0.1 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.

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.

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 also. 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

- 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.

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".

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.(More on this in next class)

Critical Region

In class : graphical representation of material is scheduled here.

Performing a Hypothesis test

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.

- μ_d mean value for the population of differences.
- $\bar{d}$ mean value for the sample of differences,
- s_d standrd deviation of the differences for the paired sample data.
- n number of pairs

Conclusions in hypothesis testing

- We always test the null hypothesis.
- We reject the null hypothesis, or
- We *fail to reject* the null hypothesis.

Test statistics for testing a claim about a mean, when the population variance is known.

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

0.2 Introduction to Hypothesis tests

> The process by which we use data to answer questions about parameters is very similar to how juries evaluate evidence about a defendant. from Geoffrey Vining, Statistical Methods for Engineers, Duxbury, 1st edition, 1998.

- Setting up and testing hypotheses is an essential part of statistical inference. In order to formulate such a test, usually some theory has been put forward, either because it is believed to be true or because it is to be used as a basis for argument, but has not been proved, for example, claiming that a new drug is better than the current drug for treatment of the same symptoms.
- In each problem considered, the question of interest is simplified into two competing claims / hypotheses between which we have a choice; the ***null hypothesis***, denoted H_0, against the ***alternative hypothesis***, denoted H_1. These two competing claims / hypotheses are not however treated on an equal basis: special consideration is given to the null hypothesis.
- We have two common situations:
  1. The experiment has been carried out in an attempt to disprove or reject a particular hypothesis, the null hypothesis, thus we give that one priority so it cannot be rejected unless the evidence against it is sufficiently strong. For example,
     H_0: there is no difference in taste between coke and diet coke against H_1: there is a difference.
  2. If one of the two hypotheses is 'simpler' we give it priority so that a more 'complicated' theory is not adopted unless there is sufficient evidence against the simpler one. For example, it is 'simpler' to claim that there is no difference in flavour between coke and diet coke than it is to say that there is a difference.
- The hypotheses are often statements about population parameters like expected value and variance; for example H_0 might be that the expected value of the height of ten year old boys in the Scottish population is not different from that of ten year old girls. A hypothesis might also be a statement about the distributional form of a characteristic of interest, for example that the height of ten year old boys is normally distributed within the Scottish population.
- The outcome of a hypothesis test test is "Reject H_0: in favour of H_1" or "Do not reject H_0".

0.2.1 Hypothesis testing: introduction

The objective of hypothesis testing is to access the validity of a claim against a counterclaim using sample data

- The claim to be proved is the alternative hypothesis(H_1).
- The competing claim is called the null hypothesis(H_0).
- One begins by assuming that H_0 is true.

If the data fails to contradict H_0 beyond a reasonable doubt, then H_0 is not rejected. However, failing to reject H_0 does not mean that we accept it as true. It simply means that H_0 cannot be ruled out as a possible explanation for the observed data. A proof by insufficient data is not a proof at all.

0.3 Steps of 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.

0.3.1 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 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).

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.

0.3.2 Introduction to Hypothesis tests

- A hypothesis test is a method of making decisions using experimental data.
- An outcome is called ***statistically significant*** if it is unlikely to have occurred by chance. (i.e. sampling fluctuation)
- 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 Testing

Normal distribution

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} = .3989 e^{-5z^2}$$

- Hypothesis Testing Important Considerations
- 1) Sample size n
  - Is it large or small?
  - "small" less than or equal to thirty.
- 2) Significance level
  - 95% confidence means = 0.05
  - 99% confidence means = 0.01
- 3) Number of tails in procedure
  - Procedures are either one tailed or two tailed. k=1 or 2
  - Confidennce intervals are always two tailed.

- 4) Standard Error Formula

  Back of exam paper

- 5) Tables to use

  - Table 3 - Normal "Z" Distribution
  - Table 7 - student's "t" Distribution
  - Table 8 - Chi Square Distribution

- 6) Degrees of freedom

  notation is sometimes

  Large samples df = Small samples df = n-1

  Chi-Square df = (r-1)x(c-1) r = number of rows c = number of columns

- 7) Hypothesis tests usually have the following format.

  Step 1 : Formally state the null and alternative hypotheses

  Step 2 : Determine the test statistic

  Step 3 : Determine the critical value

  Step 4 : Decision Rule

- 8) Hypothesis testing using p-values

  If asked to use p-value, we have a slightly diffferent approach. The first two steps are the same as in note 7.

  Step A : Formally state the null and alternative hypotheses

  Step B : Determine the test statistic

  Step C : Determine the p-value

  Step D : Decision Rule for p-values.

- 9) What is a p-value?

  $$\text{P-Value} = P(Z|T_S|)$$

  TS: Test Statistic

  P-value is found from Murdoch Barnes Tables 3

  For example, if the test statistic is 1.96, then the p-value is 0.025

- 10) How to interpret the p-value

  (see previous notes)

  is the significance value

  k is the number of tails

  If the p-value is less than k, we reject the null hypothesis. If the p-value is greater than k, we reject the null hypothesis.

- 11) The general structure of a test statistic

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

- 12) General Structure of a Confidence Interval

  Quantiles are compute the same way as critical values.

0.3.3 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***.

0.3.4 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

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

0.4 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.
  (For this module, we will mostly stick with 0.05.)

0.4.1 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).

0.4.2 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

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

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

0.4.3 2 sided test

A two-sided test is used when we are concerned about a possible deviation in either direction from the hypothesized value of the mean. The formula used to establish the critical values of the sample mean is similar to the formula for determining confidence limits for estimating the population mean, except that the hypothesized value of the population mean m0 is the reference point rather than the sample mean.

Test statistics for testing a claim about a mean, when the population variance is known.

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Conclusions in hypothesis testing

- We always test the null hypothesis.
- We reject the null hypothesis, or
- We *fail to reject* the null hypothesis.

0.5 Hypothesis Testing

0.5.1 Hypothesis Testing: Number of Tails

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

0.5.2 Inference : Structure of a Hypothesis Test (1)

- Formally write out the null and Alternative Hypothesis.
  - Denote the null as H_0 and the alternative as H_1.
  - Use the parameter values (i.e. μ and π), not the sample estimates.
  - Remember to provide a brief description of each hypothesis.

0.5.3 Inference : Structure of a Hypothesis Test (2)

- Compute the Test Statistic (TS)
  - You will need to compute the value for Standard Error (See back of exam paper).
  - The general structure is
$$\frac{\text{observed value} - \text{null value}}{\text{Standard Error}}$$
  - The p-value is computed as $P(Z \geq |TS|)$ (from Murdoch Barnes 3). N.B. p-value is for large samples only.

0.5.4 Inference : Structure of a Hypothesis Test (3)

- Determine the Critical Value
  - You will need to know the sample size (n), the significance (α) , and the number of tails (k).
  - In this module, $\alpha = 0.05$ and $k = 2$ always.
  - Depending on the sample size the degrees of freedom is $\nu = n - 1$ 9 when $n \leq 30$ or $\nu = \infty$ when $n > 30$

0.5.5 Inference : Structure of a Hypothesis Test (4)

- Making a decision (Critical Value) : Is the absolute value of the Test Statistic greater than the Critical Value?
  - If $|TS| > CV$ We reject the null hypothesis.
  - If $|TS| \leq CV$ We fail to reject the null hypothesis.

0.5.6 Inference : Structure of a Hypothesis Test (4)

- Making a decision (p-value) :
  Is the p-value less than than the critical threshold α/k.?
  - If p-value $< \alpha/k$: We reject the null hypothesis.
  - If p-value $\geq \alpha/k$: We fail to reject the null hypothesis.

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

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

- 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

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

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.

0.5.7 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 probability computed in a significance test.
- The convention in hypothesis testing is that the null hypothesis (H_0) is assumed to be true.
- 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.
- The difference between the statistic computed in the sample and the parameter specified by H_0 is computed and the probability of obtaining a difference this large or large is calculated.
- This probability value is the probability of obtaining data as extreme or more extreme than the current data (assuming H_0 is true).

Hypothesis Testing and p-values

- However, the test does not **prove** the person cannot predict better than chance; it simply fails to provide evidence that he or she can.
- The probability that the null hypothesis is true is not determined by the statistical analysis conducted as part of hypothesis testing.
- Rather, the probability computed is the probability of obtaining data as different or more different from the null hypothesis (given that the null hypothesis is true) as the data actually obtained.

0.5.8 Hypothesis Testing(Contd)

It is not the probability of the null hypothesis itself. Thus, if the probability value is 0.005, this does not mean that the probability that the null hypothesis is either true or false is .005. It means that the probability of obtaining data as different or more different from the null hypothesis as those obtained in the experiment is 0.005.

- To illustrate that the probability is not the probability of the hypothesis, consider a test of a person who claims to be able to predict whether a coin will come up heads or tails.
- One should take a rather sceptical attitude toward this claim and require strong evidence to believe in its validity.

- The null hypothesis is that the person can predict correctly half the time ($H_0 : \pi = 0.5$). In the test, a coin is flipped 20 times and the person is correct 11 times.
- If the person has no special ability (H_0 is true), then the probability of being correct 11 or more times out of 20 is 0.41.
- Would someone who was originally sceptical now believe that there is only a 0.41 chance that the null hypothesis is true?

0.5.9 Hypothesis Testing

- They almost certainly would not since they probably originally thought H_0 had a very high probability of being true (perhaps as high as 0.9999).
- There is no logical reason for them to decrease their belief in the validity of the null hypothesis since the outcome was perfectly consistent with the null hypothesis.

The proper interpretation of the test is as follows:

- A person made a rather extraordinary claim and should be able to provide strong evidence in support of the claim if the claim is to believed.
- The test provided data consistent with the null hypothesis that the person has no special ability since a person with no special ability would be able to predict as well or better more than 40% of the time.
- Therefore, there is no compelling reason to believe the extraordinary claim.

- 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 alcohol, the experimenter probably expects sleep deprivation to have a harmful effect.
- If the experimental data show a sufficiently large effect of sleep deprivationl, then the null hypothesis that sleep deprivationl has no effect can be rejected.
- Hypothesis tests are almost always made using null-hypothesis tests i.e., tests that answer the question 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, cause the null hypothesis to be rejected in favour of the alternative hypothesis.
- Selecting a suitable critical region is arbitrary.

0.5.10 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.

0.5.11 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).

0.6 Guidance on hypothesis testing

For the most part, this module will use the $p-$value approach to interpreting a hypothesis test.

For the sake of simplicity, we will

- If the p-value is less than 0.02, we reject the null hypothesis.
- If the p-value is greater than 0.02, we fail to reject the null hypothesis.

Again, the choice of 0.02 as a threshold is arbitrary. Conventionally a p-value between 0.01 and 0.05 would indicate that the testing procedure should be re-appraised, rather than being used as a basis for a decision.

However, later on, we will disgress from this approach, using the "star" system, which is directly implemented with some statistical procedures.

Important Steps in a Hypothesis Test Procedure

- Hypothesis Testing Important Considerations
- 1) Sample size n
  - Is it large or small?
  - "small" less than or equal to thirty.

- 2) Significance level
  - 95% confidence means = 0.05
  - 99% confidence means = 0.01
- 3) Number of tails in procedure
  - Procedures are either one tailed or two tailed. k=1 or 2
  - Confidennce intervals are always two tailed.
- 4) Standard Error Formula

  Back of exam paper
- 5) Tables to use
  - Table 3 - Normal "Z" Distribution
  - Table 7 - student's "t" Distribution
  - Table 8 - Chi Square Distribution
- 6) Degrees of freedom

  notation is sometimes

  Large samples df = Small samples df = n-1

  Chi-Square df = (r-1)x(c-1) r = number of rows c = number of columns
- 7) Hypothesis tests usually have the following format.

  Step 1 : Formally state the null and alternative hypotheses

  Step 2 : Determine the test statistic

  Step 3 : Determine the critical value

  Step 4 : Decision Rule
- 8) Hypothesis testing using p-values

  If asked to use p-value, we have a slightly diffferent approach. The first two steps are the same as in note 7.

  Step A : Formally state the null and alternative hypotheses

  Step B : Determine the test statistic

  Step C : Determine the p-value

  Step D : Decision Rule for p-values.
- 9) What is a p-value?

  $$\text{P-Value} = P(Z|T_S|)$$

  TS: Test Statistic

  P-value is found from Murdoch Barnes Tables 3

  For example, if the test statistic is 1.96, then the p-value is 0.025

- 10) How to interpret the p-value

  (see previous notes)

  is the significance value

  k is the number of tails

  If the p-value is less than k, we reject the null hypothesis. If the p-value is greater than k, we reject the null hypothesis.

- 11) The general structure of a test statistic

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

- 12) General Structure of a Confidence Interval

  Quantiles are compute the same way as critical values.

0.6.1 Conclusions in hypothesis testing

- We always test the null hypothesis.
- We reject the null hypothesis, or
- We *fail to reject* the null hypothesis.