Stat 13, Intro. to Statistical Methods for the Life and Health Sciences.

1. Exams back and hw1 announcement.
2. Simulation approach with paired data and baseball example.
3. Theory based approach for paired data and M&M example.

There will be no lecture or OH Mon Mar11.
Read ch7 and 10.
Hw4 is due Mon Mar11 1159pm by email to statgrader or statgrader2.
10.1.8, 10.3.14, 10.3.21, and 10.4.11.

http://www.stat.ucla.edu/~frederic/13/W24 .

If you are in Section 2c or 2d, please resubmit your hw1 to statgrader2@stat.ucla.edu by this Mon, Mar4, 1159pm. Apparently there was some kind of email error previously.

For the final exam, bring a pencil or pen, and a calculator. On the final exam, you cannot use computers or ipads or phones or anything that can surf the web or do email.

2. Simulation based Approach for Analyzing Paired Data, and rounding first base example.

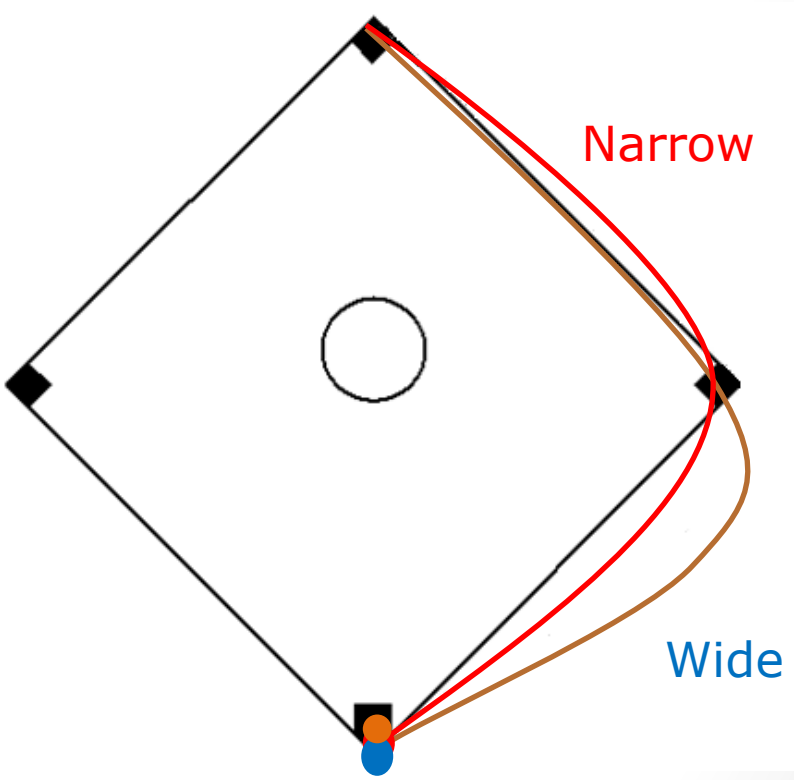
Section 7.2

Rounding First Base

Example 7.2

Rounding First Base

- Imagine you've hit a line drive and are trying to reach second base.
- Does the path that you take to round first base make much of a difference?
  - Narrow angle
  - Wide angle



Rounding First Base

- Woodward (1970) investigated these base running strategies.
- He timed 22 different runners from a spot 35 feet past home to a spot 15 feet before second.
- Each runner used each strategy (paired design), with a rest in between.
- He used random assignment to decide which path each runner should do first.
- **This paired design controls for the runner-to-runner variability.**

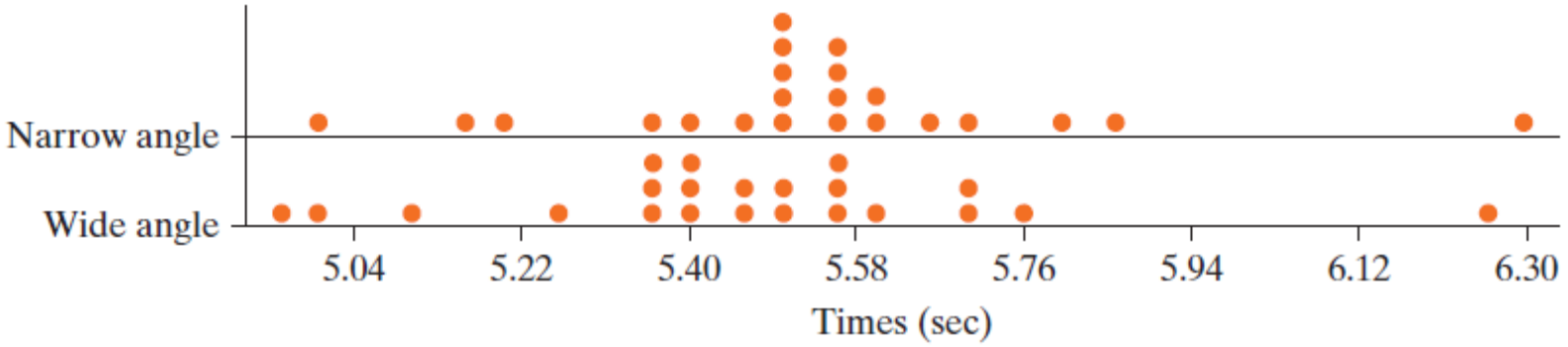
First Base

- What are the observational units in this study?
  - The runners (22 total)
- What variables are recorded? What are their types and roles?
  - Explanatory variable: base running method: wide or narrow angle (categorical)
  - Response variable: time from home plate to second base (quantitative)
- Is this an observational study or an experiment?
  - Randomized experiment.

The results

TABLE 7.1 The running times (seconds) for the first 10 of the 22 subjects

Subject	1	2	3	4	5	6	7	8	9	10	
Narrow angle	5.50	5.70	5.60	5.50	5.85	5.55	5.40	5.50	5.15	5.80	...
Wide angle	5.55	5.75	5.50	5.40	5.70	5.60	5.35	5.35	5.00	5.70	...



The Statistics

- There is a lot of overlap in the distributions and substantial variability.

	Mean	SD
Narrow	5.534	0.260
Wide	5.459	0.273

- It is difficult to detect a difference between the methods when these is so much variation.

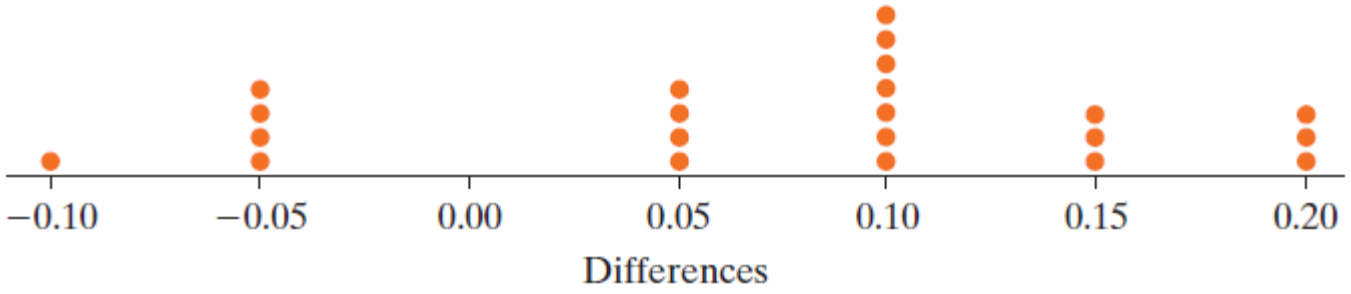
Rounding First Base

- However, these data are clearly paired.
- The paired response variable is time difference in running between the two methods and we can use this in analyzing the data.

The Differences in Times

TABLE 7.2 Last row is difference in times for each of the first 10 runners (narrow – wide)

Subject	1	2	3	4	5	6	7	8	9	10	
Narrow angle	5.50	5.70	5.60	5.50	5.85	5.55	5.40	5.50	5.15	5.80	…
Wide angle	5.55	5.75	5.50	5.40	5.70	5.60	5.35	5.35	5.00	5.70	…
Difference	−0.05	−0.05	0.10	0.10	0.15	−0.05	0.05	0.15	0.15	0.10	…

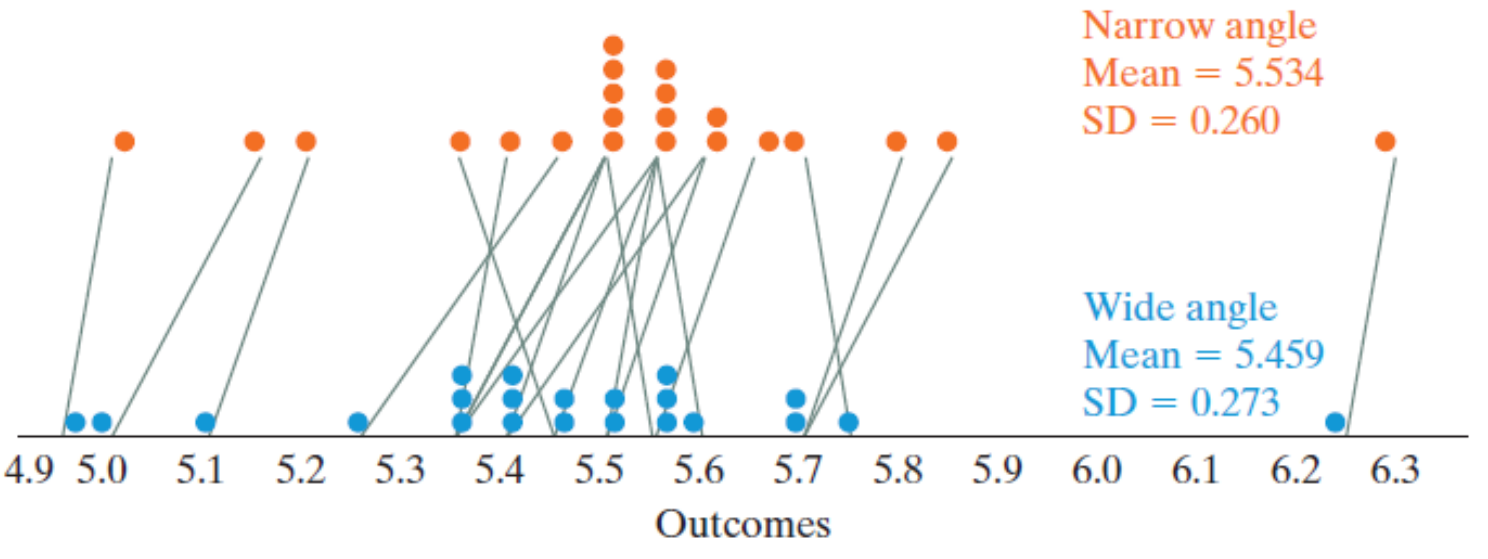


The Differences in Times

- Mean difference is $\bar{x}_d = 0.075$ seconds
- Standard deviation of the differences is $SD_d = 0.0883$ sec.
- This standard deviation of 0.0883 is smaller than the original standard deviations of the running times, which were 0.260 and 0.273.

Rounding First Base

- Below are the original dotplots with each observation paired between the base running strategies.
- What do you notice?



Rounding First Base

- Is the average difference of $\bar{x}_d = 0.075$ seconds significantly different from 0?
- The parameter of interest, μ_d, is the long run mean difference in running times for runners using the narrow angled path instead of the wide angled path. (narrow – wide)

Rounding First Base

The hypotheses:

- H_0: $\mu_d = 0$
  - The long run mean difference in running times is 0.
- H_a: $\mu_d \neq 0$
  - The long run mean difference in running times is not 0.

- The statistic $\bar{x}_d = 0.075$ is above zero.
- *How likely is it to see an average difference in running times this big or bigger by chance alone, even if the base running strategy has no genuine effect on the times?*

Rounding First Base

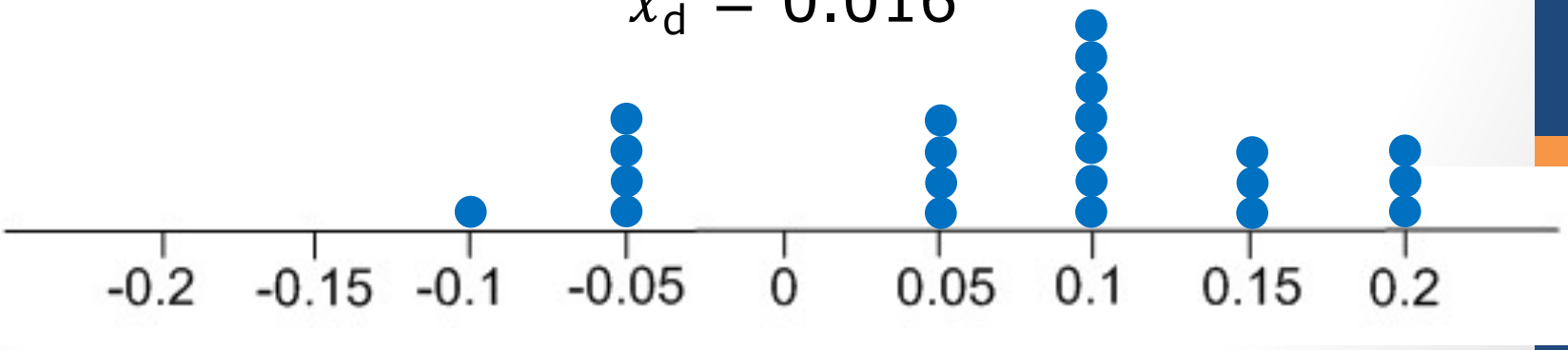
How can we use simulation-based methods to find an approximate p-value?

- The null hypothesis says the running path does not matter.
- So we can use our same data set and, for each runner, randomly decide which time goes with the narrow path and which time goes with the wide path and then compute the difference. (Notice we do not break our pairs.)
- After we do this for each runner, we then compute a mean difference.
- We will then repeat this process many times to develop a null distribution.

Random Swapping

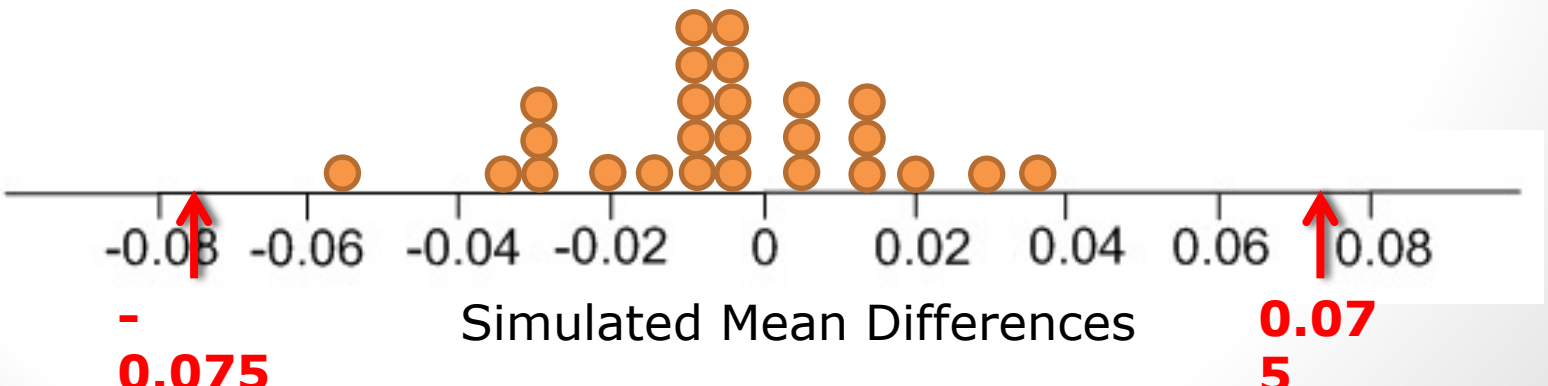
Subject	1	2	3	4	5	6	7	8	9	10	
narrow angle	5.50	5.70	5.60	5.50	5.85	5.55	5.40	5.50	5.15	5.80	…
wide angle	5.55	5.75	5.50	5.40	5.70	5.60	5.35	5.35	5.00	5.70	…
diff	0.05	-0.05	-0.10	0.10	0.15	0.05	0.05	0.15	0.15	-0.10	…

$\bar{x}_d = 0.016$



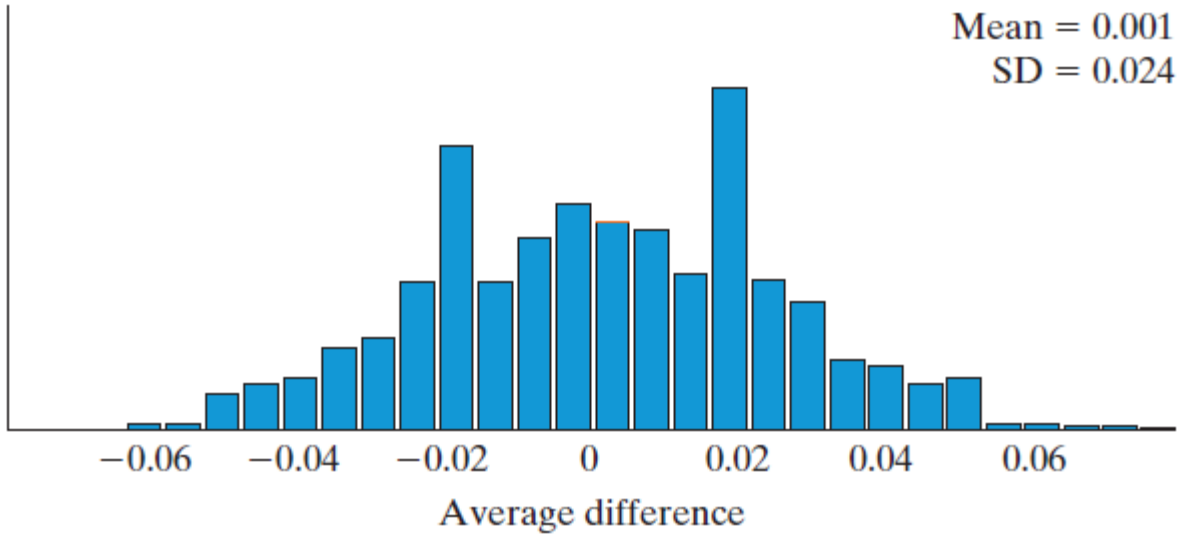
More Simulations

With 26 repetitions of creating simulated mean differences, we did not get any that were as extreme as 0.075.



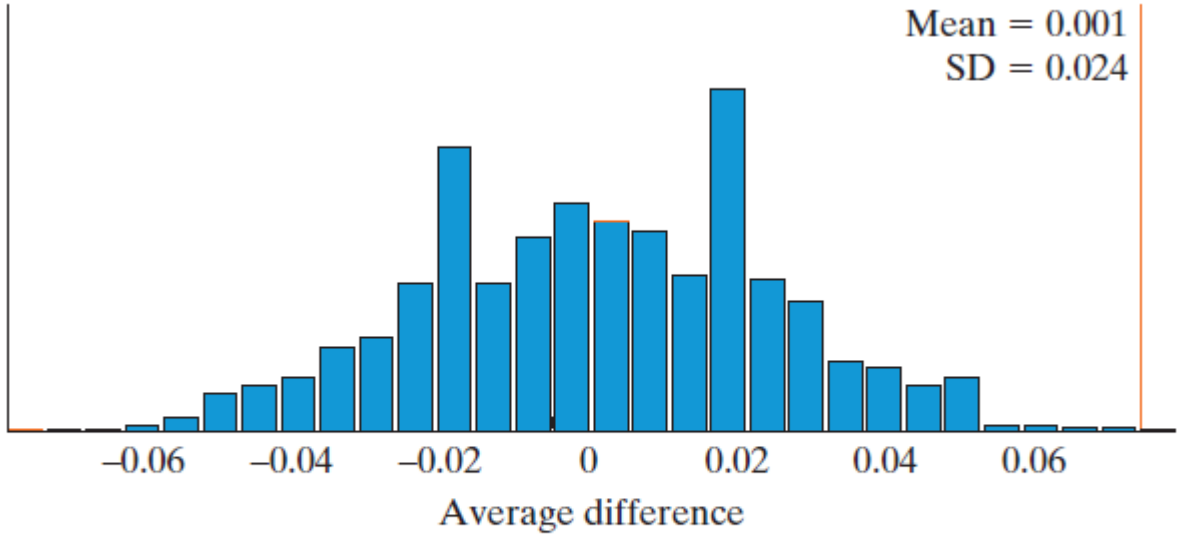
First Base

- Here is a null distribution of 1000 simulated mean differences.
- Notice it is centered at zero, which makes sense in agreement with the null hypothesis.
- Notice also the SD of these MEAN DIFFERENCES is 0.024. This is the SE.
- SD of time differences was 0.0883. SE = SD of mean time diff.s = .024.
- Where is our observed statistic of 0.075?



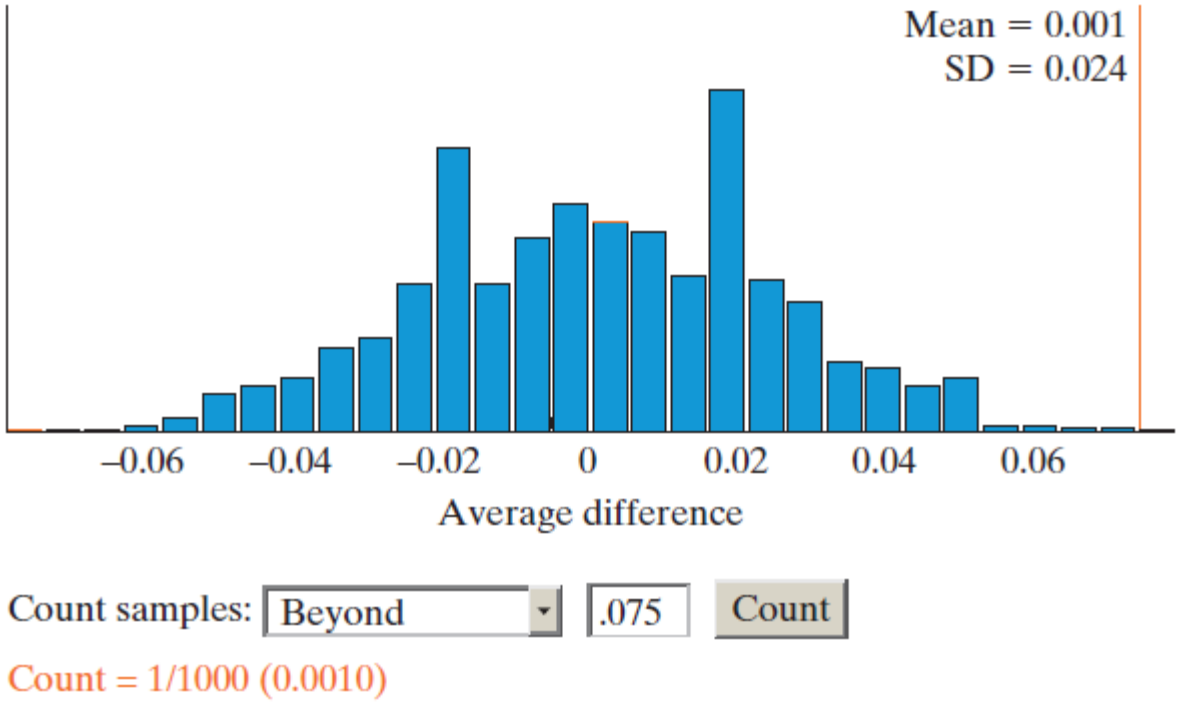
First Base

- Only 1 of the 1000 repetitions of random swappings gave a $\bar{x}_d$ value at least as extreme as 0.075.



First Base

- We can also standardize 0.075 by dividing by the SE of 0.024 to see our standardized statistic = $\frac{0.075}{0.024} = 3.125$.



Rounding First Base

- With a p-value of 0.1%, we have very strong evidence against the null hypothesis. The running path makes a statistically significant difference with the wide-angle path being faster on average.
- We can draw a cause-and-effect conclusion since the researcher used random assignment of the two base running methods for each runner.
- There was not much information about how these 22 runners were selected though so it is unclear if we can generalize to a larger population.

3S Strategy

- **Statistic:** Compute the statistic in the sample. In this case, the statistic we looked at was the observed mean difference in running times.
- **Simulate:** Identify a chance model that reflects the null hypothesis. We tossed a coin for each runner, and if it landed heads we swapped the two running times for that runner. If the coin landed tails, we did not swap the times. We then computed the mean difference for the 22 runners and repeated this process many times.
- **Strength of evidence:** We found that only 1 out of 1000 of our simulated mean differences was at least as extreme as the observed difference of 0.075 seconds.

First Base

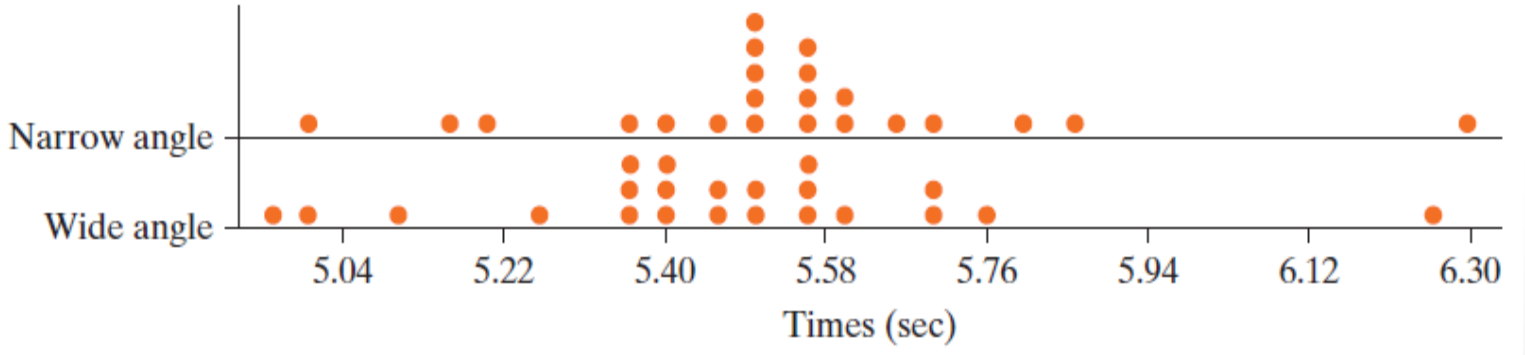
- Approximate a 95% confidence interval for μ_d:
  - 0.075 ± 1.96(0.024) seconds.
  - (0.028, 0.122) seconds.
- What does this mean?
  - We are 95% confident that, if we were to keep testing this indefinitely, the narrow angle route would take somewhere between 0.028 to 0.122 seconds longer on average than the wide angle route.

Since n = 22 here, the sample size is pretty small and the multiplier of 1.96 is not quite correct. If we assume the population of differences is normal, we should use a t multiplier, which here would be 2.08, so the 95% CI would be (.025, .125).

First Base

Alternative Analysis

- What do you think would happen if we wrongly analyzed the data using a 2 independent samples procedure? (i.e. The researcher selected 22 runners to use the wide method and an independent sample of 22 other runners to use the narrow method, obtaining the same 44 times as in the actual study.

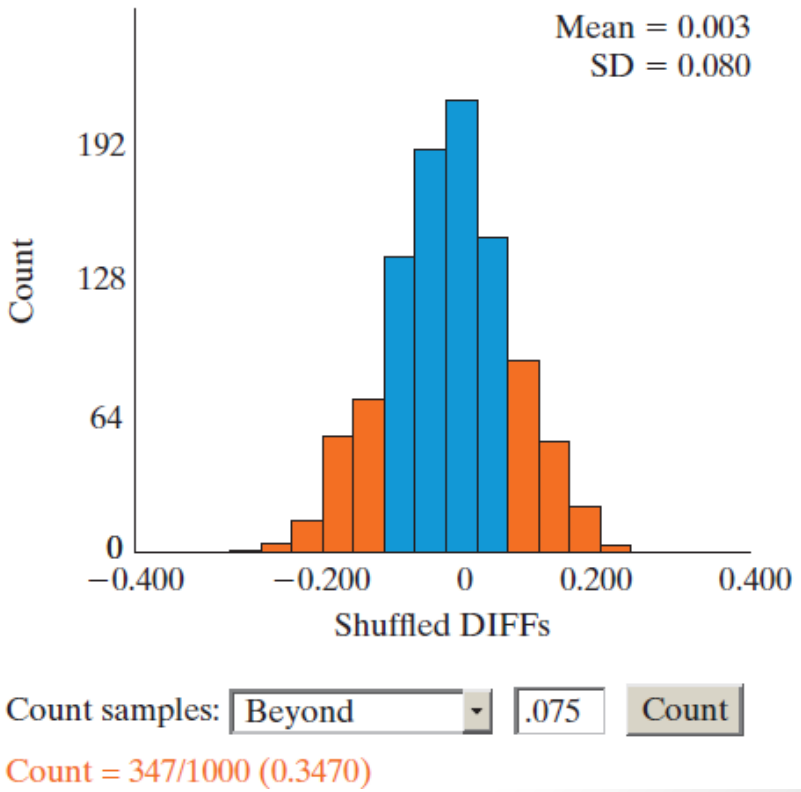


First Base

Ignoring the fact that it is paired data,

we get a p-value of 0.3470.

Does it make sense that this p-value is larger than the one we obtained earlier?



3. Theory based approach for Analyzing Data from Paired Samples, and M&Ms.

Section 7.3

How Many M&Ms Would You Like?

Example 7.3

How Many M&Ms Would You Like?

- Does your bowl size affect how much you eat?
- Brian Wansink studied this question with college students over several days.
- At one session, the 17 participants were assigned to receive either a small bowl or a large bowl and were allowed to take as many M&Ms as they would like.
- At the following session, the bowl sizes were switched for each participant.

How Many M&Ms Would You Like?

- What are the observational units?
- What is the explanatory variable?
- What is the response variable?
- Is this an experiment or an observational study?
- Will the resulting data be paired?

How Many M&Ms Would You Like?

The hypotheses:

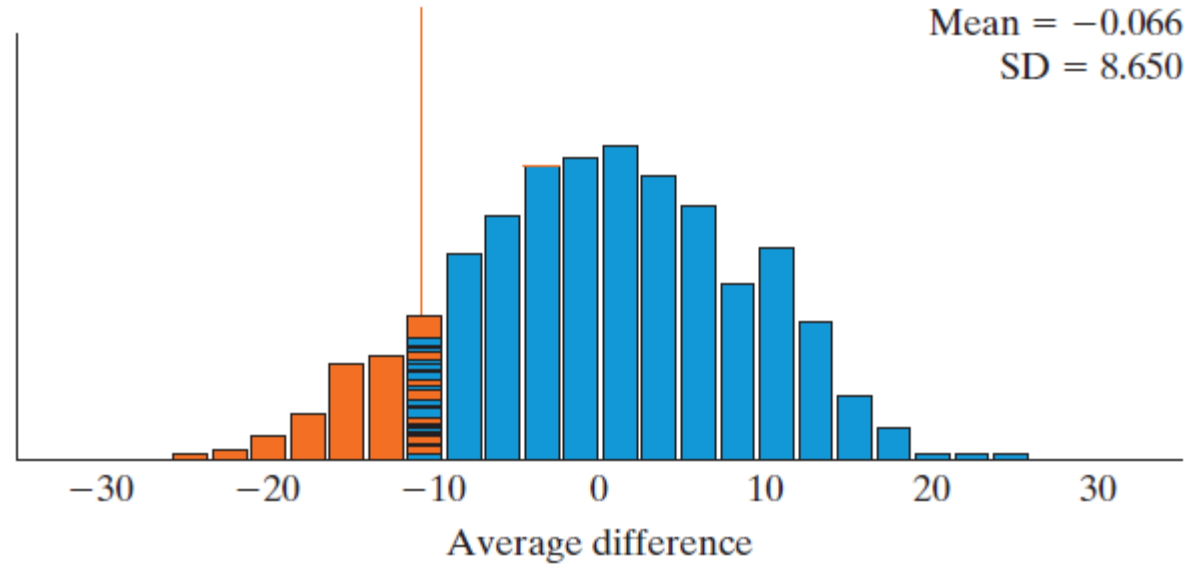
- H_0: $\mu_d = 0$
  - The long-run mean difference in number of M&Ms taken (small – large) is 0.
- H_a: $\mu_d < 0$
  - The long-run mean difference in number of M&Ms taken (small – large) is less than 0.

TABLE 7.5 Summary statistics, including the difference (small – large) in the number of M&Ms taken between the two bowl sizes

Bowl size	Sample size, *n*	Sample mean	Sample SD
Small	17	$\bar{x}_s = 38.59$	$s_s = 16.90$
Large	17	$\bar{x}_l = 49.47$	$s_l = 27.21$
Difference = small − large	17	$\bar{x}_d = -10.88$	$s_d = 36.30$

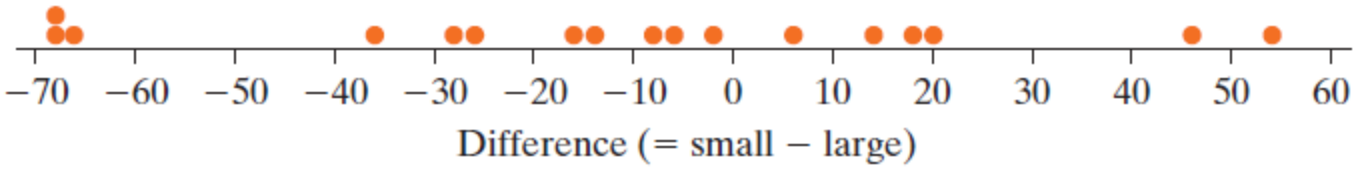
How Many M&Ms Would You Like?

- Here are the results of a simulation-based test.
- The p-value is quite large at 0.1220.



How Many M&Ms Would You Like?

- Our null distribution was centered at zero and fairly bell-shaped.
- Theory-based methods using the t distribution should be valid if σ is unknown and the population distribution of differences is normal (we can guess at this by looking at the sample distribution of differences). Alternatively, we can use the normal distribution if our sample size is at least 30.
- Our sample size was only 17, but this distribution of differences looks pretty normal, so we will proceed with a t-test.



Theory-based test

$$t = \frac{\bar{x}_d}{s_d/\sqrt{n}}$$

- This kind of test is called a paired *t*-test.

Theory-based results

Scenario: One mean

☐ Paste data

n: 17

mean, $\bar{x}$: −10.88

sample sd, s: 36.3

Calculate

☑ Confidence interval

confidence level 95 % Calculate CI

(−29.5435, 7.7835)

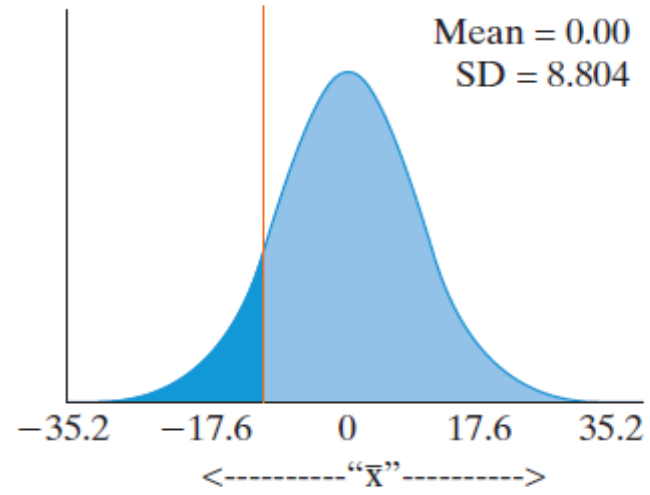
Theory-based inference

☑ Test of significance

H_0: μ = 0

H_a: μ < 0

Calculate



Standardized statistic t = −1.24 df = 16

p-value 0.1172

Conclusion

- The theory-based test gives slightly different results than simulation, 11.7% instead of 12.2% for the p-value, but we come to the same conclusion. We do not have strong evidence that the bowl size affects the number of M&Ms taken.
- We can see this in the large p-value (0.1172) and the confidence interval that included zero (-29.5, 7.8).
- The confidence interval tells us that we are 95% confident that when given a small bowl, people will take somewhere between 29.5 fewer M&Ms to 7.8 more M&Ms on average than when given a large bowl.

Why wasn't the difference statistically significant?

- There could be a number of reasons we didn’t get significant results.
  - Maybe bowl size doesn’t matter.
  - Maybe bowl size does matter and the difference was too small to detect with our small sample size.
  - Maybe bowl size does matter with some foods, like pasta or cereal, but not with a snack food like M&Ms.

Strength of Evidence

- We will have stronger evidence against the null (smaller p-value) when:
  - The sample size is increased.
  - The variability of the data is reduced.
  - The effect size, or mean difference, is farther from 0.
- We will get a narrower confidence interval when:
  - The sample size is increased.
  - The variability of the data is reduced.
  - The confidence level is decreased.