Statistics for Computing

MA4413 Lecture 7A

Kevin O'Brien

Kevin.obrien@ul.ie

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

Autumn Semester 2013

Margin of Error

- The product of the quantile and the standard error give us the width of the confidence interval
- The width of the confidence interval is known as the ***margin of error***.

$$\text{Margin of Error} = [\text{Quantile} \times \text{Standard Error}]$$

- The margin of error gives us some idea about how uncertain we are about the unknown population parameter.
- A very wide interval may indicate that more data should be collected before anything very definite can be said about the parameter.
- The only way to control the margin of error is to adjust the sample size accordingly.
- By choosing an appropriate sample size, it is possible to ensure that the margin of error does not excess a certain threshold.

Point Estimates for proportions (1)

(From Last Class)
Of a sample of 160 computer programmers, 56 reported than Python was their primary programming language.
Let π be the proportion of all programmers who regard Python as their programming language. What is the point estimate for π?

$$\hat{p} = \frac{x}{n} \times 100\% = \frac{56}{160} = 35\%$$

Confidence Interval for a proportion (2)

Refer back to our earlier example of the proportion of Python programmers. Compute a 95% confidence interval.

Determining the Quantile The confidence level is 95%. The sample size is greater than 30. Therefore the appropriate quantile is 1.96.

Computing the Standard Error

$$S.E.(\hat{p}) = \sqrt{\frac{35 \times 65}{160}} = 3.77\%$$

Confidence Interval for proportion

$$35 \pm (1.96 \times 3.77)\% = (35 \pm 7.4)\% = (27.61\%, 42.39\%)$$

Student's $t-$distribution (1)

(Revision from last class, but very important).

- We indicated that use of the normal distribution in estimating a population mean is warranted for any large sample ($n > 30$).
- For a small sample ($n \leq 30$) only if the population is normally distributed **and** σ is known, the standard normal distribution can be used compute quantiles. In practice,this case is unusual.
- Now we consider the situation in which the sample is small and the population is assumed to be normally distributed, but σ is not known.
- We use the ***Student's $t-$distribution*** for small samples.

Student's $t-$distribution (2)

- Student's $t-$distribution is a variation of the normal distribution, designed to factor in the increased uncertainty resulting from smaller samples.
- The distribution is really a family of distributions, with a somewhat different distribution associated with the degrees of freedom (df). For a confidence interval for the population mean based on a sample of size n, $df = n - 1$.

Student's $t-$distribution (3)

- With increasing sample size, the $t-$distribution approaches the form of the standard normal ('Z') distribution.
- In fact the standard normal distribution can be thought of as the $t-$distribution with ∞ degrees of freedom.
- For computing quantiles, we will consider the 'Z' distribution in this way. (i.e. from now on, we will use only the Student $t-$distribution for Inference Procedures, as it can simplify matters greatly)
- For values of n much greater then 30, the difference between using $df = n - 1$ and $df = \infty$ is negligible.
- (As this will be relevant later, remember that a confidence interval is a **two-tailed** procedure, i.e. $k = 2$.)

Confidence Interval for a Mean (Small Sample)

- The mean operating life for a random sample of $n = 10$ light bulbs is $\bar{x} = 4,000$ hours, with the sample standard deviation $s = 200$ hours.
- The operating life of bulbs in general is assumed to be approximately normally distributed.
- We estimate the mean operating life for the population of bulbs from which this sample was taken, using a 95 percent confidence interval as follows:

$$4,000 \pm (2.262)(63.3) = (3857, 4143)$$

- The point estimate is 4,000 hours. The sample standard deviation is 200 hours, and the sample size is 10. Hence

$$S.E(\bar{x}) = \frac{200}{\sqrt{10}} = 63.3$$

- From Murdoch Barnes Table 7, the $t-$quantile with $df = 9$ is 2.262. (95% confidence interval so use 0.0250 column).

Interpreting Confidence Intervals

- In the previous lectures, we looked at confidence intervals, noting that these intervals are a pair of limits defining an interval.
- Often, we can use confidence intervals to make inferences on a population parameter.
- Consider the following example: Suppose that, when considering the leaving cert points of two groups of students A and B, the difference of the sample means was found to be $\bar{x}_B - \bar{x}_A = 30$ points.
- We would surmise than the average points level for group B is higher.
- Lets suppose that the 95% confidence interval was $(-15, 75)$ points. Consider what each of the two numbers mean,

Interpreting Confidence Intervals

- The upper bound $(+75)$ suggests that those in group B could have, on average, 75 more points than those in group A.
- But the lower bound (-15) suggests that those in group A could have, on average, 15 more points than those in group B.
- Also, the confidence interval allows for the possibility of both groups having equal means (i.e. $\bar{x}_B - \bar{x}_A = 0$)
- Essentially we can not be 95% confident that group B has a higher mark than group A.

Confidence Intervals

- So far, we have studied two types of confidence interval, a confidence interval for a sample mean and for a sample proportion.
  (Later we will call these ***One Sample*** confidence Intervals.
- There are more types of confidence intervals that we will cover later in this course.
  (We shall refer to these confidence intervals as the ***Two Sample*** confidence Intervals.
- We shall turn our attention to ***Hypothesis Testing*** in the mean time.

Introduction to Hypothesis tests

- In statistics, 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.
- A statistical test procedure is comparable to a trial where a defendant is considered innocent as long as his guilt is not proven.
- The prosecutor tries to prove the guilt of the defendant. Only when there is enough charging evidence the defendant is condemned.

Hypothesis tests (Null and Alternative Hypotheses)

- 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.

The Null Hypothesis

- The null hypothesis is what the experimenter supposes the outcome before the test is performed, based on prior assumptions (note: future remarks on the Dice experiment will be based on this view).
- An alternative view is that the null hypothesis is often the reverse of what the experimenter actually believes; it is put forward to allow the data to contradict it.
- In a hypothetical experiment on the effect of sleep deprivation, the experimenter probably expects sleep deprivation to have a harmful effect.
- If the experimental data show a sufficiently large effect of sleep deprivation, then the null hypothesis, expressing that sleep deprivation has no effect, can be rejected.

The Null Hypothesis

- Hypothesis tests are almost always performed using null-hypothesis tests.
- The rationale is as follows: "Assuming that the null hypothesis is true, what is the probability of observing a value for the test statistic that is at least as extreme as the value that was actually observed?"
- The critical region of a hypothesis test is the set of all outcomes which, if they occur, will lead us to decide that there is a difference.
- That is to say, cause the null hypothesis to be rejected in favour of the alternative hypothesis.

Writing the Null Hypothesis

- The null hypothesis is denoted H_0.
- 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 '$\geq$')

Writing the Null Hypothesis

Simple examples of null hypothesis (disregard context for the time being):

- H_0: $\mu = 350$. Population mean is 350.
- H_0: $\pi = 70\%$. Population proportion is 70%.
- H_0: $\mu \leq 100$. Population mean is less than or equal to 100.
- H_0: $\pi \geq 60\%$. Population proportion is greater than or equal to 60%.

Writing the Null Hypothesis (Dice Example)

- Recall our experiment of throwing a dice 100 times and computing the result, performed using a fair die and a crooked die.
- Suppose we perform this experiment again. We do not know whether the die we are using is fair or crooked. As we have no reason to believe otherwise, we will assume the dice is fair.
- We expect a result close to 350. This can be our null hypothesis.
- We will write this as H_0: $\mu = 350$. The die is fair.

Writing the Alternative Hypothesis

- The alternative hypothesis is denoted H_1 (or H_a)
- It will express the precise opposite argument of the null hypothesis, again mathematically with a written description of the hypothesis.
- H_1 use the following relational operators; ‘ $\neq$ ’ , ‘$<$’ or ‘$>$’, depending on the null hypothesis.
- H_1 will never contain a condition of equality.

Writing the Alternative Hypothesis

Simple examples of Alternative hypothesis (based on previous example):

- H_0: $\mu = 350$. Therefore H_1: $\mu \neq 350$. (Die Throws Example)
- H_0: $\pi = 70\%$. Therefore H_1: $\pi \neq 70\%$.
- H_0: $\mu \leq 100$. Therefore H_1: $\mu > 100$.
- H_0: $\pi \geq 60\%$. Therefore H_1: $\pi < 60\%$.

Remember to provide a brief written description for both hypotheses.

Number of Tails (For Later)

- 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

- 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 results of the experiment and the null hypothesis is determined.(i.e. 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 (Dice 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 α

- 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 use a significance level of $\alpha = 0.05$ only in this module. You may assume this level unless clearly stated otherwise

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. 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.

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 more. The first two steps are the same.

- 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.

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

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

The Test Statistic (TS)

- 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).
- We will use the initials TS for the sake of brevity.
- The test statistic is therefore

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

The Critical Value (CV)

- 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.
- We will use the initials CV for the sake of brevity.
- 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.

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.
- In other words, the critical region for a one-sided test is the set of values less than the critical value of the test, or the set of values greater than the critical value of 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.)

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, H0 are located in both tails of the probability distribution.
- In other words, the critical region for a two-sided test is the set of values less than a first critical value of the test and the set of values greater than a second critical value of the test.
- 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.)

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. (For later - One tailed or two tailed : 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”.

Critical Region

- $|TS| > CV$ 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.

Performing a Hypothesis test

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

- Formally writing out the null and alternative hypothesis.
- Computing the test statistic.
- Determining the critical value.
- Using the decision rule.