Inference Worksheet

Step 1 State the Hypotheses

We begin by stating the value of a population mean that we are claiming in a null hypothesis. So, we may hypothesise that children watch TV for 3 hours per day. So it would be that the population mean is 3 hours per day. This is the starting point – like the proposition of innocence in a court room.

When a defendant is on trial, the jury starts by assuming that the defendant is innocent. The basis of the decision is to determine whether this is likely based on the evidence produced. Similarly in hypothesis testing, we start by assuming the null hypothesis is true unless we have enough evidence to dispute that.

Definition: Null hypothesis (H_0)

The null hypothesis is a statement about a population parameter, such as the population mean, that is assumed to be true.

The alternative hypothesis is what we think it might be. We have a choice of three alternatives. For the children we might think the number of hours is more than 3, or less than 3 or just not 3.

Definition: Alternative Hypothesis (H_1)

An alternative hypothesis is a statement that contradicts the null hypothesis by stating that the actual value of a population parameter is less than, greater than or not equal to the value stated in the null hypothesis.

Step 2: Set the Criteria for the Decision

For this we must state the level of significance for a test i.e., beyond reasonable doubt. We are collecting data that will either support, or not, the null hypothesis. The level of significance is typically set at 5% for behavioural studies, often lower (say 1%) for medical studies.

When the probability of obtaining the sample mean value we have is less than 5%, if the null hypothesis is true, then we conclude it is unlikely that the null hypothesis is true.

This probability will change for each sample we take so can never say we have ‘proved’ anything – there is a small chance that this value could have been obtained if the null hypothesis is true. This is determined by our choice of criteria.

Level of significance

This refers to a criterion of judgement upon which a decision is made regarding the value stated in a null hypothesis. The criterion is based on the probability of obtaining a statistic measured in a sample if the value stated in the null hypothesis were true.

The alternative hypothesis establishes where to place the level of significance.

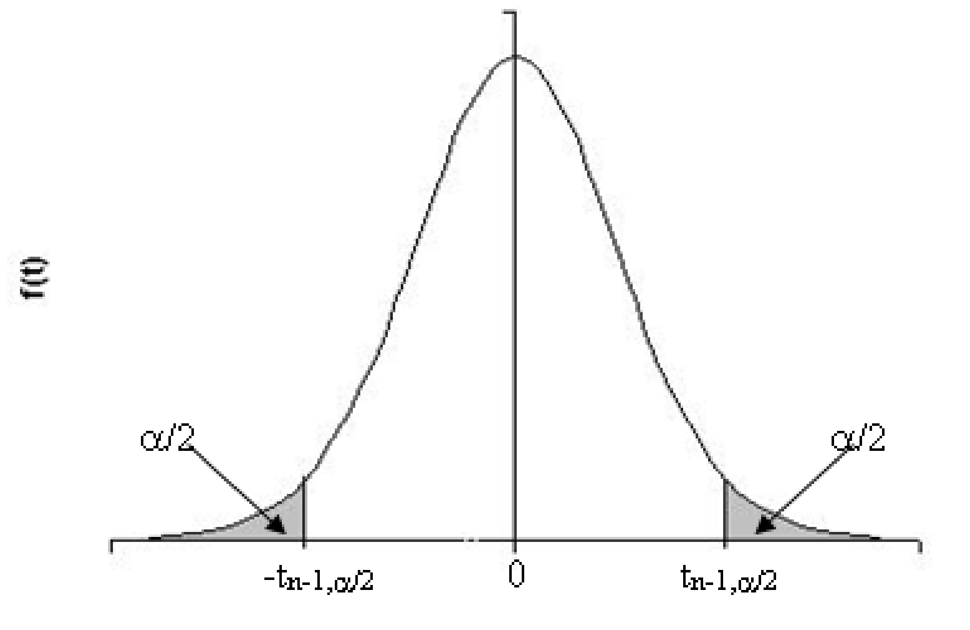

Figure 1 A normal distribution curve showing two extremities of alpha over 2.

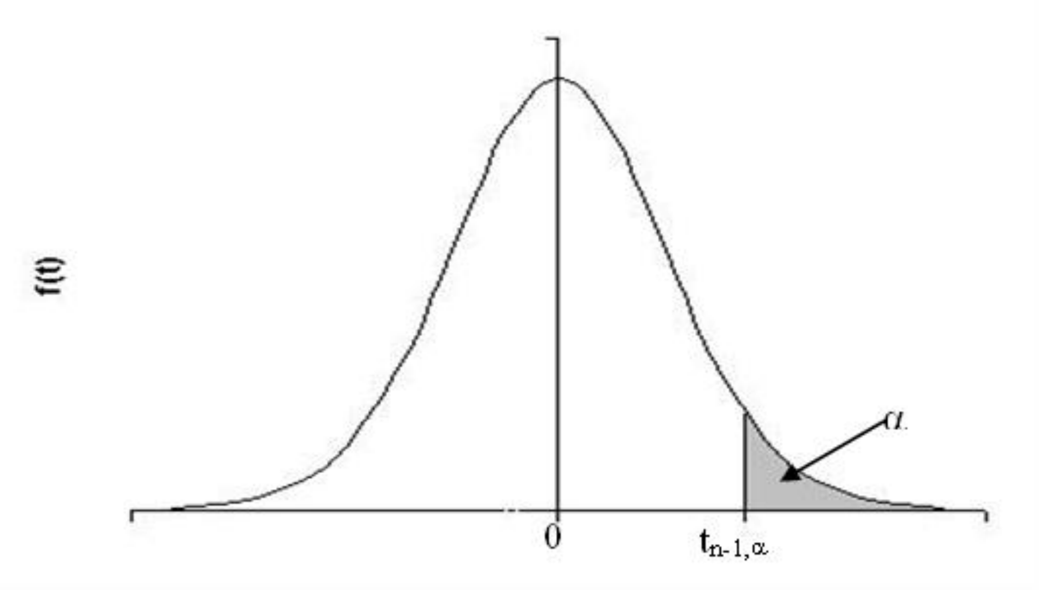

Figure 2 A normal distribution curve showing a right-most extremity of alpha.

Step 3: Compute the Test Statistic

Suppose we measure a sample mean of 4 hours for the children. To decide we need to decide how likely this outcome is if the population mean is 3. We use a test statistic to determine this likelihood. Specifically, a test statistic tells us how far a sample mean is from the population mean. It is not just a matter of distance as this will also be dependent on the standard deviation.

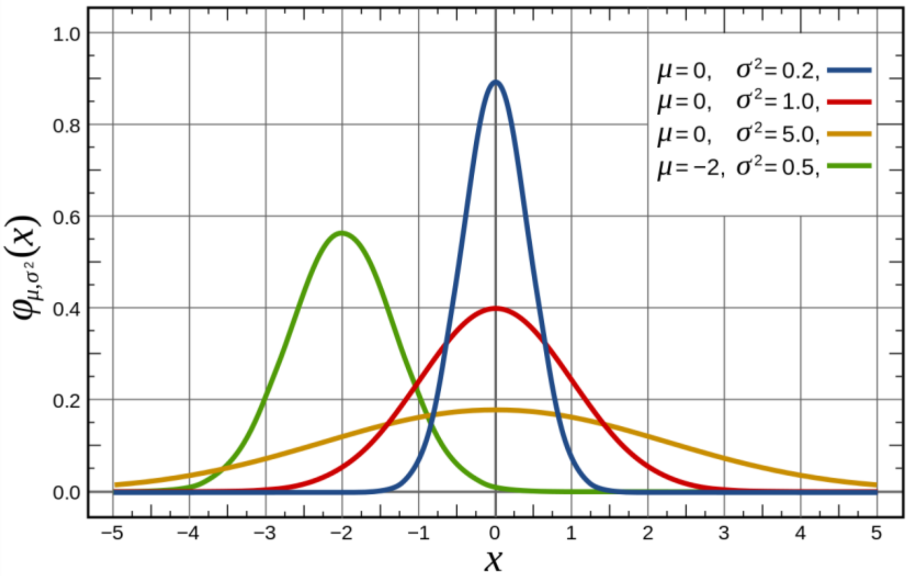

Figure 3 Figure showing different distributions.

The test statistic is a mathematical formula that allows researchers to determine the likelihood of obtaining sample outcomes if the null hypothesis were true.

Step 4: Make a Decision

We use the value of the test statistic to decide about the null hypothesis. The decision is based on the probability of obtaining a sample value, given the value in the null hypothesis is true. If the probability of obtaining a sample value is less than 5% when the null hypothesis is true, then the decision is to reject the null hypothesis. If the probability of obtaining a sample value is greater than 5% when the null hypothesis is true, then the decision is to accept the null hypothesis.

So, there are two decisions we can make:

1. Reject the null hypothesis.
2. Accept the null hypothesis.

The probability of obtaining a sample value, given that the value stated in the null hypothesis is true is stated by the p-value. This is a probability. In Step 2 we stated the criterion for decisions.

p-value: A p-value is the probability of obtaining a sample value, given that the value stated in the null hypothesis is true. The p-value for obtaining a sample value is then compared to the level of significance.

If we have set out criteria at 5% then if the:

p-value is less than 0.05 then we reject the null hypothesis.

p-value is greater than 0.05 then we accept the null hypothesis.

Hypothesis Testing and Sampling Distributions

The logic of hypothesis testing is rooted in sampling distributions. If we are considering a population mean, then we know:

1. The sample mean is an unbiased estimator of the population mean. On average, a randomly selected sample will have a mean equal to that of the population.

2. Regardless of the distribution of the population, the sampling distribution of the sample mean is normally distributed. That is why we can always use the normal distribution for the testing of means.

To locate the probability of obtaining a sample mean in a sampling distribution we must know:

1. The population mean.
2. The standard error of the mean.

In Step 4, because we are observing a sample and not a population, it is possible that the conclusion could be wrong.

The following table shows there are 4 decision alternatives:

1. Accepting the null hypothesis is correct.
2. Accepting the null hypothesis is wrong.
3. Rejecting the null hypothesis is correct.
4. Rejecting the null hypothesis is wrong.

	Accept null hypothesis	Reject null hypothesis
True	Correct 1-α	Type 1 error α
False	Type 2 error β	Correct 1-β Power

The Decision to Reject the Null Hypothesis

When we decide to reject the null hypothesis, we can be right or wrong. If it is wrong, then this is called a Type 1 error. A researcher who makes this decision reject previous notions of the truth that are in fact true. This is analogous to finding an innocent party guilty. The level of this error, α, is the level of significance or criteria that we set. Which is why this is set low in medical studies. α is the largest probability of committing a Type 1 error that we will allow.

Reporting Significance Levels

When reporting significance levels, great care must be taken not to confuse them with confidence coefficients (as has sadly happened in some recently published textbooks!) *If a significance test results in rejection of the null hypothesis at significance level a, this **emphatically does not** mean that we have 100(1 - a) % assurance that the null hypothesis is false.*

University of Essex

Online

Setting Up Hypotheses in Advance

When performing a hypothesis test, it is of particular importance that the relevant null and alternative hypotheses are set up in advance, before the data have been inspected. Only in this way is an impartial test possible.

Once the data have been scrutinised, it is possible that they will influence the investigator's judgement as to which form of test is most appropriate, especially as it is always a little "easier" to obtain a significant result on an appropriate one-tailed test.

In fact, it is quite dishonest to select a one-tailed test seeking a difference in the direction indicated by the data. A little careful thought will confirm that such a procedure is equivalent to performing a two-tailed test with the significance level a in each tail. Effectively, this is a two-tailed test with significance level 2a. The dishonesty lies in reporting the results to be significant at level a (one-tailed), when, in fact they are only significant at level 2a (two-tailed). They may not even be statistically significant at all.

Hypothesis Tests and "Proof"

Recall that the null hypothesis is rejected if the relevant test statistic is "consistent with the alternative hypothesis", and has a low probability of occurrence if H_0 is, in fact, true (i.e. a low p-value). We could only *disprov*e H_0 if the observed value of the test statistic was *impossible* if H_0 were true (i.e. it had a p-value of exactly zero). This cannot, in fact, occur in practice, so avoid claiming to have proved or disproved anything with a statistical test.

Related versus Unrelated Samples

Related samples tests are usually more powerful than their unrelated samples equivalents for a given sample size. In other words, related samples tests have a higher probability of correctly rejecting a false null hypothesis. This is because individual differences that inflate the variability of the data in unrelated samples virtually "cancel out" when taking the relevant differences in related samples.

However, if the same individuals are required to contribute two observations under different conditions, it is likely that some will "drop out" before the second observation is taken. Such dropout represents a potentially serious practical problem that should be borne in mind when choosing an appropriate study design. If in doubt, advice should be sought from a professional statistician.