**Lecture 5 – Power and sample size**

---

**Mistakes in our conclusions about the test statistic**

When making a decision about a hypothesis mistakes, or errors, can be made. These mistakes are divided into two categories; Type I and Type II errors.

- **Type I error**: $H_0$ *may be rejected when it is in fact true*. We denote the probability of committing this type of error by $\alpha$ - the *significance level* of the test. In a legal analogy, a Type I error would be finding a defendant guilty when the defendant was, in fact, innocent.

- **Type II error**: $H_0$ *may be accepted when it is in fact false*. The probability of committing this type of error is denoted $\beta$. Similarly, this type of error occurs when defendant is found innocent when he/she is actually guilty.

The probability that $H_0$ is rejected when it is false is called the ***power*** of the test, and is equal to 1 - β. The power of a test measures how sensitive the test is at detecting deviations from $H_0$ or how sure we are of detecting some sort of result.

| | | **Population (hypotheses)** $H_0$ | **Population (hypotheses)** $H_a$ |
|---|---|---|---|
| **Sample (decision)** | accept $H_0$ | Correct | type II error (β) |
| | reject $H_0$ | type I error (α) | Correct (Power = 1-β) |

We would ideally like to construct tests with α and β as small as possible. Indeed, since they are probabilities of error, we would like them, ideally, to be equal to 0. However, this can only be achieved in trivial examples of no practical value. In practice, given a fixed sample size, there exists a trade-off between these two quantities: in order to decrease α, we must increase β and vice versa.

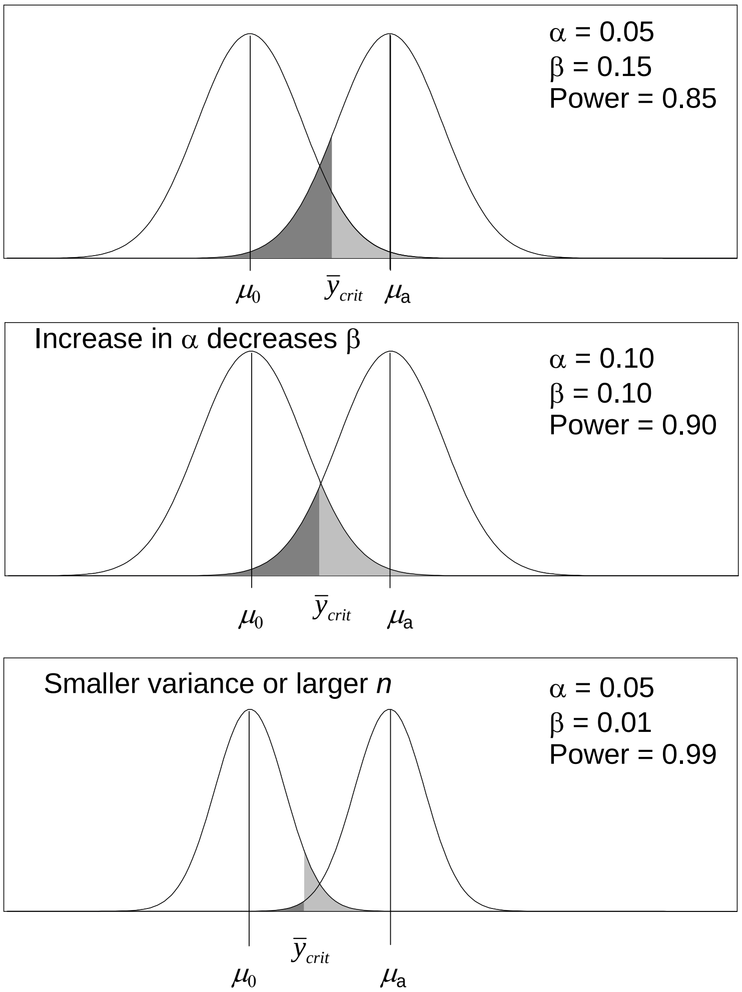


Figure 5.1 The relationships between power (1-β), Type I error (α) and Type II error (β) given a null distribution and an alternative distribution. The critical value of $\bar{y}$ under the null hypothesis is given as $\bar{y}_{crit}$. Values of $\bar{y} > \bar{y}_{crit}$ would lead to a rejection of the null hypothesis.

Now suppose that we conducted a two-tailed one-sample t-test to test the null hypothesis that the mean length of a population of fish is 16cm where $\bar{y} = 17$ cm, $n$ = 64 and $s$ = 5 cm. The test assumes a Type I error of 5%. Assume an alternative hypothesis of 18 cm.

If $H_0: \mu = 16cm$ then our test statistic is calculated as $Z = \dfrac{\bar{y} - \mu_0}{s/\sqrt{n}} = \dfrac{17-16}{5/\sqrt{64}} = 1.6$.

As the critical value for the Z-distribution is 1.96 we FAIL TO REJECT the null hypothesis. We could, however, have committed a type II error. What is the power of this test we have just conducted?

To estimate power, we need to estimate $\beta$. But to estimate $\beta$, we need to estimate $\bar{y}_{crit}$ under the null hypothesis.

Step 1 to calculate $\bar{y}_{crit}$:

As $Z = \dfrac{\bar{y} - \mu_0}{s/\sqrt{n}} > 1.96$ then $\bar{y} - \mu_0 > 1.96\left(\dfrac{s}{\sqrt{n}}\right)$ such that

$$\bar{y} > \mu_0 + 1.96\left(\frac{s}{\sqrt{n}}\right)$$

By substituting $\mu_0 = 16$ and $n$ = 64 then

$$\bar{y} > 16 + 1.96\left(\frac{5}{\sqrt{64}}\right)$$

The value $\bar{y}_{crit}$ at which to achieve $\alpha$ is $\bar{y} > 17.22$.

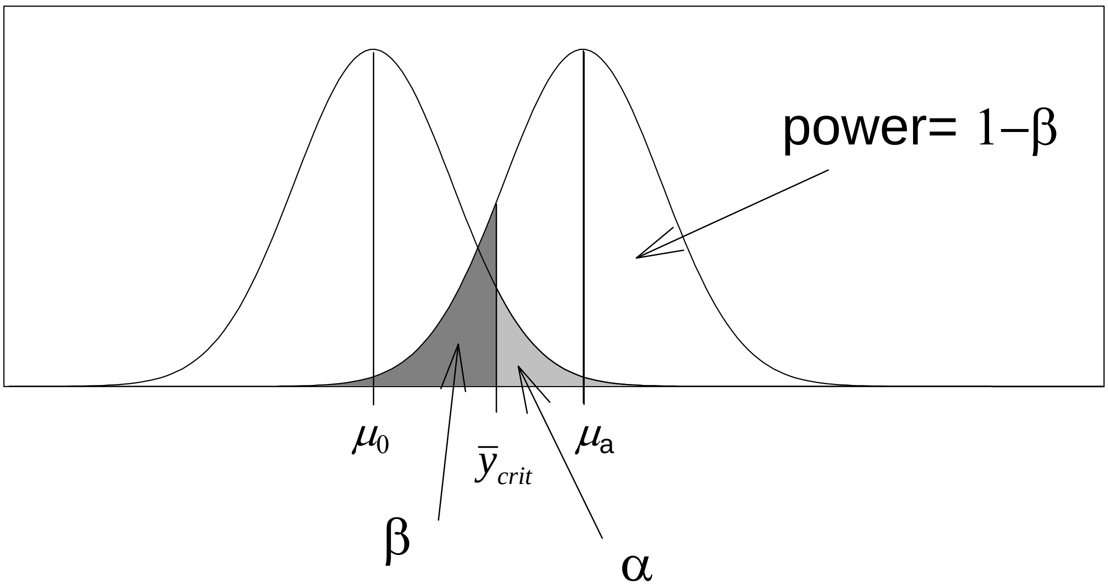


Figure 5.2.Relationship between $\alpha$, $\beta$ and the value $\bar{y}_{crit}$ – the critical value by which the null hypothesis is rejected to achieve $\alpha$.

This is the $\bar{y}_{crit}$ in the above figure. Then by definition,
$\beta = P(\bar{y} \le 17.22 \text{ when } \mu_a = 18)$ is the dark shaded region to the left of $\bar{y}_{crit}$.

Step 2 to calculate $\beta$**:**

Thus for $\mu_a = 16$

$$\beta = P\left[ Z \le \frac{\bar{y}_{crit} - \mu_a}{s/\sqrt{n}} \text{ when } \mu = \mu_a \right] = P\left( Z \le \frac{17.22 - 18}{5/\sqrt{64}} \right) = P(Z \le -1.24)$$

To get this probability, we look up from the Z table the probability associated with $P(Z \le -1.24) = 0.1075$.

Step 3 to calculate the power**:**

The power of the test is, therefore, $Power = 1 - \beta = 1 - 0.1075 = 0.8925$**.**

Therefore, we have a 90% chance of finding a significant difference if one in fact existed.

**Estimating sample size**

To estimate sample size to obtain a requisite $\alpha$ and $\beta$ we can solve two equations simultaneously.

We know that from the figure

$$\alpha = P[\bar{y} > \bar{y}_{crit} \ when \ \mu = \mu_0] = P\left[ \frac{\bar{y} - \mu_0}{s/\sqrt{n}} > \frac{\bar{y}_{crit} - \mu_0}{s/\sqrt{n}} \ when \ \mu = \mu_0 \right] = P(Z > z_{\alpha/2})$$

$$\beta = P[\bar{y} \le \bar{y}_{crit} \text{ when } \mu = \mu_a] = P\left[ \frac{\bar{y} - \mu_a}{s/\sqrt{n}} \le \frac{\bar{y}_{crit} - \mu_a}{s/\sqrt{n}} \text{ when } \mu = \mu_a \right] = P(Z \le -z_\beta)$$

From the equation for $\alpha$ we have

$\dfrac{\bar{y}_{crit} - \mu_0}{s/\sqrt{n}} = z_{\alpha/2}$ and from the equation for $\beta$ we have $\dfrac{\bar{y}_{crit} - \mu_a}{s/\sqrt{n}} = -z_\beta$

By eliminating the common $\bar{y}_{crit}$ from both equations we have the equality:

$$\mu_0 + z_\alpha \left( \frac{s}{\sqrt{n}} \right) = \mu_a - z_\beta \left( \frac{s}{\sqrt{n}} \right)$$

$$\mu_a - \mu_0 = \left(z_\alpha + z_\beta\right)\left(\frac{s}{\sqrt{n}}\right)$$

$$n = \frac{\left(z_\alpha + z_\beta\right)^2}{\left(\mu_a - \mu_0\right)^2} s^2$$ (note for a one-tailed test)

$$n = \frac{\left(z_{\alpha/2} + z_\beta\right)^2}{\left(\mu_a - \mu_0\right)^2} s^2$$ (for a two-tailed test)

*Example:*

Suppose we wish to test $H_0 : \mu = 15$ against $H_a : \mu = 16$ with $\alpha = \beta = 0.05$. Find the sample size to ensure this accuracy for both one and two sample alternatives. Assume $s^2 = 9$.

For the one-tail equivalent

Because $\alpha = \beta = 0.05$ then $z_\alpha = z_\beta = 1.645$ from the inverse table.

Then $n = \frac{\left(z_\alpha + z_\beta\right)^2}{\left(\mu_a - \mu_0\right)^2} s^2 = \frac{\left(1.645 + 1.645\right)^2}{\left(16 - 15\right)^2} 9 = 97.4$

Then 98 samples are required.

Similarly for the two tail option

$$n = \frac{\left(z_{\alpha/2} + z_\beta\right)^2}{\left(\mu_a - \mu_0\right)^2} s^2 = \frac{\left(1.96 + 1.645\right)^2}{\left(16 - 15\right)^2} 9 = 116.95$$

Either 98 or 117 samples are needed to find a difference between the two means.