Inference for Difference in 2 Means

Prof. Wells

STA 209, 4/28/23

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

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 **independently** 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 **independently** 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

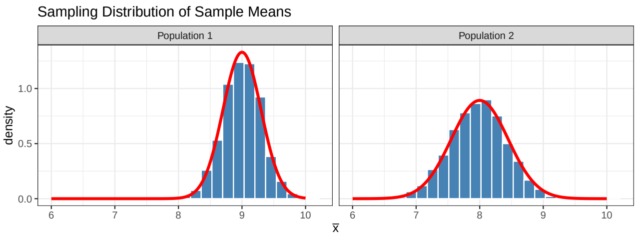
- Suppose random samples of size n_1 and n_2 are drawn **independently** from populations with means μ_1 and μ_2 and standard deviations σ_1 and σ_2, respectively.
- **Goal**: Estimate the value of the parameter $\mu_1 - \mu_2$ using the statistic $\bar{x}_1 - \bar{x}_2$.
    - We need to know the shape,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

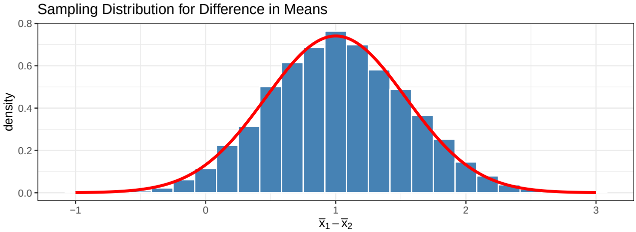
Distribution for Difference in Means

- Suppose random samples of size n_1 and n_2 are drawn **independently** 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 **independently** 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 $SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
  - The SE of the difference in means is **larger** than the SE for either mean.

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 $SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
    - The SE of the difference in means is **larger** than the SE for either mean.
    - 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 $SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
    - The SE of the difference in means is **larger** than the SE for either mean.
    - 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 $SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
  - The SE of the difference in means is **larger** than the SE for either mean.
  - 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

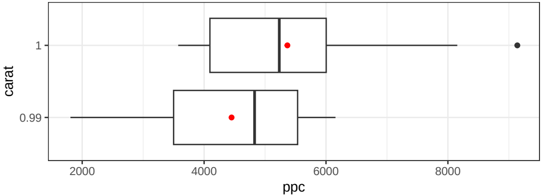
Question: Do 1.0 carat diamonds command a higher price than .99 carat diamonds, beyond what you would expect due to increase in weight?

- 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?

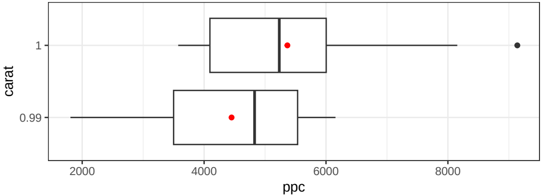
- To answer, we collect random samples of 30 1.0 and 23 .99 carat diamonds.
  - To decouple the effect of increase in size between the two groups, we divide price of the .99 carat diamonds by .99 to obtain price per carat ppc
- Here are side-by-side boxplots of the diamonds, along with summary statistics



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?

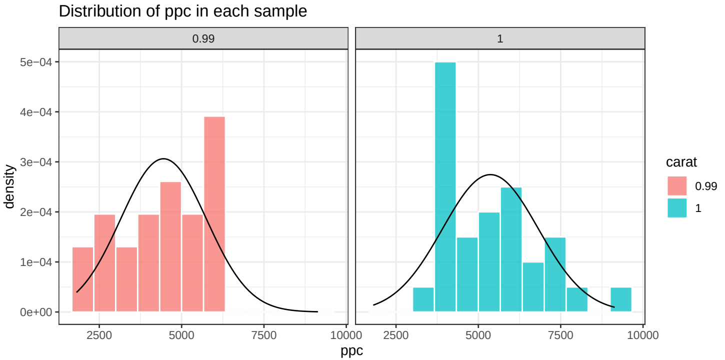
- To answer, we collect random samples of 30 1.0 and 23 .99 carat diamonds.
  - To decouple the effect of increase in size between the two groups, we divide price of the .99 carat diamonds by .99 to obtain price per carat ppc
- Here are side-by-side boxplots of the diamonds, along with summary statistics



carat	mean_ppc	sd_ppc	n
1	5366.433	1478.283	30
0.99	4450.681	1332.311	23

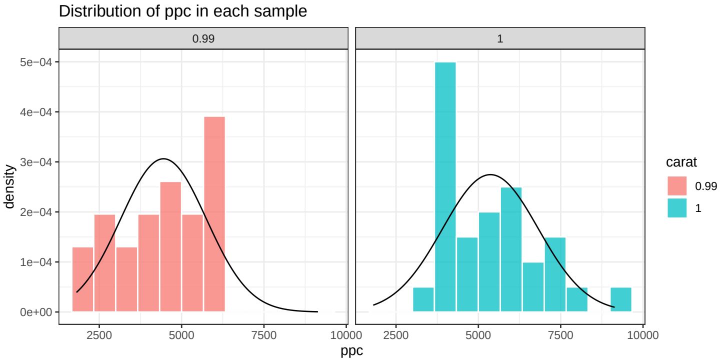
Normal Conditions

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



Normal Conditions

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



- The data does not appear to closely follow a Normal distribution (especially for .99 carat diamonds); use theory-based methods with caution!

Estimating Difference

Goal: Create a 90% confidence interval to estimate difference in mean ppc $\mu_1 - \mu_{99}$ for 1.0 and .99 carat diamonds.

Estimating Difference

Goal: Create a 90% confidence interval to estimate difference in mean ppc $\mu_1 - \mu_{99}$ for 1.0 and .99 carat diamonds.

- The formula for our confidence interval is

$$\text{statistic} \pm t^* \cdot SE$$

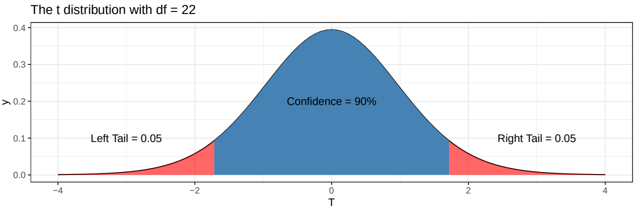
Estimating Difference

Goal: Create a 90% confidence interval to estimate difference in mean ppc $\mu_1 - \mu_{99}$ for 1.0 and .99 carat diamonds.

- The formula for our confidence interval is

$$\text{statistic} \pm t^* \cdot SE$$

- To obtain the critical value t^* for 90% confidence, use t-distributed with 22 degrees of freedom: $\mathrm{df} = \min(23-1, 30-1) = 22$




```
qt(.95, df = 22)
```

```
## [1] 1.717144
```

Confidence Interval

- Our statistic of interest is

$$\bar{x}_1 - \bar{x}_{99} = 5366 - 4450 = 916$$

Confidence Interval

- Our statistic of interest is

$$\bar{x}_1 - \bar{x}_{99} = 5366 - 4450 = 916$$

- Our standard error is

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_{99}^2}{n_{99}}} = \sqrt{\frac{1478^2}{30} + \frac{1332^2}{23}} = 387.24$$

Confidence Interval

- Our statistic of interest is

$$\bar{x}_1 - \bar{x}_{99} = 5366 - 4450 = 916$$

- Our standard error is

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_{99}^2}{n_{99}}} = \sqrt{\frac{1478^2}{30} + \frac{1332^2}{23}} = 387.24$$

- The formula for our confidence interval is

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

Confidence Interval

- Our statistic of interest is

$$\bar{x}_1 - \bar{x}_{99} = 5366 - 4450 = 916$$

- Our standard error is

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_{99}^2}{n_{99}}} = \sqrt{\frac{1478^2}{30} + \frac{1332^2}{23}} = 387.24$$

- The formula for our confidence interval is

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

- Our interval is therefore

$$916 \pm 1.717 \cdot 387.24 \qquad \text{or} \qquad (250, 1580)$$

Confidence Interval

- Our statistic of interest is

$$\bar{x}_1 - \bar{x}_{99} = 5366 - 4450 = 916$$

- Our standard error is

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_{99}^2}{n_{99}}} = \sqrt{\frac{1478^2}{30} + \frac{1332^2}{23}} = 387.24$$

- The formula for our confidence interval is

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

- Our interval is therefore

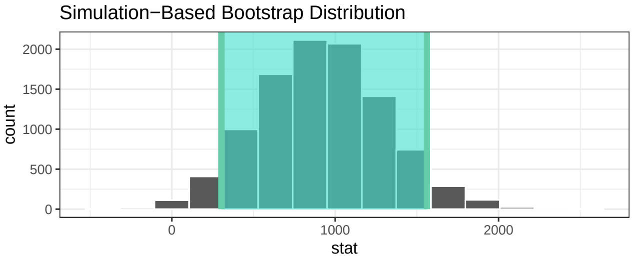
$$916 \pm 1.717 \cdot 387.24 \qquad \text{or} \qquad (250, 1580)$$

- Since this interval **does not** contain 0, it is plausible (at 90% confidence) that there is a price increase for 1.0 carat diamonds.
  - Moreover, this price increase is likely between \$250 and \$1580.

Comparison with infer

```
dd %>% specify(ppc ~ carat) %>%
  generate(reps = 10000, type = "bootstrap") %>%
  calculate(stat = "diff in means", order = c("1", "0.99")) %>%
  get_ci(level = 0.90, type = "percentile")
```

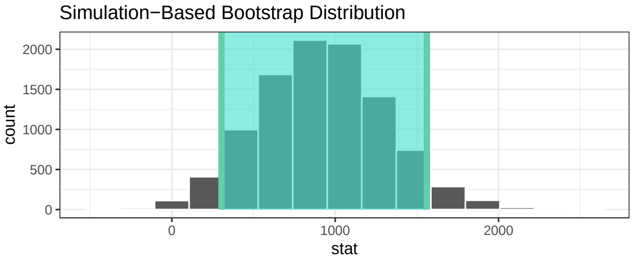
```
## # A tibble: 1 x 2
##   lower_ci upper_ci
##      <dbl>    <dbl>
## 1     304.    1561.
```



Comparison with infer

```
dd %>% specify(ppc ~ carat) %>%
  generate(reps = 10000, type = "bootstrap") %>%
  calculate(stat = "diff in means", order = c("1", "0.99")) %>%
  get_ci(level = 0.90, type = "percentile")
```

```
## # A tibble: 1 x 2
##   lower_ci upper_ci
##      <dbl>    <dbl>
## 1     304.    1561.
```



- Despite concerns about Normality of population, simulation and theory-based methods give similar confidence intervals

Section 2

Hypothesis Testing

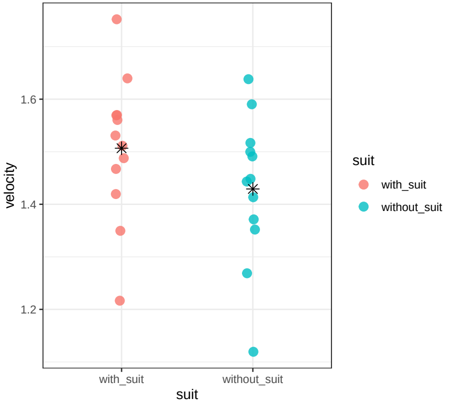
The World's Fastest Swimsuit

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:

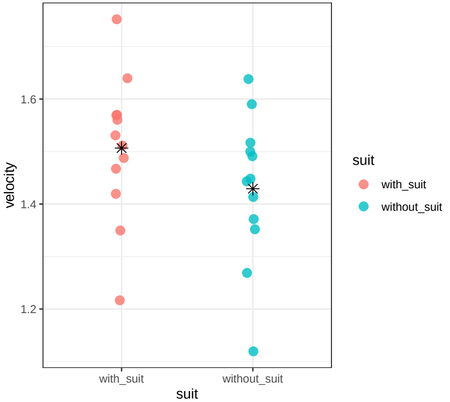
swimmer	with_suit	without_suit
1	1.57	1.49
2	1.47	1.37
3	1.42	1.35
4	1.35	1.27
5	1.22	1.12
6	1.75	1.64
7	1.64	1.59
8	1.57	1.52
9	1.56	1.50
10	1.53	1.45
11	1.49	1.44
12	1.51	1.41



Data

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

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



- Is is plausible the average velocity with the suit is greater than the average without the suit?

Hypothesis Testing

- Suppose we first treat this data as if it came from two *independent* populations.

Hypothesis Testing

- Suppose we first treat this data as if it came from two *independent* populations.
- We want to determine whether there is a positive difference in *average* max velocities for these two populations.
  - Let μ_s and μ_n be the average velocities for swimmers with **S**uit and with **N**o suit.

$$H_0 : \mu_s - \mu_n = 0 \qquad H_a : \mu_s - \mu_n > 0$$

Hypothesis Testing

- Suppose we first treat this data as if it came from two *independent* populations.
- We want to determine whether there is a positive difference in *average* max velocities for these two populations.
    - Let μ_s and μ_n be the average velocities for swimmers with **S**uit and with **N**o suit.

$$H_0 : \mu_s - \mu_n = 0 \qquad H_a : \mu_s - \mu_n > 0$$

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

Hypothesis Testing

- Suppose we first treat this data as if it came from two *independent* populations.
- We want to determine whether there is a positive difference in *average* max velocities for these two populations.
  - Let μ_s and μ_n be the average velocities for swimmers with **S**uit and with **N**o suit.

$$H_0 : \mu_s - \mu_n = 0 \qquad H_a : \mu_s - \mu_n > 0$$

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

- Compute relevant statistics

```
## # A tibble: 2 x 4
##   suit         x_bar     s     n
##   <chr>        <dbl> <dbl> <int>
## 1 with_suit     1.51 0.136    12
## 2 without_suit  1.43 0.141    12
```

Hypothesis Testing

- Suppose we first treat this data as if it came from two *independent* populations.
- We want to determine whether there is a positive difference in *average* max velocities for these two populations.
  - Let μ_s and μ_n be the average velocities for swimmers with **S**uit and with **N**o suit.

$$H_0 : \mu_s - \mu_n = 0 \qquad H_a : \mu_s - \mu_n > 0$$

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

- Compute relevant statistics

```
## # A tibble: 2 x 4
##   suit         x_bar     s     n
##   <chr>        <dbl> <dbl> <int>
## 1 with_suit     1.51 0.136    12
## 2 without_suit  1.43 0.141    12
```

- Compute Test Statistic

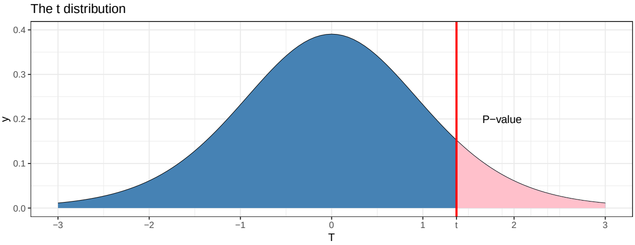
$$t = \frac{\bar{x}_s - \bar{x}_n}{\sqrt{\frac{s_s^2}{n_s} + \frac{s_n^2}{n_n}}} = \frac{0.0775}{\sqrt{\frac{0.136^2}{12} + \frac{0.141^2}{12}}} = 1.369$$

Hypothesis Testing, cont'd

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

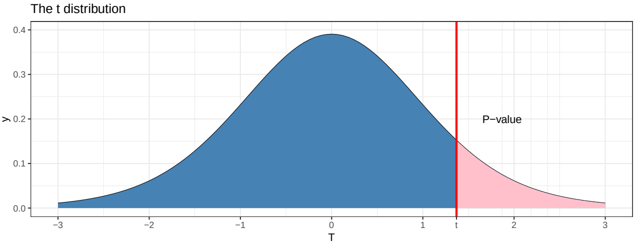
Hypothesis Testing, cont'd

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



Hypothesis Testing, cont'd

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



- Obtain p-value:

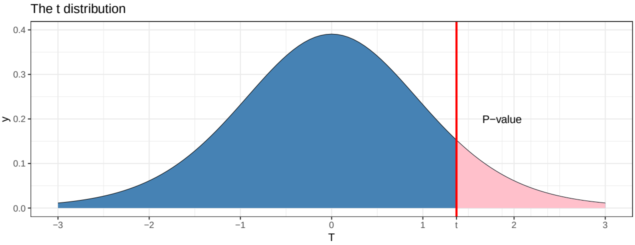
```
1-pt(1.369, df = 11)
```

```
## [1] 0.09915186
```

- The p-value of the sample 0.099 is relatively large (greater than $\alpha = 0.05$) so does not give convincing evidence to reject H_0

Hypothesis Testing, cont'd

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



- Obtain p-value:

```
1-pt(1.369, df = 11)
```

```
## [1] 0.09915186
```

- The p-value of the sample 0.099 is relatively large (greater than $\alpha = 0.05$) so does not give convincing evidence to reject H_0
  - Yet, the International Swimming Organization banned the new suit. What's going on?

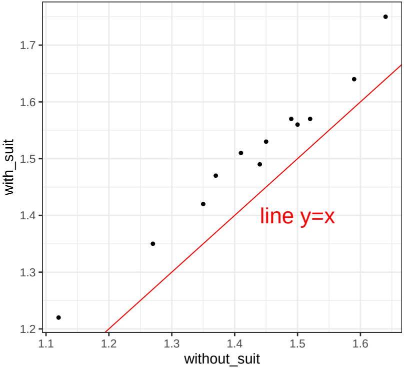
Reflections

- Consider our data. Each swimmer performed a race with and without the suit. We would expect velocities for each swimmer to be close together!

Reflections

- Consider our data. Each swimmer performed a race with and without the suit. We would expect velocities for each swimmer to be close together!

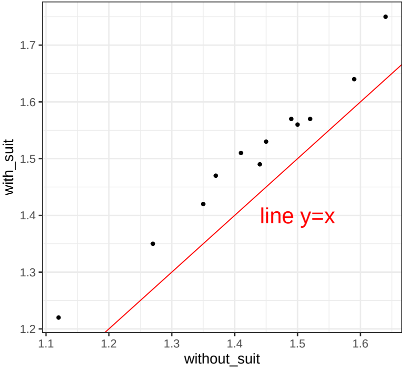
swimmer	with_suit	without_suit	difference
1	1.57	1.49	0.08
2	1.47	1.37	0.10
3	1.42	1.35	0.07
4	1.35	1.27	0.08
5	1.22	1.12	0.10
6	1.75	1.64	0.11
7	1.64	1.59	0.05
8	1.57	1.52	0.05
9	1.56	1.50	0.06
10	1.53	1.45	0.08
11	1.49	1.44	0.05
12	1.51	1.41	0.10



Reflections

- Consider our data. Each swimmer performed a race with and without the suit. We would expect velocities for each swimmer to be close together!

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



- Each swimmer's time was faster with the suit than without the suit!

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.
  - Or it is possible that a confounding variable *masks* a true effect.

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.
  - Or it is possible that a confounding variable *masks* a true effect.
- 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.
  - Or it is possible that a confounding variable *masks* a true effect.
- To reduce effect of confounding variable, you could match individuals in one sample with similar individuals in the other sample.
  - If matching is used, 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.
  - Or it is possible that a confounding variable *masks* a true effect.
- To reduce effect of confounding variable, you could match individuals in one sample with similar individuals in the other sample.
  - If matching is used, 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.
  - Or it is possible that a confounding variable *masks* a true effect.
- To reduce effect of confounding variable, you could match individuals in one sample with similar individuals in the other sample.
  - If matching is used, 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.
  - Or it is possible that a confounding variable *masks* a true effect.
- To reduce effect of confounding variable, you could match individuals in one sample with similar individuals in the other sample.
  - If matching is used, 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!

Hypothesis Testing

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

1. Write Hypotheses

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

Hypothesis Testing

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

1. Write Hypotheses

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

2. Compute relative statistics
   - Average and sd of **differences**

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

Hypothesis Testing

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

1. Write Hypotheses

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

2. Compute relative statistics
   - Average and sd of **differences**

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

3. Compute test statistic
   - Using the 1 sample procedures!

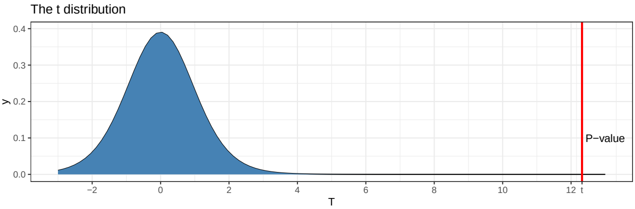
$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{0.0775}{\frac{0.022}{\sqrt{12}}} = 12.32$$

P-Value

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

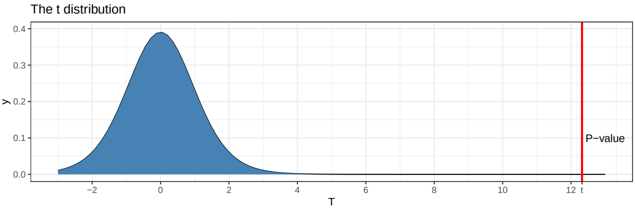
P-Value

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



P-Value

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



4. Obtain p-value:

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

```
## [1] 0.000000044
```

Conclusion

- Since the p-value of the sample is extremely small (< 0.001), this gives us very convincing evidence that suits indeed increase max race speed.

Conclusion

- Since the p-value of the sample is extremely small (< 0.001), this gives us very convincing evidence that suits indeed increase max race speed.
  - This is the opposite conclusion we reached when treating the two samples as independent.

Conclusion

- Since the p-value of the sample is extremely small (< 0.001), this gives us very convincing evidence that suits indeed increase max race speed.
  - This is the opposite conclusion we reached when treating the two samples as independent.
  - The differences between swimmers in the sample overpowered the differences within each swimmer, diminishing our ability to detect the effect of the suit

Conclusion

- Since the p-value of the sample is extremely small (< 0.001), this gives us very convincing evidence that suits indeed increase max race speed.
  - This is the opposite conclusion we reached when treating the two samples as independent.
  - The differences between swimmers in the sample overpowered the differences within each swimmer, diminishing our ability to detect the effect of the suit
  - By controlling for swimmer and analyzing differences for each swimmer, we were able to see the effect!

Conclusion

- Since the p-value of the sample is extremely small (< 0.001), this gives us very convincing evidence that suits indeed increase max race speed.
  - This is the opposite conclusion we reached when treating the two samples as independent.
  - The differences between swimmers in the sample overpowered the differences within each swimmer, diminishing our ability to detect the effect of the suit
  - By controlling for swimmer and analyzing differences for each swimmer, we were able to see the effect!
- Does the very small p-value give evidence that the suits give an extreme advantage?
  - Is a difference of 0.077 max velocity of practical significance?

Conclusion

- Since the p-value of the sample is extremely small (< 0.001), this gives us very convincing evidence that suits indeed increase max race speed.
  - This is the opposite conclusion we reached when treating the two samples as independent.
  - The differences between swimmers in the sample overpowered the differences within each swimmer, diminishing our ability to detect the effect of the suit
  - By controlling for swimmer and analyzing differences for each swimmer, we were able to see the effect!
- Does the very small p-value give evidence that the suits give an extreme advantage?
  - 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.
  - The velocity difference between 1st and 2nd place swimmers in many Olympic races is less than 1%