Two Independent Samples: Finding the p-value

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

Step 3: Finding the p-value of the test

Since our test is called the two-sample t test ,we know that the p-values are calculated under a t distribution. Indeed, it turns out that the null distribution of our test statistic is approximately t. Figuring out which one of the t distributions (in other words, how many degrees of freedom this t distribution has) is quite involved and will not be discussed here. Instead, we use a statistics package to find that the p-value in this case is 0.

Example

Here, again is the relevant output for our example:

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

According to the output the p-value of this test is less than 0.0001. How do we interpret this?

A p-value which is practically 0 means that it would be almost impossible to get data like that observed (or even more extreme) had the null hypothesis been true.

More specifically to our example, if there were no differences between females and males with respect to whether they value looks vs. personality, it would be almost impossible (probability approximately 0) to get data where the difference between the sample means of females and males is -2.596 (that difference is 10.733 - 13.329 = -2.596) or higher.

Comment: Note that the output tells us that $\overline{y_1} - \overline{y_2}$ is approximately -2.6. But more importantly, we want to know if this difference is significant. To answer this, we use the fact that this difference is 4.66 standard errors below the null value.

Step 4: Conclusion in context

As usual a small p-value provides evidence against H_0. In our case our p-value is practically 0 (which smaller than any level of significance that we will choose). The data therefore provide very strong evidence against H_0 so we reject it and conclude that the mean Importance score (of looks vs personality) of males differs from that of females. In other words, males and females differ with respect to how they value looks vs. personality.

Comments

You might ask yourself: "Where do we use the test statistic?"

It is true that for all practical purposes all we have to do is check that the conditions which allow us to use the two-sample t-test are met, lift the p-value from the output, and draw our conclusions accordingly.

However, we feel that it is important to mention the test statistic for two reasons:

1. The test statistic is what's behind the scenes; based on its null distribution and its value, the p-value is calculated.

2. Apart from being the key for calculating the p-value, the test statistic is also itself a measure of the evidence stored in the data against H_o. As we mentioned, it measures (in standard errors) how different our data is from what is claimed in the null hypothesis.
