HYPOTHESIS TESTING

We have already described how the sample information is used to estimate the parameters of the underlying population distributions. Now, we intend to use our sample information to test existing presumptions regarding the values of parameters. More generally, we are concerned with testing hypotheses regarding the state of nature.

The initial presumptions, which are formed without the benefit of the statistical evidence, constitute the null or prior hypothesis, denoted by H_0. The alternative hypothesis, denoted by H_1, is what we shall assert if we reject H_0 in the light of the statistical evidence.

Together, H_0 and H_1 ought to comprises all possibilities. Thus, the set Ω of all states of nature is partitioned as $\Omega = \{\Omega_0 \cup \Omega_1; \Omega_0 \cap \Omega_1 = \emptyset\}$ in a manner corresponding the hypotheses. The decision to maintain the null hypotheses after the evidence has been reviewed will be denoted by d_0; and the decision to reject it in favour of the alternative hypothesis will be denoted by d_1.

The procedure for testing an hypothesis will depend upon our forming a test statistic x from a random sample. The decision will depend upon the value of the statistic. Let $S = \{C \cup C^C; C \cap C^C = \emptyset\}$ be the sample space comprising all possible values of x. This is partitioned into the critical region C and its complement, which is the non-critical region C^C. Then, the decision rule is as follows:

$$x \in C^C \Longrightarrow d_0,$$

$$x \in C \Longrightarrow d_1.$$

Since x is a random variable, it has a finite probability of falling in either region, regardless of the true state of nature; and there are two ways in which the test procedure might mislead us, which are the errors indicated in the following table:

	d_0	d_1
Ω_0		*Type I error*
Ω_1	*Type II error*	

Here we have

Type I error: falsely rejecting the null hypothesis,

Type II error: falsely maintaining the null hypothesis.

We must adopt test procedures that minimise the probabilities of making the errors. We are constrained by a situation where reducing the probability of one type of error increases the probability of the other type. If we could reduce the probabilities both types of errors at the same time, then our test procedure would be an inefficient one that should replaced by another. (In the language of economics, the efficient tests constitute a Pareto-optimal choice set.)

Unless we have both a prior probability measure defined over Ω and a loss function or utility function relating to the costs of making the errors, we are bound to adopt a rule of thumb in fixing the probability of the *Type I* error before attempting to minimise the probability of the *Type II* error. Conventionally, we fix the probability of the *Type I error* at 1%, 5% or 10%, depending on the context of test procedure.

A two-tailed test for the mean. Let $x \sim N(\mu, \sigma^2)$, where σ^2 is known and $\mu \in \Omega$ lies somewhere on the real line, which means that $\Omega = (-\infty, \infty)$. Let the null hypothesis be $H_0 : \mu = \mu_0$, so that the alternative hypothesis is simply $H_1 : \mu \neq \mu_0$. Let $x_1, \ldots, x_n$ be a random sample with $x_i \sim N(\mu, \sigma^2)$ for all i. The test statistic is $\sum_i x_i/n = \bar{x} \sim N(\mu, \sigma^2/n)$. Then, under the null hypothesis

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$$

is a standard normal variate; and the test procedure is summarised as follows:

$$\left|\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right| \geq \beta \Longrightarrow d_1, \qquad \left|\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right| < \beta \Longrightarrow d_0.$$

It remains to determine the probability of the *Type I* error that we are prepared to accept. In this case, the critical region comprises the two tails of the normal distribution; and the sum of their areas, which denotes the probability of rejecting the null hypothesis, is conventionally called the *size of the critical region* or, alternatively, the *significance level of the test.* Having chosen this value, we can look in the back of the book to find the corresponding value for β. What we have described is called a two-tailed test.

A one-tailed test for the mean. Let $x \sim N(\mu, \sigma^2)$, where σ^2 is known, and let $\Omega = (\mu; \mu_0 \leq \mu < \infty)$, which is to say that the null hypothesis is $H_0 : \mu = \mu_0$ and the alternative hypothesis is $H_1 : \mu > \mu_0$. Let $x_1, \ldots, x_n$ be a random sample with $x_i \sim N(\mu, \sigma^2)$ for all i. Then, to test the hypothesis, we use $\bar{x} = \sum_i x_i/n$ as the test statistic; and we shall reject H_0 if $\bar{x}$ is significantly greater than μ_0 such as to render the hypothesis implausible. We have complete faith in our specification of Ω and, therefore, it would not make sense to reject H_0 if $\bar{x}$ were anything less than μ_0. Under the null hypothesis,

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$$

is a standard normal variate; and the test procedure is as follows:

$$\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < \beta \Longrightarrow d_0, \qquad \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \geq \beta \Longrightarrow d_1.$$

If we adopt a $Q \times 100\%$ level of significance, then we must use the table in the back of the book to find the value of β for which $P(z > \beta) = Q$.

Example. A certain type of seed grows to a plant of an expected height of 8.5 ins. A sample of 49 seeds, grown under new conditions with a different fertiliser, produces plants with an average height of 8.8 ins. Using a 5% level of significance and making the assumption that the standard deviations of the heights is $\sigma = 1$ inch and that it is unaffected by the new conditions, you are asked to determine whether there has been a significant increase in the heights. Given $\bar{x} = 8.8$, there is

$$z = \frac{\bar{x} - \mu_0}{\sigma\sqrt{n}} = \frac{8.8 - 85}{1/7} = 2.1.$$

Under the null hypothesis, this is assumed to be a value sampled from a standard normal $N(0,1)$ distribution. The critical value that isolates the upper 5% tail of the distribution is $\beta = 1.645$. This is exceeded by the value of z; and so we are inclined to reject the null hypothesis that the fertiliser has had no effect.

Type II errors. If there is a simple an a wholly specific alternative hypothesis, then we are able to calculate the probability of the *Type II* error. Let $x \sim N(\mu, \sigma^2)$ and let the null hypothesis be $H_0 : \mu = \mu_0$ and the alternative hypothesis be $H_q : \mu = \mu_1$. Then

> The probability of a *Type I* error is $P(d_1|\mu = \mu_0) = P(\bar{x} > \mu_0 + \beta\sigma/\sqrt{n})$, when $\bar{x} \sim (\mu_0, \sigma^2/n)$.
>
> The probability of a *Type II* error is $P(d_0|\mu = \mu_1) = P(\bar{x} \leq \mu_1 + \beta\sigma/\sqrt{n})$, when $\bar{x} \sim (\mu_1, \sigma^2/n)$.