Geo Factsheet

Number 74

Hypothesis Tests & Mann-Whitney U-test

Many geography projects involve a hypothesis test. Students often find difficulty in deciding on suitable hypotheses, and accordingly can waste time collecting unhelpful data. This Factsheet explains what is involved in a statistical test of a hypothesis and discusses the role of the null hypothesis and the level of significance. It also covers in detail the calculation and use of the Mann-Whitney U-test, which is applicable to many geographical situations.

Hypotheses
All statistical tests involve testing a *null hypothesis* (H_0) against an *alternative hypothesis* (H_1).

The null hypothesis can be described as "the boring case" - i.e. nothing has changed. For example:

H_0: There is no difference in population age structure between a rural village and the country as a whole

H_0: There is no difference in current velocity up and downstream a particular river

H_0: There is no correlation between lead levels and distance from the road

H_0: There is no difference between visitor numbers at two different National Parks

The null hypothesis cannot ever be of the form "something is greater than something else" or "something is related to something else". The exact form of it depends on the test you are using (see Factsheet 73 - Which Stats Test Should I Use?).

The alternative hypothesis is effectively saying the opposite of the null hypothesis - for example, for the last case above the alternative hypothesis would be:

H_1: There is a difference between visitor numbers at the two National Parks

When we carry out an investigation, we start off assuming that the null hypothesis is true, and only change our minds if the data obtained in the investigation provides strong enough evidence. This is rather analogous to the situation in a court of law, where the defendant is assumed innocent unless proven guilty!

Obviously there will always be some chance variations - if, for example, the visitor numbers only differed by one person over the course of the year at the two National Parks, we would probably feel - even without conducting any tests! - that this was not a "significant" difference.

The role of the statistical test is to give an objective definition of what constitutes sufficient evidence to reject H_0.

Choosing Hypotheses
Good hypotheses for a statistical test must:

- be specific, not vague or general
- refer to something that can be measured in an unambiguous way
- be simple, not attempt to include several variables
- include a null hypothesis

Table 1 shows some examples of "bad" hypotheses, and how they can be improved.

Doing the test
Whichever statistical test is used, we will effectively be plugging our data values collected in the experiment into some formula, and coming out with a single number. This number is what we will use to decide whether or not to reject the null hypothesis.

To make that decision, we will have to compare this number we have worked out - which is often called a *test statistic* - to the appropriate statistical table. There are different tables for different statistical tests. Table 2 (overleaf) shows an extract from a statistical table for the Mann-Whitney U-test. Statistical tables give *critical values* at various *significance levels*.

The *significance level* is a measure of how strong we are requiring the evidence to be before we reject H_0. Common significance levels used are 10% (0.1), 5% (0.05) and 1% (0.01). A 1% significance level, for example, means that we have only a 1% chance of rejecting H_o when we shouldn't have, whereas a 10% level would give us a 1 in 10 chance of rejecting H_0 when we should have accepted it.

To get an idea of what this means, imagine 100 students carrying out the same investigation into visitor numbers at two National Parks. We will imagine that there is really no significant difference in visitor numbers - in other words, the null hypothesis is true.

Table 1. Choosing a hypothesis

Original Hypothesis	What's wrong with it	Improved version
populations in villages are different to the rest of the country.	• Not specific or measurable - which aspect of the population - its age structure, sex structure, occupations? • There's no null hypothesis.	H_0: the population age structure in the village is the same as the national age structure. H_1: the population age structure in the village is not the same as the national age structure.
the closer to the road, the higher the pollution.	• Not specific - what sort of pollution? • How will it be measured? • There's no null hypothesis.	H_0: there is no correlation between lead levels and distance from the road. H_1: there is some correlation between lead levels and distance from the road.
slope affects vegetation.	• Not specific - which aspect of the slope is referred to? Is it the gradient, the altitude or the length of the slope? • Not measurable - you cannot just measure "vegetation". Should it be species diversity, or percentage cover, or biomass, or incidence of a particular species? • There's no null hypothesis.	H_0: the gradient of the slope has no effect on percentage cover. H_1: the gradient of the slope has some effect on percentage cover.

The students will all collect their data at different times, and hence get slightly different results. If they all carry out their statistical test at a 5% significance level, then on average five of them would find themselves rejecting the null hypothesis, because they happened to get "odd" data.

The *critical value* is just the value we have to compare with the number we have worked out - the test statistic - to decide whether or not we should reject H_o. Each significance level - and each sample size - has its own critical value. Critical values come from books of statistical tables.

For the Mann-Whitney U-test, we reject H_o if our value is smaller. For every test <u>except</u> Mann-Whitney, we reject H_o if our value is bigger than the critical (tables) value.

It is worth noting the way that the critical values vary with sample size; with a large sample, it is much easier to get a significant result!

Table 2. Critical values for the U-test

n_1	α	n_2 5	6	7	8
5	10%	4	5	6	8
	5%	2	3	5	6
6	10%	-	7	8	10
	5%	-	5	6	8
7	10%	-	-	11	13
	5%	-	-	8	10
8	10%	-	-	-	15
	5%	-	-	-	13

Mann-Whitney U-test

We use the U-test to compare the average of two sets of data - e.g. the species diversity on a path and off a path. We are just trying to find out whether there is a difference - e.g. whether being on a path affects the diversity.

The hypotheses to be tested are: **H_0 : there is no difference between X and Y**
H_1 : there is a difference between X and Y

(If you wish to be mathematically correct, you would use H_0: median1 = median2; H_1: $median_1 \neq median_2$)

We will **reject** the null hypothesis if the value we calculate (the *test statistic*) is **below** the value from the tables (the *critical value).*

The procedure for carrying out the test will be illustrated by applying it to data on species diversity on and off a path.

GLOSSARY

Simpson's Diversity Index is calculated using the formula

Diversity = $\frac{N(N-1)}{\Sigma n(n-1)}$

where n refers to the number of individuals of each particular species and N is the total number of individuals.

METHOD	APPLICATION
1. Write down your hypotheses	**H_0: There is no difference in species diversity on and off the path** **H_1: There is a difference in species diversity on and off the path**
2. Obtain data about the things you wish to compare - you need two sets of data Each set of data must contain at least 5 values, but they don't have to have the same number of values as each other.	The figures for species diversity in 8 path sites and 8 off-path sites are: Site 1 2 3 4 5 6 7 8 on-path 2.20 4.65 6.00 3.47 4.33 2.20 2.50 3.33 off-path 4.09 2.93 3.88 10.50 3.50 5.14 4.40 10.00
3. Consider one set of data at a time - say we start with the on-path sites. We now must calculate a score for each on-path site. A site is given: **1** for every **off-path** site that has a higher value **0.5** for every **off path** site that has an equal value.	Site 1, on-path: The following off-path sites have higher values: 1,2,3,4,5,6,7,8. Hence the score is 8 Similarly, for the other on-path sites, the scores are: Site 1 2 3 4 5 6 7 8 Score 8 3 2 7 4 8 8 7
4. Sum the scores of the on-path sites. This gives you the overall on-path score	So the total on-path score is **47**
5. Repeat steps 3) and 4) for off-path sites. This time, you will be awarding points for on-path sites with higher or equal scores.	We now find the scores for off-path sites: Site 1, off-path : The following on-path sites have higher values: 2,3,5 Hence the score is 3 Similarly, for the other off-path sites the scores are Site 1 2 3 4 5 6 7 8 Score 3 5 3 0 3 1 2 0 So the total off-path score is **17.**
6. Take the smaller of the two scores. This is the U-value	So the U-value is **17**
7. Compare your calculated U-value with the critical value in the tables at the appropriate significance level. If your value is **smaller**, reject H_0 - otherwise accept H_0.	The critical value for two samples of 8 is 13 at the 5% level of significance. Since our value is 17, we accept H_0 and conclude that there is no significant difference between species diversity on and off the path.

Projects using the Mann-Whitney U-test

- Measure the CO levels at 20 road intersections, which all have an equal volume of traffic, but half the intersections have traffic lights and half do not. Test whether or not the CO levels are significantly different.
- Measure the percentage vegetation cover at at least 5 sites on each of two slopes. Test whether or not there is a difference in % cover.
- Ask two groups of ten people - for example local residents and tourists - to award a score from 0 to 50 to indicate how favourably they view a planned change in the area (e.g. the New Forest becoming a National Park). Test whether or not there is a difference in opinion.

Acknowledgements; This Factsheet was researched and written by Cath Brown. **ISSN 1351-5136**