  n
-0.0215 -0.01065 -0.0071 -0.0027 0.0083 -0.006996078 0.006382537 51
missing
0
```

That’s why I asked for your response to four decimal places.

24.2 Results 24

I expected this question to be difficult before any of you saw the Quiz. It was.

Grades on Question 24	3 points
% receiving full credit	15.0
% of available points awarded	36.7

- I gave full credit (3 points) for both -0.0075 and 0.0075, even though I had specified the direction (2010 to 2019) for you in the question.
- I gave 2 points for 0.0074 or 0.0076, regardless of the sign.
- I gave 1 point for -0.0070 or 0.0070, even though, as mentioned, it's not the way to go. I also gave 1 point for 0.0069 or 0.0071 regardless of the sign.
- Those of you who fit confidence intervals were very much on the wrong track, but I still gave you the point you had earned for your point estimate.
- That was it for partial credit.

25 Question 25

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? (**CHECK ALL THAT APPLY**)

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

25.1 Answer 25 is c and c only.

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 c) is clearly part of the Bayesian approach, and is explicitly described in the book.

25.2 Results 25

Grades on Question 25	3 points
% receiving full credit	71.7
% of available points awarded	81.1

- In addition to including c, it was important that you **not** include e in your response, since that was the focus of Chapter 8, which I'd pointed you to.
- I assigned credit according to the following table.

Response	Grade
c only	3
b and c only	2
c and d only	2
c and e only	1
all other responses	0

26 Question 26

According to Jeff Leek in *The Elements of Data Analytic Style*, which of the following is **NOT** a good idea in creating graphs you will share with other people to describe your work? (**CHECK ALL THAT APPLY**)

- If you have multiple plots to compare, use the same scale on the vertical axis.
- Axis labels should be large, easy to read, in plain language.
- Add a third dimension, perhaps with animation.
- Include units in figure labels and legends.
- Use color and size to help communicate information, for instance to point out confounding.

26.1 Answer 26 is c, only.

If this isn't clear, take a look at Chapter 10 of Jeff's book.

26.2 Results 26

Grades on Question 26	3 points
% receiving full credit	85.0
% of available points awarded	89.2

- I assigned credit according to the following table.

Response	Grade
c alone	3
c and one other	1.5
c and two others	1
c and three others	0.5
all other responses	0

27 Question 27 (4 points)

Suppose that 200 of 260 applicants from students at private undergraduate institutions to a graduate school are accepted, while 140 of 210 from students at public undergraduate institutions are accepted. Estimate a two-sided 95% confidence interval for the relative risk of acceptance for a “private undergrad” applicant as compared to a “public undergrad” applicant. Round your response to two decimal places.

27.1 Answer 27: RR = 1.15 with 95% CI of (1.03, 1.30)

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

```
twobytwo(200, 60, 140, 70, "Private", "Public", "Accepted", "Not Accepted")
```

2 by 2 table analysis:

Outcome : Accepted

Comparing : Private vs. Public

	Accepted	Not Accepted	P(Accepted)	95% conf. interval
Private	200	60	0.7692	0.7141 0.8164
Public	140	70	0.6667	0.6002 0.7271

	95% conf. interval
Relative Risk: 1.1538	1.0269 1.2964
Sample Odds Ratio: 1.6667	1.1095 2.5035
Conditional MLE Odds Ratio: 1.6648	1.0869 2.5568
Probability difference: 0.1026	0.0210 0.1838

Exact P-value: 0.0169

Asymptotic P-value: 0.0139

27.2 Results 27

Grades on Question 27	4 points
% receiving full credit	66.7
% of available points awarded	74.4

- Point Estimate (45/60 were correct)
 - 2 points for 1.15
 - 1 point for 1.14 or 1.16 or 1.1 or 1.2
 - 0 points for anything else
- Confidence Interval (40/60 were correct)
 - 2 points for (1.03, 1.30)
 - 1.5 points for (1.02, 1.30) or (1.02, 1.29) or (1.03, 1.31)
 - 1 point for (1, 1.3)
 - 0 points for anything else

28 Question 28

Each of the 470 students described in Question 27 applied to exactly one program at the school: either Program A, B or C. Breaking down the applications, we find that

- Program A received 120 applications, and accepted 75.
- Program A accepted 35 of its 60 applicants who came from private schools.
- Program B received 125 applications in total.
- Program B accepted exactly half of its 20 applicants from private schools.
- Program C accepted 40 of its 45 applicants from public schools.

- Program C rejected 25 applicants from private schools.

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

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

28.1 Answer 28 is a, b and c.

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

- for the school, we know that 200 of 260 male and 140 of 210 female applicants were admitted.

```
twobytwo(200, 60, 140, 70, "Private", "Public", "Accepted by School", "Not Accepted by School")
```

2 by 2 table analysis:

```
-----
Outcome      : Accepted by School
Comparing    : Private vs. Public
```

	Accepted by School	Not Accepted by School	P(Accepted by School)
Private	200	60	0.7692
Public	140	70	0.6667
95% conf. interval			
Private	0.7141	0.8164	
Public	0.6002	0.7271	

	95% conf. interval		
Relative Risk:	1.1538	1.0269	1.2964
Sample Odds Ratio:	1.6667	1.1095	2.5035
Conditional MLE Odds Ratio:	1.6648	1.0869	2.5568
Probability difference:	0.1026	0.0210	0.1838

```
Exact P-value: 0.0169
Asymptotic P-value: 0.0139
-----
```

- and so the odds ratio is 1.67, which is, of course, greater than 1.
- Working through the available materials, we obtain the following results for each program...

For Program A, we have an odds ratio of 0.70.

```
twobytwo(35, 25, 40, 20, "Private", "Public", "Accepted by A", "Not Accepted by A")
```

2 by 2 table analysis:

```
-----
Outcome      : Accepted by A
Comparing    : Private vs. Public
```

	Accepted by A	Not Accepted by A	P(Accepted by A)	95% conf.
Private	35	25	0.5833	0.4559

Public	40	20	0.6667	0.5390
interval				
Private	0.7005			
Public	0.7738			

		95% conf. interval
Relative Risk:	0.8750	0.6621 1.1564
Sample Odds Ratio:	0.7000	0.3331 1.4710
Conditional MLE Odds Ratio:	0.7021	0.3111 1.5688
Probability difference:	-0.0833	-0.2481 0.0880
Exact P-value:	0.4509	
Asymptotic P-value:	0.3465	

For Program B, we have an odds ratio of 0.75.

```
twobytwo(10, 10, 60, 45, "Private", "Public", "Accepted by B", "Not Accepted by B")
```

2 by 2 table analysis:

Outcome : Accepted by B
Comparing : Private vs. Public

	Accepted by B	Not Accepted by B	P(Accepted by B)	95% conf.
Private	10	10	0.5000	0.2939
Public	60	45	0.5714	0.4753
interval				
Private	0.7061			
Public	0.6624			

		95% conf. interval
Relative Risk:	0.8750	0.5477 1.3979
Sample Odds Ratio:	0.7500	0.2878 1.9548
Conditional MLE Odds Ratio:	0.7518	0.2563 2.2031
Probability difference:	-0.0714	-0.2916 0.1508
Exact P-value:	0.6267	
Asymptotic P-value:	0.5561	

For Program C, we have an odds ratio of 0.775.

```
twobytwo(155, 25, 40, 5, "Private", "Public", "Accepted by C", "Not Accepted by C")
```

2 by 2 table analysis:

Outcome : Accepted by C
Comparing : Private vs. Public

	Accepted by C	Not Accepted by C	P(Accepted by C)	95% conf.
Private	155	25	0.8611	0.8025
Public	40	5	0.8889	0.7595
interval				
Private	0.9044			
Public	0.9530			

	95% conf. interval		
Relative Risk:	0.9688	0.8602	1.0910
Sample Odds Ratio:	0.7750	0.2791	2.1517
Conditional MLE Odds Ratio:	0.7758	0.2182	2.2438
Probability difference:	-0.0278	-0.1133	0.1034

Exact P-value: 0.8073
Asymptotic P-value: 0.6247

Private school students were more likely to be accepted than public school students overall, but were less likely to be accepted into each of the three programs. This happens because it was substantially easier to be accepted into Program C, and private school students were also far more likely to apply to Program C.

28.2 Results 28

Grades on Question 28	3 points
% receiving full credit	40.0
% of available points awarded	57.8

- I assigned credit according to the table below.

Response	Grade
a, b, and c	3
a and b	2
a and c	2
a alone	1
b alone	1
all other responses	0

29 Question 29

Suppose we wanted to do a new study of people guessing my age, at the 5% significance level. We believe we can enroll 220 people in total, 100 of whom are female, 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 clinically 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?

- Less than 80%.
- Between 80% and 84.9%.
- Between 85% and 89.9%.
- Between 90% and 94.9%.
- 95% or higher.
- It is impossible to estimate the power in this setting.

29.1 Answer 29 is e

The power is estimated to be 95.7%, using the following function.

```
pwr::pwr.t2n.test(n1 = 100, n2 = 120, d = 2.5/5,  
                 sig.level=0.05, alternative="two.sided")
```

```
t test power calculation
```

```
      n1 = 100  
      n2 = 120  
      d = 0.5  
sig.level = 0.05  
  power = 0.9569615  
alternative = two.sided
```

29.2 Results 29

Grades on Question 29	3 points
% receiving full credit	90.0
% of available points awarded	90.0

- Partial credit was not available on this question.
- The most common incorrect response was d.

30 Question 30 (4 points)

A June 2018 special article in the New England Journal of Medicine by Michelle M. Mello, Van Lieu and Steven N. Goodman entitled “Clinical Trial Participants’ Views of the Risks and Benefits of Data Sharing” is the focus of this question. You can see the whole article (although it won’t help you with the Quiz) at <https://www.nejm.org/doi/full/10.1056/NEJMsa1713258>.

In that article, the investigators describe a survey of 771 “current and recent participants from a diverse sample of clinical trials at three academic medical centers in the United States” and report an overall response rate of 79%.

The 771 respondents were asked “How likely would you be to allow your anonymous, individual clinical trial data to be shared with...” two types of people, specifically:

1. scientists in universities and other not-for-profit organizations
2. scientists in companies developing medical products, such as prescription drugs

The data in `quiz_sim_nejm.csv` show simulated responses (collapsed to address the question of whether the response in each case was “Very Likely” [indicated by Yes] or something else [indicated by No]) from 771 subjects to these items. I’ve used those data to create four analyses, labeled W, X, Y and Z, which are available to you, below.

Your job is to peruse these four analyses to find (if possible) appropriate point and 90% confidence interval estimates for the difference in the proportion of respondents who would be “Very Likely” to allow universities but not companies to use their data to share their results. Which of the following responses is correct?

- a. Point Estimate is -0.1582, with 90% CI (-0.1981, -0.1177) from Analysis W
- b. Point Estimate is 0.0892, with 90% CI (0.0251, 0.1525) from Analysis X

- c. Point Estimate is 0.1582, with 90% CI (0.1177, 0.1981) from Analysis Y
- d. Point Estimate is 0.1582, with 90% CI (0.1197, 0.1968) from Analysis Z
- e. None of these estimates are correct.

```
quiz_sim_nejm <- read.csv(here("data", "quiz_sim_nejm.csv")) %>%
  tbl_df() %>% clean_names() %>%
  mutate(university = fct_relevel(university, "Yes"),
         company = fct_relevel(company, "Yes"))

head(quiz_sim_nejm)
```

```
# A tibble: 6 x 3
  subject university company
  <fct>    <fct>    <fct>
1 S-001   Yes      Yes
2 S-002   Yes      No
3 S-003   Yes      Yes
4 S-004   Yes      Yes
5 S-005   No       Yes
6 S-006   Yes      No
```

```
sim_long <- pivot_longer(quiz_sim_nejm, -subject,
                        names_to = "type",
                        values_to = "response") %>%
  mutate(response = fct_relevel(response, "Yes"))

head(sim_long)
```

```
# A tibble: 6 x 3
  subject type      response
  <fct>    <chr>    <fct>
1 S-001   university Yes
2 S-001   company   Yes
3 S-002   university Yes
4 S-002   company   No
5 S-003   university Yes
6 S-003   company   Yes
```

30.1 Analysis W

```
sim_long %>% table(type, response)
```

```
      response
type   Yes  No
company 412 359
university 534 237
```

```
Epi::twoby2(sim_long %>% table(type, response),
            conf.level = 0.90)
```

2 by 2 table analysis:

```
-----
Outcome      : Yes
Comparing    : company vs. university
```

	Yes	No	P(Yes)	90% conf. interval
company	412	359	0.5344	0.5047 0.5638
university	534	237	0.6926	0.6646 0.7192

	90% conf. interval
Relative Risk:	0.7715 0.7209 0.8258
Sample Odds Ratio:	0.5093 0.4276 0.6067
Conditional MLE Odds Ratio:	0.5096 0.4253 0.6101
Probability difference:	-0.1582 -0.1981 -0.1177

Exact P-value: 0.0000
Asymptotic P-value: 0.0000

30.2 Analysis X

```
quiz_sim_nejm %>% table(university, company)
```

	company
university	Yes No
Yes	300 234
No	112 125

```
Epi::twoby2(quiz_sim_nejm %>% table(university, company),
  conf.level = 0.90)
```

2 by 2 table analysis:

Outcome : Yes
Comparing : Yes vs. No

	Yes	No	P(Yes)	90% conf. interval
Yes	300	234	0.5618	0.5262 0.5967
No	112	125	0.4726	0.4197 0.5260

	90% conf. interval
Relative Risk:	1.1888 1.0447 1.3528
Sample Odds Ratio:	1.4309 1.1059 1.8514
Conditional MLE Odds Ratio:	1.4302 1.0926 1.8732
Probability difference:	0.0892 0.0251 0.1525

Exact P-value: 0.0234
Asymptotic P-value: 0.0222

30.3 Analysis Y

```
y <- sim_long %>%
  mutate(type = fct_relevel(type, "university"))
y %>% table(type, response)
```

	response
--	----------

```

type      Yes  No
university 534 237
company    412 359

```

```

Epi::twoby2(y %$$ table(type, response),
             conf.level = 0.90)

```

2 by 2 table analysis:

Outcome : Yes

Comparing : university vs. company

	Yes	No	P(Yes)	90% conf. interval
university	534	237	0.6926	0.6646 0.7192
company	412	359	0.5344	0.5047 0.5638

	90% conf. interval
Relative Risk: 1.2961	1.2110 1.3872
Sample Odds Ratio: 1.9633	1.6483 2.3385
Conditional MLE Odds Ratio: 1.9625	1.6391 2.3514
Probability difference: 0.1582	0.1177 0.1981

Exact P-value: 0.0000

Asymptotic P-value: 0.0000

30.4 Analysis Z

```

quiz_sim_nejm %>% tabyl(university, company) %>%
  adorn_totals(where = c("row", "col")) %>%
  adorn_title()

```

	company		
	Yes	No	Total
university			
Yes	300	234	534
No	112	125	237
Total	412	359	771

```

PropCIs::diffpropci.Wald.mp(b = 112, c = 234, n = 771,
                             conf.level = 0.90)

```

data:

90 percent confidence interval:

0.1196754 0.1967967

sample estimates:

[1] 0.1582361

30.5 Answer 30 is d

This is a paired samples design. Analysis Z is the only one that is correct. All of the others assume these are independent samples.

30.6 Results 30

Grades on Question 30		4 points
% receiving full credit		46.7
% of available points awarded		46.7

- Partial credit was not available on this question.
- The most common incorrect response was **c**, although **b** was popular, too.

31 Answer Key

Item	Points	Correct Response	% Awarded
01	4	4.1 (-4.6, 12.8) kg/m^2	26.7
02	4	80.5% (67.8%, 91.3%)	79.8
03	3	a = 3, b = 1, c = 4, d = 2	90.6
04	3	b	88.3
05	4	d	65.8
06	3	m or n	95.0
07	3	0.40	63.3
08	4	a line of R code	72.5
09	3	a	78.3
10	3	a = 4, b = 5	over 95
11	3	d or e	90.0
12	3	e	over 95
13	3	e	78.9
14	4	1.96 (1.33, 2.88)	81.3
15	3	c	93.3
16	3	a = 1, b = 2, c = 1, d = 4	84.4
17	4	b, d and f	62.9
18	3	a = 2, b = 1	80.0
19	3	d	53.3
20	3	d	71.7
21	3	c	80.0
22	4	b	91.3
23	3	c	61.7
24	3	-0.0075	36.7
25	3	c	81.1
26	3	a	89.2
27	4	1.15 (1.03, 1.30)	74.4
28	3	a, b and c	57.8
29	3	d	90.0
30	4	d	46.7

32 Grading

Sum up your scores on the 30 items. That's your raw total.

Add 8.5 points. That's your Quiz 2 score.

Quiz 2 score	Rough "Grade"	n
95 and higher	very solid A	17
90 to 94.5	A	6
85 to 89.5	A-	12
80 to 84.5	B+	10
75 to 79.5	B	7
70 to 74.5	B-	0
below 70	weaker	some

For those of you with a Quiz 2 score below 70, I'm concerned about your score on the Quiz. Send me an email

today or tomorrow, please, once you've had the chance to review your work, and we'll discuss next steps.

- These grade groupings now incorporate the change to how I graded Question 11.