The infer package

Prof. Wells

STA 209, 4/5/23

Outline

In this lecture, we will...

Outline

In this lecture, we will...

- Discuss misinterpretations of P-values
- Utilize the infer package to create bootstrap confidence intervals and simulate the null distribution for hypothesis testing

Section 1

(Mis)Intepreting P-Values

• In the early days of statistical theory, *P*-values were introduced as an *informal* measure to indicate whether a phenomenon warrants further investigation.

- In the early days of statistical theory, *P*-values were introduced as an *informal* measure to indicate whether a phenomenon warrants further investigation.
- Later, and primarily to create simple statistical manuals for untrained practitioners, this informal measure became the unassailable rule:

- In the early days of statistical theory, *P*-values were introduced as an *informal* measure to indicate whether a phenomenon warrants further investigation.
- Later, and primarily to create simple statistical manuals for untrained practitioners, this informal measure became the unassailable rule:

"If p-value < 0.05, reject H_0 ; If p-value > 0.05, do not reject H_0 "

• As a result, many academic journals used this threshold to determine whether or not a claim is true, and therefore, publication-worthy

- In the early days of statistical theory, *P*-values were introduced as an *informal* measure to indicate whether a phenomenon warrants further investigation.
- Later, and primarily to create simple statistical manuals for untrained practitioners, this informal measure became the unassailable rule:

- As a result, many academic journals used this threshold to determine whether or not a claim is true, and therefore, publication-worthy
 - Non-technical reports (i.e. news media, pop-literature, word-of-mouth) further propagate this rule

- In the early days of statistical theory, *P*-values were introduced as an *informal* measure to indicate whether a phenomenon warrants further investigation.
- Later, and primarily to create simple statistical manuals for untrained practitioners, this informal measure became the unassailable rule:

- As a result, many academic journals used this threshold to determine whether or not a claim is true, and therefore, publication-worthy
 - Non-technical reports (i.e. news media, pop-literature, word-of-mouth) further propagate this rule
- This editorial bias also leads to the practice of "data dredging" or "p-hacking":
 - Researchers prioritize the search for phenomenon with small p-values, at the expense of otherwise noteworthy or important outcomes, and often eschewing other statistical and scientific reasoning.

- In the early days of statistical theory, *P*-values were introduced as an *informal* measure to indicate whether a phenomenon warrants further investigation.
- Later, and primarily to create simple statistical manuals for untrained practitioners, this informal measure became the unassailable rule:

- As a result, many academic journals used this threshold to determine whether or not a claim is true, and therefore, publication-worthy
 - Non-technical reports (i.e. news media, pop-literature, word-of-mouth) further propagate this rule
- This editorial bias also leads to the practice of "data dredging" or "p-hacking":
 - Researchers prioritize the search for phenomenon with small p-values, at the expense of
 otherwise noteworthy or important outcomes, and often eschewing other statistical and
 scientific reasoning.
- This may be one cause of the *Reproducibility Crisis* currently faced in the fields of Psychology and Medicine (and to some extent, other natural and social sciences)

In 2016, the American Statistical Association put forth 6 guidelines to address misconceptions about p-values:

1 P-Values indicate how incompatible the data are with a specific statistical model.

- **1** P-Values indicate how incompatible the data are with a specific statistical model.
- P-values do not measure the probability that the null hypothesis is true; or that the data were produced by random chance alone.

- **1** P-Values indicate how incompatible the data are with a specific statistical model.
- P-values do not measure the probability that the null hypothesis is true; or that the data were produced by random chance alone.
- Scientific, business, or policy decisions should not be based only on whether a p-value is less than a specific threshold.

- **1** P-Values indicate how incompatible the data are with a specific statistical model.
- P-values do not measure the probability that the null hypothesis is true; or that the data were produced by random chance alone.
- Scientific, business, or policy decisions should not be based only on whether a p-value is less than a specific threshold.
- **4** Proper statistical inference requires full reporting and transparency

- **1** P-Values indicate how incompatible the data are with a specific statistical model.
- P-values do not measure the probability that the null hypothesis is true; or that the data were produced by random chance alone.
- Scientific, business, or policy decisions should not be based only on whether a p-value is less than a specific threshold.
- **4** Proper statistical inference requires full reporting and transparency
- **6** A p-value does not measure the size of an effect, or the importance of a result.

- **1** P-Values indicate how incompatible the data are with a specific statistical model.
- P-values do not measure the probability that the null hypothesis is true; or that the data were produced by random chance alone.
- Scientific, business, or policy decisions should not be based only on whether a p-value is less than a specific threshold.
- **4** Proper statistical inference requires full reporting and transparency
- **6** A p-value does not measure the size of an effect, or the importance of a result.
- **(b)** By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.

• By definition, a P-value is the probability of observing data as extreme as the data collected **if the null hypothesis were true**

- By definition, a P-value is the probability of observing data as extreme as the data collected **if the null hypothesis were true**
 - P-values are one measure indicating how incompatible the data is with a specified hypothesis (Small p-value suggests greater incompatibility)

- By definition, a P-value is the probability of observing data as extreme as the data collected **if the null hypothesis were true**
 - P-values are one measure indicating how incompatible the data is with a specified hypothesis (Small *p*-value suggests greater incompatibility)
- But p-values are **NOT** the probability that the null hypothesis is true.

- By definition, a P-value is the probability of observing data as extreme as the data collected **if the null hypothesis were true**
 - P-values are one measure indicating how incompatible the data is with a specified hypothesis (Small *p*-value suggests greater incompatibility)
- But p-values are **NOT** the probability that the null hypothesis is true.
- Consider the following *hypothetical* example:
 - Pro basketball player Stephen Curry and I each take five 3-point shots. Stephen Curry makes all 5, while I make 2.

- By definition, a P-value is the probability of observing data as extreme as the data collected **if the null hypothesis were true**
 - P-values are one measure indicating how incompatible the data is with a specified hypothesis (Small *p*-value suggests greater incompatibility)
- But p-values are **NOT** the probability that the null hypothesis is true.
- Consider the following *hypothetical* example:
 - Pro basketball player Stephen Curry and I each take five 3-point shots. Stephen Curry makes all 5, while I make 2.
 - The null hypothesis that we have the same shot probability, while the alternative is that Stephen Curry has higher probability.

- By definition, a P-value is the probability of observing data as extreme as the data collected **if the null hypothesis were true**
 - P-values are one measure indicating how incompatible the data is with a specified hypothesis (Small *p*-value suggests greater incompatibility)
- But p-values are **NOT** the probability that the null hypothesis is true.
- Consider the following *hypothetical* example:
 - Pro basketball player Stephen Curry and I each take five 3-point shots. Stephen Curry makes all 5, while I make 2.
 - The null hypothesis that we have the same shot probability, while the alternative is that Stephen Curry has higher probability.
 - The p-value for this experiment (i.e. probability of a result as extreme or more) is 0.17.

- By definition, a P-value is the probability of observing data as extreme as the data collected **if the null hypothesis were true**
 - P-values are one measure indicating how incompatible the data is with a specified hypothesis (Small *p*-value suggests greater incompatibility)
- But p-values are **NOT** the probability that the null hypothesis is true.
- Consider the following *hypothetical* example:
 - Pro basketball player Stephen Curry and I each take five 3-point shots. Stephen Curry makes all 5, while I make 2.
 - The null hypothesis that we have the same shot probability, while the alternative is that Stephen Curry has higher probability.
 - The p-value for this experiment (i.e. probability of a result as extreme or more) is 0.17.
 - Is it reasonable to conclude that there is a 17% chance that Stephen Curry and I are equally good shooters?

- By definition, a P-value is the probability of observing data as extreme as the data collected **if the null hypothesis were true**
 - P-values are one measure indicating how incompatible the data is with a specified hypothesis (Small *p*-value suggests greater incompatibility)
- But p-values are **NOT** the probability that the null hypothesis is true.
- Consider the following *hypothetical* example:
 - Pro basketball player Stephen Curry and I each take five 3-point shots. Stephen Curry makes all 5, while I make 2.
 - The null hypothesis that we have the same shot probability, while the alternative is that Stephen Curry has higher probability.
 - The p-value for this experiment (i.e. probability of a result as extreme or more) is 0.17.
 - Is it reasonable to conclude that there is a 17% chance that Stephen Curry and I are equally good shooters?
 - No. We would also need to take into account our prior beliefs about the likelihood of this hypothesis.

• A small *p*-value indicates a result that is unlikely to occur due to chance, if the null hypothesis were true.

- A small *p*-value indicates a result that is unlikely to occur due to chance, if the null hypothesis were true.
 - But the size of the p-value gives NO indication about the actual size of the effect measured;
 - Moreover, the p-value gives no indication about whether the observed difference is of practical importance.

- A small *p*-value indicates a result that is unlikely to occur due to chance, if the null hypothesis were true.
 - But the size of the p-value gives NO indication about the actual size of the effect measured;
 - Moreover, the p-value gives no indication about whether the observed difference is of practical importance.
 - A large sample is able to detect extremely minuscule differences between populations, producing very small p-values.

- A small *p*-value indicates a result that is unlikely to occur due to chance, if the null hypothesis were true.
 - But the size of the p-value gives NO indication about the actual size of the effect measured;
 - Moreover, the p-value gives no indication about whether the observed difference is of practical importance.
 - A large sample is able to detect extremely minuscule differences between populations, producing very small p-values.
- The effect size is the difference between the true value of the parameter and the null value.

- A small *p*-value indicates a result that is unlikely to occur due to chance, if the null hypothesis were true.
 - But the size of the p-value gives NO indication about the actual size of the effect measured;
 - Moreover, the p-value gives no indication about whether the observed difference is of practical importance.
 - A large sample is able to detect extremely minuscule differences between populations, producing very small p-values.
- The effect size is the difference between the true value of the parameter and the null value.
 - Effect size determines whether a result is *practically significant* (i.e. is noteworthy or worth changing behavior over).

- A recent *Nature* study of 19,000 people found that those who met their spouses online...
 - Are less likely to divorce (*p*-value < 0.002)
 - Are more likely to have high marital satisfaction (p-value < 0.001)

- A recent *Nature* study of 19,000 people found that those who met their spouses online...
 - Are less likely to divorce (p-value < 0.002)
 - Are more likely to have high marital satisfaction (p-value < 0.001)
- BUT! The estimated *effect sizes* are tiny (and perhaps not practically significant)

- A recent *Nature* study of 19,000 people found that those who met their spouses online...
 - Are less likely to divorce (*p*-value < 0.002)
 - Are more likely to have high marital satisfaction (p-value < 0.001)
- BUT! The estimated *effect sizes* are tiny (and perhaps not practically significant)
 - Divorce rates are 5.96% for those who met online, versus 7.07% for those who met in-person (Effect Size = 1.11)

- A recent *Nature* study of 19,000 people found that those who met their spouses online...
 - Are less likely to divorce (*p*-value < 0.002)
 - Are more likely to have high marital satisfaction (p-value < 0.001)
- BUT! The estimated *effect sizes* are tiny (and perhaps not practically significant)
 - Divorce rates are 5.96% for those who met online, versus 7.07% for those who met in-person (Effect Size = 1.11)
 - On a 7 point scale, happiness values were 5.64 for those who met online, versus 5.48 for those who met in-person (Effect Size = 0.16)

- A recent *Nature* study of 19,000 people found that those who met their spouses online...
 - Are less likely to divorce (*p*-value < 0.002)
 - Are more likely to have high marital satisfaction (*p*-value < 0.001)
- BUT! The estimated *effect sizes* are tiny (and perhaps not practically significant)
 - Divorce rates are 5.96% for those who met online, versus 7.07% for those who met in-person (Effect Size = 1.11)
 - On a 7 point scale, happiness values were 5.64 for those who met online, versus 5.48 for those who met in-person (Effect Size = 0.16)
- Does this provide compelling evidence that those seeking spouses should do so online?

- A recent *Nature* study of 19,000 people found that those who met their spouses online...
 - Are less likely to divorce (*p*-value < 0.002)
 - Are more likely to have high marital satisfaction (*p*-value < 0.001)
- BUT! The estimated *effect sizes* are tiny (and perhaps not practically significant)
 - Divorce rates are 5.96% for those who met online, versus 7.07% for those who met in-person (Effect Size = 1.11)
 - On a 7 point scale, happiness values were 5.64 for those who met online, versus 5.48 for those who met in-person (Effect Size = 0.16)
- Does this provide compelling evidence that those seeking spouses should do so online?
 - Are the estimated effect sizes meaningful?
Effect Size and Practical Significance

- A recent *Nature* study of 19,000 people found that those who met their spouses online...
 - Are less likely to divorce (p-value < 0.002)
 - Are more likely to have high marital satisfaction (*p*-value < 0.001)
- BUT! The estimated *effect sizes* are tiny (and perhaps not practically significant)
 - Divorce rates are 5.96% for those who met online, versus 7.07% for those who met in-person (Effect Size = 1.11)
 - On a 7 point scale, happiness values were 5.64 for those who met online, versus 5.48 for those who met in-person (Effect Size = 0.16)
- Does this provide compelling evidence that those seeking spouses should do so online?
 - Are the estimated effect sizes meaningful?
 - Can we deduce causal relationships from this investigation? (This is unrelated to significance and effect size)

• Despite their issues, p-values are still quite popular and widely used (although are perhaps over-used)

- Despite their issues, p-values are still quite popular and widely used (although are perhaps over-used)
- The following quote, attributed to George Cobb of Mount Holyoke College, summarizes this as:

- Despite their issues, p-values are still quite popular and widely used (although are perhaps over-used)
- The following quote, attributed to George Cobb of Mount Holyoke College, summarizes this as:
 - *Q*: Why do so many calleges and grad schools teach p = 0.05?
 - A: Because that's what the scientific community and journal editors use.
 - *Q*: Why do the scientific community and journal editors still use p = 0.05?
 - A: Because that's what they were taught in college or grad school.

- Despite their issues, p-values are still quite popular and widely used (although are perhaps over-used)
- The following quote, attributed to George Cobb of Mount Holyoke College, summarizes this as:

Q: Why do so many caolleges and grad schools teach p = 0.05? A: Because that's what the scientific community and journal editors use. Q: Why do the scientific community and journal editors still use p = 0.05? A: Because that's what they were taught in college or grad school.

• Understanding p-values and interpreting p-values in context is an important goal for STA 209

- Despite their issues, p-values are still quite popular and widely used (although are perhaps over-used)
- The following quote, attributed to George Cobb of Mount Holyoke College, summarizes this as:
 - Q: Why do so many calleges and grad schools teach p = 0.05? A: Because that's what the scientific community and journal editors use. Q: Why do the scientific community and journal editors still use p = 0.05? A: Because that's what they were taught in college or grad school.
- Understanding p-values and interpreting p-values in context is an important goal for STA 209
- Determining an appropriate significance level that balances the rate of Type I and Type II error, for your specific research question, is also an important goal for STA 209.

- Despite their issues, p-values are still quite popular and widely used (although are perhaps over-used)
- The following quote, attributed to George Cobb of Mount Holyoke College, summarizes this as:
 - Q: Why do so many calleges and grad schools teach p = 0.05? A: Because that's what the scientific community and journal editors use. Q: Why do the scientific community and journal editors still use p = 0.05? A: Because that's what they were taught in college or grad school.
- Understanding p-values and interpreting p-values in context is an important goal for STA 209
- Determining an appropriate significance level that balances the rate of Type I and Type II error, for your specific research question, is also an important goal for STA 209.
- Determining whether a given number is less than 0.05 is not an important goal for STA 209

Section 2

The infer package

The infer Package

• The infer package makes efficient use of the %>% operator perform statistical inference.

The infer Package

- The infer package makes efficient use of the %>% operator perform statistical inference.
- The infer package makes use of several verbs-like functions:
 - specify, generate, calculate, visualize, get_ci

The infer Package

- The infer package makes efficient use of the %>% operator perform statistical inference.
- The infer package makes use of several verbs-like functions:
 - specify, generate, calculate, visualize, get_ci



COVID Incubation Time

The Infectious Disease Dynamics Group at Johns Hopkins University collected data between Dec 2019 and Jan 2020 on exposure and symptom onset for COVID-19.

COVID Incubation Time

The Infectious Disease Dynamics Group at Johns Hopkins University collected data between Dec 2019 and Jan 2020 on exposure and symptom onset for COVID-19.

The distribution of Incubation times for 64 patients is shown below:



COVID Incubation Time

The Infectious Disease Dynamics Group at Johns Hopkins University collected data between Dec 2019 and Jan 2020 on exposure and symptom onset for COVID-19.

The distribution of Incubation times for 64 patients is shown below:



- What is the population of interest? What is the parameter?
- What is the sample? What is the statistic?

• Every statistical investigation begins with a sample data frame (i.e. covid)

- Every statistical investigation begins with a sample data frame (i.e. covid)
- The sample may contain many variables of interest

- Every statistical investigation begins with a sample data frame (i.e. covid)
- The sample may contain many variables of interest
- We must first specify which variable(s) will be the focus of our investigation by designating a response variable

- Every statistical investigation begins with a sample data frame (i.e. covid)
- The sample may contain many variables of interest
- We must first specify which variable(s) will be the focus of our investigation by designating a response variable
- To investigate the infection rate

covid %>%
 specify(response = Incubation)

• In order to create a bootstrap distribution, we need to resample many times from the OG sample

- In order to create a bootstrap distribution, we need to resample many times from the OG sample
- After selecting variables, pipe results into the generate function to create replicates

```
covid %>%
  specify(response = Incubation) %>%
  generate( reps = 2000, type = "bootstrap")
```

- In order to create a bootstrap distribution, we need to resample many times from the OG sample
- After selecting variables, pipe results into the generate function to create replicates

```
covid %>%
  specify(response = Incubation) %>%
  generate( reps = 2000, type = "bootstrap")
```

• We need to indicate how many replicates we want, and what type of method we'll use to create them.

- In order to create a bootstrap distribution, we need to resample many times from the OG sample
- After selecting variables, pipe results into the generate function to create replicates

```
covid %>%
  specify(response = Incubation) %>%
  generate( reps = 2000, type = "bootstrap")
```

- We need to indicate how many replicates we want, and what type of method we'll use to create them.
- For bootstrap confidence intervals, choose type = "bootstrap", and almost always use at least reps = 2000

- In order to create a bootstrap distribution, we need to resample many times from the OG sample
- After selecting variables, pipe results into the generate function to create replicates

```
covid %>%
  specify(response = Incubation) %>%
  generate( reps = 2000, type = "bootstrap")
```

- We need to indicate how many replicates we want, and what type of method we'll use to create them.
- For bootstrap confidence intervals, choose type = "bootstrap", and almost always use at least reps = 2000
- The resulting data frame has a number of rows equal reps \times sample_size

• Once we have our bootstrap samples, we need to compute the corresponding statistics

- Once we have our bootstrap samples, we need to compute the corresponding statistics
- Use the calculate function, whose first argument is stat

- Once we have our bootstrap samples, we need to compute the corresponding statistics
- Use the calculate function, whose first argument is stat
- Many statistics are available: "mean", "sum", "sd", "median", "prop", "diff in mean, "correlation", "slope", and more!

- Once we have our bootstrap samples, we need to compute the corresponding statistics
- Use the calculate function, whose first argument is stat
- Many statistics are available: "mean", "sum", "sd", "median", "prop", "diff in mean, "correlation", "slope", and more!

```
covid %>%
specify(response = Incubation) %>%
generate( reps = 2000, type = "bootstrap") %>%
calculate(stat = "mean")
```

- Once we have our bootstrap samples, we need to compute the corresponding statistics
- Use the calculate function, whose first argument is stat
- Many statistics are available: "mean", "sum", "sd", "median", "prop", "diff in mean, "correlation", "slope", and more!

```
covid %>%
  specify(response = Incubation) %>%
  generate( reps = 2000, type = "bootstrap") %>%
  calculate(stat = "mean")
```

• After applying calculate the resulting data frame consists of one bootstrap statistic for each replicate (saved to the variable stat)

Sample Statistic

• Suppose you want to just calculate summary statistics of the OG sample

Sample Statistic

- Suppose you want to just calculate summary statistics of the OG sample
- By using specify and calculate (and omitting generate) we can do just that, paralleling similar calculation for the bootstrap statistics

```
covid_stat<- covid %>%
   specify(response = Incubation) %>%
   calculate(stat = "mean")
covid_stat
```

```
## Response: Incubation (numeric)
## # A tibble: 1 x 1
## stat
## <dbl>
## 1 3.03
```

Sample Statistic

- Suppose you want to just calculate summary statistics of the OG sample
- By using specify and calculate (and omitting generate) we can do just that, paralleling similar calculation for the bootstrap statistics

```
covid_stat<- covid %>%
   specify(response = Incubation) %>%
   calculate(stat = "mean")
covid_stat
```

Response: Incubation (numeric)
A tibble: 1 x 1
stat
<dbl>
1 3.03

Note: we saved the value of this calculation as covid_stat so we could use it later

Save the bootstrap too

• Since we also will want to make frequent use of the bootstrap statistics, it's worth saving them as a variable too:

Save the bootstrap too

• Since we also will want to make frequent use of the bootstrap statistics, it's worth saving them as a variable too:

```
covid_boot<- covid %>%
  specify(response = Incubation) %>%
  generate( reps = 2000, type = "bootstrap") %>%
  calculate(stat = "mean")
```

covid_boot

```
## Response: Incubation (numeric)
## # A tibble: 2.000 x 2
##
      replicate stat
##
          <int> <dbl>
##
              1 2.95
    1
    2
              2 2.89
##
##
    3
              3 2.94
##
    4
              4 3.29
##
    5
              5 3.06
              6 3.18
##
    6
##
    7
              7 3.36
              8 3.33
##
    8
    9
              9 2.43
##
##
  10
             10
                 2.93
```

visualize Bootstrap Distribution

• In order to perform any statistical inference, we need to ensure appropriate shape conditions on bootstrap distribution are met

visualize Bootstrap Distribution

- In order to perform any statistical inference, we need to ensure appropriate shape conditions on bootstrap distribution are met
- Use the visaulize verb to quickly generate a reasonably nice-looking histogram of the bootstrap distribution.

visualize Bootstrap Distribution

- In order to perform any statistical inference, we need to ensure appropriate shape conditions on bootstrap distribution are met
- Use the visaulize verb to quickly generate a reasonably nice-looking histogram of the bootstrap distribution.

covid_boot %>% visualize()
visualize Bootstrap Distribution

- In order to perform any statistical inference, we need to ensure appropriate shape conditions on bootstrap distribution are met
- Use the visaulize verb to quickly generate a reasonably nice-looking histogram of the bootstrap distribution.

covid_boot %>% visualize()



Simulation-Based Bootstrap Distribution

• To compute a confidence interval, pipe the calculated data frame into get_confidence_interval (you can use get_ci for brevity)

- To compute a confidence interval, pipe the calculated data frame into get_confidence_interval (you can use get_ci for brevity)
- We need to specify the type of interval we want (either "percentile" or "se"), along with the confidence level

- To compute a confidence interval, pipe the calculated data frame into get_confidence_interval (you can use get_ci for brevity)
- We need to specify the type of interval we want (either "percentile" or "se"), along with the confidence level
- It's useful to save the resulting data frame for later use

- To compute a confidence interval, pipe the calculated data frame into get_confidence_interval (you can use get_ci for brevity)
- We need to specify the type of interval we want (either "percentile" or "se"), along with the confidence level
- It's useful to save the resulting data frame for later use

```
percentile_ci<-covid_boot %>%
  get_ci(level = .95, type = "percentile")
percentile_ci
```

```
## # A tibble: 1 x 2
## lower_ci upper_ci
## <dbl> <dbl>
## 1 2.49 3.63
```

- To compute a confidence interval, pipe the calculated data frame into get_confidence_interval (you can use get_ci for brevity)
- We need to specify the type of interval we want (either "percentile" or "se"), along with the confidence level
- It's useful to save the resulting data frame for later use

```
percentile_ci<-covid_boot %>%
  get_ci(level = .95, type = "percentile")
percentile_ci
```

```
## # A tibble: 1 x 2
## lower_ci upper_ci
## <dbl> <dbl>
## 1 2.49 3.63
```

• When using the percentile type, the first value printed is the lower and the second is the upper bound.

Shade Confidence Intervals

• Once you've used get_ci to obtain endpoints of the confidence interval, you can shade the sampling distribution with the confidence interval region.

```
covid_boot %>% visualize()+shade_ci(endpoints = percentile_ci)
```



Simulation-Based Bootstrap Distribution

Prof. Wells

The infer package

• We can also use nearly identical code in the infer package to perform hypothesis testing.

- We can also use nearly identical code in the infer package to perform hypothesis testing.
- Using the same covid data as before, let's assess the claim that the average incubation time μ is 3 days.

- We can also use nearly identical code in the infer package to perform hypothesis testing.
- Using the same covid data as before, let's assess the claim that the average incubation time μ is 3 days.
- Hypotheses:

$$H_0: \mu = 3$$
 $H_a: \mu \neq 3$

- We can also use nearly identical code in the infer package to perform hypothesis testing.
- Using the same covid data as before, let's assess the claim that the average incubation time μ is 3 days.
- Hypotheses:

$$H_0: \mu = 3$$
 $H_a: \mu \neq 3$

• As before, we use the specify verb to indicate the variable(s) of interest. But now, we also encode our hypotheses:

```
covid %>% specify(response = Incubation) %>%
    hypothesize(null = "point", mu = 3)
```

- We can also use nearly identical code in the infer package to perform hypothesis testing.
- Using the same covid data as before, let's assess the claim that the average incubation time μ is 3 days.
- Hypotheses:

$$H_0: \mu = 3$$
 $H_a: \mu \neq 3$

• As before, we use the specify verb to indicate the variable(s) of interest. But now, we also encode our hypotheses:

```
covid %>% specify(response = Incubation) %>%
hypothesize(null = "point", mu = 3)
```

• In this case, our null hypothesis is the claim that the parameter is single point, and the value of this point is mu = 3

- We can also use nearly identical code in the infer package to perform hypothesis testing.
- Using the same covid data as before, let's assess the claim that the average incubation time μ is 3 days.
- Hypotheses:

$$H_0: \mu = 3$$
 $H_a: \mu \neq 3$

• As before, we use the specify verb to indicate the variable(s) of interest. But now, we also encode our hypotheses:

```
covid %>% specify(response = Incubation) %>%
    hypothesize(null = "point", mu = 3)
```

- In this case, our null hypothesis is the claim that the parameter is single point, and the value of this point is mu = 3
 - If we are instead comparing the means of two different population, we would instead use null = "independence" to indicate that the mean is independent of the population studied.

Simulating the Null Distribution

• To simulate the distribution of statistics under the null hypothesis, we first generate many replicates of the sample.

Simulating the Null Distribution

- To simulate the distribution of statistics under the null hypothesis, we first generate many replicates of the sample.
- To do so, we need to think about the **method** used to new samples.
 - Since we have a single response variable which is quantitative, the only way to do so is by *bootstrapping*.

```
covid %>% specify(response = Incubation) %>%
hypothesize(null = "point", mu = 3) %>%
generate(reps = 2000, type = "bootstrap")
```

Simulating the Null Distribution

- To simulate the distribution of statistics under the null hypothesis, we first generate many replicates of the sample.
- To do so, we need to think about the **method** used to new samples.
 - Since we have a single response variable which is quantitative, the only way to do so is by *bootstrapping*.

```
covid %>% specify(response = Incubation) %>%
hypothesize(null = "point", mu = 3) %>%
generate(reps = 2000, type = "bootstrap")
```

• Once we've generated simulated samples, we compute the null statistic for each:

```
covid %>% specify(response = Incubation) %>%
hypothesize(null = "point", mu = 3) %>%
generate(reps = 2000, type = "bootstrap") %>%
calculate(stat = "mean")
```

• As with confidence intervals, we plot the histogram of the resulting distribution using visualize

• As with confidence intervals, we plot the histogram of the resulting distribution using visualize

```
covid %>% specify(response = Incubation) %>%
hypothesize(null = "point", mu = 3) %>%
generate(reps = 2000, type = "bootstrap") %>%
calculate(stat = "mean") %>%
visualize()
```



Simulation-Based Null Distribution

Prof. Wells

• We can also visualize the p-value using shade_p_value, in which case we also indicate our two-sided alt. hypothesis

 We can also visualize the p-value using shade_p_value, in which case we also indicate our two-sided alt. hypothesis

```
covid %>% specify(response = Incubation) %>%
hypothesize(null = "point", mu = 3) %>%
generate(reps = 2000, type = "bootstrap") %>%
calculate(stat = "mean") %>%
visualize()+ shade_p_value(obs_stat = covid_stat, direction = "two-sided")
```



Simulation-Based Null Distribution

Prof. Wells

24 / 25

 To compute the *p*-value, we return to the data set of null statistics, and then use the get_p_value function; Again, we indicate the direction of our alternative hypothesis.

 To compute the p-value, we return to the data set of null statistics, and then use the get_p_value function; Again, we indicate the direction of our alternative hypothesis.

```
covid %>% specify(response = Incubation) %>%
hypothesize(null = "point", mu = 3) %>%
generate(reps = 2000, type = "bootstrap") %>%
calculate(stat = "mean") %>%
get_p_value(obs_stat = covid_stat, direction = "two-sided")
```

```
## # A tibble: 1 x 1
## p_value
## <dbl>
## 1 0.916
```

 To compute the *p*-value, we return to the data set of null statistics, and then use the get_p_value function; Again, we indicate the direction of our alternative hypothesis.

```
covid %>% specify(response = Incubation) %>%
hypothesize(null = "point", mu = 3) %>%
generate(reps = 2000, type = "bootstrap") %>%
calculate(stat = "mean") %>%
get_p_value(obs_stat = covid_stat, direction = "two-sided")
```

A tibble: 1 x 1
p_value
<dbl>
1 0.916

• What conclusion do we draw based on this p-value?

 To compute the *p*-value, we return to the data set of null statistics, and then use the get_p_value function; Again, we indicate the direction of our alternative hypothesis.

```
covid %>% specify(response = Incubation) %>%
hypothesize(null = "point", mu = 3) %>%
generate(reps = 2000, type = "bootstrap") %>%
calculate(stat = "mean") %>%
get_p_value(obs_stat = covid_stat, direction = "two-sided")
```

```
## # A tibble: 1 x 1
## p_value
## <dbl>
## 1 0.916
```

- What conclusion do we draw based on this p-value?
 - If the null hypothesis were true, we would observe a statistic as extreme as this one about 92% of the time.

 To compute the *p*-value, we return to the data set of null statistics, and then use the get_p_value function; Again, we indicate the direction of our alternative hypothesis.

```
covid %>% specify(response = Incubation) %>%
hypothesize(null = "point", mu = 3) %>%
generate(reps = 2000, type = "bootstrap") %>%
calculate(stat = "mean") %>%
get_p_value(obs_stat = covid_stat, direction = "two-sided")
```

A tibble: 1 x 1
p_value
<dbl>
1 0.916

- What conclusion do we draw based on this p-value?
 - If the null hypothesis were true, we would observe a statistic as extreme as this one about 92% of the time.
 - At most significance levels ($\alpha = 0.1, 0.05, 0.01$), we wouldn't reject the null hypothesis.

 To compute the *p*-value, we return to the data set of null statistics, and then use the get_p_value function; Again, we indicate the direction of our alternative hypothesis.

```
covid %>% specify(response = Incubation) %>%
hypothesize(null = "point", mu = 3) %>%
generate(reps = 2000, type = "bootstrap") %>%
calculate(stat = "mean") %>%
get_p_value(obs_stat = covid_stat, direction = "two-sided")
```

```
## # A tibble: 1 x 1
## p_value
## <dbl>
## 1 0.916
```

- What conclusion do we draw based on this p-value?
 - If the null hypothesis were true, we would observe a statistic as extreme as this one about 92% of the time.
 - At most significance levels ($\alpha = 0.1, 0.05, 0.01$), we wouldn't reject the null hypothesis.
- **BUT** this does not mean we have evidence that the null hypothesis is true (only that we failed to find evidence that it was false)