

## community project

encouraging academics to share statistics support resources

All stcp resources are released under a Creative Commons licence

stcp-gilchristsamuels-1

The following resource is associated:  
Checking Normality for Parametric Tests worksheet

### Statistical Hypothesis Testing

**Research question:** Exploring a comparison, association or relationship

**Level of measurement of data:** Categorical or scale (continuous)

From your research question you should formulate one or more hypotheses each leading to two possible outcomes:

**Null Hypothesis** (NH or  $H_0$ ):

Assumes no difference, association or relationship between the variables

**Alternative Hypothesis** (AH or  $H_1$  or  $H_A$ ):

Assumes a difference, association or relationship between the variables

The use of an appropriate statistical hypothesis test will allow you to decide between these two outcomes by examining the probability (or p) value of the test statistic.

#### Probability or p-value

A decision between the two hypotheses is made by viewing the 'p-value' (often referred to as the 'Sig-value' by SPSS), which is the probability (or chance) of the collected data or a more extreme event happening under the assumption of the null hypothesis.

If this probability is small,  $H_0$  is rejected in favour of  $H_1$ , which is termed a 'statistically significant result'; otherwise we 'fail to reject  $H_0$ ', which is termed a 'non-statistically significant result'.

#### What is 'small'?

Conventionally,  $p = 0.05$  is used as the cut of value, hence:

- If  $p \leq 0.05$  (i.e. p is less than or equal to 0.05),  $H_0$  is rejected in favour of  $H_1$
- If  $p > 0.05$  (i.e. p is greater than 0.05),  $H_0$  is not rejected

**[Note:** This is not always quite the same as accepting  $H_0$ . It is more like *acquitting a defendant when there is insufficient evidence to secure a conviction.*]



Alternatively, the p-value can be interpreted in more detail using the table below:

p-value of test statistic	Statistically significant?	Formal action	Informal interpretation
Greater than 0.1	No	Do not reject $H_0$	<b>No evidence</b> to reject $H_0$
Between 0.1 and 0.05	No	Do not reject $H_0$	<b>Weak evidence</b> to reject $H_0$
Between 0.05 and 0.01	Yes: at 95%	Reject $H_0$ at 95% confidence	<b>Evidence</b> to reject $H_0$
Between 0.01 and 0.001	Yes: at 99%	Reject $H_0$ at 99% confidence	<b>Strong evidence</b> to reject $H_0$
Less than 0.001	Yes: at 99.9%	Reject $H_0$ at 99.9% confidence	<b>Very strong evidence</b> to reject $H_0$

You should always relate back the outcome of the hypothesis testing to particular variables in your study; don't just conclude with 'reject the null hypothesis'. You will need to discuss your results and conclusions in the context of the work of others, and the potential for further studies. Note that statistical significance does not necessarily indicate practical usefulness of a result.

Several features in the design of a study affect the value of p, which is used to make a statistical decision. These features also affect the ability of a test to reject a null hypothesis when it is false. The probability of correctly rejecting a false null hypothesis is called the **power** of a test. See table below:

Decision	Reality	
	$H_0$ True	$H_0$ False
Accept $H_0$	✓ $1 - \alpha$	Type II error $\beta$
Reject $H_0$	Type I error $\alpha$	✓ Power ( $1 - \beta$ )

A decision to reject the null hypothesis might be wrong, with probability equal to the chosen significance level (often referred to as  $\alpha$ ). This incorrect decision is called a **type I error**. This is equivalent to *convicting an innocent person*.

Setting the significance level at 0.1% would reduce the risk of a type I error to a chance of 1 in 1000 hypothesis tests. Other things being equal, however, this action would increase the probability of not rejecting a null hypothesis when it is false. This incorrect decision is associated with a statistically non-significant result and is known as a **type II error**. This is equivalent to *acquitting a guilty person*. The probability of a type II error (often referred to as  $\beta$ ) is related to the **power** of a test (which is  $1 - \beta$ ).



## Common hypothesis tests

Level of measurement of data	Test of association (relationship)	Test of difference (comparison)	
		2 groups	>2 groups
Nominal	Chi-squared	Chi-squared	
Ordinal	Spearman	Mann-Whitney U (unrelated)	Kruskall-Wallis (unrelated)
		Wilcoxon (related)	Friedman (related)
Scale (continuous)	Pearson	Independent t-test (unrelated)	One-way ANOVA (unrelated)
		Paired t-test (related)	Repeated measures ANOVA (related)

### Limitations

For your project/dissertation, you may be restricted in the data you are able to collect. For example, you may not have a true 'random sample', but rather a 'convenience' sample. Always try to get as big a sample as possible, but you may only be able to obtain a relatively small number of cases (e.g. less than 30). This may affect the results of statistical testing. Another influence on outcomes is underlying assumptions necessary for the validity of the tests.

### Assumptions

When performing parametric tests (in general, those for scale data), e.g. t-tests and ANOVA, the data are assumed to be representative of the population from which they have been drawn and come from a normal distribution. See the **checking normality for parametric tests** worksheet.

### Robustness

Although parameter-based statistical tests make certain assumptions in their underlying models, they often work well in other situations where these assumptions are violated. This is known as robustness.

### Violation of assumptions

If assumptions are still in doubt, you can use the equivalent non-parametric test, such as the Mann-Whitney U test instead of an unpaired t-test. These are found in SPSS under the Analyze menu.

