UNIVERSITY OF LONDON

Probability and Statistics: To p, or not to p?

Module Leader: Dr James Abdey

5.4 Testing a population mean claim

We consider the **hypothesis test of a population mean** in the context of a claim made by a manufacturer.

As an example, the amount of water in mineral water bottles exhibits slight variations attributable to the bottle-filling machine at the factory not putting in *identical* quantities of water in each bottle. The labels on each bottle may state '500 ml' but this equates to a **claim about the average contents** of all bottles produced (in the population of bottles).

Let X denote the quantity of water in a bottle. It would seem reasonable to assume a normal distribution for X such that:

$$X \sim N(\mu, \sigma^2)$$

and we wish to test:

$$\mathrm{H}_0 : \mu = 500\text{ml} \quad \text{vs.} \quad \mathrm{H}_1 : \mu \neq 500\text{ml}.$$

Suppose a random sample of $n = 100$ bottles is to be taken, and let us assume that $\sigma = 10$ ml. From our work in Section 4.5 we know that:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) = N\left(\mu, \frac{(10)^2}{100}\right) = N(\mu, 1).$$

Further suppose that the sample mean in our random sample of 100 is $\bar{x} = 503$ ml. Clearly, we see that:

$$\bar{x} = 503 \neq 500 = \mu$$

where 500 is the claimed value of μ being tested in H_0.

The question is whether the difference between $\bar{x} = 503$ and the claim $\mu = 500$ is:

(a) **due to sampling error** (and hence H_0 is true)?

(b) **statistically significant** (and hence H_1 is true)?

Determination of the p-value will allow us to choose between explanations (a) and (b).

We proceed by **standardising** $\bar{X}$ such that:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,\, 1)$$

acts as our **test statistic**. Note the test statistic includes the effect size, $\bar{X} - \mu$, as well as the sample size, n.

Using our sample data, we now obtain the test statistic value (noting the **influence of both the effect size and the sample size**, and hence ultimately the influence on the p-value):

$$\frac{503 - 500}{10/\sqrt{100}} = 3.$$

The p-value is the probability of our test statistic value or a more extreme value conditional on H_0. Noting that $H_1 : \mu \neq 500$, 'more extreme' here means a z-score > 3 and < -3. Due to the symmetry of the standard normal distribution about zero, this can be expressed as:

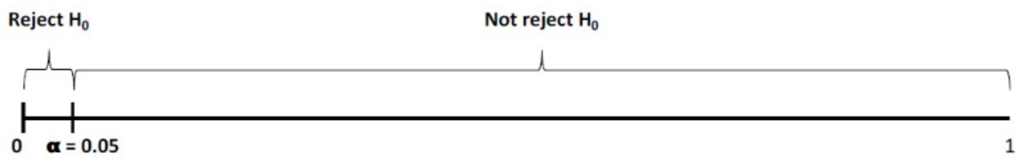
$$p\text{-value} = P(Z \geq |3|) = 0.0027.$$

Note this value can easily be obtained using Microsoft Excel, say, as:

`=NORM.S.DIST(-3)*2` **or** `=(1-NORM.S.DIST(3))*2`

where the function `NORM.S.DIST(z)` returns $P(Z \leq z)$ for $Z \sim N(0,\, 1)$.

Recall the **p-value decision rule**, shown below for $\alpha = 0.05$:



Therefore, since $0.0027 < 0.05$ we reject H_0 and conclude that the result is 'statistically significant' at the 5% significance level (and also, of course, at the 1% significance level). Hence there is (strong) evidence that $\mu \neq 500$. Since $\bar{x} > \mu$ we might go further and suppose that $\mu > 500$.

Finally, recall the possible **decision space**:

		Decision made	
		H_0 not rejected	H_0 rejected
True state of nature	H_0 true	Correct decision	Type I error
	H_1 true	Type II error	Correct decision

As we have rejected H_0 this means one of two things:

- **we have correctly rejected H_0**
- **we have committed a Type I error**.

Although the p-value is very small, indicating it is *highly unlikely* that this is a Type I error, unfortunately we cannot be *certain* which outcome has actually occurred!