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

1. Simulation approach with paired data and baseball example, continued.
2. Theory based approach with paired data and bowl size example.
3. t versus normal and assumptions.
4. When to do what.

NO LECTURE THU NOV 3! Review for the midterm will be in class Nov 1.
Recall there is also no lecture or office hour Tue Nov 8.
Bring a PENCIL and CALCULATOR and any books or notes you want
to the midterm and final.
HW3 is due Tue Nov 1. 4.CE.10, 5.3.28, 6.1.17, and 6.3.14.
In 5.3.28d, use the theory-based formula. You do not need to use an applet.
The midterm Thu Nov 10 will be on ch1-7.
http://www.stat.ucla.edu/~frederic/13/F16 .

1. Paired data and rounding first base example.

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

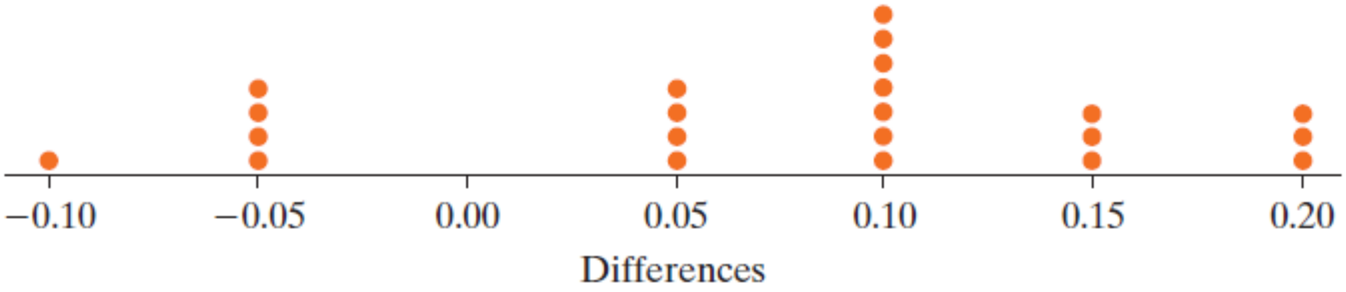
	Mean	SD
Narrow	5.534	0.260
Wide	5.459	0.273

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

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	...

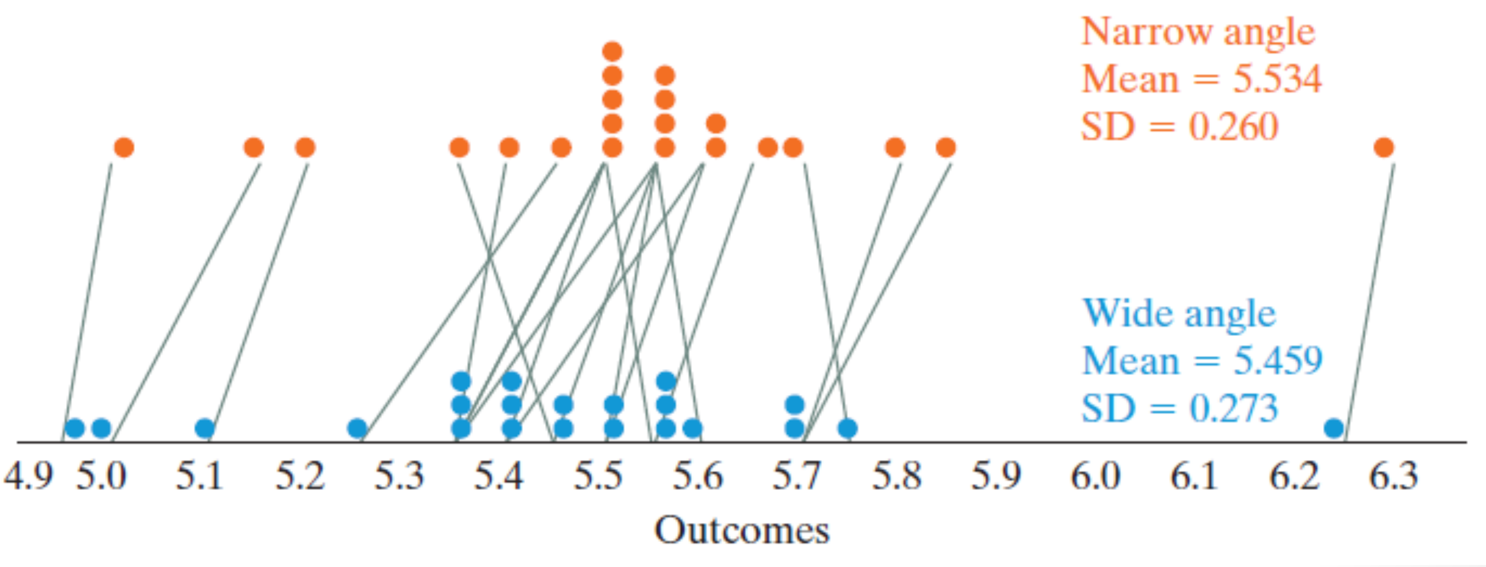


The Differences in Times

- 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

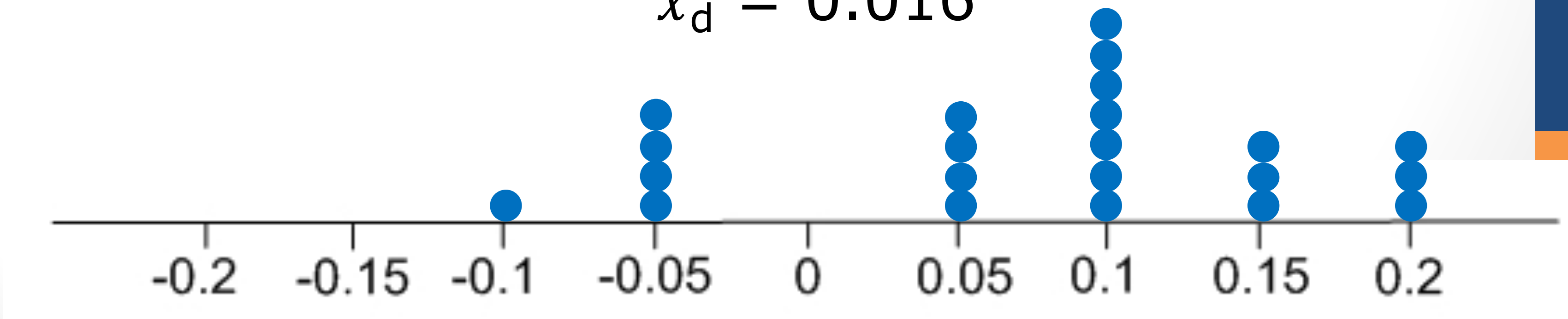
How can we use simulation-based methods to find an approximate p-value?

- 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	…

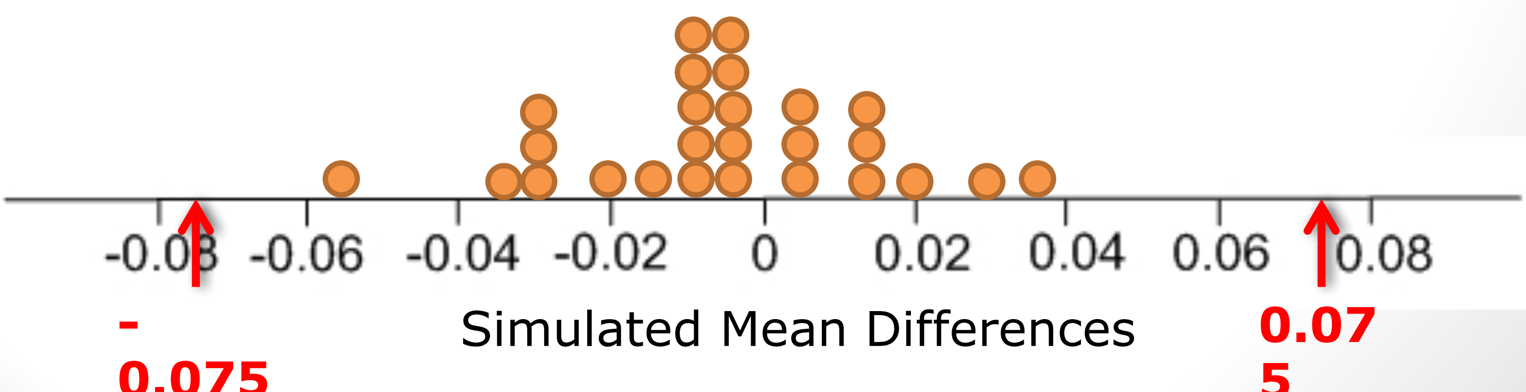
$\bar{x}_{\text{d}} = 0.016$



More Simulations

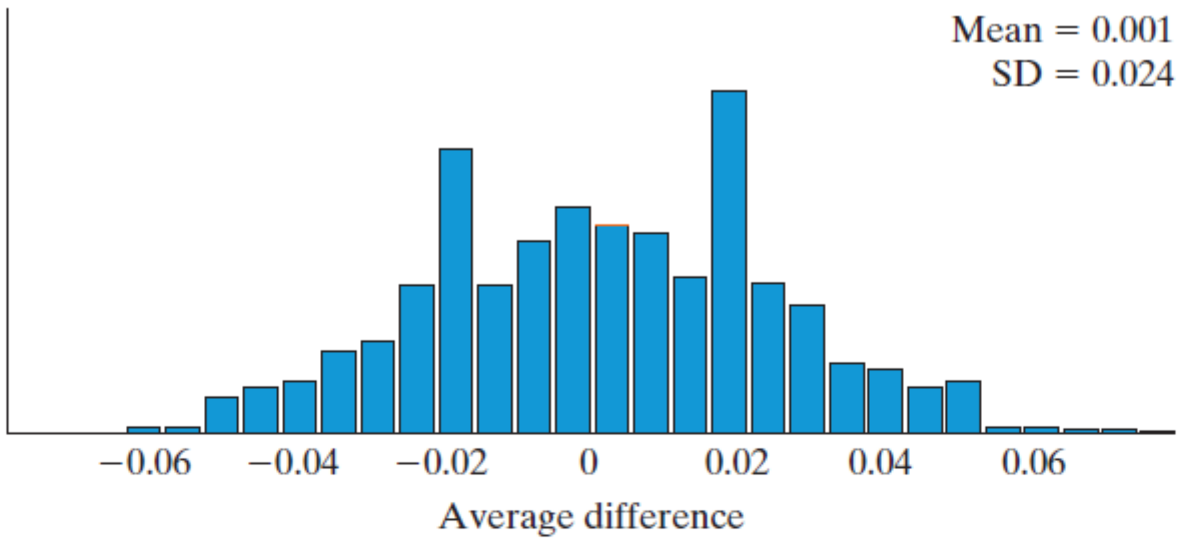
0.002 -0.002 0.030 -0.011 -0.007
-0.002 -0.016 0.016 -0.007
-0.067 0.002 0.020 -0.007 -0.002
0.007 -0.030 -0.034 -0.016 0.067 0.002
0.020
-0.002 -0.002 -0.025 0.056

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



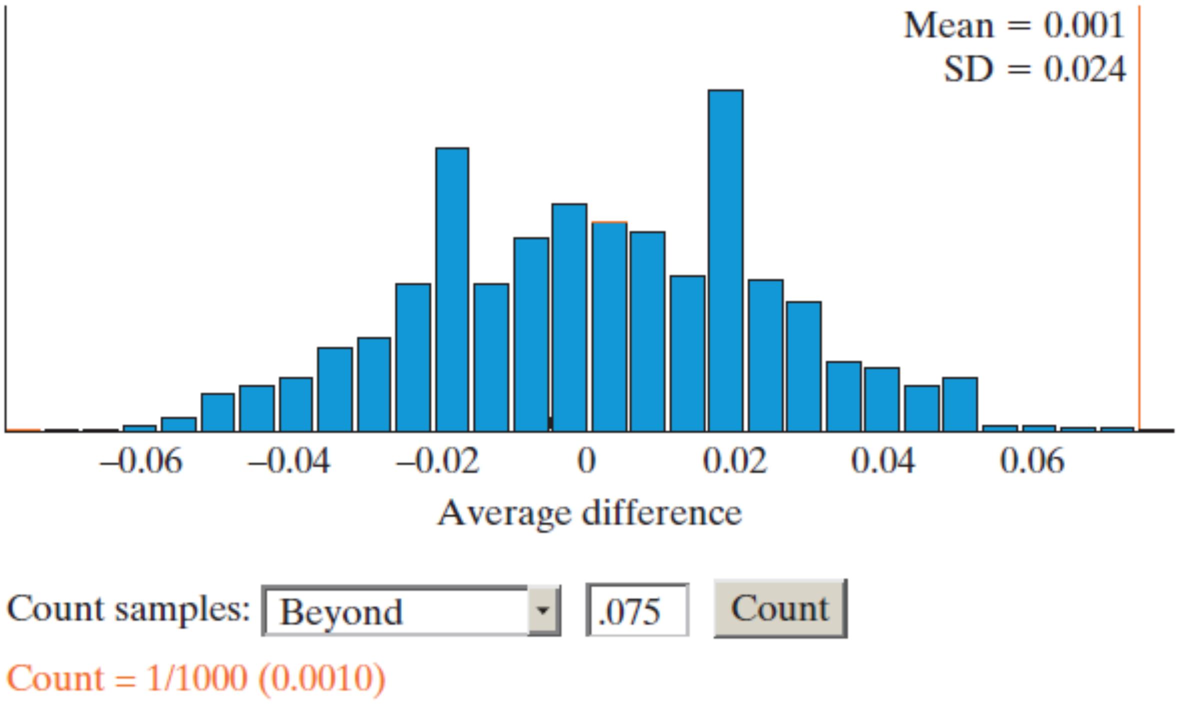
First Base

- 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 = SE.
  SD of time differences was 0.0883. SD of mean time diff.s = .024.
- Where is our observed statistic of 0.075?



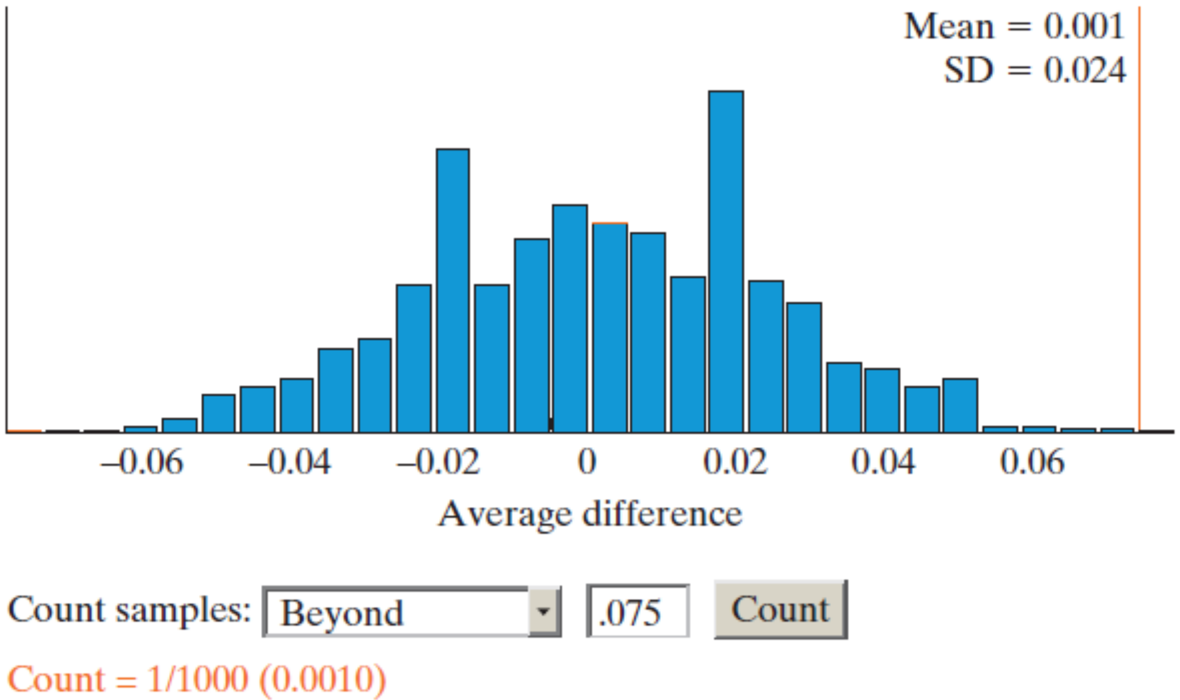
First Base

- 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.

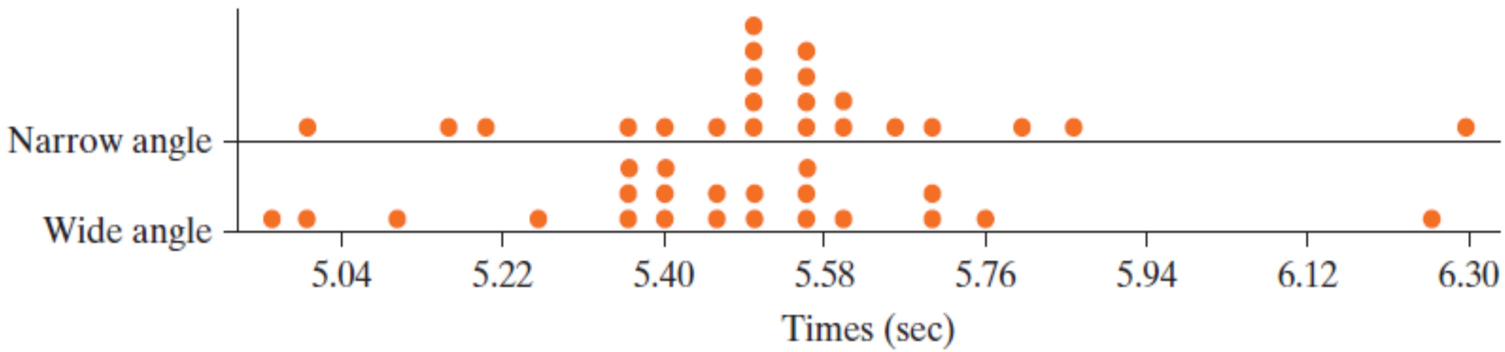
First Base

- 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.

First Base

Alternative Analysis

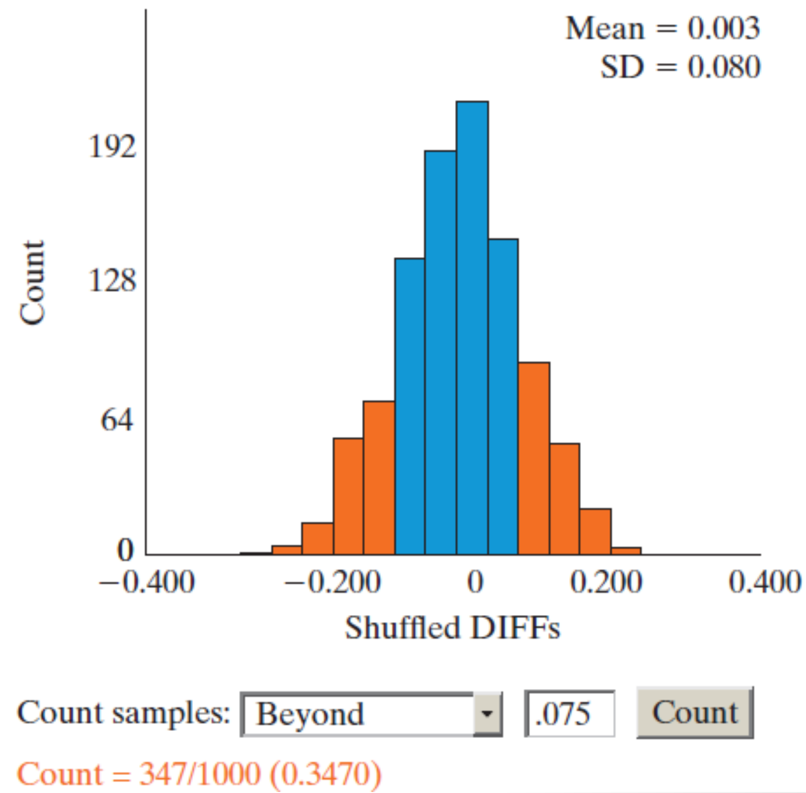
- What do you think would happen if we wrongly analyzed the data using a 2 independent samples procedure? (i.e. The researcher selected 22 runners to use the wide method and an independent sample of 22 other runners to use the narrow method, obtaining the same 44 times as in the actual study.



First Base

Using an applet which tests a difference between these two means, 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?



2. 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:

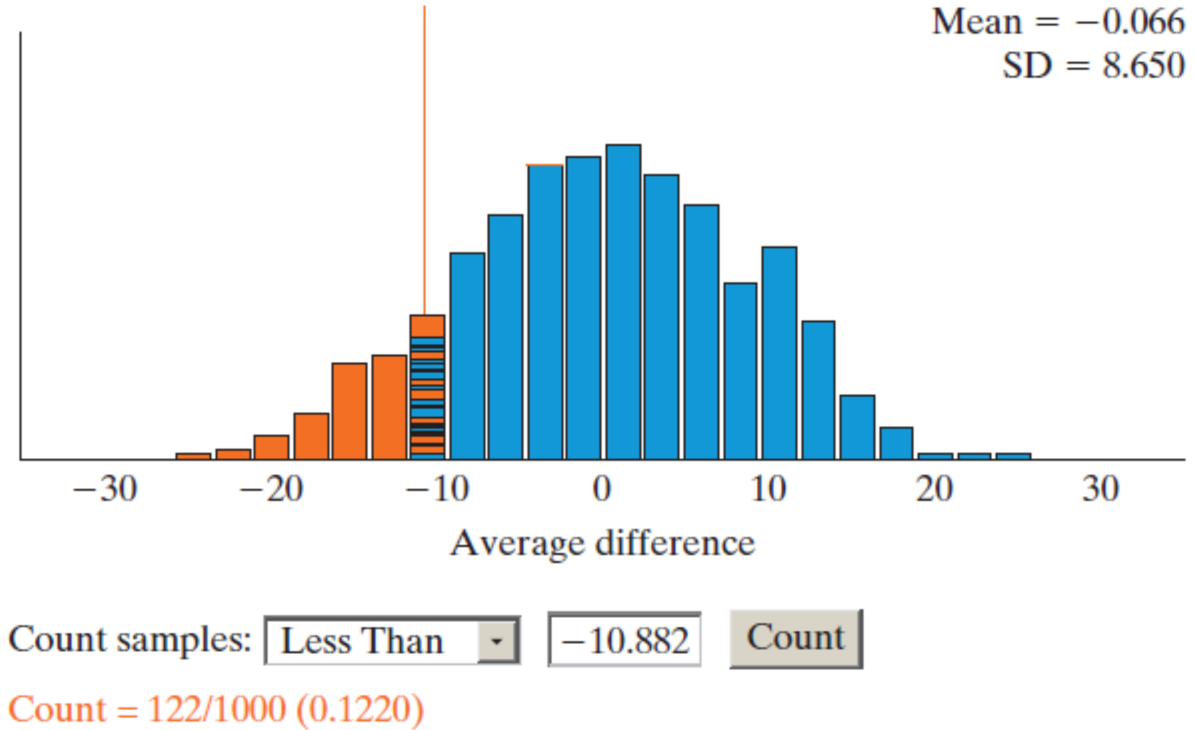
- H_0: $\mu_d = 0$
  - The long-run mean difference in number of M&Ms taken (small – large) is 0.
- H_a: $\mu_d < 0$
  - The long-run mean difference in number of M&Ms taken (small – large) is less than 0.

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	$\bar{x}_d = -10.88$	$s_d = 36.30$

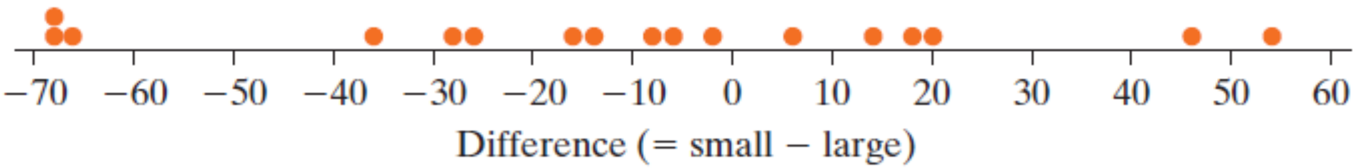
How Many M&Ms Would You Like?

- 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.
- This can all be predicted (along with the variability) using theory-based methods.
- Theory-based methods should be valid if the population distribution of differences is symmetric (we can guess at this by looking at the sample distribution of differences) or our sample size is at least 20.
- Our sample size was only 17, but this distribution of differences is fairly symmetric, so we will proceed with a theory-based test.



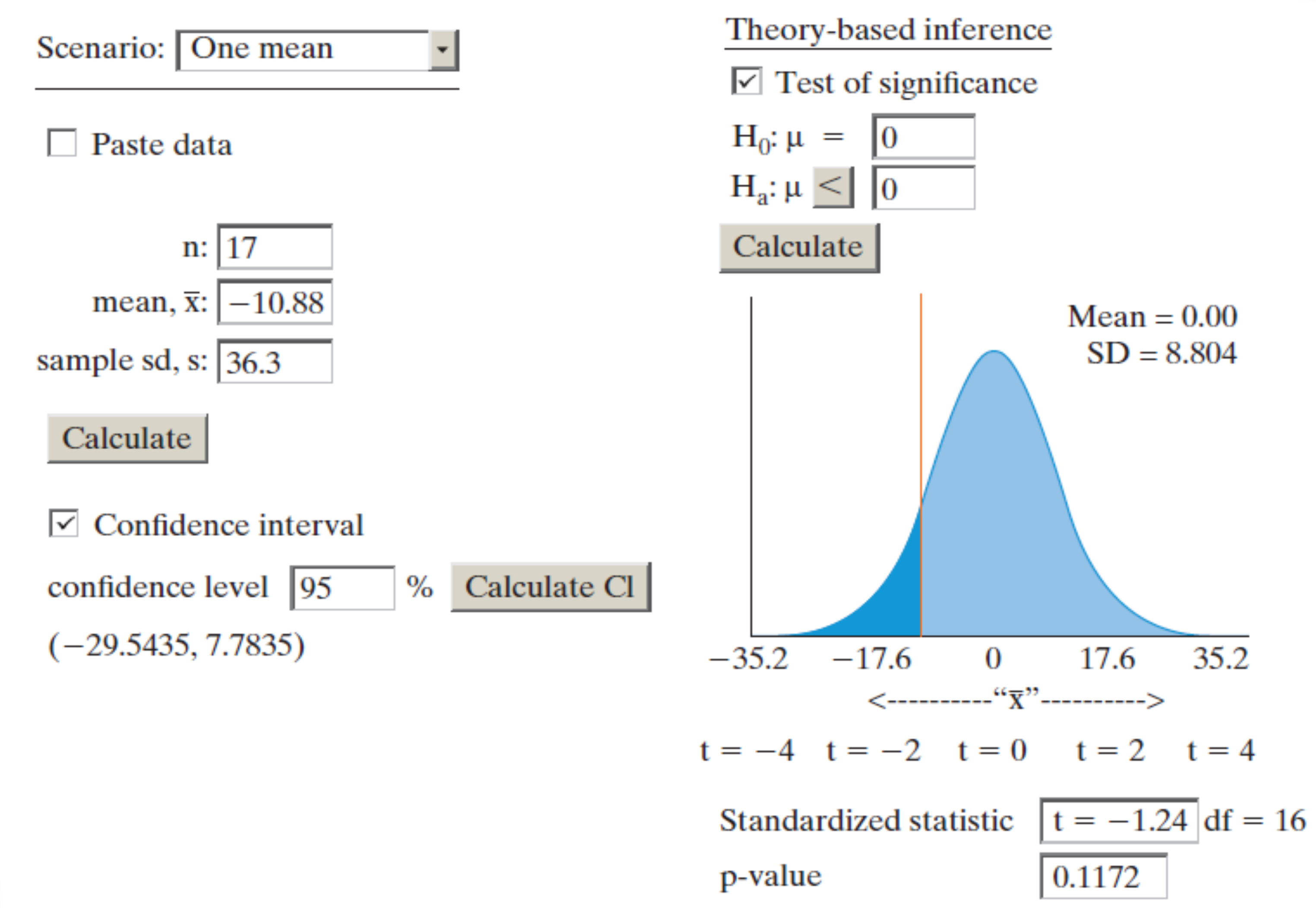
Theory-based test

- We can do theory-based methods with the applet we used last time or the theory-based applet.
- With the applet we used last time, we need to calculate the t-statistic:

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

- With the theory-based applet, we just need to enter the summary statistics and use a **test for a one mean.**
- This kind of test is called a paired *t*-test.

Theory-based results



Conclusion

- The theory-based model gives slightly different results than simulation, 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.
  - Other ideas?

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.

3. t versus normal and assumptions.

Why do we sometimes use the t distribution and sometimes the normal distribution in testing and confidence intervals?

The central limit theorem states that, for any iid random variables $X_1, \ldots, X_n$ with mean μ and SD σ, $(\bar{x} - \mu) \div (\sigma/\sqrt{n}) \to$ standard normal, as $n \to \infty$.

iid means independent and identically distributed, like draws from the same large population.

standard means mean 0 and SD 1.

3. t versus normal and assumptions.

CLT: $(\bar{x}$ - μ) ÷ (σ/√n) -> standard normal.

If Z is std. normal, then P(|Z| < 1.96) = 95%.

So, if n is large, then

P(|$(\bar{x}$ - μ) ÷ (σ/√n)| < 1.96) ~ 95%.

Mult. by (σ/√n) and get

P(|$\bar{x}$ - μ| < 1.96 σ/√n) ~ 95%.

P(μ $-$ $\bar{x}$ is in the range 0 +/- 1.96 σ/√n) ~ 95%.

P(μ is in the range $\bar{x}$ +/- 1.96 σ/√n) ~ 95%.

This all assumes n is large. What if n is small?

3. t versus normal and assumptions.

CLT: ($\bar{x}$ - μ) ÷ (σ/√n) -> standard normal.

What about if n is small?

A property of the normal distribution is that the sum of independent normals is also normal, and from this it follows that if X_1, ..., X_n are iid and normal, then ($\bar{x}$ - μ) ÷ (σ/√n) is standard normal.

So again P(μ is in the range $\bar{x}$ +/- 1.96 σ/√n) = 95%.

This assumes you know σ. What if σ is unknown?

3. t versus normal and assumptions.

Suppose $X_1, \ldots, X_n$ are iid with mean μ and SD σ.

CLT: $(\bar{x} - \mu) \div (\sigma/\sqrt{n})$ ~ std. normal.

If $X_1, \ldots, X_n$ are normal, then $(\bar{x} - \mu) \div (\sigma/\sqrt{n})$ is std. normal.

σ is the SD of the population from which $X_1, \ldots, X_n$ are drawn. s is the SD of the sample, $X_1, \ldots, X_n$.

Gosset (1908) showed that replacing σ with s,

if $X_1, \ldots, X_n$ are normal, then $(\bar{x} - \mu) \div (s/\sqrt{n})$ is t distributed.

So we need the multiplier from the t distribution.

3. t versus normal and assumptions.

To sum up,

if the observations are iid and n is large, then

$P(\mu \text{ is in the range } \bar{x} +/- 1.96\ \sigma/\sqrt{n}) \sim 95\%$.

If the observations are iid and normal, then

$P(\mu \text{ is in the range } \bar{x} +/- 1.96\ \sigma/\sqrt{n}) \sim 95\%$.

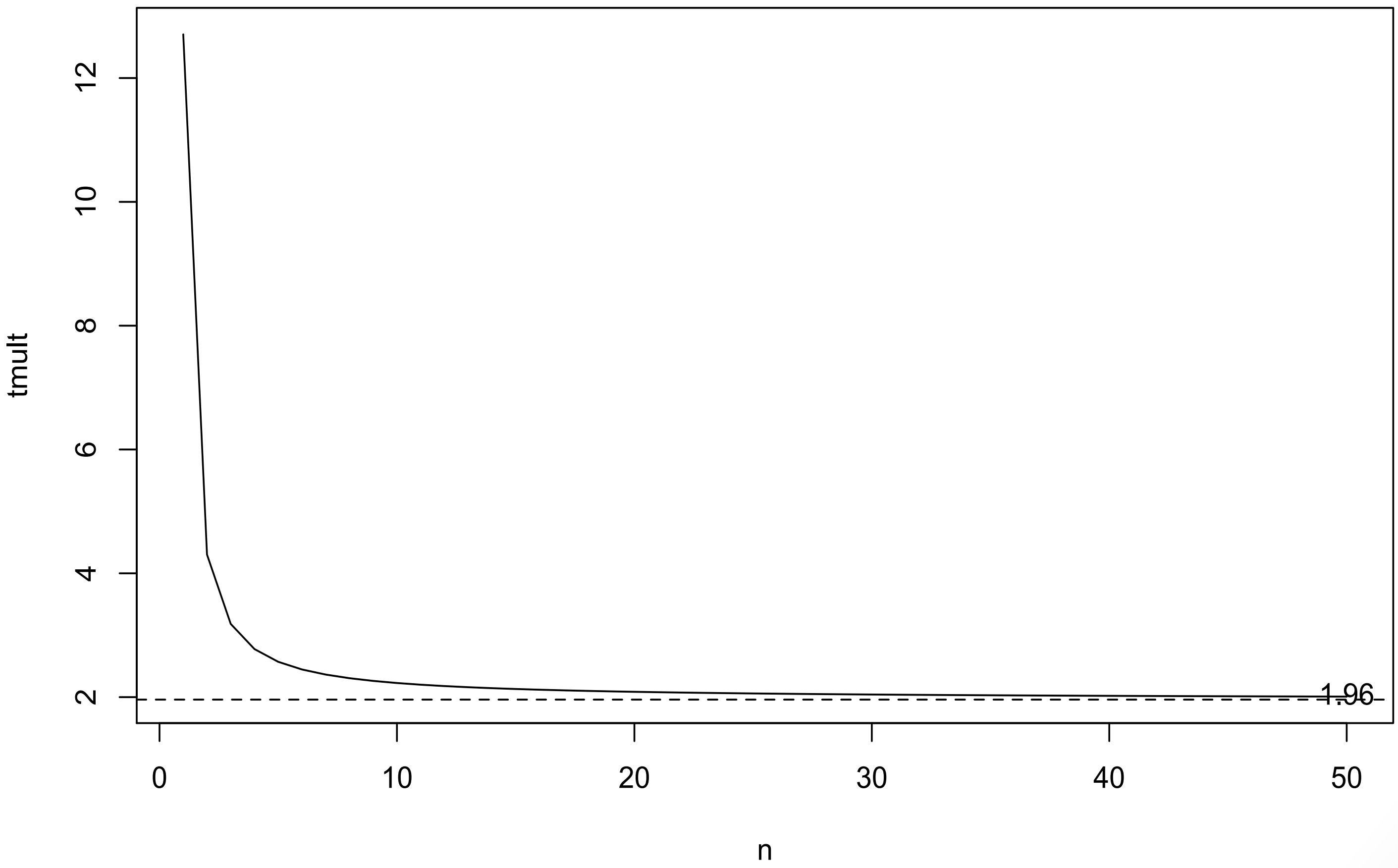
If the obs. are iid and normal and σ is unknown, then

$P(\mu \text{ is in the range } \bar{x} +/- t_{mult}\ s/\sqrt{n}) \sim 95\%$.

where t_{mult} is the multiplier from the t distribution.

This multiplier depends on n.

3. t versus normal and assumptions.



4. When to use which formula.

a. 1 sample numerical data, iid observations, want a 95% CI for μ.

- If n is large and σ is known, use $\bar{x}$ +/- 1.96 $\sigma/\sqrt{n}$.
- If n is small, draws are normal, and σ is known, use $\bar{x}$ +/- 1.96 $\sigma/\sqrt{n}$.
- If n is small, draws are normal, and σ is unknown, use $\bar{x}$ +/- t_{mult} $s/\sqrt{n}$.
- If n is large and σ is unknown, t_{mult} ~ 1.96, so we can use $\bar{x}$ +/- 1.96 $s/\sqrt{n}$.

n ≥ 30 is often considered large enough to use 1.96.

In practice, we typically do not know the draws are normal, but if the distribution looks roughly symmetrical without enormous outliers, the t formula may be reasonable.

b. 1 sample binary data, iid observations, want a 95% CI for π.

View the data as 0 or 1, so sample percentage p = $\bar{x}$, and

$s = \sqrt{[p(1-p)]}$, $\sigma = \sqrt{[\pi(1-\pi)]}$.

4. When to use which formula.

a. 1 sample numerical data, iid observations, want a 95% CI for μ.

- If n is large and σ is known, use $\bar{x}$ +/- 1.96 $\sigma/\sqrt{n}$.
- If n is small, draws are normal, and σ is known, use $\bar{x}$ +/- 1.96 $\sigma/\sqrt{n}$.
- If n is small, draws ~ normal, and σ is unknown, use $\bar{x}$ +/- t_{mult} $s/\sqrt{n}$.
- If n is large and σ is unknown, t_{mult} ~ 1.96, so we can use $\bar{x}$ +/- 1.96 $s/\sqrt{n}$.

b. 1 sample binary data, iid observations, want a 95% CI for π.

View the data as 0 or 1, so sample percentage p = $\bar{x}$, and

$s = \sqrt{[p(1-p)]}$, $\sigma = \sqrt{[\pi(1-\pi)]}$.

If n is large and π is unknown, use $\bar{x}$ +/- 1.96 $s/\sqrt{n}$.

Here large n means ≥ 10 of each type in the sample.

4. When to use which formula.

What if n is small and the draws are not normal?

That is a situation outside the scope of this course, but some techniques have been developed, such as the bootstrap, which are sometimes useful in these situations.

4. When to use which formula.

c. Numerical data from 2 samples, iid observations, want a 95% CI for $\mu_1 - \mu_2$.

If n is large and σ is unknown, use $\bar{x}_1 - \bar{x}_2$ +/- 1.96 $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$.

As with one sample, if σ_1 is known, replace s_1 with σ_1, and the same for σ_2. And as with one sample, if σ_1 and σ_2 are unknown, the sample sizes are small, and the distributions are roughly normal, then use t_{mult} instead of 1.96. If the sample sizes are small, the distributions are normal, and σ_1 and σ_2 are known, then use 1.96.

d. Binary data from 2 samples, iid observations, want a 95% CI for $\pi_1 - \pi_2$.

same as in c above, with $p_1 = \bar{x}_1$, $s_1 = \sqrt{[p_1 (1-p_1)]}$, $\sigma_1 = \sqrt{[\pi_1 (1-\pi_1)]}$.

Large for binary data means sample has ≥ 10 of each type.

4. When to use which formula.

e. Matched pairs data, iid observations, want a 95% CI for μ.

Look at differences (score with treatment minus score with control) and treat differences as ordinary numerical data according to parts a or b.

- If n is large and σ is known, use $\bar{x}$ +/- 1.96 $\sigma/\sqrt{n}$.
- If n is small, draws are normal, and σ is known, use $\bar{x}$ +/- 1.96 $\sigma/\sqrt{n}$.
- If n is small, draws are normal, and σ is unknown, use $\bar{x}$ +/- t_{mult} $s/\sqrt{n}$.
- If n is large and σ is unknown, t_{mult} ~ 1.96, so we can use $\bar{x}$ +/- 1.96 $s/\sqrt{n}$.

$n \geq 30$ is often considered large enough to use 1.96.

In practice, we typically do not know the draws are normal, but if the distribution looks roughly symmetrical without enormous outliers, the t formula may be reasonable.