Wednesday, March 20, 2019

- Warm-up
  - Use the given computer output:
    - Identify:
      - Confidence interval
      - Mean

One-Sample Test

	Test Value = 50					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
writing score	4.140	199	.000	2.7750	1.4533	4.0967

    - Using the mean and CI, calculate the Margin of Error
    - Using ME and t*, calculate the Standard Error
    - Using df – calculate n
- Check Homework
- Quiz
- 2 Sample CI

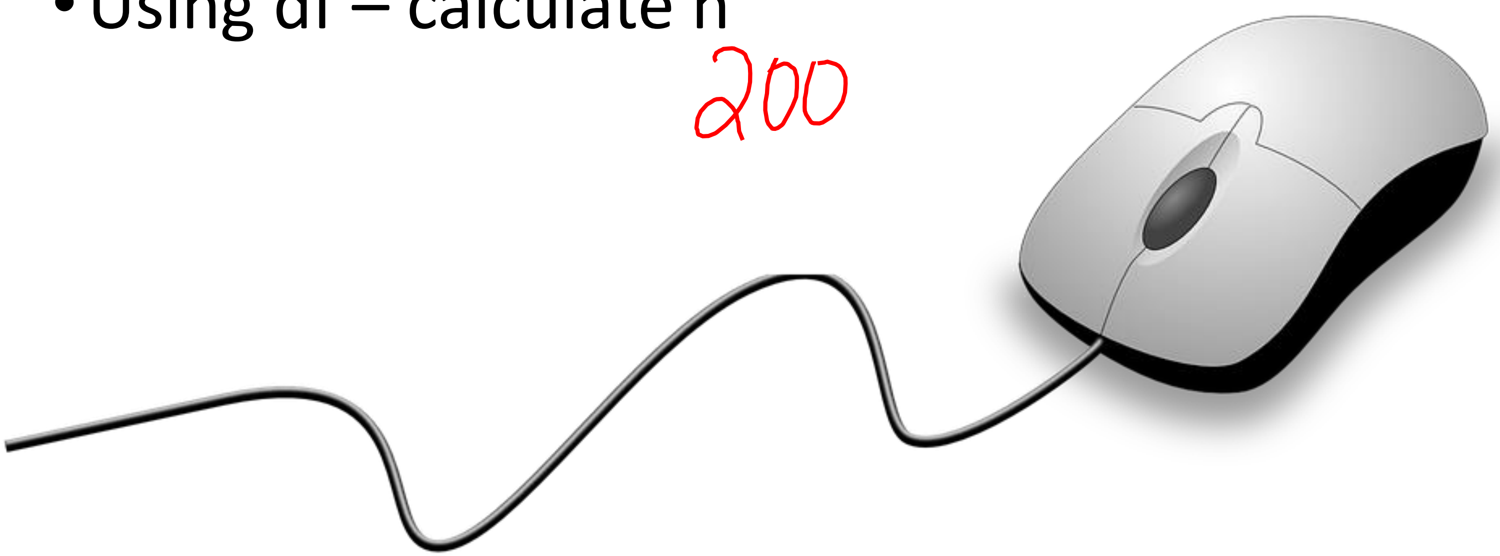
- Identify the:
  - Confidence interval
    (1.453, 4.097)
  - Mean
    2.775

One-Sample Test

	Test Value = 50					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
writing score	4.140	199	.000	2.7750	1.4533	4.0967

1.984

- Using the mean and CI, calculate the Margin of Error
  4.097 − 2.775 = 1.3217   2.775 − 1.453
- Using ME and t*, calculate the Standard Error   ME = t* SE
  $\frac{1.3217}{1.984} = \frac{1.984 \cdot SE}{1.984}$ → SE = 0.666 $SE = \frac{SD}{\sqrt{n}}$
- Using df – calculate n
  200



Objectives

- Content Objective: I will use the t-distribution to compare means of different samples.
- Social Objective: I will listen and not cause distractions for myself or others.
- Language Objective: I will take clear notes that I can understand when I refer to them later.


QUIZ

Question – how many shoes do you own?

Sometimes we want to compare two means…

Males -> 5, 6, 25, 7, 12, 2, 4, 3, 3, 3, 4, 15

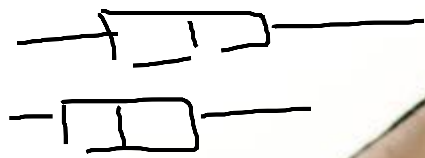
Females -> 7, 16, 16, 8, 21, 27, 22, 12, 8, 41, 7, 37, 22

$H_0: \mu_m = \mu_F$

$H_A: \mu_m < \mu_F$

m ->

F ->



2 sample t-test

2 sample CI
90%
(-18.73, -1.11)

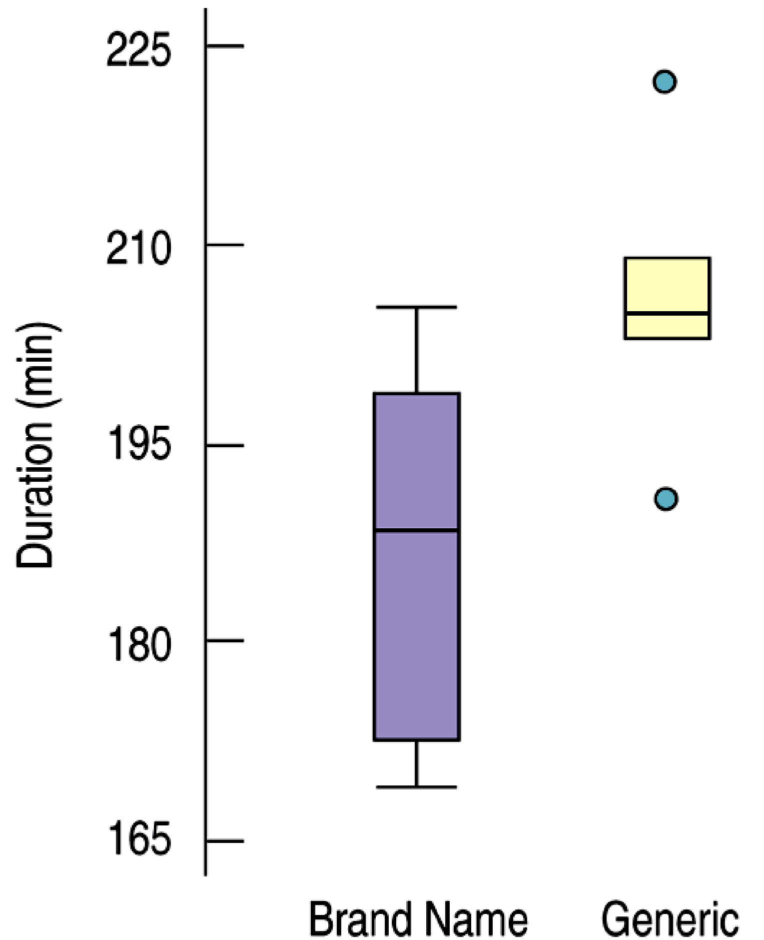
$t_{21.43} = -2.38$

p-value: 0.013

Due to a low p-value of 0.013, we reject the null. There is sufficient evidence that the average number of shoes for females is greater than males.

Comparing Two Means

- Once we have examined the side-by-side boxplots, we can turn to the comparison of two means.
- Comparing two means is not very different from comparing two proportions.
- This time the parameter of interest is the difference between the two means, $\mu_1 - \mu_2$.



Comparing Two Means

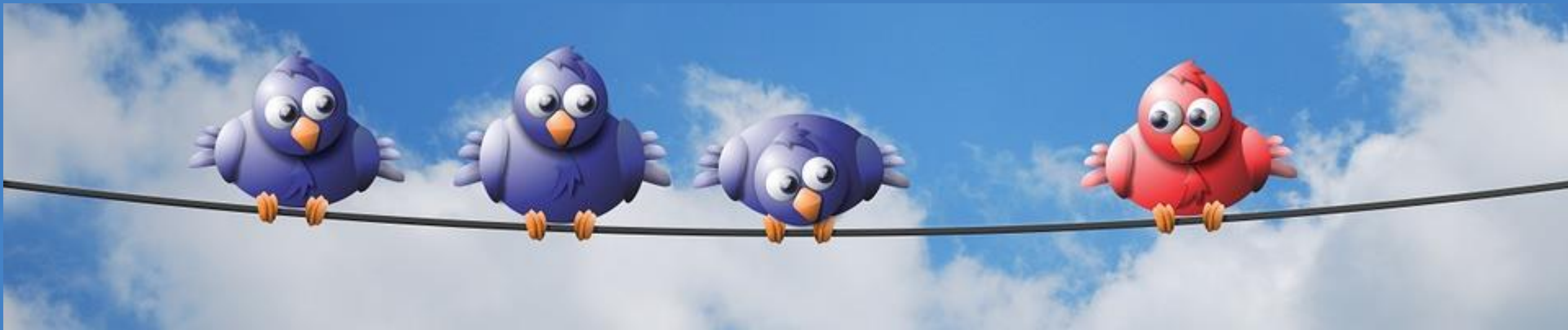
- Because we are working with means and estimating the standard error of their difference using the data, we shouldn't be surprised that the sampling model is a Student's *t*.
  - The confidence interval we build is called a **two-sample *t*-interval** (for the difference in means).
  - The corresponding hypothesis test is called a **two-sample *t*-test**.

Assumptions and Conditions

- Independence Assumption (Each condition needs to be checked for both groups.):
  - Randomization Condition: Were the data collected with suitable randomization (representative random samples or a randomized experiment)?
  - 10% Condition: We don't usually check this condition for differences of means. We will check it for means only if we have a very small population or an extremely large sample.

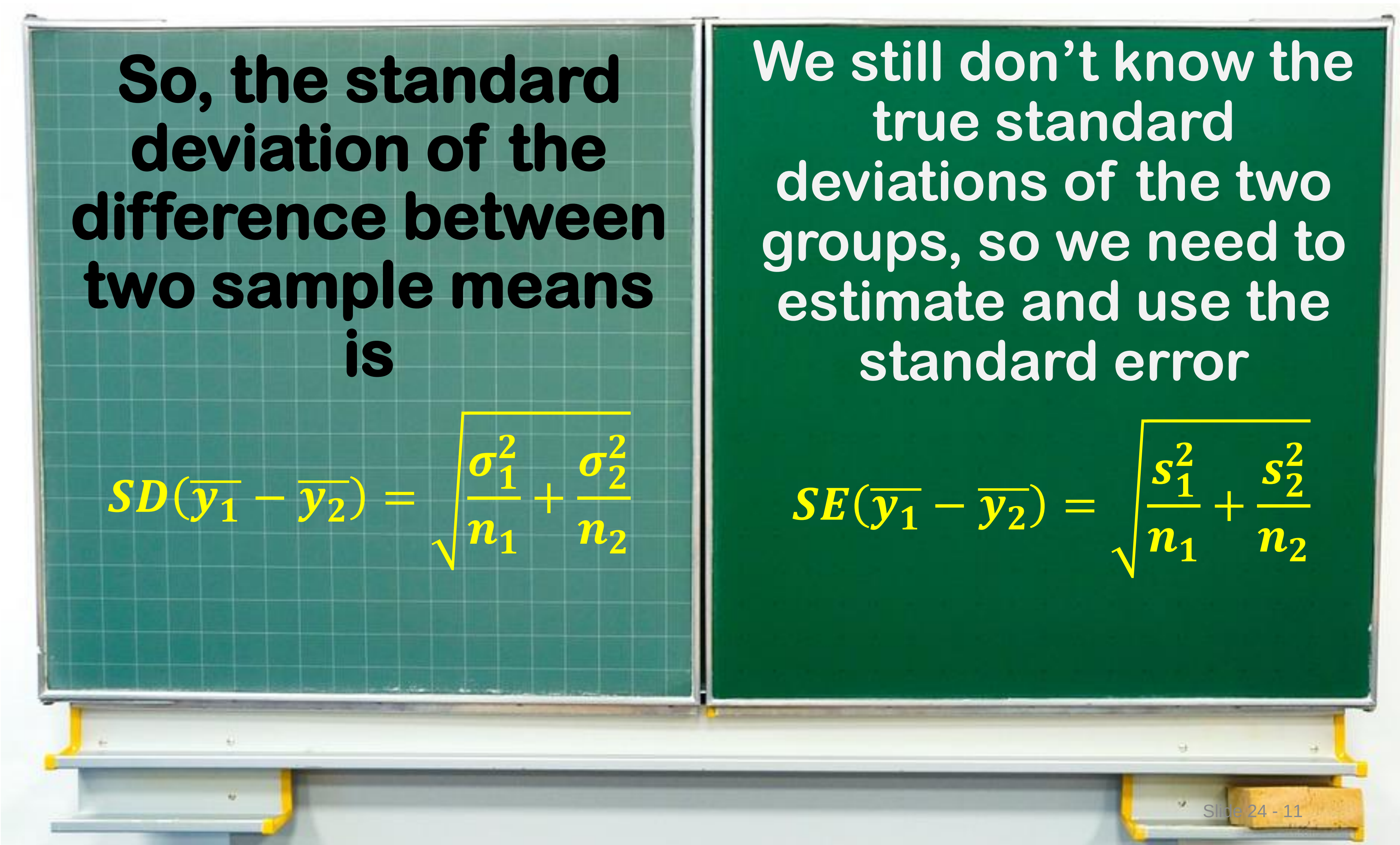
Assumptions and Conditions

- **Normal Population Assumption**:
  - **Nearly Normal Condition:** This must be checked for *both* groups. A violation by either one violates the condition.
- **Independent Groups Assumption**: The two groups we are comparing must be independent of each other.



Formulas

Remember that, for independent random quantities, variances add.



So, the standard deviation of the difference between two sample means is

$$SD(\overline{y_1} - \overline{y_2}) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

We still don't know the true standard deviations of the two groups, so we need to estimate and use the standard error

$$SE(\overline{y_1} - \overline{y_2}) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Two-Sample *t*-Interval

When the conditions are met, we are ready to find the confidence interval for the difference between means of two independent groups, $\mu_1 - \mu_2$.

The confidence interval is $\left(\bar{y}_1 - \bar{y}_2\right) \pm t^*_{df} \times SE\left(\bar{y}_1 - \bar{y}_2\right)$

where the standard error of the difference of the means is

$$SE\left(\bar{y}_1 - \bar{y}_2\right) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

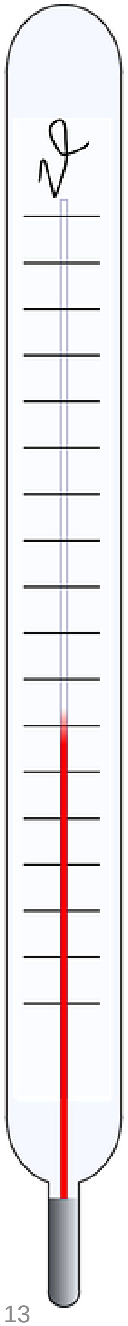
The critical value t^*_{df} depends on the particular confidence level, *C*, that you specify and on the number of degrees of freedom, which we get from the sample sizes and a special formula.

Degrees of Freedom

- **The special formula for the degrees of freedom for our *t* critical value is a bear:**

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$

- **Because of this, we will let technology calculate degrees of freedom for us!**




Homework
P 579
(3-6)