Two Independent Samples: Summary

Learning Objective: In a given context, carry out the inferential method for comparing groups and draw the appropriate conclusions.

Comment

As we saw in previous tests, as well as in the two-samples case, the 95% confidence interval for $\mu_1 - \mu_2$ can be used for testing in the two-sided case ($H_o : \mu_1 - \mu_2 = 0$ vs. $H_a : \mu_1 - \mu_2 \neq 0$):

If the null value, 0, falls outside the confidence interval, H_o is rejected

If the null value, 0, falls inside the confidence interval, H_o is not rejected

Example

Let's go back to our leading example of the looks vs. personality score where we had a two-sided test.

Two Sample T – Test and CI: Score (Y), Gender (X)

Summary statistics for Score (Y):

Gender (X)	n	Mean	Std. Dev.	Std. Err.
Female	150	10.733334	4.254751	0.347399
Male	85	13.3294115	4.0189676	0.43591824

Hypothesis test results:
μ_1 : mean of Score (Y) where X = Female
μ_2 : mean of Score (Y) where X = Male
$\mu_1 - \mu_2$: mean difference
$H_0 : \mu_1 - \mu_2 = 0$
$H_A : \mu_1 - \mu_2 \neq 0$

Difference	Sample Mean	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	-2.5960784	0.55741435	182.97267	-4.657358	<0.0001

95% confidence interval results:

Difference	Sample Mean	Std. Err.	DF	L. Limit	U. Limit
$\mu_1 - \mu_2$	-2.5960784	0.55741435	182.97267	-3.6958647	-1.4962921

We used the fact that the p-value is so small to conclude that Ho can be rejected. We can also use the confidence interval to reach the same conclusion since 0 falls outside the confidence interval. In other words, since 0 is not a plausible value for $\mu_1 - \mu_2$ we can reject H_o, which claims that $\mu_1 - \mu_2 = 0$.

Scenario: Sugar Content in Fruit Juice and Soda

The purpose of this activity is to help you practice the connection between the two branches of formal inference (significance testing and confidence intervals) in the context of two independent samples.

Background:

Fruit juice is often marketed as being a healthier alternative to soda. And although juice does contain vitamins, juice can also be surprisingly high in sugar. Since excess sugar from any source can play a role in diseases like obesity and diabetes, it is important to be quantitatively informed about the beverages we consume.

To compare the **sugar content** (in grams) between **soda** and **100% fruit juice**, an investigation was made of **34** representative popular U.S. brands of 100% bottled juice (such as Dole, Minute Maid, Motts, Juicy Juice, Ocean Spray, Tree Top Apple, V8 Fusion, and Welch's Grape), and **45** representative popular

U.S. brands of soda pop (such as 7-Up, A&W Root Beer, Coca-Cola, Crush, Dr. Pepper, Fanta, Hawaiian Punch, Pepsi Cola, RC Cola, Sierra Mist, Schweppes Ginger Ale, and Sprite).

(Note, these are real data.)

Hypotheses:

If we let μ_1 represent the mean sugar content of the population of all bottled 100% fruit juices on the market, and if we let μ_2 represent the mean sugar content of the population all sodas on the market, then the significance test of interest are the hypotheses:

Null hypothesis: $\mu_1 - \mu_2 = 0$

(in other words, that there is no difference between the overall mean sugar content of on-the-market juices and on-the-market sodas, i.e., that the two beverage categories have the same overall mean sugar content).

Alternative hypothesis: $\mu_1 - \mu_2 \neq 0$

(in other words, that there is a difference between the overall mean sugar content of on-the-market juices and on-the-market sodas, i.e., that the two beverage categories don't have the same overall mean sugar content).

Summary statistics:

The summary of sugar content (in grams) for the two samples is as follows (remember, these are real data):

category	N	Mean	StDev	SE Mean
Fruit Juice	34	30.38	7.12	1.2
Soda	45	28.69	3.53	0.53

We see that the sample mean sugar content of the 34 juices was actually higher, at 30.38 grams, while the sample mean sugar content of the 45 sodas was only 28.69 grams.

The inferential question of interest is whether the slight difference is statistically significant.

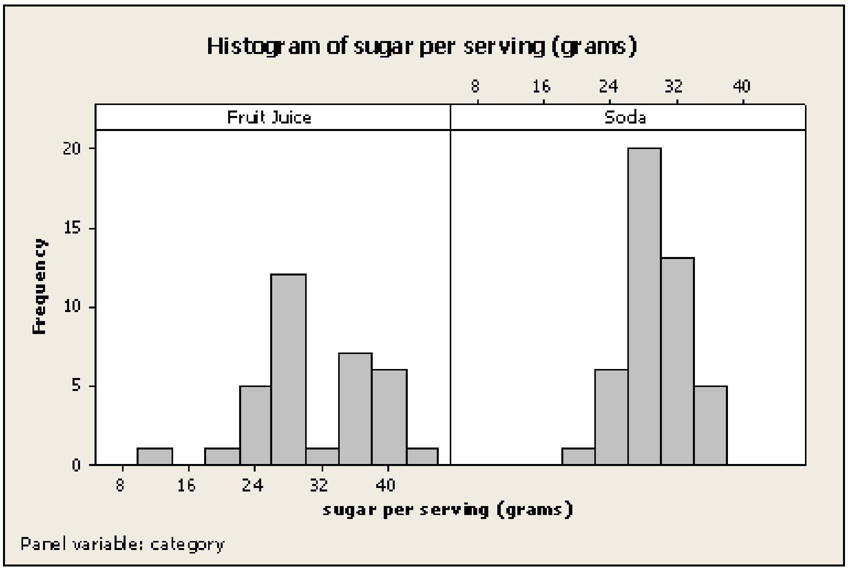
Checking conditions for inference:

For the purpose of statistical inference, we will consider the drinks in the study to be **random samples** of all such drinks on the market.

There are two independent groups (fruit juice versus soda); and since the σs (the standard deviations of the populations) are unknown to us, our required test will be a **t-test for two independent means**.

To check that a t-test for two independent means is reasonably justified, we should consider the shape of the histograms as well as the sample sizes in the study.

The histograms of sugar content for the two groups are as follows:



The histogram of sugar content for the **soda** sample (the graph on the right) is clearly unimodal and symmetric without any outliers; that shape helps to justify the desired inference procedure. The histogram of sugar content for the **juice** sample (the graph on the left) isn't quite as nice from the standpoint of justifying inference, since it's less clearly unimodal (although still clearly symmetric) and it has one possible outlier on its left side (although not too severe an outlier); but since the sample sizes in the study were each relatively large (n_1 = **34** brands of juice and n_2 = **44** brands of soda) the t-test is still justified, despite the possible bimodality and possible (and not-too-severe) outlier of the juice sample.

Result of Inference:

Here is the output from the formal t-test for two independent means:

```
Difference = mu (Fruit Juice) - mu (Soda)
Estimate for difference:  1.69

T-Test of difference = 0 (vs not =): T-Value = 1.27  P-Value = 0.210  DF = 45
```

Learn By Doing (1/1 point)

Based on the output, state the appropriate formal conclusion of the test of the hypotheses Ho: $\mu 1 - \mu 2 = 0$ Ha: $\mu 1 - \mu 2 \neq 0$ and then briefly state the meaning of the conclusion in the context of the question regarding whether there is reason to believe that the average sugar content of all 100% fruit juices sold is any different than the average sugar content of all sodas sold.

Your Answer:

Not enough evidence to reject Ho, because our p-value was 0.21 (meaning there was a 21% chance to encounter the results that we encountered).

Our Answer:
Based on the magnitude of the p-value shown in the output (p-value = 0.210), we do not reject the null hypothesis (since the p-value is relatively large). This means that the study does not provide sufficient evidence that the average sugar content of all fruit juices sold is any different than the average sugar content of all sodas. Follow up remarks: The conclusion is interesting for two reasons. First, we might have initially suspected (prior to the study) that sodas would have a higher average sugar content than juices; but the formal inference shows that this supposition is not supported. Second, after seeing the summary statistics (but not the inference), we saw that the sample of juices actually had slightly higher average sugar content than the sodas, so we might have then wondered if juices overall actually have a higher average sugar content than sodas; but the inferential conclusion shows that this is not supported either.

Resubmit Reset

Learn By Doing

1/1 point (graded)
Based on the output and the corresponding conclusion, which of the following is a plausible 95% confidence interval for the difference $\mu_1 - \mu_2$?

- ( ) (0.35, 3.03)
- ( ) (-3.03, -0.35)
- (•) (-0.99, 4.37) ✔

Answer
Correct:
The confidence interval most likely contains the true value of the difference being hypothesized about, i.e., the confidence interval should contain the value of $\mu_1 - \mu_2$. So, since we didn't reject the null hypothesis, the interval in this case should contain zero, because we believe the null hypothesis (i.e., we believe that $\mu_1 - \mu_2 = 0$). The interval here is the only interval that contains zero, since the left-hand endpoint of the interval is negative while the right-hand endpoint of the interval is positive.

Submit

Did I Get This

Below you'll find three sample outputs of the two-sided two-sample t-test:

$$H_0 : \mu_1 - \mu_2 = 0 \quad vs.$$
$$H_a : \mu_1 - \mu_2 \neq 0$$

However, only one of the outputs could be correct (the other two contain an inconsistency). Your task is to decide which of the following outputs is the correct one.

- **Output A:**
  - p-value: 0.289
  - 95% Confidence Interval: (-5.93090, -1.78572)
- **Output B:**
  - p-value: 0.003
  - 95% Confidence Interval: (-13.97384, 2.89733)
- **Output C:**
  - p-value: 0.223
  - 95% Confidence Interval: (-9.31432, 2.20505)

Did I Get This

1/1 point (graded)

Which of the following is the correct output?

- ○ A
- ○ B
- ◉ C ✔

Answer

Correct:

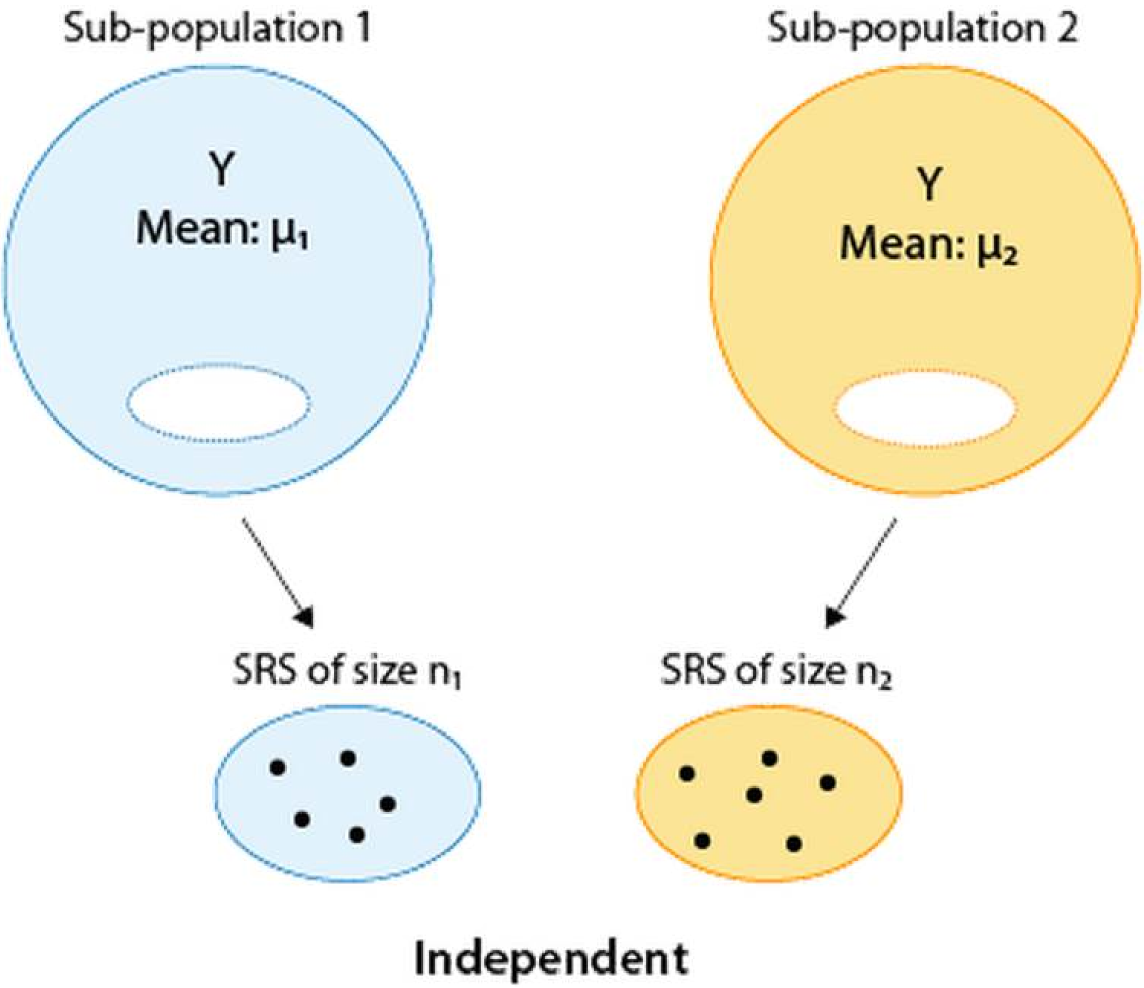
Indeed, this is the correct output, since it is the only one out of the three in which both the confidence interval and p-value lead us to the same conclusion (as it should be). Note that 0 falls inside the 95% confidence interval for $\mu_1 - \mu_2$, which means that H_o cannot be rejected. Also, the p-value is large (0.223) indicating that H_o cannot be rejected.

Submit

Let's Summarize

We have completed our discussion of the two-sample t-test for comparing two populations' means when the samples are independent. Let's summarize what we have learned.

- The two sample t-test is used for comparing the means of a quantitative variable (Y) in two populations (which we initially called sub-populations).



- Our goal is comparing μ_1 and μ_2 (which in practice is done by making inference on the difference $\mu_1 - \mu_2$). The null hypotheses is
  - Ho: $\mu_1 - \mu_2 = 0$

  and the alternative hypothesis is one of the following (depending on the context of the problem):
  - Ha: $\mu_1 - \mu_2 < 0$
  - Ha: $\mu_1 - \mu_2 > 0$
  - Ha: $\mu_1 - \mu_2 \neq 0$
- The two-sample t-test can be safely used when the samples are independent and at least one of the following two conditions hold:
  - The variable Y is known to have a normal distribution in both populations
  - The two sample sizes are large.

When the sample sizes are not large (and we therefore need to check the normality of Y in both population), what we do in practice is look at the histograms of the two samples and make sure that there are no signs of non-normality such as extreme skewedness and/or outliers.

- The test statistic is as follows and has a t distribution when the null hypothesis is true:

$$t = \frac{(\overline{y_1} - \overline{y_2}) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- P-values are obtained from the output, and conclusions are drawn as usual, comparing the p-value to the significance level alpha.
- If H_o is rejected, a 95% confidence interval for μ_1 - μ_2 can be very insightful and can also be used for the two-sided test.