Answer Sketch for 431 Quiz 2 with Results

Thomas E. Love

Version: 2021-10-27 14:02:47

Packages Loaded

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

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

theme_set(theme_bw())
```

1 Answer 01 is 1a, 2b, 3a, 4d

Increasing the level of confidence (say from 90% to 95%), 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 as compared to a t-based procedure, our interval might get wider and it might get narrower, so we cannot tell.

1.1 Results 01 (4 points)

- We awarded 1 point per correct response, across the four rows.
- Part 4 was the tough one for students. As is often the case, people were unwilling to select "we cannot tell" but that was the best response.

Question 01	% Correct
part 1	over 90
part 2	over 90

Question 01	% Correct
part 3	over 90
part 4	65.5

Question 01	Max Points = 4
% Correct Responses	56.9
% of Available Points Awarded	87.9

2 Answer 02 is b

`fct_relevel` lets you specify a new order for factor levels "by hand" which is what we need, and `facet_wrap` is the easiest approach to getting the individual stratified (faceted) histograms we need.

2.1 Results 02 (4 points)

- There was no partial credit available on this item.

Question 02	Max Points = 4
% Correct Responses	84.5
% of Available Points Awarded	84.5

3 Answer 03 is 1.32 with CI (1.12, 1.56)

We can calculate the resulting confidence interval using the `twobytwo` function in `Love-boost.R`. Our two-by-two table is as follows:

```
twobytwo(95, 155-95, 160, 345-160,
         "Private", "Public", "Accepted", "Not Accepted",
         conf = 0.95)
```

```
2 by 2 table analysis:
------------------------------------------------------
Outcome   : Accepted
Comparing : Private vs. Public

        Accepted Not Accepted    P(Accepted) 95% conf. interval
Private       95           60         0.6129    0.5340   0.6863
Public       160          185         0.4638    0.4117   0.5166

                                   95% conf. interval
             Relative Risk: 1.3216    1.1162   1.5647
         Sample Odds Ratio: 1.8307    1.2441   2.6940
Conditional MLE Odds Ratio: 1.8285    1.2226   2.7489
    Probability difference: 0.1491    0.0545   0.2388

             Exact P-value: 0.0027
        Asymptotic P-value: 0.0022
```

```
-----------------------------------------------------
```

We did not ask you to use a Bayesian augmentation here, but if you had, it would have looked like this:

```r
twobytwo(95+2, 155-95+2, 160+2, 345-160+2,
         "Private", "Public", "Accepted", "Not Accepted",
         conf = 0.95)
```

```
2 by 2 table analysis:
-----------------------------------------------------
Outcome   : Accepted
Comparing : Private vs. Public

        Accepted Not Accepted    P(Accepted) 95% conf. interval
Private       97           62         0.6101    0.5322   0.6827
Public       162          187         0.4642    0.4124   0.5167

                                   95% conf. interval
             Relative Risk: 1.3143    1.1113   1.5543
         Sample Odds Ratio: 1.8060    1.2327   2.6457
Conditional MLE Odds Ratio: 1.8039    1.2119   2.6980
    Probability difference: 0.1459    0.0523   0.2348

             Exact P-value: 0.0030
        Asymptotic P-value: 0.0024
-----------------------------------------------------
```

3.1 Results 03 (5 points)

- 2 points for the correct point estimate (1.32). I also gave 2 points for 1.31, assuming people giving that response included the Bayesian augmentation.
  - There was no obvious pattern to the incorrect responses.
- 3 points for the correct confidence interval endpoints, whether (1.12, 1.56) without the augmentation or (1.11, 1.55) with it.
  - I gave 2 points here to students who made a rounding error and presented something that was off by 0.01, for example (1.12, 1.57) or (1.11, 1.56).
  - I gave no credit to a response that had the upper endpoint right but the lower endpoint seriously wrong.
- I didn't penalize people who wrote something besides the number (like *point estimate: 1.32*), even though I asked that you not do that in the answer sheet.
- I also didn't penalize people for formatting their result in ways other than I'd provided in the answer sheet.

Question 03	% Correct
point estimate	over 90
confidence interval	82.8

Question 03	Max Points = 5
% Correct Responses	82.8
% of Available Points Awarded	89.7

4 Answer 04 is b, alone.

- Working through the available materials, we obtain the following results for each program…

For Program A, we have an odds ratio of 3.50.

```
twobytwo(35, 25, 40, 100, "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. interval
Private            35                25              0.5833    0.4559   0.7005
Public             40               100              0.2857    0.2170   0.3660

                                   95% conf. interval
             Relative Risk: 2.0417    1.4559   2.8631
         Sample Odds Ratio: 3.5000    1.8626   6.5767
Conditional MLE Odds Ratio: 3.4758    1.7761   6.9113
    Probability difference: 0.2976    0.1484   0.4323

             Exact P-value: 0.0001
        Asymptotic P-value: 0.0001
------------------------------------------------------
```

For Program B, we have an odds ratio of 0.63.

```
twobytwo(10, 10, 80, 50, "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. interval
Private            10                10              0.5000    0.2939   0.7061
Public             80                50              0.6154    0.5291   0.6949

                                    95% conf. interval
             Relative Risk:  0.8125    0.5135   1.2856
         Sample Odds Ratio:  0.6250    0.2429   1.6081
Conditional MLE Odds Ratio:  0.6270    0.2168   1.8107
    Probability difference: -0.1154   -0.3311   0.1029

             Exact P-value: 0.3383
        Asymptotic P-value: 0.3297
------------------------------------------------------
```

For Program C, we have an odds ratio of 1.75.

```
twobytwo(50, 25, 40, 35, "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. interval
Private            50                25              0.6667    0.5531   0.7637
Public             40                35              0.5333    0.4206   0.6427

                                           95% conf. interval
             Relative Risk: 1.2500    0.9586   1.6299
         Sample Odds Ratio: 1.7500    0.9040   3.3876
Conditional MLE Odds Ratio: 1.7434    0.8592   3.5779
    Probability difference: 0.1333   -0.0230   0.2808

             Exact P-value: 0.1333
        Asymptotic P-value: 0.0968
------------------------------------------------------
```

4.1 Results 04 (4 points)

- 4 points for a correct response (b, alone.)
- 2 points for one incorrect (check / no check) decision among a, b and c.
    - For example, a and b got you 2 points, as did d.
- 1 point for two incorrect (check / no check) decisions among a, b and c.
- 0 points if you checked e

The most common incorrect response was `d`.

Question 04	Max Points = 4
% Correct Responses	69.0
% of Available Points Awarded	79.7

5 Answer 05 is 0.011

We know that…

- The overall probability of a woman having breast cancer in this population is 0.025, which we can express as 2500/1,000,000.
- The sensitivity of the mammogram is 0.85, so this means that 85% of the women who are actually "breast cancer positive" will be identified as "positive" by the mammogram.
- The specificity of the mammogram is 0.80, so this means that 80% of the women who are actually "breast cancer negative" will be identified as such by the mammogram.

So if we have a population of 1,000,000 people on which we do this test, we then have 2500 who actually had breast cancer and 997,500 who did not.

- Of the 2500 who actually have breast cancer, 2125 (85% of 2500) will test positive, and the remaining 375 will test negative.
- Of the 997,500 who actually do not have breast cancer, 798,000 (80% of 997,500) will test negative, and the remaining 199,500 will test positive.

That produces the following two-by-two table describing one million women.

–	Breast Cancer present	Breast Cancer absent	Total
Mammogram positive	2125	199500	201625

–	Breast Cancer present	Breast Cancer absent	Total
Mammogram negative	375	798000	798375
Total	2500	997500	1000000

We can now see that if a woman tests positive (so she's one of the 201,625 women in the top row), she has a 2125/201625 = 0.01053937 probability of actually having breast cancer. We then round this to three decimal places, as requested. See Chapter 6 in Spiegelhalter for more on sensitivity and specificity.

- If you're interested, what I was asking you for is sometimes called the *positive predictive value*.

5.1 Results 05 (4 points)

- This was one of the two questions I anticipated would be tricky before anyone took the Quiz. The other was Question 17.
- You received 3 points for improper rounding to more than three decimal places (like to 0.01054.)
- You received 2 points for improper rounding to less than three decimal places (like to 0.01.)
- You also received 2 points for mistaken rounding, like 0.012 or 0.013.
- No lost credit for presenting the probability as a percentage, rather than a proportion, so long as you presented the correct percentage, rounded properly, so that would be 1.054%.
- You received 0 points for a response like 0.01%, or 1.17%, though, since that is simply incorrect.

Question 05	Max Points = 4
% Correct Responses	65.5
% of Available Points Awarded	70.7

6 Answer 06 is e.

- R^2 cannot be negative so a and b and g and h are nonsense.
- The models proposed in g, h, i, j, k and l each get the slope and intercept backwards
- The Y-X slope is clearly negative (as X increases, Y decreases) so c and d are incorrect
- The cloud of points is tight around the line, and R^2 of 0.82 is far more plausible than 0.32 as a result.

6.1 Results 06 (4 points)

- 4 points for the correct response, e.
- I gave 1 point to people who responded with choice a or f, since at least they identified the correct regression line, although I felt like that was the easier part of the question.

The most common incorrect response was `a`, but again, an R^2 cannot possibly be negative.

Question 06	Max Points = 4
% Correct Responses	86.2
% of Available Points Awarded	89.7

7 Answer 07 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            0.678          0.913
```

Remember that the SAIFS procedure automatically includes a Bayesian adjustment, so `saifs.ci(34, 43, conf.level = 0.90)` or `saifs.ci(35, 45, conf.level = 0.90)` would *not* be appropriate.

7.1 Results 07 (5 points)

- 3 points for the correct point estimate
  - We accepted 80.5 or 80.5%.
  - 2 points out of 3 for 0.805, since we'd explicitly discussed the importance of showing a percentage, and not a proportion.
  - 1 point out of 3 for 0.81, or 81%1
- 2 points for the correct confidence interval, which is (67.8, 91.3)
  - 1 point for (65.6, 93.3), which is the *95%* CI
  - 1 point for insufficient decimals, like (68%, 91%) or (70%, 90%).
  - No credit for expressing the result as a proportion and rounding incorrectly, like (0.7, 0.9)

Question 07	% Correct
point estimate	79.3
confidence interval	63.8

Question 07	Max Points = 5
% Correct Responses	62.1
% of Available Points Awarded	74.8

8 Answer 08 is a short essay, pointing us to Result 2, and concluding that the CI for the observed weight loss is strictly positive. (6 points)

The comparison in this design involves paired samples, specifically before-after comparisons of the same individuals. Note that Results 1 and 2 are for paired samples, and that Results 3-5 are for independent samples comparisons.

The large outliers encourage us to use the bootstrap rather than a t test, and that leads us to Result 2, which shows the bootstrap result for a paired samples comparison. The 99% confidence interval for the population mean of the paired differences shown in Result 2 is (0.059, 13.402) and that does not include zero, and is in fact strictly positive, suggesting a loss of weight from before to after, on average. The point estimate is a loss of 6.8 pounds from the before to the after period.

8.1 Results 08 (6 points)

Here are a few answers (lightly edited from student submissions) that would have received full credit.

1. Result 2 is the most appropriate assessment for these paired samples. The bootstrap method is preferred over the t-test because normality is not an assumption and the distribution of the data appears to have some outliers. The mean population weight loss is estimated to be 6.8 lbs (99% CI: 0.059 lbs - 13.402 lbs) which supports the alternative hypothesis that the true mean difference is not equal to zero.
2. Result 2 is the most appropriate for this analysis because this data involves paired samples (before and after measures for each subject) and while the distribution of paired mean differences is approximately symmetric, there are outliers that would be better dealt with by generating a bootstrap confidence interval. The point estimate for the mean change in weight was 6.8 pounds (indicating weight lost after completing the diet) with a two-sided 99% confidence interval of (0.059, 13.402) pounds for the population mean weight change. While this confidence interval includes values fairly close to zero (which would indicate that there was no mean change in weight with this diet), it does not include zero, so we can therefore conclude that this diet does cause a change in weight and, more specifically, a loss.
3. We should use paired sample testing since we are measuring before and after weights in the same subjects, and we should not assume normality since the QQ plot of the paired mean differences in weight suggest that the distribution could have more outliers than what is expected for a normal model. Therefore, result 2 which uses paired samples testing and a bootstrap method for confidence interval calculation is the most appropriate. Since the confidence interval obtained here goes from 0.059 to 13.4 pounds and thus does not cross zero, we can reject the null hypothesis of the paired mean weight loss being zero with 99% confidence.
4. Result 2 using the one-sample bootstrap procedure on the `diff` variable (paired before - after differences) is the most appropriate choice as this study involves paired samples and from the Q-Q plot and the boxplot of these differences, we are unwilling to assume population normality thanks to the substantial outliers. The associated confidence interval implies that there is 99% confidence that the true population mean weight loss through this regimen is between 0.059 and 13.402 pounds. Since zero is not included in this confidence interval, we can reject the null hypothesis and conclude that there is statistically detectable evidence of average weight loss under the strict "after" regime.
5. Since we are comparing the results from the same group before and after an intervention, the appropriate inferential procedure will use paired samples. The distributions shown in Plot 8A show meaningful outliers, leading us to a bootstrap comparison, as shown in Result 2. Since the 99% confidence interval exceeds zero, we reject the null hypothesis and conclude the intervention reduced weight on average to a statistically detectable extent.

8.1.1 Step 1 (Paired Samples)

In grading, the first step was to distinguish between the people who chose a result based on paired samples (result 1 or 2) vs. those who chose a result based on independent samples (results 3-5).

- Everyone who chose Result 3, 4 or 5 lost 3 points for doing so.
- Everyone who didn't make a clear choice lost 3 points.
- Everyone who chose Result 1 lost a point for that.

8.1.2 Step 2 (Quality of Response)

In addition to the Step 1 decisions specified above, students lost points in the following ways, down to a minimum of 0.

Points	Description of Issue
-1 to -3	No reason for (or poor reason for) choice of result
-1 to -3	Misinterpretation of which results corresponded to paired vs. independent samples
-1 to -3	for lack of clarity
-1 to -2	improper specification of the point estimate (should be 6.8 pounds)
-2	failure to compare chosen CI to zero and obtain an appropriate conclusion about the before/after comparison based on your choice

Points	Description of Issue
-2	misinterpretation of the confidence interval's meaning
-1	failure to specify the direction (the subjects lost weight, on average, from before to after)
-1 to -3	using the wrong plots (plot 8A is only appropriate for paired samples, plot 8B is only appropriate for independent samples)
-1	mistakes in transposing results from the PDF to your answer
-1	for each spelling, grammar or syntax error
-2 to -4	no actual presentation of the estimates from your choice or interpretation of them
-1 to -5	stating something ancillary that is incorrect
-1 to -3	lack of clarity

Question 08	Max Points = 6
% Completely Correct Responses	3.4
% of Available Points Awarded	59.8

9 Answer 09 is d and f, although d or f alone is also fine.

These are paired samples, paired by "state" comparing 2010 to 2014.

- The `quiz_beds2` data is in a useful format for a paired comparison, while the `quiz_beds1` data are set up for independent samples comparisons. So we'll rule out options a-c, which use `quiz_beds1` and perform independent samples comparisons.
- Option e performs a 95% confidence interval (since the computer ignores the `sig.level` in a t test and sticks with its default) so that's not useful.
- Option g also performs a 95% confidence interval (default for `smean.cl.boot()`) so we can ignore that.
- Thus, we're down to options `d` and `f`. Looking at the results, I see very little difference between the intervals, so I gave full credit to anyone selecting either response, or both of them, although in practice, I would probably go with the bootstrap (option g) given the outliers in the paired differences.

9.1 Results 09 (6 points)

- Full credit (6 points) for selecting d and f, or d alone, or f alone.
- I gave 4 points to the people who selected d and e and f.
- Otherwise, no credit.

This, frankly, was a disaster, and one I wasn't expecting. Way too many people (more than half) chose an independent samples approach, and data set `quiz_beds1`. I'll have to discuss some related issues again, and ask about this sort of thing again in the future.

Question 09	Max Points = 6
% Correct Responses	36.2
% of Available Points Awarded	38.5

10 Answer 10 is c.

The Box-Cox plot suggests a λ of 0.5, pretty clearly. That's indicating that we should take the square root of the raw outcome, so that's choice `c`.

10.1 Results 10 (4 points)

- I couldn't find a way to award partial credit on this question.
- None of the incorrect choices were chosen by more than a couple of students.

Question 10	Max Points = 6
% Correct Responses	over 90
% of Available Points Awarded	over 90

11 Answer 11 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  feminine 16.52893 19.15335 21.77778 24.66299 27.54821 21.95164 5.511699  3
2 masculine 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 between group feminine and group masculine is not equal
95 percent confidence interval:
 -16.36724   8.18197
sample estimates:
 mean in group feminine mean in group masculine
               21.95164                26.04427
```

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

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

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

```
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, "masculine"))

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.

11.1 Results 11 (5 points)

- Point Estimate
  - 3 points for 4.1 with the correct units (kg/m^2)
  - 2.5 points for 4.09 (incorrect rounding) but correct units.
  - 2 point for 4.1 without units.
  - 1.5 points for 4.09 (incorrect rounding) and no units.
  - 1.5 point for 4.1 with incorrect units (like BMI or $kg/(cm)^2$).
  - 1 point for -4.1 with correct units.
  - 0 points for anything else, including 4.05 kg/m^2
- Confidence Interval
  - 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

This, too, was a disaster, and one I wasn't expecting. I'll have to discuss some related issues again, and ask about this sort of thing again in the future. Obviously, this is also potentially relevant to Analysis 2 in Project A, if you're working with a binary categorical predictor.

Question 11	% Correct
point estimate	46.6
confidence interval	48.3

Question 11	Max Points = 5
% Correct Responses	27.6
% of Available Points Awarded	60.2

12 Answer 12 is a.

Since the raw data appear right skewed in their histogram, and the logged data are left skewed, something in between seems the best choice. On the ladder of power transformations, the square root (transformation using power $\lambda = 0.5$) falls between the raw data ($\lambda = 1$) and the log ($\lambda = 0$).

12.1 Results 12 (4 points)

- No partial credit available on this Question.

Question 12	Max Points = 4
% Correct Responses	82.8
% of Available Points Awarded	82.8

13 Answer 13 is a, c, d and f.

The appropriate procedures for estimating a confidence interval around a single proportion do not include a linear regression or rank-based procedure, but all of the others are OK.

13.1 Results 13 (4 points)

- 4 points for making all six decisions (yes/no) correctly.
- 1 point lost (down to a minimum of 0) for each incorrect choice.

Question 13	Max Points = 4
% Correct Responses	53.4
% of Available Points Awarded	81.5

14 Answer 14 is d

Let's run the four appropriate choices.

Method	90% CI	Width
Agresti-Coull	(0.3870, 0.5522)	0.1652
Clopper-Pearson	(0.3814, 0.5576)	0.1762
"Plus4"	(0.3879, 0.5521)	0.1642
Score	(0.3870, 0.5522)	0.1652

So the "Plus4" procedure provides the narrowest of these four confidence intervals.

Detailed results follow.

```
binom.test(x = 45, n = 96, conf.level = 0.90,
           ci.method = "Agresti-Coull")
```

```
    Exact binomial test (Agresti-Coull CI)
```

```
data:  45 out of 96
number of successes = 45, number of trials = 96, p-value = 0.6101
alternative hypothesis: true probability of success is not equal to 0.5
90 percent confidence interval:
 0.3869794 0.5522337
sample estimates:
probability of success
               0.46875
```

```r
binom.test(x = 45, n = 96, conf.level = 0.90,
           ci.method = "Clopper-Pearson")
```

```
data:  45 out of 96
number of successes = 45, number of trials = 96, p-value = 0.6101
alternative hypothesis: true probability of success is not equal to 0.5
90 percent confidence interval:
 0.3814253 0.5575554
sample estimates:
probability of success
               0.46875
```

```r
binom.test(x = 45, n = 96, conf.level = 0.90,
           ci.method = "Plus4")
```

```
    Exact binomial test (Plus 4 CI)

data:  45 out of 96
number of successes = 45, number of trials = 96, p-value = 0.6101
alternative hypothesis: true probability of success is not equal to 0.5
90 percent confidence interval:
 0.3879055 0.5520945
sample estimates:
probability of success
               0.46875
```

```r
binom.test(x = 45, n = 96, conf.level = 0.90,
           ci.method = "Score")
```

```
    Exact binomial test (Score CI without continuity correction)

data:  45 out of 96
number of successes = 45, number of trials = 96, p-value = 0.6101
alternative hypothesis: true probability of success is not equal to 0.5
90 percent confidence interval:
 0.3869838 0.5522294
sample estimates:
probability of success
               0.46875
```

14.1 Results 14 (4 points)

- There was no partial credit available on this question, as I couldn't figure out how someone would select anything other than the correct response if they worked through each of these calculations, other than making an error in calculation.

- Of course, some people had incorrectly identified which procedures were appropriate in Question 13.
  - For the few people who'd given an answer in Question 13 which specifically excluded the "plus4" approach as a possibility, but included either Agresti-Coull or Score or both, I gave 3 points here if you identified the Agresti-Coull or Score procedure.
  - If your Question 13 choices included "plus4", though, then you only got credit here if your answer was "plus4".
- If you got a different answer than me, without just making an error in your calculations, code or something silly, please let me know via email with the subject line Quiz 2 Question 14, especially if this sketch doesn't straighten out your problem. I'm not understanding how so many folks missed this item, and I'd like to have a better understanding.

Question 14	Max Points = 4
% Correct Responses	48.3
% of Available Points Awarded	50.9

15 Answer 15 is in two parts.

15.1 In 15a, we specify the number as 32 subjects.

15.2 In 15b, we specify our code, as follows.

Here is the code I used:

```
power.t.test(delta = 30, sd = 50,
             sig.level = 0.05, power = 0.90,
             type = "paired")
```

```
     Paired t test power calculation 

              n = 31.17169
          delta = 30
             sd = 50
      sig.level = 0.05
          power = 0.9
    alternative = two.sided

NOTE: n is number of *pairs*, sd is std.dev. of *differences* within pairs
```

We need at least 32 patients, each of which provides two measurements. Since this is an even number, we can include as many male subjects (16) as female subjects (also 16), so that's OK for meeting the other criterion specified in bold in the question.

15.3 Results 15 (6 points, 3 for part a, and 3 for part b)

- The correct number is worth 3 points.
  - I gave full credit if you wrote 32, or 32 patients. I would have given full credit to 64 measurements (32 patients), as well.
  - If you wrote 31, I gave 1.5 points for bad rounding.
  - If you wrote 64, I gave 1 point.
- The correct code is also worth 3 points.

  - Alternatives I accepted included:
    * `power.t.test(power = 0.9, delta = 30, sd = 50, sig.level = 0.05, type="one.sample", alternative="two.sided")`

My hope here was that I could figure out why people made mistakes by looking at the code they presented.

- Some people miscalculated what the `sd` should be. That was a very common problem.
- Some people used a pooled sample test (only appropriate for independent samples) for example: `power.t.test(delta = 30, sd = 30/.6, sig.level = 0.05, power = .9, alternative = 'two.sided')`. That was the second most common problem.
- Some people were comparing proportions, with `power.prop.test`, rather than comparing means, which was definitely not the way to go.
- Some people didn't specify the correct significance level.
- Some people failed to use `power.t.test` properly, perhaps by calculating `d = delta/sd`, rather than individually specifying `delta` and `sd`. If you did that, you'd have to specify `delta` as 0.6, and `sd` as 1. Look at the output your code produces.

Question 15	% Correct
correct number	53.4
R code	60.3

Question 15	Max Points = 6
% Correct Responses	53.4
% of Available Points Awarded	58.3

16 Answer 16 is 1d, 2e, 3b, 4e

1. Model D has the largest R^2 value. I assume that some of the people who chose option B were looking at adjusted R^2, but that's not a fraction of anything.
2. No information about assumption violations is available in the summaries provided. Lots of people chose Model D. I don't know how, or why.
3. Model B has the smallest AIC. Most people got that right.
4. No information about the impact of a transformation of the outcome can be inferred from the summaries provided. Many, many people chose Model A here, but there's no reason to do so.

16.1 Results 16 (4 points)

- We awarded 1 point for each correct response.
- I expected parts 2 and 4 to perhaps be a little rough. Not this rough, though. Overall, this, too, is an area I clearly need to review with you a bit more.

Question 16	% Correct	Common incorrect responses
part 1	56.9	b, then a
part 2	17.2	d, then b
part 3	over 90	d, then a
part 4	46.6	a

Question 16	Max Points = 4
% Correct Responses	6.9
% of Available Points Awarded	53.0

17 Answer 17 is c, and only c.

- d is not correct, because a Wilcoxon rank sum test does not compare means.
- While options a and b could work in that they do compare means (a using independent samples and b paired samples), each also produces a symmetric confidence interval, with the point estimate precisely halfway between the two endpoints. That isn't the case here, so we know we don't have such an interval.
- c is the correct response. A bootstrap does compare means and can certainly create an asymmetric confidence interval. See (essentially) any of our examples.

17.1 Results 17 (4 points)

- This was one of the two questions I anticipated would be tricky before anyone took the Quiz. The other was Question 05.

Response	Points
c	4
a, c	3
a, b, c	3
b, c	2
c, d	2
a, b	1.5
a, b, c, d	0.5
All other responses	0

Question 17	Max Points = 4
% Correct Responses	19.0
% of Available Points Awarded	58.2

18 Answer 18 is e.

None of these approaches will produce what we want.

We can test it out as follows:

```
mydata <- tibble(sbp = c(100, 119, 123, 120, 135))

mydata
```

```
# A tibble: 5 x 1
    sbp
  <dbl>
1   100
2   119
```

```
3    123
4    120
5    135
```

What we want is a new variable `badbp` included into `mydata` that takes the values TRUE for SBPs of 120 or more, and FALSE otherwise. That would indicate FALSE, FALSE, TRUE, TRUE, TRUE should be the set of values.

Trying each option, we obtain no joy.

Option	Result
a. `mydata %>% badbp <- sbp >= 120`	error message
b. `mydata$badbp <- ifelse(mydata$sbp >= 120, "YES", "NO")`	NO and YES rather than FALSE and TRUE
c. `badbp <- mydata %>% filter(sbp >= 120)`	creates new tibble called `badbp` with 3 SBP values
d. `mydata %>% mutate(badbp = sbp < 120)`	gets TRUE and FALSE backwards

So e is the correct response.

18.1 Results 18 (4 points)

- There is no partial credit available on this Question.

Question 18	Max Points = 4
% Correct Responses	65.5
% of Available Points Awarded	65.5

19 Answer 19 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.

19.1 Results 19 (4 points)

- You lost 1 point for each incorrect decision (check or don't check) for options a-e, down to a minimum of 0.
- Selecting option f was worth no points.

The most common incorrect response was both `a` and `e`. Again, there's nothing to suggest this is a random sample. Samples can absolutely be representative without being drawn at random.

Question 19	Max Points = 4
% Correct Responses	51.7
% of Available Points Awarded	84.9

20 Answer 20 is e

The power is estimated to be 96%, using the following function, after noting that a two-sided comparison is the default.

```
pwr.t2n.test(n1 = 100, n2 = 125, d = 2.5/5,
             sig.level = 0.05)
```

```
     t test power calculation

             n1 = 100
             n2 = 125
              d = 0.5
      sig.level = 0.05
          power = 0.9600028
    alternative = two.sided
```

20.1 Results 20 (4 points)

- There is no partial credit available for this question.

Question 20	Max Points = 4
% Correct Responses	86.2
% of Available Points Awarded	86.2

21 Answer 21 odds ratio is 1.98, with 95% CI (1.34, 2.92).

First, let's obtain the 2x2 table from the data.

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

quiz_hosp %>% tabyl(sex, statin)
```

```
    sex NO YES
 FEMALE 97 277
   MALE 46 260
```

Now, we'll adjust that to standard epidemiological format, with "MALE, YES" at the top left, producing the following…

```
twobytwo(260, 46, 277, 97,
         "Male", "Female", "Statin", "No Statin")
```

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

       Statin No Statin    P(Statin) 95% conf. interval
Male      260        46       0.8497    0.8051   0.8855
Female    277        97       0.7406    0.6938   0.7825
```

```
                                   95% conf. interval
             Relative Risk: 1.1472    1.0630   1.2381
         Sample Odds Ratio: 1.9793    1.3407   2.9221
Conditional MLE Odds Ratio: 1.9774    1.3205   2.9921
    Probability difference: 0.1090    0.0481   0.1678

             Exact P-value: 0.0006
        Asymptotic P-value: 0.0006
------------------------------------------------------
```

We simply take the results for the Sample Odds Ratio and round to two decimal places, as requested.

We didn't want you to use a Bayesian augmentation here, but if you did, it would have looked like this:

```
twobytwo(260+2, 46+2, 277+2, 97+2,
         "Male", "Female", "Statin", "No Statin")
```

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

       Statin No Statin    P(Statin) 95% conf. interval
Male      262        48       0.8452    0.8005   0.8813
Female    279        99       0.7381    0.6914   0.7800

                                   95% conf. interval
             Relative Risk: 1.1451    1.0606   1.2363
         Sample Odds Ratio: 1.9368    1.3196   2.8428
Conditional MLE Odds Ratio: 1.9350    1.3001   2.9085
    Probability difference: 0.1071    0.0461   0.1659

             Exact P-value: 0.0007
        Asymptotic P-value: 0.0007
------------------------------------------------------
```

21.1 Results 21 (5 points)

- 2 points for the correct odds ratio point estimate (1.98).
  - We gave 1 point for the Bayesian augmented estimate (1.94)
  - We gave 1 point for 1.97, assuming you rounded badly.
  - We also gave 1 point for the reciprocal of the correct answer, suggesting that you compared female to male instead of what we asked for, and got 0.51. We didn't give credit for 0.5.
  - We gave no credit if you provided the probability difference or the relative risk instead of the odds ratio.
- 3 points for the correct confidence interval (1.34, 2.92).
  - We gave 2 points for the reciprocals of the correct responses, which would be (0.34, 0.75) assuming you had either 0.5 or 0.51 as your point estimate.

Question 21	% Correct
point estimate	74.1
confidence interval	75.9

Question 21	Max Points = 5
% Correct Responses	70.7
% of Available Points Awarded	79.0

22 Answer 22 is a sentence.

Any of these six responses would have received the full 6 points.

1. Our point estimate (Commercial - Medicare) is 0.89 percentage points with 90% CI (-0.53, 2.21) from a bootstrap comparison of the percentage of high school graduates, using our sample of 301 Commercial and 257 Medicare patients.
2. The difference in the mean percentage of adults in the patient's home neighborhood with at least a high school diploma is estimated to be 0.89 percentage points comparing Commercial to Medicare patients (with Commercial higher than Medicare), with a 90% bootstrap confidence interval ranging from -0.53 to 2.21 percentage points.
3. The point estimate for percentage of adults in the patient's home neighborhood who have at least a high school diploma in medicare patients minus that same percentage in commercial patients is -0.89 with a bootstrapped 90% confidence interval of -2.21 to 0.53.
4. We estimate the population mean of percentage of adults in the patient's home neighborhood who have at least a high school diploma to be 0.89 percentage points higher for those who have commercial insurance compared to those who have Medicare, and the endpoints of the 90% bootstrap confidence interval are (-0.53, 2.21).
5. We estimated the population mean (Commercially insured minus Medicare) difference in the percentage of adults in patients' home neighborhoods who have at least a high school diploma, obtaining a point estimate of 0.89 percentage points and a 90% confidence interval using the bootstrap of (-0.53, 2.21) percentage points.
6. The mean percentage of high school graduates in the Commercial insurance group is estimated to be 0.89 percentage points higher (90% CI -0.53, 2.21) as compared to the "Medicare" insurance group, using our sample of 558 subjects and a bootstrap comparison.

In brief, your sentence should tell us that the point estimate for the COMMERCIAL - MEDICARE difference is 0.89, and that the 90% confidence interval for that difference is (-0.53, 2.21).

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

quiz_hosp %>% tabyl(insurance)
```

```
  insurance   n    percent
 COMMERCIAL 301 0.44264706
   MEDICAID  96 0.14117647
   MEDICARE 257 0.37794118
  UNINSURED  26 0.03823529
```

The two largest insurance categories are COMMERCIAL and MEDICARE, so those are the two we want to compare.

We next filter the data to include only those two categories.

```
quiz_22 <- quiz_hosp %>%
  filter(insurance %in% c("COMMERCIAL", "MEDICARE"))
```

Next, we obtain the means and sample sizes in each of the two insurance groups for our outcome (`hsgrads`).

```r
quiz_22 %>% group_by(insurance) %>%
  summarise(n = n(), mean = mean(hsgrads))
```

```
# A tibble: 2 x 3
  insurance      n  mean
  <chr>      <int> <dbl>
1 COMMERCIAL   301  86.9
2 MEDICARE     257  86.0
```

We're instructed to use a bootstrap procedure to compare the means. The COMMERCIAL - MEDICARE difference should have a positive point estimate, it seems.

```r
set.seed(4312021)
quiz_22 %$% bootdif(hsgrads, insurance, conf.level = 0.90)
```

```
Mean Difference            0.05            0.95
     -0.8879067      -2.2051428       0.5339050
```

Our sentence should include the point estimate for the COMMERCIAL - MEDICARE difference (0.89), and the confidence interval for that difference (-0.53, 2.21).

If, for some reason, you set insurance to be a factor, with MEDICARE first, you would get a different result...

```r
quiz_22 <- quiz_22 %>%
    mutate(insur = fct_relevel(factor(insurance),
                               "MEDICARE"))

set.seed(4312021)
quiz_22 %$% bootdif(hsgrads, insur,
                    conf.level = 0.95)
```

```
Mean Difference           0.025           0.975
      0.8879067      -0.7544918       2.4981256
```

- Ideally you would provide an indication that the point estimate suggests that COMMERCIAL `hsgrad` rates are higher on average than MEDICARE ones, but the confidence interval covers values both above and below zero, and that the confidence interval was developed with the bootstrap.

22.1 Results 22 (6 points)

- I planned to award 6 points to anyone who:
  - compared Commercial to Medicare
  - obtained the appropriate point estimate (0.89) which is in percentage points, as distinct from percent.
  - indicated the direction (Commercial > Medicare)
  - used a 90% bootstrap confidence interval
  - obtained the correct endpoints, either (-0.53, 2.21) or (-0.50, 2.26) would be fine.
  - wrote a clear English sentence

No one was completely successful in doing this, although many were close. See above for several examples of lightly edited versions of some student responses to make them 6-point responses.

Points	Response Details
-2	Describing the outcome incorrectly.
-2	Not indicating the direction of effect clearly.
-2	Incorrect point estimate.

Points	Response Details
-2	Actually incorrect confidence interval endpoints.
-1	Using a 95% bootstrap CI, obtaining either (-0.79, 2.46) or (-0.76, 2.50) if you factorized `insurance`.
-1	Incorrect rounding.
-1	Stating the estimate as 0.89% instead of 0.89 percentage points.
-1	No units on the estimates (percentage points) or description of the actual outcome.
-1	Not indicating the confidence level.
-1	Leaving out a key word (like "more" or "bootstrap" or "population")
-1	Grammatical, syntax or spelling error (-1 per error)
-1	Reporting that, for instance, the commercial patients more often graduated from high school, when all we know is that they live in neighborhoods with (slightly) higher graduation rates.

I didn't penalize people for including some sort of statement about statistical significance, though that wasn't a part of what I was looking for. Comparing the confidence interval to zero could have been helpful in this regard, even though my answers above do not include such an explicit comparison.

Question 22	Max Points = 6
% Completely Correct Responses	0
% of Available Points Awarded	61.5

23 Final Score

23.1 Calculating Your Raw and Final Scores

Add up your points from the 22 items. That's your raw score.

Now add 10 points to that score, and that's your final score. So the possible maximum score was 110, but we'll pretend it was 100.

23.2 Distribution of Final Scores

Final Score Range	Interpretation	# of Students
95 or higher	strong A	12
90 or 94.5	A	4
85 to 89.5	possible A	9
80 to 84.5	strong B	9
75 to 79.5	B	4
70 to 74.5	weak B	9
below 70	problematic	11

- The median final score was 82.25
- The mean score was just under 81
- The standard deviation of final scores was enormous. I hope we can tighten that up a bit in Quiz 3.
- If you've scored below 70 on both Quiz 1 and Quiz 2, I'm getting concerned, and this has to improve a bit. My best advice is to work on making Project A as strong as you can, right now.