Statistics for Computing

MA4413 Lecture 6B

Kevin O'Brien

Kevin.obrien@ul.ie

Dept. of Mathematics & Statistics,
University *of* Limerick

Autumn Semester 2013

Hypothesis Testing : Revision from Last Class

- A hypothesis test is a method of making decisions using experimental data.
- A result is called ***statistically significant*** if it is unlikely to have occurred by chance.
- The null hypothesis (which we will denoted H_0) is an hypothesis about a population parameter, such as the population mean μ.
- The purpose of hypothesis testing is to test the viability of the null hypothesis in the light of experimental data.
- The alternative hypothesis H_1 expresses the exact opposite of the null hypothesis.
- Depending on the data, the null hypothesis either will or will not be **rejected** as a viable possibility in favour of the alternative hypothesis.

Writing the Null Hypothesis

- It will often express it's argument in the form of a mathematical relation, with a written description of the hypothesis (we will do it this way).
- H_0 will always refer to the population parameter (i.e. never the observed value) and must contain a condition of equality. (i.e. ' = ' , '$\leq$' or '$\geq$')

Number of Tails

Inference Procedures are either **One-tailed** or **Two-Tailed**.
Confidence Intervals are always two-tailed . Hypothesis tests can either be one-tailed or two-tailed. It is important to know how determine correctly the number of tails.

- The alternative hypothesis indicates the number of tails.
- A rule of thumb is to consider how many alternative to the H_0 is offered by H_1.
- When H_1 includes either of these relational operators;'$>$' ,'$<$' , only one alternative is offered.
- When H_1 includes the $\neq$ relational operators, two alternatives are offered (i.e.'$>$' or '$<$').

Significance Level (α)

- In hypothesis testing, the significance level α is the criterion used for rejecting the null hypothesis.
- The significance level is used in hypothesis testing as follows: First, the difference between the result (i.e. observed statistic or point estimate) of the experiment and the **null value**
- The null value is the expected value of this statistic, assuming that the null hypothesis is true), is determined.(i.e.we denote this ***Observed - Null***).
- Then, assuming the null hypothesis is true, the probability of a difference that large or larger is computed .
- Finally, this probability is compared to the significance level.
- If the probability is less than or equal to the significance level, then the null hypothesis is rejected and the outcome is said to be statistically significant.

Hypothesis Testing

The inferential step to conclude that the null hypothesis is false goes as follows: The data (or data more extreme) are very unlikely given that the null hypothesis is true. This means that:

(1) a very unlikely event occurred or

(2) the null hypothesis is false.

The inference usually made is that the null hypothesis is false. Importantly it doesn't prove the null hypothesis to be false.

Significance (Die Throw Example)

- Suppose that the outcome of the die throw experiment was a sum of 401. In previous lectures, a simulation study found that only in approximately 1.75% of cases would a fair die yield this result.
- However, in the case of a crooked die (i.e. one that favours high numbers) this result would not be unusual.
- A reasonable interpretation of this experiment is that the die is crooked, but importantly the experiment doesn't prove it one way or the other.
- We will discuss the costs of making a wrong decision later (Type I and Type II errors).

Significance Level

- Traditionally, experimenters have used either the 0.05 level (sometimes called the 5% level) or the 0.01 level (1% level), although the choice of levels is largely subjective.
- The lower the significance level, the more the data must diverge from the null hypothesis to be significant.
- Therefore, the 0.01 level is more conservative than the 0.05 level.
- The Greek letter alpha (α) is sometimes used to indicate the significance level.
- We will $\alpha = 0.05$ in this module. (i.e. 5%)

Hypothesis Testing and p-values

- In hypothesis tests, the difference between the observed value and the parameter value specified by H_0 is computed and the probability of obtaining a difference this large or large is calculated.
- The probability of obtaining data as extreme, or more extreme, than the expected value under the null hypothesis is called the ***p-value***.
- There is often confusion about the precise meaning of the p-value probability computed in a significance test. It is not the probability of the null hypothesis itself.
- Thus, if the probability value is 0.0175, this does not mean that the probability that the null hypothesis is either true or false is 0.0175.
- It means that the probability of obtaining data as different or more different from the null hypothesis as those obtained in the experiment is 0.0175.

Significance Level

- The significance level of a statistical hypothesis test is a fixed probability of wrongly rejecting the null hypothesis H_0, if it is in fact true.
- Equivalently, the significance level (denoted by α) is the probability that the test statistics will fall into the ***critical region***, when the null hypothesis is actually true. (We will discuss the critical region shortly).
- Common choices for α are 0.05 and 0.01

The Hypothesis Testing Procedure

We will use both of the following four step procedures for hypothesis testing. The level of significance must be determined in advance.

Procedure 1

The first procedures is as follows:

- Formally write out the null and alternative hypotheses (already described).
- Compute the ***test statistic*** - a standardized value of the numerical outcome of an experiment.
- Compute the p-value for that test statistic.
- Make a decision based on the p-value. (smaller than α or $\alpha/2$? - reject null)

(We will re-visit this approach later in the course).

The Hypothesis Testing Procedure

The second procedures is very similar to the first, but is more practicable for written exams, so we will use this one also. The first two steps are the same.

Procedure 2

- Formally write out the null and alternative hypotheses (already described).
- Compute the test statistic
- Determine the ***critical value*** (described shortly)
- Make a decision based on the critical value. (We call this step the **decision rule** step, and shall discuss it in depth shortly).

(We will mostly use this approach to hypothesis testing).

Test Statistics

- A test statistic is a quantity calculated from our sample of data. Its value is used to decide whether or not the null hypothesis should be rejected in our hypothesis test.
- The choice of a test statistic will depend on the assumed probability model and the hypotheses under question.
- The general structure of a test statistic is

$$\mathrm{TS} = \frac{\text{Observed Value} - \text{Null Value}}{\text{Std. Error}}$$

- Recall: The "Null Value" is the expected value, assuming that the null hypothesis is true.

The Test Statistic

- In our dice experiment, we observed a value of 401. Under the null hypothesis, the expected value was 350.
- The standard error is of the same form as for confidence intervals. $\frac{s}{\sqrt{n}}$.
- (For this experiment the standard error is 17.07).
- The test statistic is therefore

$$\text{TS} = \frac{401 - 350}{17.07} = 2.99$$

The Critical Value

- The critical value(s) for a hypothesis test is a threshold to which the value of the test statistic in sample is compared to determine whether or not the null hypothesis is rejected.
- The critical value for any hypothesis test depends on the significance level at which the test is carried out, and whether the test is one-sided or two-sided.
- The critical value is determined the exact same way as quantiles for confidence intervals; using Murdoch Barnes table 7.

Critical Region

In class : graphical representation of material on the black-board is scheduled here.

Important: A critical value is any value that separates the critical region (where we reject the null hypothesis) for thatvalues of the test statistic that do not lead to a rejection of the null hypothesis.

One Tailed Hypothesis test

- A one-sided test is a statistical hypothesis test in which the values for which we can reject the null hypothesis, H_0 are located entirely in one tail of the probability distribution (either the upper tail, or the lower tail, but not both).
- In other words, the critical region for a one-sided test is the set of values beyond than the critical value for the test
- A one-sided test is also referred to as a one-tailed test of significance.
- A rule of thumb is to consider the alternative hypothesis. If only one alternative is offered by H_1 (i.e. a $`<'$ or a $`>'$ is present, then it is a one tailed test.)
- (When computing quantiles from Murdoch Barnes table 7, we set $k = 1$)

Two Tailed Hypothesis test

- A two-sided test is a statistical hypothesis test in which the values for which we can reject the null hypothesis, H_0 are located in both tails of the probability distribution, on an equal basis.
- A two-sided test is also referred to as a two-tailed test of significance.
- A rule of thumb is to consider the alternative hypothesis. If only one alternative is offered by H_1 (i.e. a ‘ $\neq'$ is present, then it is a two tailed test.)
- (When computing quantiles from Murdoch Barnes table 7, we set $k = 2$)

Determining the Critical value

- The critical value for a hypothesis test is a threshold to which the value of the test statistic in a sample is compared to determine whether or not the null hypothesis is rejected.
- The critical value for any hypothesis test depends on the significance level at which the test is carried out, and whether the test is one-sided or two-sided.

Determining the Critical value

- A pre-determined level of significance α must be specified. Usually it is set at 5% (0.05).
- The number of tails must be known. (k is either 1 or 2).
- Sample size will be also be an issue. We must decide whether to use $n-1$ degrees of freedom or ∞ degrees of freedom, depending on the sample size in question.
- The manner by which we compute critical value is identical to the way we compute quantiles.We will consider this in more detail during tutorials.
- For the time being we will use 1.96 as a critical value.

Decision Rule: The Critical Region

- The critical region CR (or rejection region RR) is a set of values of the test statistic for which the null hypothesis is rejected in a hypothesis test.
- That is, the sample space for the test statistic is partitioned into two regions; one region (the critical region) will lead us to reject the null hypothesis H_0, the other will not.
- A test statistic is in the critical region if the absolute value of the test statistic is greater than the critical value.
- So, if the observed value of the test statistic is a member of the critical region, we conclude "Reject H_0"; if it is not a member of the critical region then we conclude "Do not reject H_0".
- (Demonstration on BlackBoard).

Critical Region

Remark: the absolute value fuction of some value x is denoted $|x|$. It is the magnitude of the value without consideration of whether the value is positive or negative.

- Let TS denote Test Statistic and CV denoted Critical Value.
- $|TS| > CV$ Then we reject null hypothesis.
- $|TS| \leq CV$ Then we **fail to reject** null hypothesis.
- For our die-throw example; TS = 2.99, CV = 1.96.
- Here $|2.99| > 1.96$ we reject the null hypothesis that the die is fair.
- Consider this in the context of "proof". (More on this in next class)

Performing a Hypothesis test

Important To summarize: a hypothesis test can be considered as a four step process

1. Formally writing out the null and alternative hypothesis.
2. Computing the test statistic.
3. Determining the critical value.
4. Using the decision rule.

Conclusions in Hypothesis Testing

Important

- We always test the null hypothesis.
- We either ***reject*** the null hypothesis, or
- We ***fail to reject*** the null hypothesis.
- our conclusion is always one of these two.

p-values

(Mentioned Previously, but discussed again)

- The p-value (or P-value or probability value) is the probability of getting a value of the test statistic that is at least as extreme as the one representing the sample data, assuming the null hypothesis is true.
- The null hypothesis is rejected if the p-value is very small, such as less than 0.05.
- When performing an inference procedure on a computer, it is much more common to use the p-value as a basis for decision, rather than the critical value

p-values

- The less likely the obtained results (or more extreme results) under the null hypothesis, the more confident one should be that the null hypothesis is false.The null hypothesis should not be rejected once and for all. The possibility that it was falsely rejected is always present, and, all else being equal, the lower the p-value, the lower this possibility.
- According to this view, research reports should not contain the p-value, only whether or not the values were significant (at or below the significance level).
- However it is much more reasonable to just report the p-values. That way each reader can make up his or her mind about just how convinced they are that the null hypothesis is false.

Hypothesis Testing

Recall: the inferential step to conclude that the null hypothesis is false goes as follows: The data (or data more extreme) are very unlikely given that the null hypothesis is true. This means that:

(1) a very unlikely event occurred or

(2) the null hypothesis is false.

The inference usually made is that the null hypothesis is false. Importantly it doesnt prove the null hypothesis to be false.

Type I and II errors

There are two kinds of errors that can be made in hypothesis testing:

(1) a true null hypothesis can be incorrectly rejected

(2) a false null hypothesis can fail to be rejected.

The former error is called a ***Type I error*** and the latter error is called a ***Type II error***.

The probability of Type I error is always equal to the level of significance α (alpha) that is used as the standard for rejecting the null hypothesis .

Type II Error

- The probability of a Type II error is designated by the Greek letter beta (β).
- A Type II error is only an error in the sense that an opportunity to reject the null hypothesis correctly was lost.
- It is not an error in the sense that an incorrect conclusion was drawn since no conclusion is drawn when the null hypothesis is not rejected.

Types of Error

- A Type I error, on the other hand, is an error in every sense of the word. A conclusion is drawn that the null hypothesis is false when, in fact, it is true.
- Therefore, Type I errors are generally considered more serious than Type II errors.
- The probability of a Type I error (α) is set by the experimenter.
- There is a trade-off between Type I and Type II errors. The more an experimenter protects himself or herself against Type I errors by choosing a low level, the greater the chance of a Type II error.

Types of Error

- Requiring very strong evidence to reject the null hypothesis makes it very unlikely that a true null hypothesis will be rejected.
- However, it increases the chance that a false null hypothesis will not be rejected, thus increasing the likelihood of Type II error.
- The Type I error rate is almost always set at 0.05 or at 0.01, the latter being more conservative since it requires stronger evidence to reject the null hypothesis at the 0.01 level then at the 0.05 level.
- **Important** In this module, the significance level α can be assumed to be 0.05, unless explicitly stated otherwise.

Type I and II errors

These two types of errors are defined in the table below.

	True State: H0 True	True State: H0 False
Decision: Reject H0	Type I error	Correct
Decision: Do not Reject H0	Correct	Type II error

Type I and Type II errors - Die Example

- Recall our die throw experiment example.
- Suppose we perform the experiment twice with two different dice.
- We don't not know for sure whether or not either of the dice is fair or crooked (favouring high values).
- Suppose we get a sum of 401 from one die, and 360 from the other.

Type I and Type II errors - Die Example

- For our first dice (sum 401), we feel that it is likely that the die is crooked.
- A Type I error describes the case when in fact that dice was fair, and what happened was just an unusual result.
- For our second dice (sum 360), we feel that it is likely that the die is fair.
- A Type II error describes the case when in fact that dice was crooked, favouring high values, and what happened was, again, just an unusual result.