

SPSS & Hypothesis Testing

Hypothesis Testing

Hypothesis testing is a decision making process for evaluating claims about a population.

The researcher must:

- Define the population under study
- State the hypothesis that is under investigation
- Give the significance level
- Select a sample from the population
- Collect the data
- Perform the statistical test
- Reach a conclusion

Q1: Give at least three examples of a population.

1) _____

2) _____

3) _____

Q2: For the population listed in 1), give an example of a sample from the population.

Q3: Can you make up some hypothesis about the population in 1).

Examples of hypothesis tests include t-test, Chi- Square, and correlation analysis to name a few.

My goal is to give you enough information to use SPSS to perform some different hypothesis tests without getting into the nitty gritty detail. To use this tool properly, you must have a statistics background; otherwise, chances are very good that an incorrect test can be applied to a given set of data.

Hypothesis testing is the formal statistical technique of collecting data to answer questions through the use of a statistical model.

Each question asked needs to be posed as a null hypothesis; the null hypothesis is that there are no differences of the dependent variables of your model that can be explained by the independent variables.

ONE-SAMPLE T-TEST:

This is the easiest of the statistical tests to understand. Specifically, this test compares an *observed mean* (computed from a set of observed values) to a *hypothesized mean* and determines the likelihood that the difference between the means occurs by chance. The chance is reported as the p-value.

The p-value measures the probability that the difference we see between the hypothesized mean and sample mean occurs due to chance. A small p-value means that the difference is unlikely to be the result of chance; a large p-value means the difference is likely to be the result of chance.

What do we mean by random chance? Where does this randomness come from? Relate this back to the populations we discussed earlier.

Thus, the lower the p-value the more certain that we can be that there is a **statistically significant** difference between the observed and hypothesized mean. Most disciplines look for a p-value of 0.05 or less; that is, if the p-value is less than 0.05 then the difference is regarded as statistically significant.

- if $p < 0.05$, reject the null hypothesis
- if $p \geq 0.05$, accept the null hypothesis
-

Exercise 1

The blood lipid data set in the **CS130 Public->DataFiles** directory represents a blood lipid screening of medical students.

a) Grab this Excel file, open it up in SPSS and save the file as lipid.sav.

b) What is the mean Cholesterol value? _____

c) Is the cholesterol level significantly greater than 190? Can you tell by looking at the data? What do you think?

d) Our first objective is to perform a one-sample t-test on data from blood lipid screening of medical students. Specifically, we will test whether the mean cholesterol level is greater than 190 in a *statistically significant* way, the point at which cholesterol levels may be unhealthy.

Null hypothesis: _____

- Open Lipid Data.
- From the Analyze menu, select Compare Means and then One Sample t-test .
- Select your Test Variable which is Cholesterol .
- Enter the Test Value which is 190.
- In the variable browser, select Cholesterol and click ADD

One-Sample Test

	Test Value = 190					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Cholesterol	.336	94	.737	1.23158	-6.0356	8.4988

Notice: In SPSS, the **p-value** is given in the box labelled **Sig. (2-tailed)**.

You will notice that the mean is slightly higher than the hypothesized value of 190. However, although the mean is in fact higher, this difference is well within the range of sampling variance. In particular, the p-value of 0.7373 indicates you would see a difference of this magnitude by chance more than 73% of the time. Thus the mean cholesterol level is not significantly greater than 190.

PAIRED T-TEST:

The most common use of the paired t-test is the comparison of two measurements (typically one measurement occurs before a "treatment" and the other after the "treatment") from the same

individual or group. This test is used primarily to determine if the "treatment" had a statistically significant effect. As in the case of the one-sample t-test, the primary statistic of concern is the p-value, and thankfully the p-value has the same interpretation here as it did in the case of the one-sample t-test.

Exercise 2

- a) What is the mean for Triglycerides? _____
- b) What is the mean for Trig-3yrs? _____
- c) Does it look like there is a statistically significant difference between Triglycerides and Trig-3yrs?
- d) Perform a paired t-test using the Lipid Data file to see if there is a statistically significant difference between Triglycerides and Trig-3yrs.

What is the Null hypothesis?

- Open the Lipid Data file used in example one.
- From the Analyze menu, select Compare Means and then Paired Samples t-test.
- Select the paired variables and set your confidence interval at 95%

Q4: Should we accept the null hypothesis in this case? Why?

State your conclusion
