

Power and Sample Size for Testing Means and Proportion

Type I & Type II Error

Type I Error: reject the null hypothesis when it is true. The probability of a Type I Error is denoted by α .

Type II Error: accept the null hypothesis when it is false and the alternative hypothesis is true. The probability of a Type II Error is denoted by β .

Decision	Null hypothesis, H_0	
	True	False
Reject H_0	Type I Error α	Correct Decision $1 - \beta$
Do Not Reject H_0	Correct Decision $1 - \alpha$	Type II Error β

Power of the Test (Power = 1 - β)

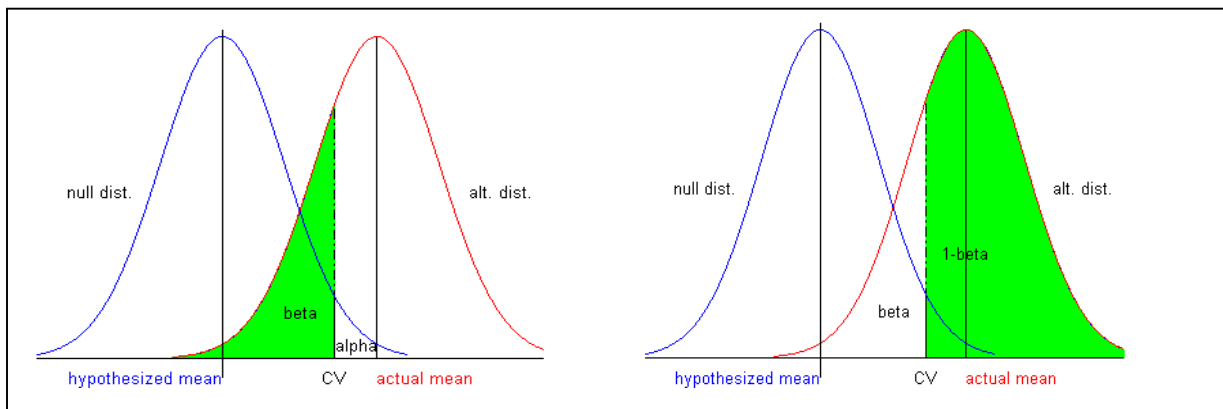
The *power of the statistical test* is the ability of the study as designed to distinguish between the hypothesized value and some specific alternative value. For a statistician, the power of a test is the probability that the test will reject the hypothesis tested when a specific alternative hypothesis is true. To calculate the power of a given test, it is necessary to specify α (the probability that the test will lead to the rejection of the hypothesis tested when that hypothesis is true) and to specify a specific alternative hypothesis H_A . Stated another way,

$$\text{Power} = P(\text{reject } H_0 \mid H_0 \text{ is false}) \text{ or } P(\text{reject } H_0 \mid H_A \text{ is true})$$

For example, suppose a medication is about to be marketed that claims to lower blood pressure. Suppose the medication lowers the subject's blood pressure by 5 mm Hg. The probability that the null hypothesis of no change in blood pressure will be rejected based on the results of the sample study given that the real reduction is 5 mm Hg is the power of the test.

The power of the test is affected by four factors:

1. **The level of significance.** As α decreases, the power decreases.
2. **The sample size.** As n increases, the power increases.
3. **The numerical difference between the null hypothesis value and the specific alternative value.** As this difference increases, the power increases. (It is very difficult to reject a null hypothesis of no difference in favor of an alternative that there is a difference if that difference is very small.)



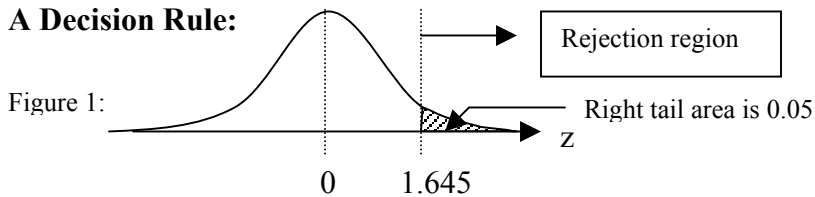
Power and Sample Size Estimation

Example: Suppose we know that the serum cholesterol levels for all 20- to 24-year-old males in the United States is normally distributed with a **mean of 180 mg/100ml** and the **standard deviation is 46 mg/100ml**. We would expect that the mean cholesterol level of a special diet group in this population to be higher than 180 mg/100ml. (Assuming that the cholesterol levels are normally distributed and have the same standard deviation, i.e., 46 mg/100ml.) To test whether the mean cholesterol level of this special diet group to be higher than 180 mg/100ml, we would conduct a one-sided test with the hypotheses:

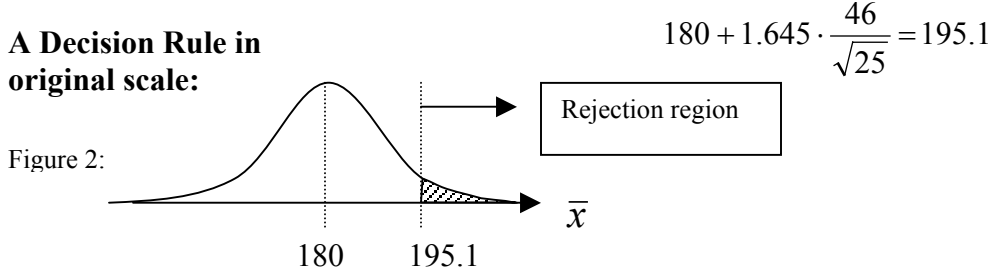
$$H_0: \mu = 180 \text{ mg/100ml (or } \mu \leq 180 \text{ mg/100ml)}$$

$$H_A: \mu > 180 \text{ mg/100ml (H}_1 \text{ is sometimes used instead of H}_A)$$

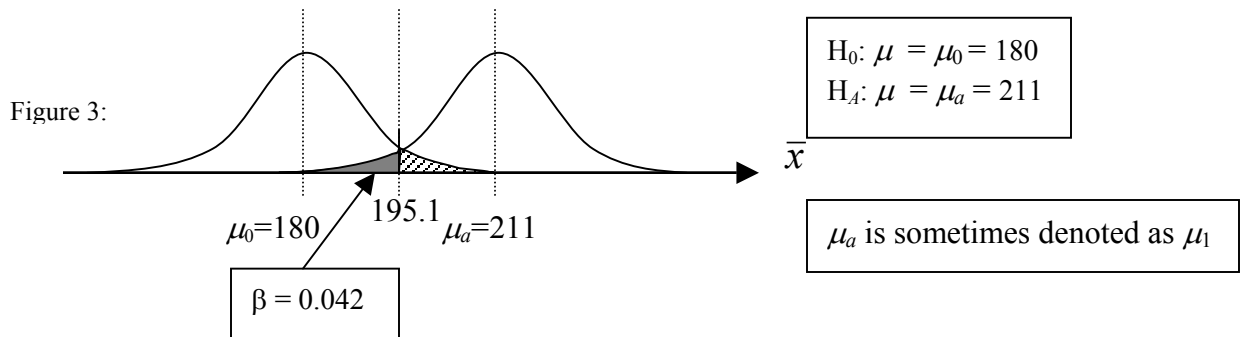
Assuming that we use a **sample of size 25** and test the hypothesis with a level of significant $\alpha = 0.05$, what would be the β and the power of the test, i.e. $1 - \beta$?



Since the standard deviation is known and sample is from a normal population, a z-test can be used. If we take the critical value approach, the critical value would be 1.645. At the level of significant $\alpha = 0.05$, H_0 would be rejected if $z \geq 1.645$. If a sample of size 25 is selected, what would be the critical value in terms of the mean cholesterol level?



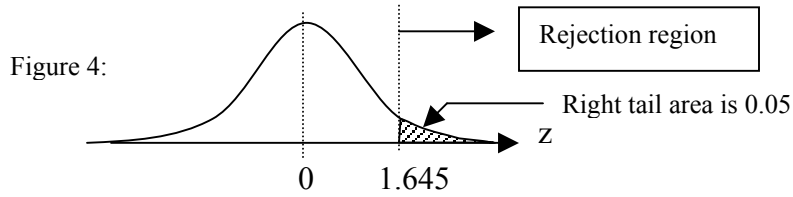
Let assume that the alternative hypothesis is $H_a: \mu = \mu_a = 211 \text{ mg/100ml}$. If H_a is true, what is the probability of accepting H_0 , i.e. β , and what would be the power of the test? That is to say “how powerful can this test detect a 31 mg/100ml increase in average cholesterol level?”



Power of the test when sample size is 25 would be: $\text{Power} = 1 - \beta = 1 - 0.042 = \mathbf{0.958}$.

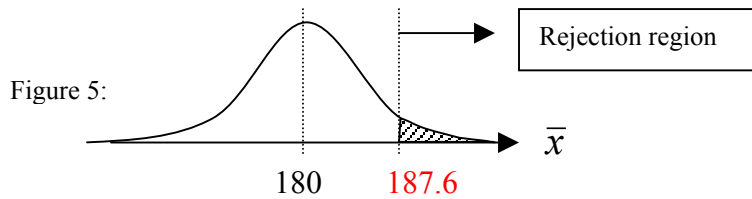
Power and Sample Size Estimation

Assuming that we use a **sample of size 100** and test the hypothesis with a level of significant $\alpha = 0.05$, what would be the β and the power of the test, i.e. $1 - \beta$?

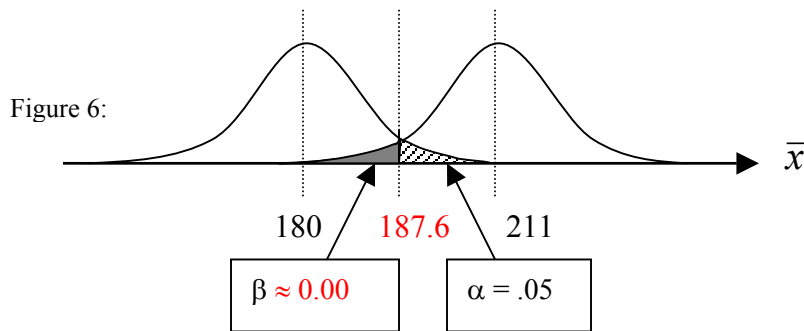


If a sample of size 100 is selected, what would be the critical value in terms of the mean cholesterol level?

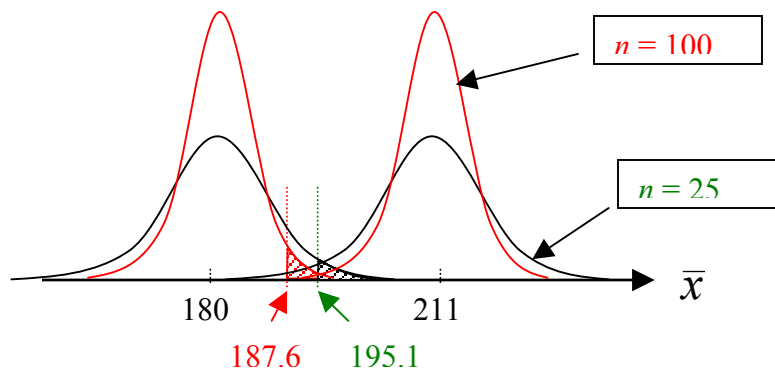
$$\text{c.v.} = 180 + 1.645 \cdot \frac{46}{\sqrt{100}} = 187.6$$



If $H_a: \mu = \mu_a = 211 \text{ mg}/100\text{ml}$ is true, what is the probability of accepting H_0 , i.e. β ?

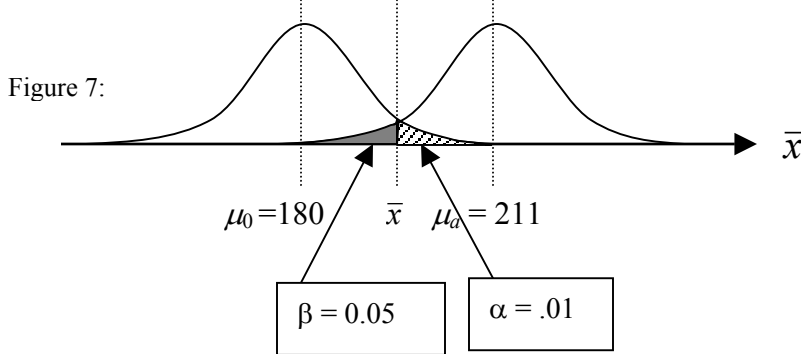


Power of the test when sample size is 100 would be: $\text{Power} = 1 - \beta \approx 1 - 0.0 = \mathbf{1.0}$.



Sample Size Estimation

Suppose we wish to test the hypothesis **Ho: $\mu = 180$ mg/100ml** (or $\mu \leq 180$ mg/100ml) v.s. **Ha: $\mu > 180$ mg/100ml**, at the **level of significance of 0.01**. If we want to risk a **5%** (or say, with the **power of the test 0.95**) chance of failing to reject the null hypothesis in case of that the true mean is as large as 211 mg/100ml. how large a sample do we need? (Power of 0.95 can also mean we want to have a 95% chance of rejecting the null hypothesis if the true mean is as large as 211 mg/100ml.) Assume the standard deviation is 46.



If $\alpha = .01$, then the critical value in z-score would be 2.32. The critical value is cholesterol level would be $\bar{x} = 180 + 2.32 \cdot \frac{46}{\sqrt{n}}$.

If the true mean is 211 mg/100ml, with a power of 0.95, i.e. $\beta = .05$, we would reject the null hypothesis when the sample average is less than $\bar{x} = 211 - 1.645 \cdot \frac{46}{\sqrt{n}}$

Set the two equations equal to each other: $180 + 2.32 \cdot \frac{46}{\sqrt{n}} = 211 - 1.645 \cdot \frac{46}{\sqrt{n}}$

Solve for n : $n = \left[\frac{(2.32 + 1.645)(46)}{211 - 180} \right]^2 = 34.6 \approx 35$

Formula: The estimated sample size for a **one-sided test** for mean with level of significance α and power $1 - \beta$ is,

$$n = \left[\frac{(z_\alpha + z_\beta) \cdot \sigma}{\mu_a - \mu_0} \right]^2$$

For a **two-sided test**,

$$n = \left[\frac{(z_{\alpha/2} + z_\beta) \cdot \sigma}{\mu_a - \mu_0} \right]^2$$

* Check the following web page for power calculation: <http://calculators.stat.ucla.edu/powercalc/>

Try this problem using the formula above:

For testing hypothesis that

$$H_0: \mu = \mu_0 = 70 \quad \text{v.s.} \quad H_A: \mu \neq 70 \quad (\text{Two-sided Test})$$

with a level of significance of $\alpha = .05$. Find the sample size so that one can have a power of $1 - \beta = .90$ to reject the null hypothesis if the actual mean is $\mu_a = 80$. (Another way to say it that if the actual mean is 80, you want the test to have a power of 90% to reject the null hypothesis and accept that the mean is significantly different from 70.) The standard deviation, σ , is approximately equal **15**.

Solution:

Testing one proportion:

As in the case of proportion, the specification of α for p_0 and β for p_1 would lead to the following two equations, ignoring the continuity correction:

$$z_\alpha = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \quad z_\beta = \frac{\hat{p} - p_1}{\sqrt{p_1(1-p_1)/n}}$$

Solving for n , eliminating \hat{p} , we get

$$n = \left[\frac{z_\alpha \sqrt{p_0(1-p_0)} + z_\beta \sqrt{p_1(1-p_1)}}{p_1 - p_0} \right]^2$$

Replace z_α by $z_{\alpha/2}$ if for two-sided test.

Example: Consider the planning of a survey to find out how smoking behavior changed while students were in college. A comprehensive survey four years ago found that 30% of freshmen smoked. The investigator wants to know how many seniors to be sampled. He wants to perform a two-tailed test at the five percent level. The null hypothesis is $p_0=0.3$. This indicates $z_\alpha = 1.96$. He also states that if the proportion is changed as much as 5 percentage points, then he wishes to risk 10% chance of failing to reject the null hypothesis. This indicates $p_1 = .35$ and $z_\beta = 1.28$ (as mentioned earlier, one-tailed z value is used for z_β). Then the required sample size is

$$n = \left[\frac{1.96\sqrt{(0.3)(0.7)} + 1.28\sqrt{(0.35)(0.65)}}{0.35 - 0.3} \right]^2 = 910.48 \approx 911$$

Testing the difference of two means:

We can determine the sample size for testing the difference of two independent means in the same manner. We assume equal variance in two groups ($\sigma_1^2 = \sigma_2^2$) and equal division of sample size between the two groups ($n_1 = n_2$). Specifying Type I error

for the null hypothesis ($\mu_1 - \mu_2 = \Delta_0$), we have $z_\alpha = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{\sqrt{2\sigma_1^2/n_1}}$.

Specifying Type II for the alternative hypothesis ($\mu_1 - \mu_2 = \Delta_1$), we have

$z_\beta = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_1}{\sqrt{2\sigma_1^2/n_1}}$. By solving these two equations for n_1 , eliminating $(\bar{x}_1 - \bar{x}_2)$, we

get $n_1 = 2 \left[\frac{(z_\alpha - z_\beta)\sigma_1}{\Delta_1 - \Delta_0} \right]^2$. Since Δ_0 is zero in most application, the denominator is

usually Δ_1 , the value specified in the alternative hypothesis. Note that n is the sample size in each group and the total sample size for the study is $2n_1$.

Let us consider a case of designing a clinical nutritional study of special diet regimen to lower blood pressure among hypertensive adult males (diastolic blood pressure over 90 mmHg). The investigator expects to demonstrate the new diet would reduce diastolic blood pressure by 4 mmHg in three months. He is willing to risk Type I error of 5% and Type II error of 10% for a one-sided test. The NHANES data show that standard deviation of diastolic blood pressure among hypertensive males is around 5.6 mmHg. The required sample size in each group can be calculated by

$n_1 = 2 \left(\frac{5.6(1.645 + 1.28)}{4} \right)^2 = 33.5 \approx 34$. The proposed study would need a total of 68 subjects, allocated randomly and equally between the treatment and the control group.

Formula: The estimated sample size for a **one-sided test**, with level of significance α and power $1 - \beta$ is, for $n_1 = n_2 = n$, where $\Delta = |\mu_1 - \mu_2|$ is sometimes a difference that is considered practically significant, is

$$n = 2 \left[\frac{(z_\alpha + z_\beta) \cdot \sigma}{\Delta} \right]^2$$

For a two-sided test,

$$n = 2 \left[\frac{(z_{\alpha/2} + z_\beta) \cdot \sigma}{\Delta} \right]^2$$