Packet 11: Two Sample Inferences for the Mean

After completing this material, you should be able to:

- identify whether two samples are *independent* or *dependent*. unpaired paired
- conduct a hypothesis test for the mean difference (μ_d) when *dependent* samples are taken.
- calculate a confidence interval estimating the mean difference (μ_d) when *dependent* samples are taken.
- conduct a hypothesis test for the difference in the means (μ_1 - μ_2) when *independent* samples are taken.
- calculate confidence interval estimating the difference in the means (μ_1 - μ_2) when *independent* samples are taken.

To celebrate the Cubs winning the World Series … a baseball enthusiast wants to compare the mean batting average for the Chicago Cubs & the Cleveland Indians.

What variable will be recorded for the sampled players? Is the variable quantitative or categorical?

Batting Averages – quantitative

To make the comparison, he needs to gather 2016 batting averages from members of each team. This can be done using one of the following two scenarios:

Scenario 1: The individual randomly samples 5 players from the Cubs' roster and records each player's batting average. Then, he samples 5 players from the Indians' roster and records each player's batting average.

"Unpaired scenario" – risk is that we randomly get 5 pitchers from one team & 5 sluggers from the other. These two samples are unrelated, so it's possible... Independent of each other.

Scenario 2: The individual randomly samples 5 positions in the line-up from 9 possible and records the batting average for the player batting in that position.

"Paired" scenario – We're carefully controlling for ability to hit. This presumes the managers are using the same strategy for the lineup.

1
2
3
4
Best hitters

5
6
7
8
9
Worse hitters

In both sampling scenarios described, the sports enthusiast ends up with 5 batting averages for the Cubs and 5 batting averages for the Indians. But, the way in which these samples were taken were fundamentally different. One scenario employed the use of *dependent* (or paired) samples, while the other used *independent* samples. Let's define these two sampling techniques:

Dependent samples: Paired samples: so that obs. units in sample 1 match the obs. units in sample 2 in some important & relevant way (we "control" for a variable).

Independent samples: Unpaired samples: no attempt is made to pair or relate individual observational units across the two samples.

Let's look at additional examples and determine the type of sample selected:

Example 1: Three hundred registered voters were selected at random to participate in a study on attitudes about how well the president is performing his job. They were each asked to answer a short multiple-choice questionnaire and then they watched a 20-minute video that presented information about the job description of the president. After watching the video, the same 300 selected voters were asked to answer a follow-up multiple-choice questionnaire. The investigator of this study will have two sets of data: the initial questionnaire scores and the follow-up questionnaire scores. Is this a paired or independent samples design?

= unpaired

outcome - opinion about President's performance

Circle one: Dependent Independent

Explain: Video is the treatment; & we're considering the outcome on the same person from sample 1 to sample 2.

Example 2: Thirty dogs were selected at random from those residing at the humane society last month. The 30 dogs were split at random into two groups. The first group of 15 dogs was trained to perform a certain task using a reward method. The second group of 15 dogs was trained to perform the same task using a reward-punishment method. The investigator of this study will have two sets of data: the learning times for the dogs trained with the reward method and the learning times for the dogs trained with the reward-punishment method. Is this a paired or independent samples design?

Circle one: Dependent Independent

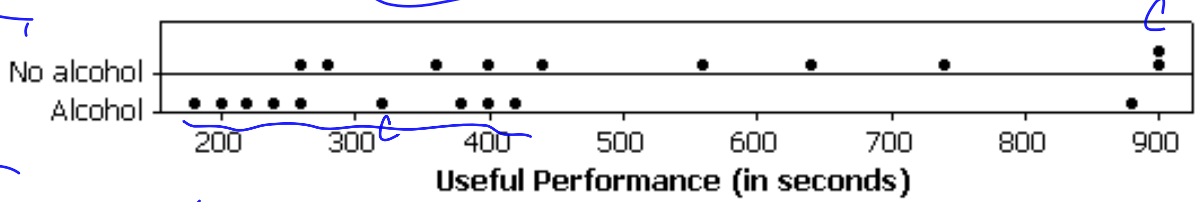
Explain: The two sets unpaired of dogs were unrelated - no attempt was made to pair a dog in one sample with a dog in the other.

Inferences for Paired (Dependent) Samples

Example: Ten pilots performed tasks at a simulated altitude of 25,000 feet. Each pilot performed the tasks in a completely sober condition and, three days later, after drinking alcohol. At the completion of each simulation, the administrator recorded the time (in seconds) of useful performance of the tasks for each condition. The longer the pilot spends on useful performance, the better. The researchers would like to know if their useful performance decreases with alcohol use.

Two dotplots of the data are given below:

We can't see who's who from the dot plot...



What type of samples were taken in this scenario? Explain. This is a paired design; we're comparing pilots in one sample to pilots in another sample - (but they're the same pilots!)

The data collected in the experiment is given below:

No alcohol	261	565	900	630	280	365	400	735	430	900
Alcohol	185	375	310	240	215	420	405	205	255	875
d = Alc - NAlc	-76	-190	-590	-390	-65	55	5	-530	-175	-25

(only have 10 data points - so we should be a little nervous moving forward)

We can summarize the differences by calculating their sample mean and sample standard deviation in StatCrunch:

$n_d = 10$ $\quad \bar{y} = -198.1 = \bar{d}$

$s_d = 228.794$

(mean difference $\bar{d}$ is the new focus of our attention)

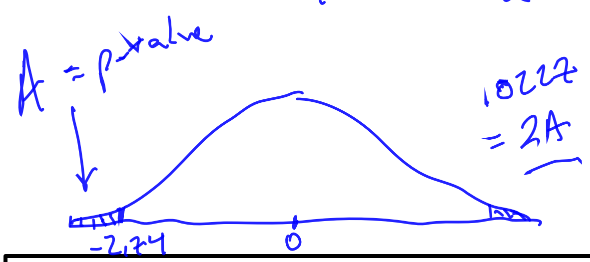
Based on the sample data, can we conclude that useful performance decreases with alcohol use? Use $\alpha = 0.05$. Note – the *two-sided* p-value for this test is 0.0227.

(alcohol has adversely affected performance)

① $H_0: \mu_d = 0$; $H_a: \mu_d < 0$

② $\alpha = .05$; reject the null in favor of the alternative $\mu_d < 0$ if p-value $< \alpha$.

③ T.S. $t_{df=n_d-1} = \dfrac{\bar{d} - \mu_d}{s_d/\sqrt{n_d}} = \dfrac{-198.1 - 0}{228.794/\sqrt{10}} \approx -2.74$ ← that looks big....



So A = p-value for this test $\approx .01135 < \alpha$

④ With $\alpha = .05$ we reject the null of no difference in favor of $\mu_d < 0$ & alcohol has a negative effect on performance.

Formulas & Assumptions for Dependent (Paired) Samples

When the samples are paired, then all inferences are about differences (d) between the pairs. Create variable $d = v_2 - v_1$ & then do the usual one sample stuff (hypothesis tests or confidence intervals): T.S. $t = \dfrac{\bar{d} - \mu_d}{s_d/\sqrt{n_d}}$

C.I.: $\bar{d} \pm t_{crit,\,df} \cdot \boxed{\dfrac{s_d}{\sqrt{n_d}}}$ ← called the "standard error" in Statcrunch & statistics.

Example: An article in the New York Times compared the prices of some common food items at the Whole Foods Market and at Fairway Supermarket in New York City. Prices were determined for the same ten items (Half-gallon milk, 64 oz. of orange juice, etc.) at each of the two stores. The data is available on StatCrunch.

Explain why the samples of prices are **dependent samples**. paired samples - same item compared in both stores, on price

Use the StatCrunch output below to estimate the mean difference in prices at the two stores with 90% confidence.

90% confidence interval results:

$\mu_1 - \mu_2$: mean of the **PAIRED** difference between Fairway and Whole Foods

Difference	Sample Diff.	Std. Err.	DF	Critical Pt.
Fairway - Whole Foods	-0.68	0.2076161	9	1.833

t_{crit}

We're going to build a 90% confidence interval for the mean difference μ_d, d = Fairway − WF

$$\bar{d} \pm t_{crit} \cdot \frac{s_d}{\sqrt{n_d}}$$

df = 9
90% conf.

$$-.68 \pm 1.833 \cdot SE$$
$$\pm 1.833 \cdot 0.2076$$

C.I. = [−1.06, −0.299] for the mean difference μ_d

So our conclusion is that μ_d is negative - means that WF is more expensive than Fairway. - with 90% confidence. (Small sample size makes me nervous!)

What happens if the limits of the interval have different signs (ie: a negative lower limit & a positive upper limit)?

Ours didn't; but if instead we'd seen [−1.06, 0.299], then 0 would be a possibility - 0 is in that interval, so it's possible that $\mu_d = 0$ - i.e., there's no difference in prices.

Example: A distributor of soft drinks knows from experience that the number of drinks purchased from a machine each day varies according to the location of the machine. At a school, two machines are placed in what the distributor believes to be two optimal locations. Both of the machines are observed for a random sample of 13 days, and the number of drinks sold each day is recorded. Using the output below, determine if there is a difference in the mean number of drinks sold at the two locations using a significance level of 0.10.

$\alpha = .10$ $n = 13$

Hypothesis test results:
$\mu_1 - \mu_2$: mean of the **PAIRED** difference between Location 1 and Location 2

Difference	Sample Diff.	Std. Err.	DF
Loc1 - Loc2	1.0769231	2.3134544	12

The area ***below*** the test statistic is 0.6751.

Data are daily sales numbers, matched to two vending machines.

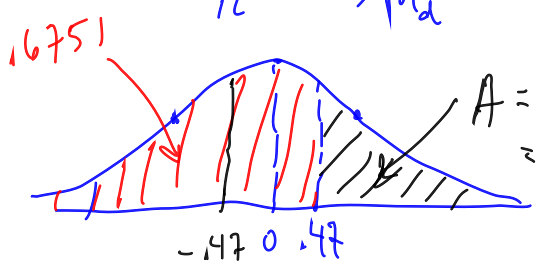
Same days for the two machines.

(1) $H_0: \mu_d = 0$; $H_a: \mu_d \neq 0$.

(2) $\alpha = .10$; reject H_0 in favor of H_a if p-value < α.

(3) T.S. $t_{df=12} = \dfrac{\bar{d} - \mu_d}{s_d / \sqrt{n_d}} = \dfrac{\bar{d} - \mu_d}{SE} = \dfrac{1.0769 - 0}{2.3135} = 0.47$

less than a standard deviation



$A = 1 - .6751$
$= .3249$

Now double it (two-sided H_a): p-value = .6498

(4) At the .10 significance level, p-value >> α, so we fail to reject a null of no difference in sales: $\mu_d = 0$.

Inferences for Independent Samples Unpaired Textbook pages: 617 - 629

Example: A travel guide is interested in comparing the mean price per person for a meal in NYC and the mean price per person for a meal in Long Island restaurants. In order to do this, a random sample of 50 restaurants in NYC is selected and a random sample of 50 restaurants in Long Island is taken – the prices at these 100 restaurants are available on StatCrunch. At the 0.10 level, is there sufficient evidence to conclude the mean price per person for a meal *differs* between the two locations? $\alpha = .10$

$\neq$ alt.

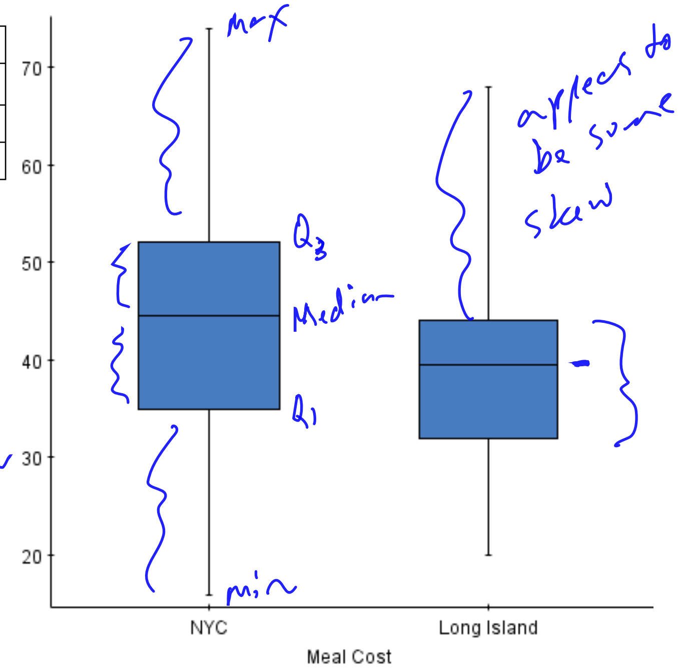
— Explain why the samples collected were *independent*. The obs. units – restaurants in the two locations – have nothing to do with each other. They are unpaired.

— Boxplots and summary statistics comparing the prices are shown below (an asterisk has been added to the boxplot denoting the mean). Compare and contrast the two distributions.

NYC	Long Island
$\bar{y} = 44.26$	$\bar{y} = 39.64$
$s = 12.8901$	$s = 10.3938$
$n = 50$	$n = 50$

close to the median symmetric (mainly)

also close to the median – but this distribution doesn't look as symmetric



Range is bigger for NYC; IQR bigger as well.

More variation in NYC. Appears that NYC has higher mean & median – is it significantly so?

When the data is collected from two independent samples, there is no way to reduce the two samples down to one sample – in other words, we can no longer analyze the differences. This is going to require different formulas for the test statistic and the confidence interval: We can't construct d – the difference variable – in any sensible way.

Formulas & Assumptions for Independent Samples

We'll express parameters & statistics in terms of sample 1 & sample 2:

$H_0: \mu_1 - \mu_2 = 0$ (or $\mu_1 = \mu_2$) $\quad H_a: \mu_1 - \mu_2 \left\{ \begin{matrix} \neq \\ < \\ > \end{matrix} \right\} 0$

n_1 may be different from n_2

T.S. $$t = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

standard error, as before.

C.I. $$\bar{y}_1 - \bar{y}_2 \pm t_{crit, df} \underbrace{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}_{SE}$$

Back to the example: The following StatCrunch output was obtained. Use this to determine if the mean price per person for a meal *differs* between the two locations using a significance level of 0.10. $\alpha = .10$

puzzle to solve

NYC	Long Island
$\bar{y}_1 = 44.26$	$\bar{y}_2 = 39.64$
$s_1 = 12.8901$	$s_2 = 10.3938$
$n_1 = 50$	$n_2 = 50$

```
Possible p-values: 0.97425, 0.0515, 0.02575
```

(1) $H_0: \mu_1 - \mu_2 = 0$

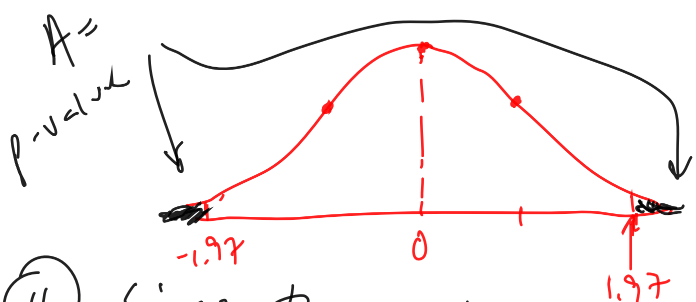
$H_a: \mu_1 - \mu_2 \neq 0.$

(2) $\alpha = .10$; reject H_0 in favor of H_a if p-value $< \alpha$.

(3) $$t = \frac{\bar{y}_1 - \bar{y}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{44.26 - 39.64 - (0)}{\sqrt{\frac{12.8901^2}{50} + \frac{10.3938^2}{50}}} = 1.97$$

pretty big

~ 2 std. dev.



$A = 0.0515$ = p-value $< \alpha$

(4) Since the p-value $.0515 < .10 = \alpha$ we reject a null of no difference in favor of the alternative: $\mu_1 \neq \mu_2$.

In fact, it looks like NYC costs more.

Example: The consumption of caffeine to benefit alertness is a common activity. Often caffeine is used in order to replace the need for sleep. One recent study was undertaken to determine if there was a difference in students' ability to recall memorized information after either the consumption of caffeine or a brief sleep. A random sample of 24 adults (between the ages of 18 and 39) were randomly divided into two groups of 12 participants each and verbally given a list of 24 words to memorize. During the break, one group takes a 90 minute nap while the other group is given a caffeine pill. After the break, each participant is asked to recall as many of the 24 words as possible. Researchers record the number of words each participant recalled.

— Identify the variable recorded for this study. Classify the variable as categorical or quantitative.

# of words recalled; quantitative

— Explain why the collected samples are **independent**. No connection made – no pairing – between individual observational units (i.e. students) in the two samples.

No way of creating a meaningful difference variable d.

Unpaired.

— The following StatCrunch output was obtained. Test the claim that average number of words recalled differs for the two groups using a significance level of 0.10.

$\alpha = .10$

μ_1 : Mean of Words where Group = Sleep
μ_2 : Mean of Words where Group = Caffeine

Difference	Sample Diff.	Std. Err.	DF
$\mu_1 - \mu_2$	3	1.3994046	21.893844

Possible p-values: 0.0217, 0.0434, 0.9783

① $H_0: \mu_1 - \mu_2 = 0$ $H_a: \mu_1 - \mu_2 \neq 0$

($H_0: \mu_1 = \mu_2$ $H_a: \mu_1 \neq \mu_2$) also okay

② $\alpha = 0.10$; reject the null of equality in favor of an alternative of a difference if p-value < α

③ T.S. $t = \dfrac{\bar{y}_1 - \bar{y}_2 - (\mu_1 - \mu_2)}{SE} = \dfrac{3 - 0}{1.3994} = 2.14$ pretty big!

df = 21.89...

p-value = .0434

A = .0434 < .10 = α

④ Since pvalue < .10 = α, we reject the null of equivalence; It appears that naps beat caffeine on the memory test!

— Using the StatCrunch output above, estimate the difference in the mean number of recalled words for the two groups. Use a confidence level of 95%.

Sample 1 - sleep
Sample 2 - caffeine
Estimate $\mu_1 - \mu_2$
(with a CI, a 95% CI)

we want this one

corresponds to 95% = .95 confidence

Two-tail probability	0.20	0.10	0.05	0.02	0.01
One-tail probability	0.10	0.05	0.025	0.01	0.005
df					
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797

closest to true df: 22

t_{crit} is in here somewhere...

C.I.: 95% $\bar{y}_1 - \bar{y}_2 \pm t_{crit}\, SE$

$3 \pm 2.508 \cdot 1.3994 \Rightarrow [.099, 5.903]$

95% CI for $\mu_1 - \mu_2$

Upshot: it's positive; so naps beat caffeine! $0 \notin$ CI - so 0 is not an option (with 95% confidence). Equality is not an option.

Example: Researchers speculate that drivers who do not wear a seatbelt are more likely to speed than drivers who do wear one. The following data were collected on a random sample of 20 drivers who were clocked to see how fast they were driving (mph).

Seatbelt	62	60	68	64	72	75	63	60	64	80
No Seatbelt	72	85	72	62	84	76	66	63	65	64

$n_1 = 10$

$n_2 = 10$

~ no relationship between these two drivers.

What type of samples were selected – independent or dependent (paired)? Explain.

unpaired – There is no pairing attempted between observational units in the two samples,

StatCrunch was used to analyze the data, and the output for both types of samples (independent and dependent) is given below. Using the *appropriate* output, determine if the mean speed is higher for those who do not wear seatbelts than for drivers that do at a significance level of 0.05. $\alpha = .05$

Hypothesis test results:
μ_1 : mean of Seatbelt
μ_2 : mean of No seatbelt

Difference	Sample Mean	Std. Err.	DF
$\mu_1 - \mu_2$	-4.1	3.4359214	17.185564

The two-sided p-value is 0.2492.

Hypothesis test results:
$\mu_1 - \mu_2$: mean of the **PAIRED** difference between Seatbelt and No seatbelt

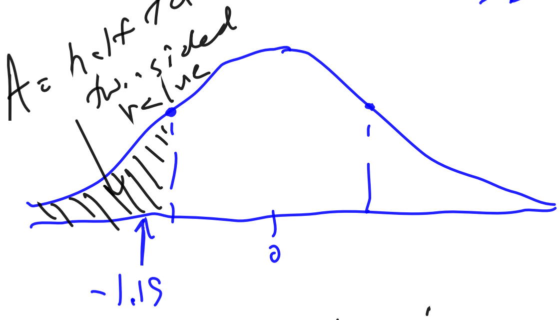
Difference	Sample Diff.	Std. Err.	DF
Seatbelt - None	-4.1	3.3281627	9

The two-sided p-value is 0.2632.

1. $H_0: \mu_1 - \mu_2 = 0$ $\quad H_a: \mu_1 - \mu_2 < 0 \ (\mu_1 < \mu_2)$

2. $\alpha = .05$; reject a null of no difference in favor of the alternative of non-seatbelt drivers drive faster if p-value $< \alpha$.

3. T.S. $t = \dfrac{\bar{y}_1 - \bar{y}_2 - (\mu_1 - \mu_2)}{SE} = \dfrac{-4.1 - 0}{3.4359} = -1.19$



A = p-value $= \frac{1}{2}(.2492) = .1246 > \alpha$

4. Since the p-value is greater than α, we fail to reject a null of equality: as far as we can tell, there is no significant difference in speeds.