Inference for Means

Nate Wells

Math 141, 4/15/22

Outline

In this lecture, we will. . .

Outline

In this lecture, we will. . .

- Investigate the theoretical distribution for difference in two means.
- Create confidence intervals and perform hypothesis tests using t distribution for differences in means.
- Compare inference procedures for two independent samples vs. paired samples

Section 1

t-distribution vs Normal

Theory vs. Simulation for 1 Mean

- In order to compute accurate confidence intervals and p-values, we need to ensure that appropriate conditions are met.
  - The strictness of these conditions depends on sample size.

Theory vs. Simulation for 1 Mean

- In order to compute accurate confidence intervals and p-values, we need to ensure that appropriate conditions are met.
  - The strictness of these conditions depends on sample size.
- Theory-based can be used if...
  - $n \geq 2$, and the population is Normal (or nearly so)
  - $n \geq 30$, and the population appears at most moderately skewed
  - $n \geq 60$, and the population is not extremely skewed

Theory vs. Simulation for 1 Mean

- In order to compute accurate confidence intervals and p-values, we need to ensure that appropriate conditions are met.
  - The strictness of these conditions depends on sample size.
- Theory-based can be used if. . .
  - $n \geq 2$, and the population is Normal (or nearly so)
  - $n \geq 30$, and the population appears at most moderately skewed
  - $n \geq 60$, and the population is not extremely skewed
- Simulation methods can be used if
  - $n \geq 15$, and the population is Normal or at most slightly skewed
  - $n \geq 30$, and the population appears at most moderately skewed
  - $n \geq 60$, and the population is not extremely skewed

Theory vs. Simulation for 1 Mean

- In order to compute accurate confidence intervals and p-values, we need to ensure that appropriate conditions are met.
  - The strictness of these conditions depends on sample size.
- Theory-based can be used if...
  - $n \geq 2$, and the population is Normal (or nearly so)
  - $n \geq 30$, and the population appears at most moderately skewed
  - $n \geq 60$, and the population is not extremely skewed
- Simulation methods can be used if
  - $n \geq 15$, and the population is Normal or at most slightly skewed
  - $n \geq 30$, and the population appears at most moderately skewed
  - $n \geq 60$, and the population is not extremely skewed
- In general, for small sample sizes, neither method should be used if population does not appear Normal. But if it is Normal, theory-based methods will be more accurate.

Theory vs. Simulation for 1 Mean

- In order to compute accurate confidence intervals and p-values, we need to ensure that appropriate conditions are met.
  - The strictness of these conditions depends on sample size.
- Theory-based can be used if. . .
  - $n \geq 2$, and the population is Normal (or nearly so)
  - $n \geq 30$, and the population appears at most moderately skewed
  - $n \geq 60$, and the population is not extremely skewed
- Simulation methods can be used if
  - $n \geq 15$, and the population is Normal or at most slightly skewed
  - $n \geq 30$, and the population appears at most moderately skewed
  - $n \geq 60$, and the population is not extremely skewed
- In general, for small sample sizes, neither method should be used if population does not appear Normal. But if it is Normal, theory-based methods will be more accurate.
- For moderate sample sizes with moderate skew, simulation-based methods will be more accurate

t- versus z-procedures

It is important to use the t-distribution (rather than the Normal distribution) for confidence intervals and hypothesis tests when the sample size is small.

t- versus z-procedures

It is important to use the *t*-distribution (rather than the Normal distribution) for confidence intervals and hypothesis tests when the sample size is small.

- To verify, we'll create 1000 95% confidence intervals using (a) the t-distribution and (b) the Normal distribution, and see how many contain the true population mean.

t- versus z-procedures

It is important to use the *t*-distribution (rather than the Normal distribution) for confidence intervals and hypothesis tests when the sample size is small.

- To verify, we'll create 1000 95% confidence intervals using (a) the t-distribution and (b) the Normal distribution, and see how many contain the true population mean.
- Suppose we have the following population distribution for measurement errors

10000 Samples

The following code collects 10000 samples from the population, each of size 5. It then computes the mean and standard deviation of each sample.

```
set.seed(1023)
samps<-population %>%
  rep_sample_n(size = 5, reps = 10000) %>%
  group_by(replicate) %>%
  summarize(avg = mean(error), st_dev = sd(error))
```

```
## # A tibble: 6 x 3
##   replicate       avg  st_dev
##       <int>     <dbl>   <dbl>
## 1         1 -0.00742  0.00554
## 2         2  0.00133  0.00885
## 3         3  0.00465  0.00840
## 4         4 -0.00598  0.00924
## 5         5  0.000320 0.00738
## 6         6 -0.00148  0.00855
```

The Confidence Intervals

- The critical value for a 95% confidence interval using...
  - the standard Normal distribution is $z^* = 1.96$.
  - the t distribution with 4 df is $t^* = 2.776$.

The Confidence Intervals

- The critical value for a 95% confidence interval using...
  - the standard Normal distribution is $z^* = 1.96$.
  - the t distribution with 4 df is $t^* = 2.776$.
- The following code creates confidence intervals for each sample:

```
samps <- samps %>% mutate(
  lower_z = avg - 1.96*st_dev/sqrt(5), upper_z = avg + 1.96*st_dev/sqrt(5),
  lower_t = avg - 2.776*st_dev/sqrt(5), upper_t = avg + 2.776*st_dev/sqrt(5))
```

```
## # A tibble: 6 x 7
##   replicate      avg  st_dev  lower_z  upper_z  lower_t   upper_t
##       <int>    <dbl>   <dbl>    <dbl>    <dbl>    <dbl>     <dbl>
## 1         1 -0.00742  0.00554 -0.0123  -0.00257 -0.0143  -0.000545
## 2         2  0.00133  0.00885 -0.00643  0.00908 -0.00966  0.0123
## 3         3  0.00465  0.00840 -0.00272  0.0120  -0.00578  0.0151
## 4         4 -0.00598  0.00924 -0.0141   0.00212 -0.0174   0.00549
## 5         5  0.000320 0.00738 -0.00615  0.00679 -0.00884  0.00949
## 6         6 -0.00148  0.00855 -0.00897  0.00601 -0.0121   0.00913
```

Which intervals contain the true mean?

- Since we **know** the population has mean 0, we can determine whether each interval contains the true mean.

```
samps<-samps %>% mutate(
  z_success = ifelse(( lower_z < 0 & upper_z > 0 ) , "yes", "no"),
  t_success = ifelse(( lower_t < 0 & upper_t > 0 ) , "yes", "no"))
```

```
## # A tibble: 6 x 7
##   replicate  lower_z  upper_z  lower_t   upper_t z_success t_success
##       <int>    <dbl>    <dbl>    <dbl>     <dbl> <chr>     <chr>
## 1         1 -0.0123  -0.00257 -0.0143  -0.000545 no        no
## 2         2 -0.00643  0.00908 -0.00966  0.0123   yes       yes
## 3         3 -0.00272  0.0120  -0.00578  0.0151   yes       yes
## 4         4 -0.0141   0.00212 -0.0174   0.00549  yes       yes
## 5         5 -0.00615  0.00679 -0.00884  0.00949  yes       yes
## 6         6 -0.00897  0.00601 -0.0121   0.00913  yes       yes
```

Which intervals contain the true mean?

- Since we **know** the population has mean 0, we can determine whether each interval contains the true mean.

```
samps<-samps %>% mutate(
  z_success = ifelse(( lower_z < 0 & upper_z > 0 ) , "yes", "no"),
  t_success = ifelse(( lower_t < 0 & upper_t > 0 ) , "yes", "no"))
```

```
## # A tibble: 6 x 7
##   replicate  lower_z  upper_z  lower_t   upper_t z_success t_success
##       <int>    <dbl>    <dbl>    <dbl>     <dbl> <chr>     <chr>
## 1         1 -0.0123  -0.00257 -0.0143  -0.000545 no        no
## 2         2 -0.00643  0.00908 -0.00966  0.0123   yes       yes
## 3         3 -0.00272  0.0120  -0.00578  0.0151   yes       yes
## 4         4 -0.0141   0.00212 -0.0174   0.00549  yes       yes
## 5         5 -0.00615  0.00679 -0.00884  0.00949  yes       yes
## 6         6 -0.00897  0.00601 -0.0121   0.00913  yes       yes
```

- What proportion of z- and t-intervals contain 0?

```
## # A tibble: 1 x 2
##   z_rate t_rate
##    <dbl>  <dbl>
## 1  0.879  0.949
```

Section 2

Inference for 2 Means

Differences in Means

- Consider the following questions:

Differences in Means

- Consider the following questions:
  - Are variations in house prices in two different towns in the same state just due to sampling variability, or do they suggest an underlying difference?

Differences in Means

- Consider the following questions:
  - Are variations in house prices in two different towns in the same state just due to sampling variability, or do they suggest an underlying difference?
  - Does daily consumption of coffee improve performance on a standardized test compared to a control?

Differences in Means

- Consider the following questions:
  - Are variations in house prices in two different towns in the same state just due to sampling variability, or do they suggest an underlying difference?
  - Does daily consumption of coffee improve performance on a standardized test compared to a control?
  - Is there an appreciable difference in price between .99 and 1.0 carat diamonds?

Differences in Means

- Consider the following questions:
  - Are variations in house prices in two different towns in the same state just due to sampling variability, or do they suggest an underlying difference?
  - Does daily consumption of coffee improve performance on a standardized test compared to a control?
  - Is there an appreciable difference in price between .99 and 1.0 carat diamonds?
- Each of these questions can be answered by analyzing the difference in means between samples taken from two populations.

Differences in Means

- Consider the following questions:
  - Are variations in house prices in two different towns in the same state just due to sampling variability, or do they suggest an underlying difference?
  - Does daily consumption of coffee improve performance on a standardized test compared to a control?
  - Is there an appreciable difference in price between .99 and 1.0 carat diamonds?
- Each of these questions can be answered by analyzing the difference in means between samples taken from two populations.
- Groups could be formed from...

Differences in Means

- Consider the following questions:
  - Are variations in house prices in two different towns in the same state just due to sampling variability, or do they suggest an underlying difference?
  - Does daily consumption of coffee improve performance on a standardized test compared to a control?
  - Is there an appreciable difference in price between .99 and 1.0 carat diamonds?
- Each of these questions can be answered by analyzing the difference in means between samples taken from two populations.
- Groups could be formed from...
  - Two different populations.
  - Two subsets within the same sample distinguished by levels of a categorical variable.
  - Two treatment groups in an experiment.

Distribution for Difference in Means

- Suppose random samples of size n_1 and n_2 are drawn **independentally** from populations with means μ_1 and μ_2 and standard deviations σ_1 and σ_2, respectively.

Distribution for Difference in Means

- Suppose random samples of size n_1 and n_2 are drawn **independentally** from populations with means μ_1 and μ_2 and standard deviations σ_1 and σ_2, respectively.
- **Goal**: Estimate the value of the parameter $\mu_1 - \mu_2$ using the statistic $\bar{x}_1 - \bar{x}_2$.

Distribution for Difference in Means

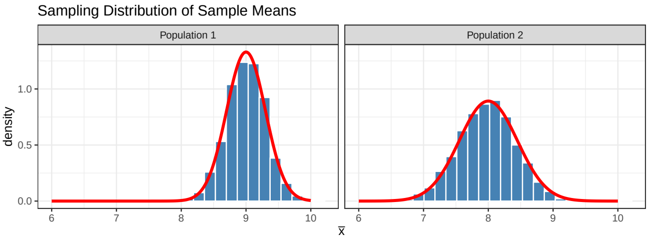
- Suppose random samples of size n_1 and n_2 are drawn **independentally** from populations with means μ_1 and μ_2 and standard deviations σ_1 and σ_2, respectively.
- **Goal**: Estimate the value of the parameter $\mu_1 - \mu_2$ using the statistic $\bar{x}_1 - \bar{x}_2$.
    - We need to know the shape, center, and spread of distribution of $\bar{x}_1 - \bar{x}_2$.

Distribution for Difference in Means

- Suppose random samples of size n_1 and n_2 are drawn **independentally** from populations with means μ_1 and μ_2 and standard deviations σ_1 and σ_2, respectively.
- **Goal**: Estimate the value of the parameter $\mu_1 - \mu_2$ using the statistic $\bar{x}_1 - \bar{x}_2$.
  - We need to know the shape, center, and spread of distribution of $\bar{x}_1 - \bar{x}_2$.
- By CLT, the distributions of $\bar{x}_1$ and $\bar{x}_2$ are approximately Normal.

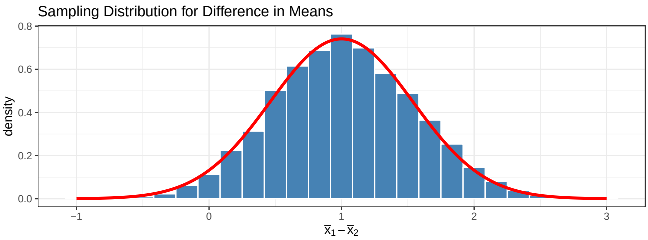
Distribution for Difference in Means

- Suppose random samples of size n_1 and n_2 are drawn **independentally** from populations with means μ_1 and μ_2 and standard deviations σ_1 and σ_2, respectively.
- **Goal**: Estimate the value of the parameter $\mu_1 - \mu_2$ using the statistic $\bar{x}_1 - \bar{x}_2$.
  - We need to know the shape, center, and spread of distribution of $\bar{x}_1 - \bar{x}_2$.
- By CLT, the distributions of $\bar{x}_1$ and $\bar{x}_2$ are approximately Normal.



Distribution for Difference in Means

- Suppose random samples of size n_1 and n_2 are drawn **independentally** from populations with means μ_1 and μ_2 and standard deviations σ_1 and σ_2, respectively.
- **Goal**: Estimate the value of the parameter $\mu_1 - \mu_2$ using the statistic $\bar{x}_1 - \bar{x}_2$.
  - We need to know the shape, center, and spread of distribution of $\bar{x}_1 - \bar{x}_2$.
- The distribtuion of the difference $\bar{x}_1 - \bar{x}_2$ is approximately Normal also



Practical Considerations

- By the Central Limit Theorem, as both n_1 and n_2 get larger, the distribution of difference in sample means $\bar{x}_1 - \bar{x}_2$ becomes more Normally distributed, with mean $\mu_1 - \mu_2$ and standard error $SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Practical Considerations

- By the Central Limit Theorem, as both n_1 and n_2 get larger, the distribution of difference in sample means $\bar{x}_1 - \bar{x}_2$ becomes more Normally distributed, with mean $\mu_1 - \mu_2$ and standard error $SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
- In practice, we estimate the parameters σ_1 and σ_2 with the sample statistics s_1 and s_2

Practical Considerations

- By the Central Limit Theorem, as both n_1 and n_2 get larger, the distribution of difference in sample means $\bar{x}_1 - \bar{x}_2$ becomes more Normally distributed, with mean $\mu_1 - \mu_2$ and standard error $SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
- In practice, we estimate the parameters σ_1 and σ_2 with the sample statistics s_1 and s_2
- Consider the standardized difference in sample means:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\mathrm{SE}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Practical Considerations

- By the Central Limit Theorem, as both n_1 and n_2 get larger, the distribution of difference in sample means $\bar{x}_1 - \bar{x}_2$ becomes more Normally distributed, with mean $\mu_1 - \mu_2$ and standard error $SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
- In practice, we estimate the parameters σ_1 and σ_2 with the sample statistics s_1 and s_2
- Consider the standardized difference in sample means:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\mathrm{SE}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Theorem

The standardized difference t is approximately t-distributed with degrees of freedom $df = \min\{n_1 - 1, n_2 - 1\}$.

This approximation is appropriate either when both sample sizes are large (i.e. $n_1, n_2 \geq 30$*), or when both populations are approximately Normally distributed.*

Diamonds are Forever

Question: Do 1.0 carat diamonds command a higher price than .99 carat diamonds, beyond what you would expect due to increase in weight?

- To answer, we collect random samples of 30 1.0 and 23 .99 carat diamonds.

Diamonds are Forever

Question: Do 1.0 carat diamonds command a higher price than .99 carat diamonds, beyond what you would expect due to increase in weight?

- To answer, we collect random samples of 30 1.0 and 23 .99 carat diamonds.
  - To decouple the effect of increase in size between the two groups, we divide price of the .99 carat diamonds by .99 to obtain price per carat ppc

Diamonds are Forever

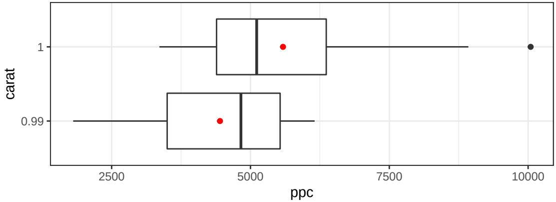
Question: Do 1.0 carat diamonds command a higher price than .99 carat diamonds, beyond what you would expect due to increase in weight?

- To answer, we collect random samples of 30 1.0 and 23 .99 carat diamonds.
    - To decouple the effect of increase in size between the two groups, we divide price of the .99 carat diamonds by .99 to obtain price per carat ppc
- Here are side-by-side boxplots of the diamonds, along with summary statistics

Diamonds are Forever

Question: Do 1.0 carat diamonds command a higher price than .99 carat diamonds, beyond what you would expect due to increase in weight?

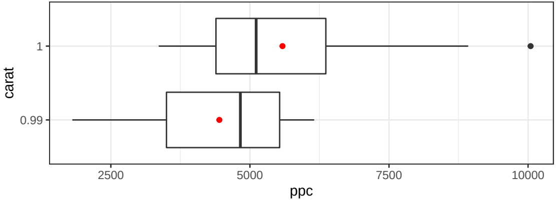
- To answer, we collect random samples of 30 1.0 and 23 .99 carat diamonds.
  - To decouple the effect of increase in size between the two groups, we divide price of the .99 carat diamonds by .99 to obtain price per carat ppc
- Here are side-by-side boxplots of the diamonds, along with summary statistics



Diamonds are Forever

Question: Do 1.0 carat diamonds command a higher price than .99 carat diamonds, beyond what you would expect due to increase in weight?

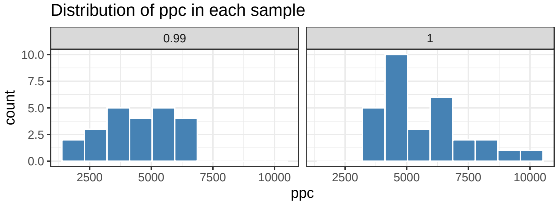
- To answer, we collect random samples of 30 1.0 and 23 .99 carat diamonds.
  - To decouple the effect of increase in size between the two groups, we divide price of the .99 carat diamonds by .99 to obtain price per carat ppc
- Here are side-by-side boxplots of the diamonds, along with summary statistics

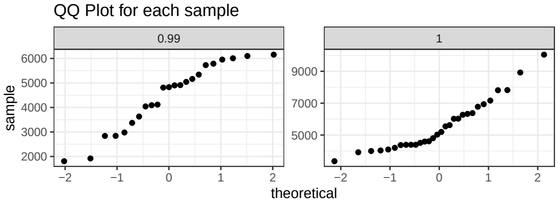


carat	mean_ppc	sd_ppc	n
1	5585.467	1613.572	30
0.99	4450.681	1332.311	23

Normal Conditions

- Our sample sizes are near the minimum conditions to use the Normal approximation. Are ppc Normally distributed for each carat value?





Hypothesis Test

1. Our null and alternate hypotheses:

$$H_0 : \mu_1 - \mu_{99} = 0 \qquad H_0 : \mu_1 - \mu_{99} > 0$$

Hypothesis Test

1. Our null and alternate hypotheses:

$$H_0 : \mu_1 - \mu_{99} = 0 \qquad H_0 : \mu_1 - \mu_{99} > 0$$

2. Assess conditions:

- Both samples showed some signs of non-normality, suggesting population may not be Normal. Sample size for each is moderate ($n_1 = 30$, $n_{99} = 23$). Proceed with caution.

Hypothesis Test

1. Our null and alternate hypotheses:

$$H_0 : \mu_1 - \mu_{99} = 0 \qquad H_0 : \mu_1 - \mu_{99} > 0$$

2. Assess conditions:

- Both samples showed some signs of non-normality, suggesting population may not be Normal. Sample size for each is moderate ($n_1 = 30$, $n_{99} = 23$). Proceed with caution.
- By construction, the two samples are independent. And observations within each sample are independent as well.

Hypothesis Test

1. Our null and alternate hypotheses:

$$H_0 : \mu_1 - \mu_{99} = 0 \qquad H_0 : \mu_1 - \mu_{99} > 0$$

2. Assess conditions:

- Both samples showed some signs of non-normality, suggesting population may not be Normal. Sample size for each is moderate ($n_1 = 30$, $n_{99} = 23$). Proceed with caution.
- By construction, the two samples are independent. And observations within each sample are independent as well.

3. We compute our test statistic

Hypothesis Test

1. Our null and alternate hypotheses:

$$H_0 : \mu_1 - \mu_{99} = 0 \qquad H_0 : \mu_1 - \mu_{99} > 0$$

2. Assess conditions:

- Both samples showed some signs of non-normality, suggesting population may not be Normal. Sample size for each is moderate ($n_1 = 30$, $n_{99} = 23$). Proceed with caution.
- By construction, the two samples are independent. And observations within each sample are independent as well.

3. We compute our test statistic

$$t = \frac{\bar{x}_1 - \bar{x}_{99}}{SE} = \frac{5585 - 4451}{\sqrt{\frac{1614^2}{30} + \frac{1332^2}{23}}} = 2.802$$

Hypothesis Test II

4. Calculate the P-Value.

Hypothesis Test II

4. Calculate the P-Value.

- The theory-based method says t is approximately t-distributed with 22 degrees of freedom: $\mathrm{df} = \min(23-1, 30-1) = 22$

Hypothesis Test II

4. Calculate the P-Value.

- The theory-based method says *t* is approximately *t*-distributed with 22 degrees of freedom: $\mathrm{df} = \min(23 - 1, 30 - 1) = 22$

```
p_value<-1 - pt( 2.802, df = 22)
p_value
```

```
## [1] 0.005194137
```

Hypothesis Test II

4. Calculate the P-Value.

- The theory-based method says t is approximately t-distributed with 22 degrees of freedom: $\mathrm{df} = \min(23 - 1, 30 - 1) = 22$

```
p_value<-1 - pt( 2.802, df = 22)
p_value
```

```
## [1] 0.005194137
```

5. Conclude

Hypothesis Test II

4. Calculate the P-Value.

- The theory-based method says t is approximately t-distributed with 22 degrees of freedom: $\mathrm{df} = \min(23-1, 30-1) = 22$

```
p_value<-1 - pt( 2.802, df = 22)
p_value
```

```
## [1] 0.005194137
```

5. Conclude

- At the $\alpha = 0.01$ significance level, we reject the null hypothesis. This sample suggests 1.0 carat diamonds command a higher price than is explained by increase in weight alone.

Hypothesis Test II

4. Calculate the P-Value.

- The theory-based method says *t* is approximately *t*-distributed with 22 degrees of freedom: $\mathrm{df} = \min(23-1, 30-1) = 22$

```
p_value<-1 - pt( 2.802, df = 22)
p_value
```

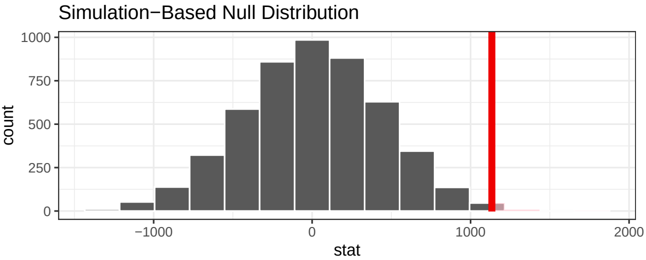
```
## [1] 0.005194137
```

5. Conclude

- At the $\alpha = 0.01$ significance level, we reject the null hypothesis. This sample suggests 1.0 carat diamonds command a higher price than is explained by increase in weight alone.
  - However, we cannot be certain the Normal condition is satisfied. The p-value may not be accurate.

Comparison with infer

```
set.seed(101)
diamonds_null <- diamonds %>% specify(ppc ~ carat) %>%
  hypothesize(null = "independence") %>%
  generate(reps = 5000, type = "permute") %>%
  calculate(stat = "diff in means", order = c("1", "0.99"))
diamonds_null %>% visualize()+shade_p_value(obs_stat = 1135, direction = "right")
```




```
diamonds_null %>% get_p_value(obs_stat = 1135, direction = "right" )
```

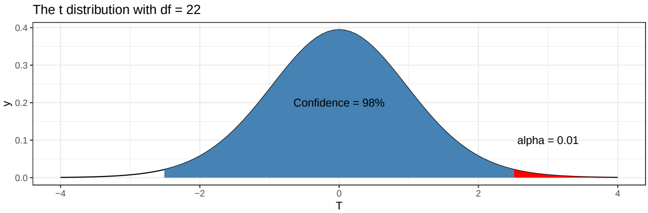
```
## # A tibble: 1 x 1
##   p_value
##     <dbl>
## 1  0.0042
```

An Equivalent Confidence Interval

Goal: Create a confidence interval that corresponds to **one-sided** $\alpha = 0.01$ significance

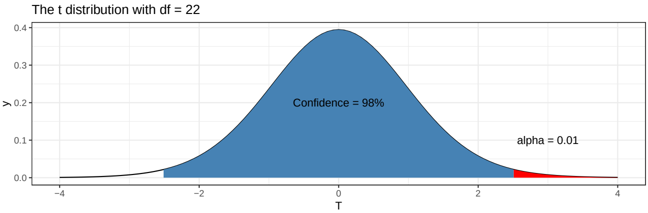
An Equivalent Confidence Interval

Goal: Create a confidence interval that corresponds to **one-sided** $\alpha = 0.01$ significance



An Equivalent Confidence Interval

Goal: Create a confidence interval that corresponds to **one-sided** $\alpha = 0.01$ significance



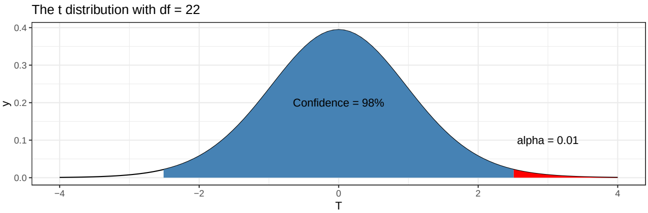
- What is the t^* critical value for 98% confidence?

```
t_star<- qt(.99, df = 22)
t_star
```

```
## [1] 2.508325
```

An Equivalent Confidence Interval

Goal: Create a confidence interval that corresponds to **one-sided** $\alpha = 0.01$ significance



- What is the t^* critical value for 98% confidence?

```
t_star<- qt(.99, df = 22)
t_star
```

```
## [1] 2.508325
```

- Note that our observed t statistic was $t = 2.802$, which was more extreme than the critical value for 98% confidence

Confidence Interval

- Create a 98% confidence interval to estimate the difference $\mu_1 - \mu_{99}$

Confidence Interval

- Create a 98% confidence interval to estimate the difference $\mu_1 - \mu_{99}$
- The formula for our confidence interval is

$$(\bar{x}_1 - \bar{x}_{99}) \pm t^* \cdot SE$$

Confidence Interval

- Create a 98% confidence interval to estimate the difference $\mu_1 - \mu_{99}$
- The formula for our confidence interval is

$$(\bar{x}_1 - \bar{x}_{99}) \pm t^* \cdot SE$$

Giving an interval of

$$(5585 - 4451) \pm 2.508 \cdot \sqrt{\frac{1614^2}{30} + \frac{1332^2}{23}}$$

which is $(118.42, 2149.58)$.

Confidence Interval

- Create a 98% confidence interval to estimate the difference $\mu_1 - \mu_{99}$
- The formula for our confidence interval is

$$(\bar{x}_1 - \bar{x}_{99}) \pm t^* \cdot SE$$

Giving an interval of

$$(5585 - 4451) \pm 2.508 \cdot \sqrt{\frac{1614^2}{30} + \frac{1332^2}{23}}$$

which is $(118.42, 2149.58)$.

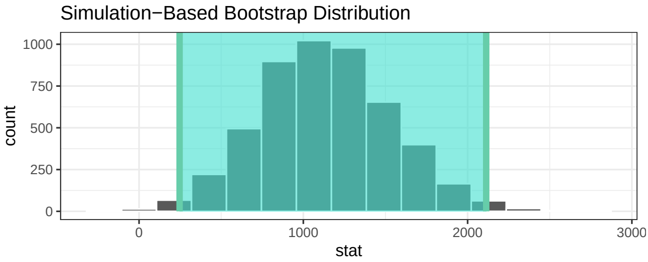
- Since this interval **does not** contain 0, we conclude that there IS a price increase for 1.0 carat diamonds.
  - Moreover, this price increase is likely between \$120 and \$2150.

Comparison with infer

```
diamonds_ci <- diamonds_boot %>% get_ci(level = 0.98, type = "percentile")
diamonds_ci
```

```
## # A tibble: 1 x 2
##   lower_ci upper_ci
##      <dbl>    <dbl>
## 1     247.    2113.
```

```
diamonds_boot %>%visualize()+ shade_ci(endpoints = diamonds_ci)
```



Section 3

Inference for Paired Samples

Matched Pairs

- Suppose you intend to design an experiment to determine whether the mean of two populations are equal.

Matched Pairs

- Suppose you intend to design an experiment to determine whether the mean of two populations are equal.
- You could obtain an SRS from each population, compute means for each sample, take the difference, and assess variability based on previous procedures.

Matched Pairs

- Suppose you intend to design an experiment to determine whether the mean of two populations are equal.
- You could obtain an SRS from each population, compute means for each sample, take the difference, and assess variability based on previous procedures.
    - It *is* possible that any observed effect is not due to the explanatory variable, but to some confounding variable present in one sample but not other.

Matched Pairs

- Suppose you intend to design an experiment to determine whether the mean of two populations are equal.
- You could obtain an SRS from each population, compute means for each sample, take the difference, and assess variability based on previous procedures.
    - It *is* possible that any observed effect is not due to the explanatory variable, but to some confounding variable present in one sample but not other.
- To reduce effect of confounding variable, you could match individuals in one sample with similar individuals in the other sample.

Matched Pairs

- Suppose you intend to design an experiment to determine whether the mean of two populations are equal.
- You could obtain an SRS from each population, compute means for each sample, take the difference, and assess variability based on previous procedures.
  - It *is* possible that any observed effect is not due to the explanatory variable, but to some confounding variable present in one sample but not other.
- To reduce effect of confounding variable, you could match individuals in one sample with similar individuals in the other sample.
  - But if matching is used in sample design, it is **not** appropriate to use the 2 sample procedures. (Why?)

Matched Pairs

- Suppose you intend to design an experiment to determine whether the mean of two populations are equal.
- You could obtain an SRS from each population, compute means for each sample, take the difference, and assess variability based on previous procedures.
  - It *is* possible that any observed effect is not due to the explanatory variable, but to some confounding variable present in one sample but not other.
- To reduce effect of confounding variable, you could match individuals in one sample with similar individuals in the other sample.
  - But if matching is used in sample design, it is **not** appropriate to use the 2 sample procedures. (Why?)
- You can create a new variable recording the **difference** in measurements in each pair of individuals

Matched Pairs

- Suppose you intend to design an experiment to determine whether the mean of two populations are equal.
- You could obtain an SRS from each population, compute means for each sample, take the difference, and assess variability based on previous procedures.
  - It *is* possible that any observed effect is not due to the explanatory variable, but to some confounding variable present in one sample but not other.
- To reduce effect of confounding variable, you could match individuals in one sample with similar individuals in the other sample.
  - But if matching is used in sample design, it is **not** appropriate to use the 2 sample procedures. (Why?)
- You can create a new variable recording the **difference** in measurements in each pair of individuals
- This new variable can be used to perform statistical inference using the 1-sample procedures for mean.

Matched Pairs

- Suppose you intend to design an experiment to determine whether the mean of two populations are equal.
- You could obtain an SRS from each population, compute means for each sample, take the difference, and assess variability based on previous procedures.
  - It *is* possible that any observed effect is not due to the explanatory variable, but to some confounding variable present in one sample but not other.
- To reduce effect of confounding variable, you could match individuals in one sample with similar individuals in the other sample.
  - But if matching is used in sample design, it is **not** appropriate to use the 2 sample procedures. (Why?)
- You can create a new variable recording the **difference** in measurements in each pair of individuals
- This new variable can be used to perform statistical inference using the 1-sample procedures for mean.
  - Rather than looking at the difference in means, we look at the mean of differences!

The World's Fastest Swimsuit

In the 2008 Olympics, controversy erupted over whether a new swimsuit design provided an unfair advantage to swimmers. Eventually, the International Swimming Organization banned the new suit. But can a certain suit really make a swimmer faster?

Data

A study analyzed max velocities for 12 pro swimmers with and without the suit:

swimmer	with_suit	without_suit	difference
1	1.57	1.49	0.08
2	1.47	1.37	0.10
3	1.42	1.35	0.07
4	1.35	1.27	0.08
5	1.22	1.12	0.10
6	1.75	1.64	0.11
7	1.64	1.59	0.05
8	1.57	1.52	0.05
9	1.56	1.50	0.06
10	1.53	1.45	0.08
11	1.49	1.44	0.05
12	1.51	1.41	0.10

Data

A study analyzed max velocities for 12 pro swimmers with and without the suit:

swimmer	with_suit	without_suit	difference
1	1.57	1.49	0.08
2	1.47	1.37	0.10
3	1.42	1.35	0.07
4	1.35	1.27	0.08
5	1.22	1.12	0.10
6	1.75	1.64	0.11
7	1.64	1.59	0.05
8	1.57	1.52	0.05
9	1.56	1.50	0.06
10	1.53	1.45	0.08
11	1.49	1.44	0.05
12	1.51	1.41	0.10

- Without performing any statistical inference, what is the likely conclusion to draw from this data?

Hypothesis Testing

We want to determine whether the *average* difference in max velocity (with - without) is positive. Let μ be the average difference.

1. State Hypotheses:

$$H_0 : \mu = 0 \quad H_a : \mu > 0$$

2. Check Conditions:

Hypothesis Testing

We want to determine whether the *average* difference in max velocity (with - without) is positive. Let μ be the average difference.

1. State Hypotheses:

$$H_0 : \mu = 0 \quad H_a : \mu > 0$$

2. Check Conditions:

- With only 12 data points, we can't use the sample's visualization to assess Normality.
- However, analysis of other swim speed data (outside of this experiment) does suggest max race swim speed is approximately Normal

Hypothesis Testing

We want to determine whether the *average* difference in max velocity (with - without) is positive. Let μ be the average difference.

1. State Hypotheses:

$$H_0 : \mu = 0 \quad H_a : \mu > 0$$

2. Check Conditions:

- With only 12 data points, we can't use the sample's visualization to assess Normality.
- However, analysis of other swim speed data (outside of this experiment) does suggest max race swim speed is approximately Normal

3. Compute relevant statistics

```
## # A tibble: 1 x 3
##    x_bar      s     n
##    <dbl>  <dbl> <int>
## 1 0.0775 0.0218    12
```

Hypothesis Testing, cont'd

3. Compute Test Statistic

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{0.0775}{\frac{0.022}{\sqrt{12}}} = 12.32$$

Hypothesis Testing, cont'd

3. Compute Test Statistic

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{0.0775}{\frac{0.022}{\sqrt{12}}} = 12.32$$

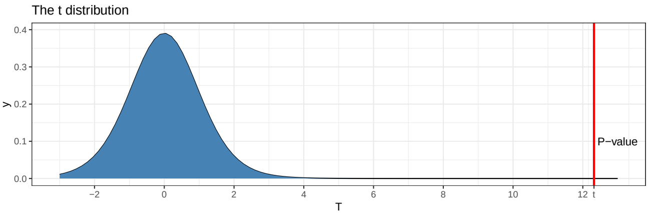
- Plotting our t-statistic against the theoretical t-distribution with $df = 11$

Hypothesis Testing, cont'd

3. Compute Test Statistic

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{0.0775}{\frac{0.022}{\sqrt{12}}} = 12.32$$

- Plotting our t-statistic against the theoretical t-distribution with $df = 11$

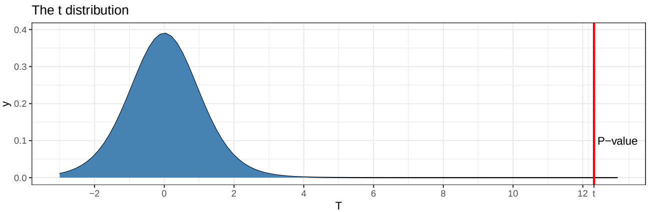


Hypothesis Testing, cont'd

3. Compute Test Statistic

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{0.0775}{\frac{0.022}{\sqrt{12}}} = 12.32$$

- Plotting our t-statistic against the theoretical t-distribution with $df = 11$



4. Obtain p-value:

```
1-pt(12.32, df = 11)
```

```
## [1] 4.435835e-08
```

Conclusion

5. Conclusion?

Conclusion

5. Conclusion?

- Compare to $\alpha = 0.01$. Since the p-value of the sample (p-value $= 4.4 \times 10^{-8}$) is less than α, we reject H_0. This sample gives evidence that suits indeed increase max race speed.

Conclusion

5. Conclusion?

- Compare to $\alpha = 0.01$. Since the p-value of the sample (p-value $= 4.4 \times 10^{-8}$) is less than α, we reject H_0. This sample gives evidence that suits indeed increase max race speed.
- Does the very, very small p-value give evidence that the suits give an extreme advantage?

Conclusion

5. Conclusion?

- Compare to $\alpha = 0.01$. Since the p-value of the sample (p-value $= 4.4 \times 10^{-8}$) is less than α, we reject H_0. This sample gives evidence that suits indeed increase max race speed.
- Does the very, very small p-value give evidence that the suits give an extreme advantage?
  - No. Only that this result was very unlikely to have occured just due to chance.

Conclusion

5. Conclusion?

- Compare to $\alpha = 0.01$. Since the p-value of the sample (p-value $= 4.4 \times 10^{-8}$) is less than α, we reject H_0. This sample gives evidence that suits indeed increase max race speed.
- Does the very, very small p-value give evidence that the suits give an extreme advantage?
  - No. Only that this result was very unlikely to have occured just due to chance.
- To determine whether the observed difference is of practical concern, we need to use domain knowlege.

Conclusion

5. Conclusion?

- Compare to $\alpha = 0.01$. Since the p-value of the sample (p-value $= 4.4 \times 10^{-8}$) is less than α, we reject H_0. This sample gives evidence that suits indeed increase max race speed.
- Does the very, very small p-value give evidence that the suits give an extreme advantage?
  - No. Only that this result was very unlikely to have occured just due to chance.
- To determine whether the observed difference is of practical concern, we need to use domain knowlege.
- Is a difference of 0.077 max velocity of practical significance?

Conclusion

5. Conclusion?

- Compare to $\alpha = 0.01$. Since the p-value of the sample (p-value $= 4.4 \times 10^{-8}$) is less than α, we reject H_0. This sample gives evidence that suits indeed increase max race speed.
- Does the very, very small p-value give evidence that the suits give an extreme advantage?
  - No. Only that this result was very unlikely to have occured just due to chance.
- To determine whether the observed difference is of practical concern, we need to use domain knowlege.
- Is a difference of 0.077 max velocity of practical significance?
  - If the average max velocity is about 1.4, this is about a 5% increase in speed.

Conclusion

5. Conclusion?

- Compare to $\alpha = 0.01$. Since the p-value of the sample (p-value $= 4.4 \times 10^{-8}$) is less than α, we reject H_0. This sample gives evidence that suits indeed increase max race speed.
- Does the very, very small p-value give evidence that the suits give an extreme advantage?
  - No. Only that this result was very unlikely to have occured just due to chance.
- To determine whether the observed difference is of practical concern, we need to use domain knowlege.
- Is a difference of 0.077 max velocity of practical significance?
  - If the average max velocity is about 1.4, this is about a 5% increase in speed.
- In the 2008 Beijing Olympics, swimmers wearing the suit. . .

Conclusion

5. Conclusion?

- Compare to $\alpha = 0.01$. Since the p-value of the sample (p-value $= 4.4 \times 10^{-8}$) is less than α, we reject H_0. This sample gives evidence that suits indeed increase max race speed.
- Does the very, very small p-value give evidence that the suits give an extreme advantage?
  - No. Only that this result was very unlikely to have occured just due to chance.
- To determine whether the observed difference is of practical concern, we need to use domain knowlege.
- Is a difference of 0.077 max velocity of practical significance?
  - If the average max velocity is about 1.4, this is about a 5% increase in speed.
- In the 2008 Beijing Olympics, swimmers wearing the suit. . .
  - Were awarded 98% of all medals (including 33 of 36 gold medals).
  - Represented 23 of the total 25 world records broken.