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*.
- 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?

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 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.

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.

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:

Independent 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?

Circle one: Dependent Independent

Explain:

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?

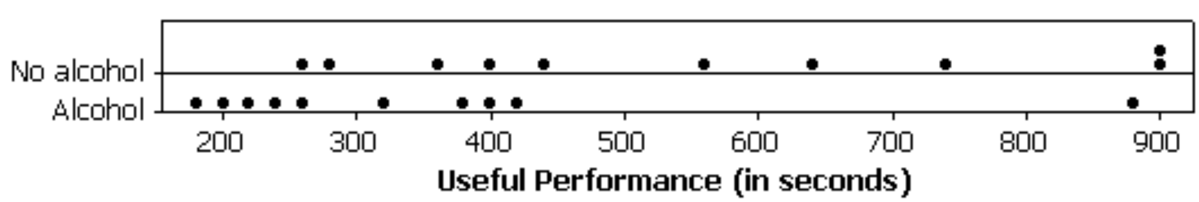
Circle one: Dependent Independent

Explain:

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:



What type of samples were taken in this scenario? Explain.

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

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

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.

Formulas & Assumptions for Dependent (Paired) Samples

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**.

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

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

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.

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.

Inferences for Independent Samples

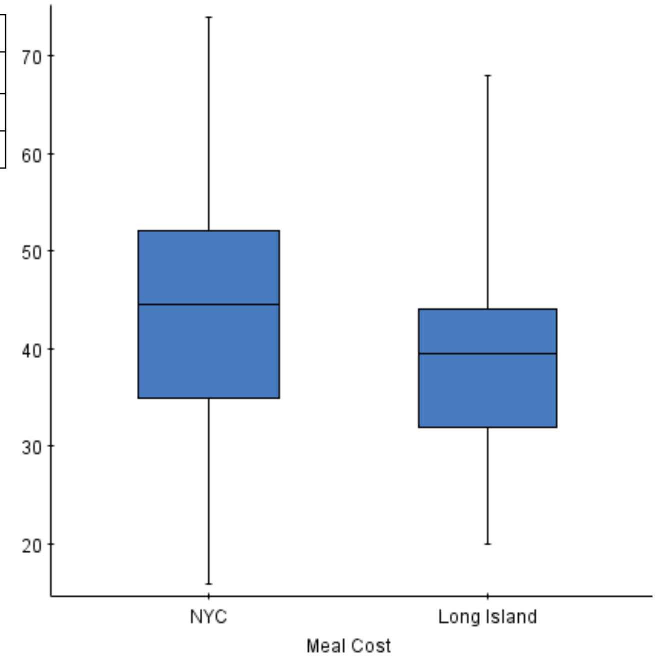
Textbook pages: 617 - 629

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?

— Explain why the samples collected were *independent*.

— 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
$\bar{y} = 44.26$	$\bar{y} = 39.64$
$s = 12.8901$	$s = 10.3938$
$n = 50$	$n = 50$



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:

Formulas & Assumptions for Independent Samples

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.

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

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

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.

— Explain why the collected samples are **independent**.

— 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.

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

— 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%.

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

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

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

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.

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.