

Module 4: T - test for Comparison of Means

Definition

- The t-test is a statistic that we use to compare the means (averages) of two groups of data to see if they are really different or if they are similar and could have been due to random chance.

Types of Statistics

- The standard deviation is a **descriptive statistic**, that describes the amount of variation in a data set.
- The t-test is an **inferential statistic** that allows us to make generalizations or predictions based on our data onto the general populations or larger segments of the population.

Comparing two groups

- So now we have two different data sets, such as the popcorn project that had two different brands to compare, called **independent samples**.
- Or we have two sets of data on one population that represents the data before and after we did an experiment, called **paired samples**.

Different or not

- The question is “Are these two data sets really different?” called **statistically significant**, or “Do the results just represent random chance?”
- Or the question is “Did our experiment really work and the difference between the two data sets was due to our experiment and not due to random chance?”

How to find out

- To answer this question we use a t-test which is a measure of the difference between the two data sets.
- This allows us to make a judgment of whether the difference is real, that is statistically significant, or just random chance.

How it works

- The t-test measures the variance between the groups and compares it to the variance within the groups.

$$t = \frac{\text{variance between groups}}{\text{variance within groups}}$$

How much difference

- A large t value means that the two groups are really significantly different.
- A small t value means that there really is no significant difference between the groups and that this data could be due to chance alone.



Probability of Significance

- The p value is a statistical term that describes the probability that the results are significant
- Most science projects and research studies use a p value of <0.05
- This means that the probability that the results are due to chance alone is $< 5\%$

p value and t value

- The t value and the p value are related
- For every t value there is a corresponding p value, but the relationship varies
- The relationship depends on the degrees of freedom, the sample size, and whether the samples are independent or paired samples.

Manual calculation

- To do this calculation manually you need to calculate the mean and variance for each sample, calculate the degree of freedom, and know if your samples are independent or paired samples
- Then you calculate the t statistic and look at the t statistic table in the appropriate row and column and compare the table value with your calculated value

Final judgment

- Look in a statistics book for a t statistic table or look at a website, for example: <http://www.stat.tamu.edu/~lzhou/stat302/T-Table.pdf>
- If the t statistic that you calculate is greater than the value listed in the table then your results are statistically significantly different
- That is, your groups are really different, or there was a significant difference due to the experiment that you conducted.

Concluding statement

- You can state your conclusion in this format:
- The mean of group A, aa, compared to the mean of group B, bb, had a t-value of tt with dd degrees of freedom, and was statistically significant at p value of < 0.05 .

Using Excel

- You can do all of this in Excel using the data analysis feature
- You will need to enter your two data sets, the degree of freedom, the desired p value, and whether your samples are independent or paired samples
- Excel will do all the calculations and give you your final t-statistic



Science Fair Popcorn Project Using the t-test on the data

The Data

- Percentage of kernels popped

Percentage of Kernels Popped					
Brand A			Brand C		
Kernels	Popped	%	Kernels	Popped	%
108	86	79.6%	92	81	88.0%
98	82	83.7%	91	79	86.8%
102	84	82.4%	94	82	87.2%
92	78	84.8%	87	77	88.5%
106	91	85.8%	88	75	85.2%
			96	84	87.5%

Excel Output

Statistic	Brand A	Brand C
Mean	83.3%	87.2%
Standard Deviation	0.02	0.01
Number of Samples, n	5	6
Deg. of Freedom = $n_a + n_c - 2$	9	
Calculated t-statistic	3.57	

Steps for t-test

- The following slide shows a portion of the t-test table.
- Look at the row for 9 degrees of freedom
- Look at the column for the P value of 0.05 which is the probability we look for
- Find the value where the two intersect



t-test table (portion)

df	P=0.20	P=0.10	P=0.05	P=0.01	P=0.005
7	0.896	1.415	1.895	2.998	3.499
8	0.889	1.397	1.86	2.896	3.355
9	0.883	1.383	1.833	2.821	3.25
10	0.879	1.372	1.812	2.764	3.169
11	0.876	1.363	1.796	2.718	3.106

Steps for t-test

- The critical value from the table is 1.833
- The t-test value that was calculated is 3.57
- Since 3.57 is greater than 1.833 the t-test shows there is a significant difference between the two brands of popcorn.

Conclusion

- The mean of group C, with 87.2% kernels popped compared to the mean of group A with 83.3% kernels popped had a t-value of 3.57 with 9 degrees of freedom and was statistically significantly better at a P value of <math><0.05</math>.