Significance Testing

Testing a mean: In essence, a significance test (or *hypothesis test*) for the mean of a sample tests whether the sample mean is *significantly* different from the population mean. How do we measure significance? Suppose we set a 5% significance level, then we want to know if there is less than a 5% chance of the sample mean being what it is, given a particular population mean.

Typical examples:

- Patients have a 60% chance of recovering from a particular illness. A new drug is tested on 100 patients and 68 recover. Could this have happened by chance? If we find it is very unlikely then we can accept the fact that the drug is having a positive effect.
- A cereal manufacturer claims that his packets contain a mean weight of cereal of 200g. A rival samples 50 packets and finds the mean weight is only 190g. If this is significantly less than the nominal mean, then this will cast doubt on the manufacturer's claim.

Known σ: The following example takes you through the process stage by stage:

The IQ scores of a population are Normally distributed with a mean of 100 and a standard deviation of 15. A psychologist wishes to test the theory that eating chocolate improves your score. A random sample of 60 people are selected and they are each given a bar of chocolate before sitting an IQ test. Their mean score is 103. Test the psychologist's theory at the 5% level.

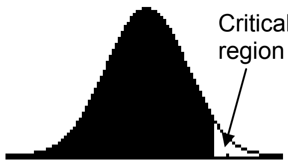
Stage 1: ***Set up the hypotheses.***

The *null* hypothesis is the status quo; in this case, that the mean IQ of the 60 people is 100. The *alternative* hypothesis is that the chocolate is having an effect, and that the mean score has increased. Write the hypotheses like this:

H_0: $\mu = 100$

H_1: $\mu > 100$

Stage 2: ***Calculate the probability.***



The diagram on the right shows the "unlikely" region (ie the top 5%) of a Normal curve. The value where this occurs is called the *critical value*, and the region is called the *critical region*. We could find this critical value and see whether 103 was in the critical region or not, but it is easier to find P(mean score > 103) and see whether this is greater than or less than 5%. Remembering that we are testing a distribution of means, we must use the standard error of the mean. So with $\mu = 100$ and $\frac{\sigma}{\sqrt{n}} = 1.936$ we find P(mean score >103) = 0.0607, or about 6.1%.

Stage 3: ***State the conclusion.***

103 is *not* in the critical region (a mean that high could have happened by chance), so we accept H_0 and conclude that there is no evidence to support the psychologist's claim.

One-tailed or two-tailed? In the example above, the claim was that chocolate increased IQ, so we were only testing if the mean was higher than normal. This was a *one-tailed test.* However, if the claim was that chocolate affected IQ (either way), then H_1 would