Learn By Doing –Two-Sample T-test and Related Confidence Interval

The purpose of these activities is to give you more practice with this process.

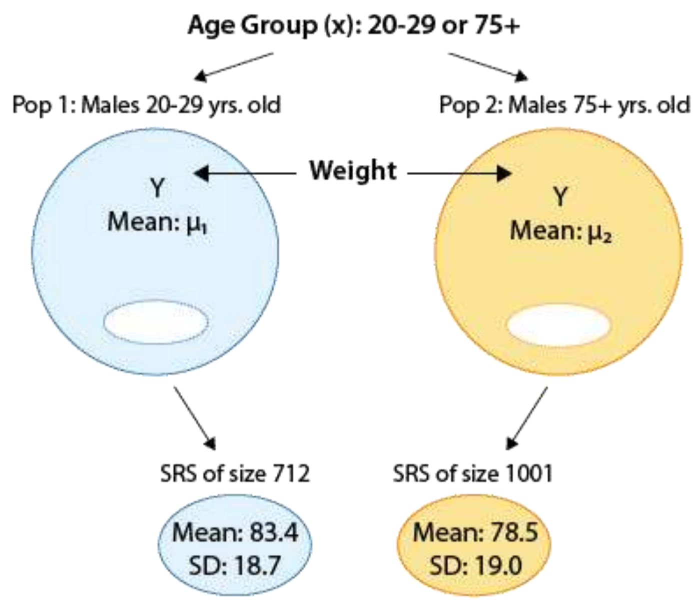
Scenario #1 – Weight vs. Age Groups (among males)

According to the National Health And Nutrition Examination Survey (NHANES) sponsored by the U.S. government, a random sample of 712 males between 20 and 29 years of age and a random sample of 1,001 males over the age of 75 were chosen, and the weight of each of the males was recorded (in kg). Here is a summary of the results (source: http://www.cdc.gov/nchs/data/ad/ad347.pdf):

We wish to conduct a test to determine if there is a difference in the population mean weight between these two age groups. Here are the summary statistics:

	n	$\overline{Y}$	S
20-29 yrs old	712	83.4	18.7
75+ yrs old	1001	78.5	19.0

And here is a figure summarizing this scenario.



Note: We defined the younger age group and the older age group as population 1 and population 2, respectively, and μ_1 and μ_2 as the mean weight of population 1 and population 2, respectively.

Our hypotheses are:

Ho: $\mu_1 - \mu_2 = 0$.

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

Now complete the following questions regarding parts of the process of conducting the two-sample t-test.

Scenario #1 – Weight vs. Age Groups (among males) – Step 2:

Can we use the two-sample t-test?

Your answer

OUR ANSWER

We can safely use the two-sample t-test in this case since:

The samples are independent, since each of the samples was chosen at random.

Both sample sizes are very large (712 and 1,001), and therefore we can proceed regardless of whether the populations are normal or not.

Scenario #1 – Weight vs. Age Groups (among males) – Step 3:

Suppose the t-statistic is 5.31 with a p-value of 0.000.

What can we learn from the p-value? (Answers may vary from our solution – we are not asking yet for a conclusion to the test but more what does the p-value itself mean in this scenario).

Your answer

OUR ANSWER

The p-value is essentially 0, indicating that it would be nearly impossible to observe a absolute value of the difference between the sample mean weights of 4.9 (or more extreme) if the mean weights in the age group populations were the same (i.e., if H_o were true).

The t-value is quite large, and the p-value correspondingly small, indicating that our data are very different from what is claimed in the null hypothesis.

Scenario #1 – Weight vs. Age Groups (among males) – Step 4:

What is the conclusion to this test in context?

Your answer

OUR ANSWER

Since the p-value is 0 (or very close to it), this indicates that the data provide strong evidence against H_o, so we reject it.

The Conclusion: There is enough evidence that the mean weight of males 20-29 years old is higher than the mean weight of males 75 years old and older.

Scenario #1 – Weight vs. Age Groups (among males) – Interpret Confidence Interval

Here are the results from MINITAB:

```
Two-Sample T-Test and CI

Sample     N  Mean  StDev  SE Mean
1        712  83.4   18.7     0.70
2       1001  78.5   19.0     0.60


Difference = mu (1) - mu (2)
Estimate for difference:  4.90000
95% CI for difference:  (3.08970, 6.71030)
T-Test of difference = 0 (vs not =):
      T-Value = 5.31
      P-Value = 0.000  DF = 1545
```

Provide a correct interpretation of the confidence interval for the difference between population means in context, given above as (3.09, 6.71). Remember that group 1 represents younger men age 20-29 and group 2 represents older men age 75 and old.

Your answer

OUR ANSWER

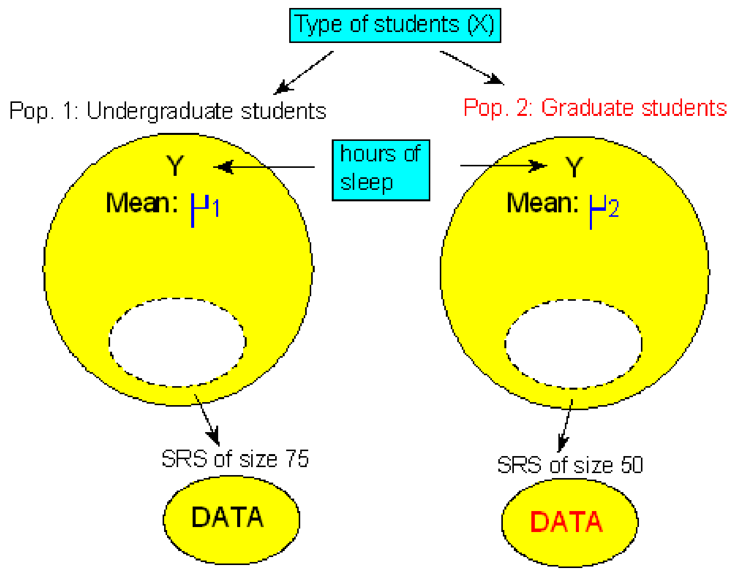
Interpretation: With 95% confidence, we can say that the population mean weight for males who are 20-29 years is between 3.1 to 6.7 kilograms more than that for males who are 75 and older.

Scenario #2 – Sleep compared between Undergraduate and Graduate students

A study was conducted at a large state university in order to compare the sleeping habits of undergraduate students to those of graduate students.

Random samples of 75 undergraduate students and 50 graduate students were chosen and each of the subjects was asked to report the number of hours he or she sleeps in a typical day.

The following figure summarizes the problem:



Note that we defined:

μ_1—the mean number of hours undergraduate students sleep in a typical day

μ_2—the mean number of hours graduate students sleep in a typical day

According to the National Health And Nutrition Examination Survey (NHANES) sponsored by the U.S. government, a random sample of 712 males between 20 and 29 years of age and a random sample of 1,001 males over the age of 75 were chosen, and the weight of each of the males was recorded (in kg). Here is a summary of the results (source: http://www.cdc.gov/nchs/data/ad/ad347.pdf):

We wish to conduct a test to determine if there is a difference in the population mean weight between these two age groups. Here are the summary statistics:

Our hypotheses are:

Ho: $\mu_1 - \mu_2 = 0$.

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

Now complete the following questions regarding parts of the process of conducting the two-sample t-test.

Scenario #2 – Sleep in College Students – Step 2:

Can we use the two-sample t-test?

Your answer

OUR ANSWER

We can safely use the two-sample t-test in this case since:

Both samples are random, and therefore independent.

The sample sizes (75 and 50) are quite large, and therefore we can proceed regardless of whether the populations are normal or not.

Scenario #2 – Sleep in College Students – Step 3:

Suppose the t-statistic is -1.23 with a p-value of 0.22.

	N	Mean	StDev	SE Mean
undergraduate	75	6.19	1.05	0.12
graduate	50	6.42	1.03	0.15

```
Difference = mu (undergraduate) - mu (graduate)
Estimate for difference:  -0.233333
```

What can we learn from the p-value? (Answers may vary from our solution – we are not asking yet for a conclusion to the test but more what does the p-value itself mean in this scenario).

Your answer

OUR ANSWER

The p-value not small at 0.22, indicating that it would not be unlikely to observe a absolution value of the difference between the sample mean hours of sleep of 0.233 (or more extreme) if the mean hours of sleep in the two group populations were the same (i.e., if H_o were true).

The t-value is quite small, and the p-value correspondingly large, indicating that our data are not very different from what is claimed in the null hypothesis.

Scenario #2 – Sleep in College Students – Step 4:

What is the conclusion to this test in context?

Your answer

OUR ANSWER

The p-value is not small (in particular, it is larger than 0.05), indicating that it is still reasonably likely (probability 0.22) to get data like those observed, or even more extreme data, under the null hypothesis (i.e., assuming that undergraduate and graduate students have the same mean sleeping hours).

Therefore, the data do not provide evidence to reject $H_{o.}$

The Conclusion: There is NOT enough evidence to conclude that the population mean hours of sleep differs between undergraduate students and graduate students.

Scenario #3 – Sugar in Juice and Soda:

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.

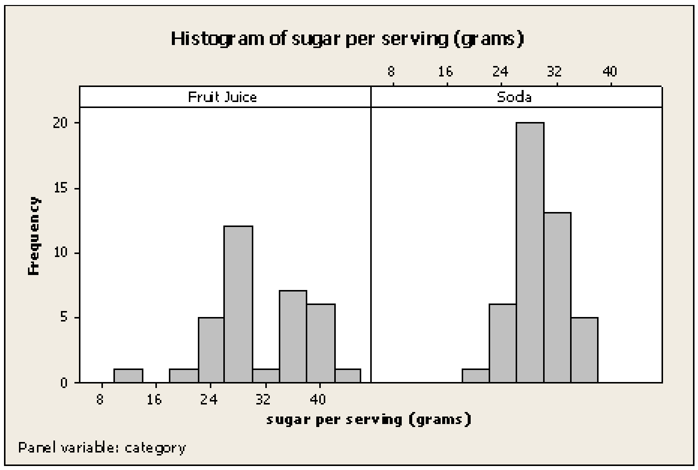
Checking conditions for inference:

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 standard deviations of the samples are very different we will apply a **t-test for two independent means assuming unequal variances**.

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.

Scenario #3 – Sugar in Juice and Soda – 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
```

Based on the output, state the appropriate formal conclusion of the test of the hypotheses

Ho: $\mu1 - \mu2 = 0$

Ha: $\mu1 - \mu2 \neq 0$

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

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.

Scenario #3 – Sugar in Juice and Soda – Possible Confidence Interval

Based on the output and the corresponding conclusion, which of the following is a plausible 95% confidence interval for the difference $\mu1 - \mu2$?

(a) (0.35, 3.03)

Incorrect. 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 $\mu1 - \mu2$. 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 $\mu1 - \mu2 = 0$). But notice that the interval in (a) does not contain zero, because both endpoints of the interval are positive.

(b) (-3.03, -0.35)

Incorrect. 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 $\mu1 - \mu2$. 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 $\mu1 - \mu2 = 0$). But notice that the interval in (b) does not contain zero, because both endpoints of the interval are negative.

(c) (-0.99, 4.37)

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 $\mu1 - \mu2$. 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 $\mu1 - \mu2 = 0$). The interval in (c) 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.

Scenario #4 – Relation between P-values and Confidence Intervals

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

(**Hint:** No calculations are necessary in order to answer this question. Instead pay attention to the p-value and confidence interval).

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)

Which of the following is the correct output?

A

Incorrect. Note that 0 falls outside the 95% confidence interval for $\mu_1 - \mu_2$, which means that $H_o$can be rejected in favor of the two-sided alternative hypothesis. On the other hand, the p-value is large (0.289) indicating that H_o cannot be rejected. ... Something is wrong here.

B

Incorrect. Note that 0 falls inside the 95% confidence interval for $\mu_1 - \mu_2$, which means that $H_o$cannot be rejected. On the other hand, the p-value is small, (0.003) indicating that H_o can be rejected. ... Something is wrong here.

C

Good job! 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.