Lecture 3 – Inference through hypothesis testing

There are many ways to draw inference in statistics. The most popular is hypothesis testing, as advocated by Sir Ronald Fisher.

In many ways, the formal procedure of hypothesis testing is much like the hypothetico-deductive scientific method advocated by Sir Karl Popper.

The scientist:
- observed nature,
- formulates a theory,
- and then tests this theory against observation.

In a statistical context, the scientist may pose a theory concerning one or more population parameters – for example that that they equal specified values. Then the scientist samples the population and compares observation against the theory. If the observations disagree with the theory then the scientist rejects the theory, else the scientist concludes that either the theory is true or that sample did not detect the difference between the real and hypothesized values of the population parameters. In this instance, the theory is a *hypothesis*.

The hypothesis
With hypothesis testing, we divide all possible interpretations of an experiment into two alternatives: a null hypothesis and an alternative hypothesis.

- A null hypothesis (H_0) is the statement that there is "no effect" or of the "status quo" or simply that what quantities we have observed are drawn by random chance.

- The alternative hypothesis (H_a) in turn is that there is an effect of some factor influencing the population or that the status quo is not in fact true. (Always bear in mind that there can be an infinite set of alternate hypotheses).

For example, if we wish to test that men and women have different mean heights, the null hypothesis would state that "Men and women are the same height on average." The alternative hypothesis would be "Men and women are NOT on average the same height."

This null hypothesis can be expressed mathematically as:

$$H_0 : \mu_{males} = \mu_{females}$$

where the alternative hypothesis is

$$H_a : \mu_{males} \neq \mu_{females}$$

We use the null hypothesis to generate the distribution of what sample statistics would look like, if the null hypothesis were true. We then compare the actual statistic we measured to this distribution of possible sample statistics to see how unusual it would be, if this null hypothesis were true.

Example:

What about a situation where you want to assess whether or not the mean size of a fish species is 110 mm. If we sampled the population and found that the mean was 105 mm then from Figure 3.1 below this would not be an unexpected result given that the distribution under the null hypothesis. Alternatively, if we found that the sample mean was 130 mm, then this would be odd and we would possibly reject the notion that the sample was from the null distribution.

So if we get a test statistic which would be very unlikely if the null hypothesis were true, we then can infer that the null hypothesis is probably <u>not</u> true.
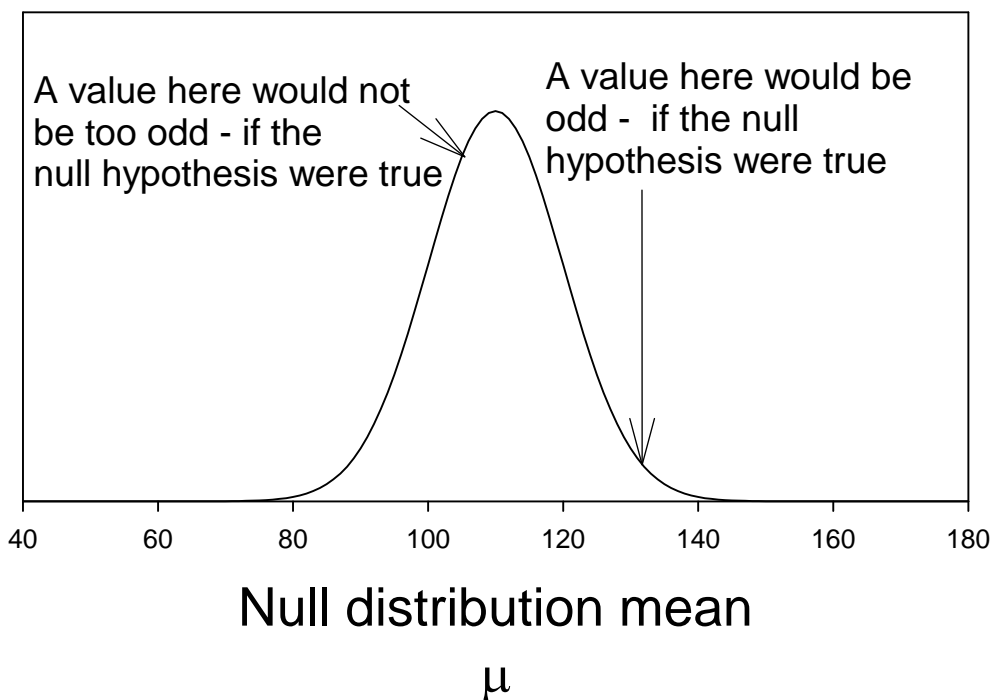

Figure 3.1. The null distribution and observed sample means. Are the means odd or not given you hypothesis about the population?
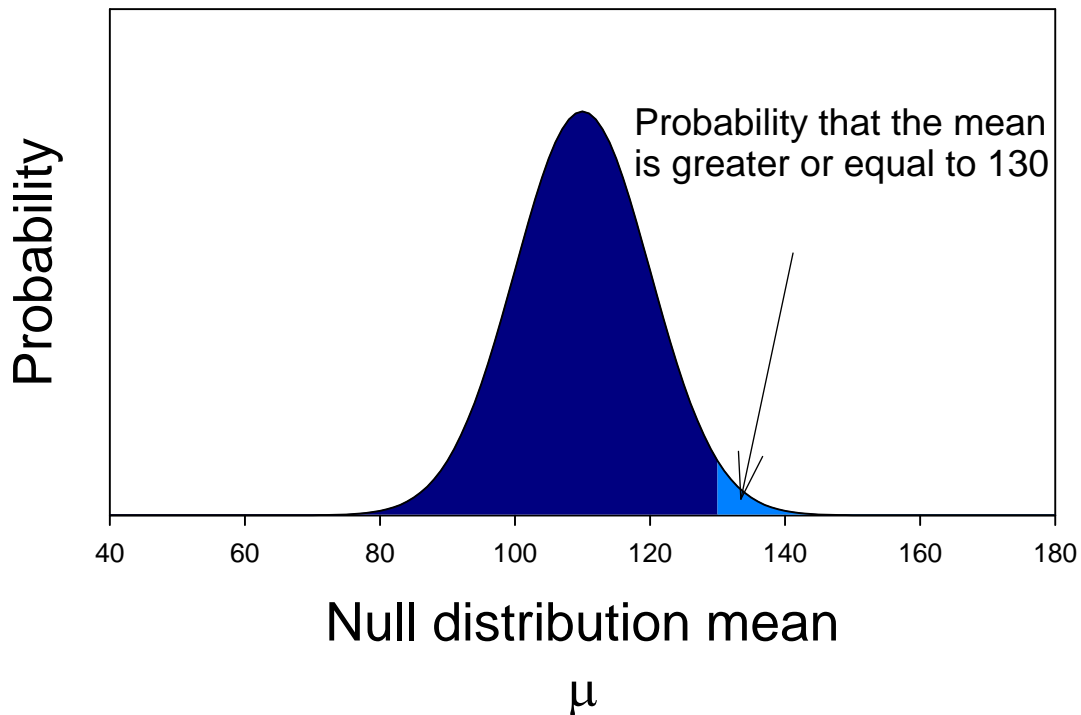
Figure 3.2. The probability, assuming that the null hypothesis is true, of getting a sample mean of at least 130. The light grey area represents 2% of the grey area – the area of the probability distribution. Finding this sample mean would therefore probably only happen in 2% of many, many samples.

Let us be more specific about what we mean by "odd" or what is the probability of getting a result as extreme or more in the possible range of results statistics as the one we actually got? This illustrated in Figure 3.2 where the probability of getting a sample mean of 130 or more is in fact 2%. Therefore we can conclude that we are 2% sure that the sample mean is probably NOT from the null distribution and reject the null hypothesis of 110 mm. This is because if we conducted this experiment over and over and over again, the sample means would be different from the null hypothesized mean about 98% of the time. We call the 2%, the level of significance. In most instances a significance level of 5% would suffice.

Towards a general statistic

The height example would be one possible option of the infinite number of data set available. We therefore need a simple distribution against which to compare our average.

If the null hypothesis were true, then we can often describe the probability distribution of different types of samples. This would be through our T statistic -
$T = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$ which has a t-distribution with (n-1) degrees of freedom.

If we go back to Figure 2.2 then we are assessing the difference between averages. Therefore, as we know that the averages of many, many samples are Normally distributed then we can compare and average to a null distribution constructed from averages.

Hypothesis testing recipe

The steps involved in conducting a hypothesis test are therefore:

Before data collection/observation:
- State the hypotheses H_0 and H_a;
- Choose and fix the significance level of the test (or α);
- Establish the critical region of the test corresponding to α where one would consider a test statistic to be odd.

After data collection/observation:
- Calculate the observed test statistic from the sample;
- Compare the calculated statistic with the null distribution
- Make a decision about the hypotheses.