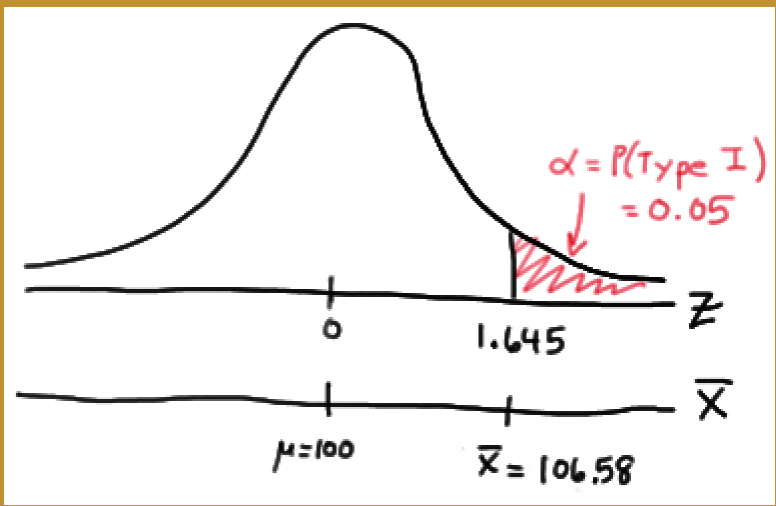
<u>**Power of a Test**</u>

Example

Let X denote the IQ of a randomly selected adult American. Assume, a bit unrealistically, that X is normally distributed with unknown mean μ and standard deviation 16. Take a random sample of n = 16 students, so that, after setting the probability of committing a Type I error at α = 0.05, we can test the null hypothesis H_0: μ = 100 against the alternative hypothesis that H_A: μ > 100.

What is the power of the hypothesis test if the true population mean were μ = 108?

Solution. Setting α, the probability of committing a Type I error, to 0.05, implies that we should reject the null hypothesis when the test statistic $Z \geq 1.645$, or equivalently, when the observed sample mean is 106.58 or greater:



because we transform the test statistic Z to the sample mean by way of:

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \Rightarrow \bar{X} = \mu + z\left(\frac{\sigma}{\sqrt{n}}\right) \Rightarrow \bar{X} = 100 + 1.645\left(\frac{16}{\sqrt{16}}\right) = 106.58$$

Now, that implies that the power, that is, the probability of rejecting the null hypothesis, when μ = 108 is 0.6406 as calculated here (recalling that $\Phi(z)$ is standard notation for the cumulative distribution function of the standard normal random variable):

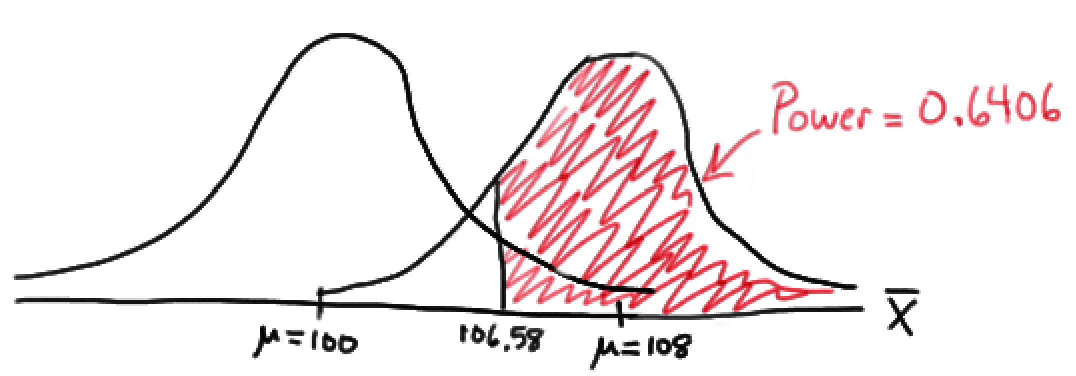
$$\text{Power} = P(\bar{X} \geq 106.58 \text{ when } \mu = 108) = P\left(Z \geq \frac{106.58 - 108}{16/\sqrt{16}}\right)$$

$$= P(Z \geq -0.36) = 1 - P(Z < -0.36)$$

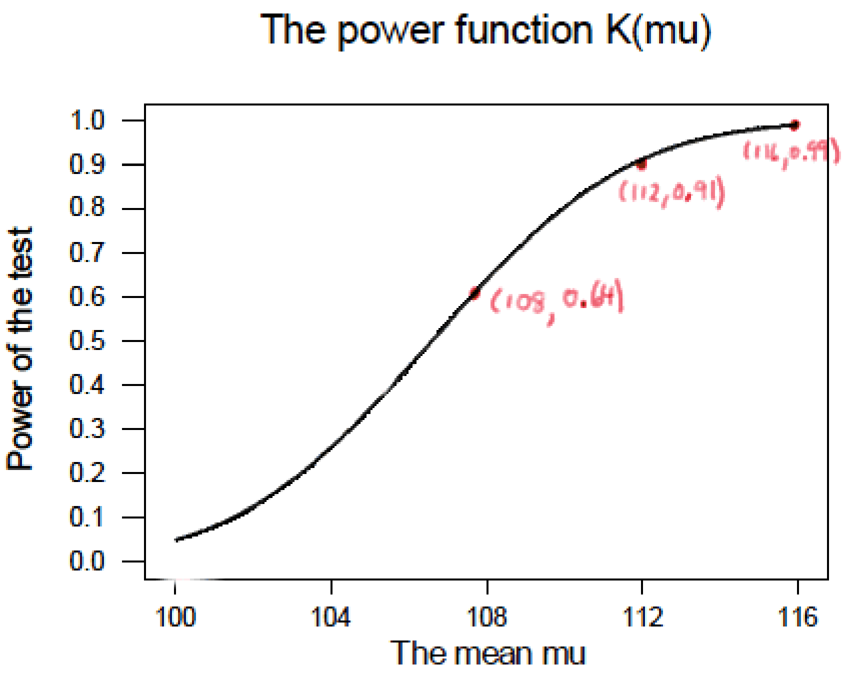
$$= 1 - \Phi(-0.36)$$

$$= 1 - 0.3594 = 0.6406$$

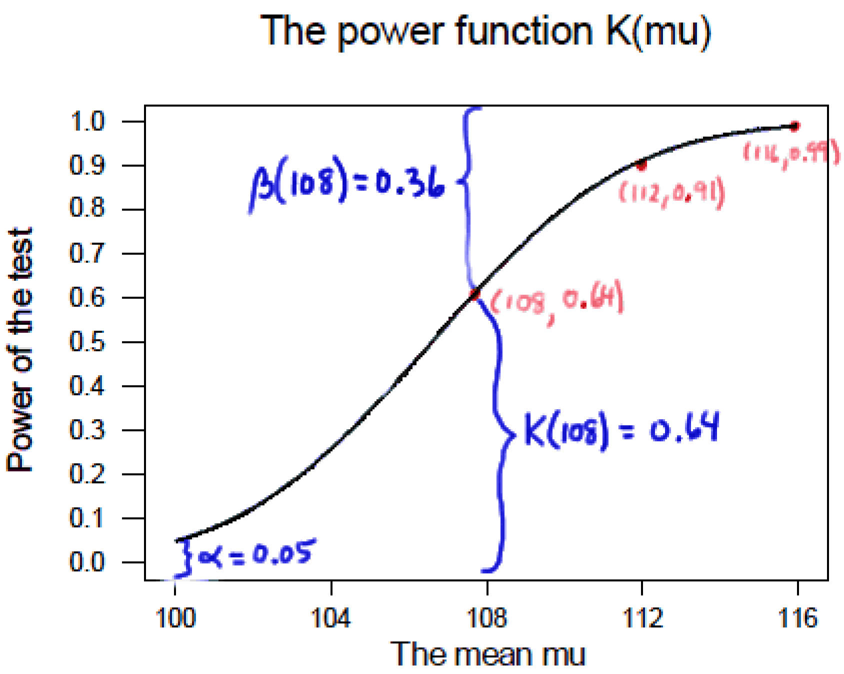
and illustrated here:



In summary, we have determined that we have (only) a 64.06% chance of rejecting the null hypothesis H_0: $\mu = 100$ in favor of the alternative hypothesis H_A: $\mu > 100$ if the true unknown population mean is in reality $\mu = 108$.



represented on a power function plot, as illustrated here:

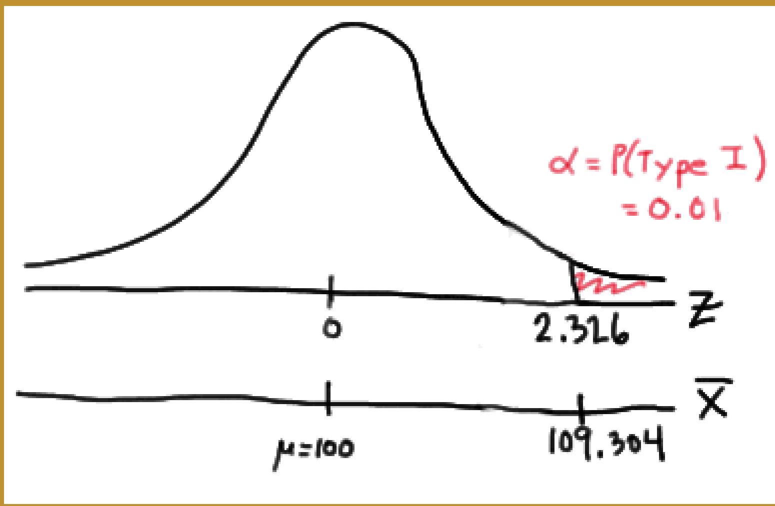


Example (continued)

Let X denote the IQ of a randomly selected adult American. Assume, a bit unrealistically, that X is normally distributed with unknown mean μ and standard deviation 16. Take a random sample of $n = 16$ students, so that, after setting the probability of committing a Type I error at $\boldsymbol{\alpha} = \mathbf{0.01}$, we can test the null hypothesis H_0: $\mu = 100$ against the alternative hypothesis that H_A: $\mu > 100$.

What is the power of the hypothesis test if the true population mean were $\mu = 108$?

Solution. Setting α, the probability of committing a Type I error, to 0.01, implies that we should reject the null hypothesis when the test statistic $Z \geq 2.326$, or equivalently, when the observed sample mean is 109.304 or greater:



because:

$$\bar{x} = \mu + z \left(\frac{\sigma}{\sqrt{n}} \right) =100 + 2.326\left(\frac{16}{\sqrt{16}} \right)=109.304$$

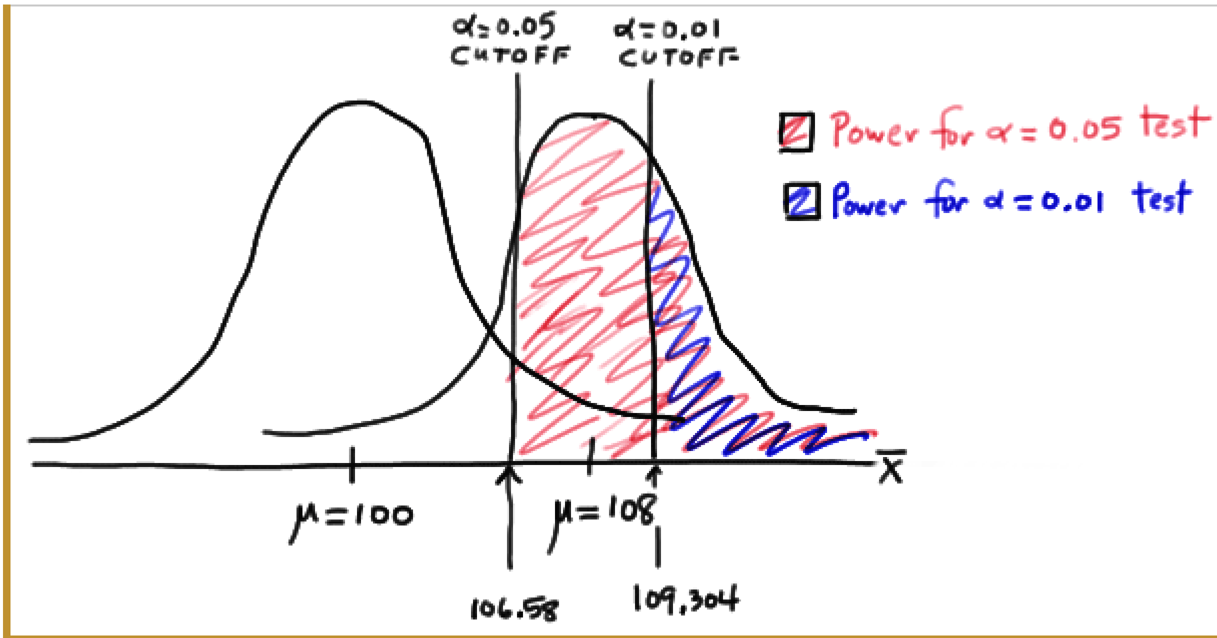
That means that the probability of rejecting the null hypothesis, when $\mu = 108$ is 0.3722 as calculated here:

$$\text{Power} = P(\bar{X} \geqslant 109.304 \text{ when } \mu = 108) = P\left(Z \geqslant \frac{109.304 - 108}{16/\sqrt{16}}\right)$$

$$= P(Z \geqslant 0.326) = 1 - P(Z < 0.326)$$

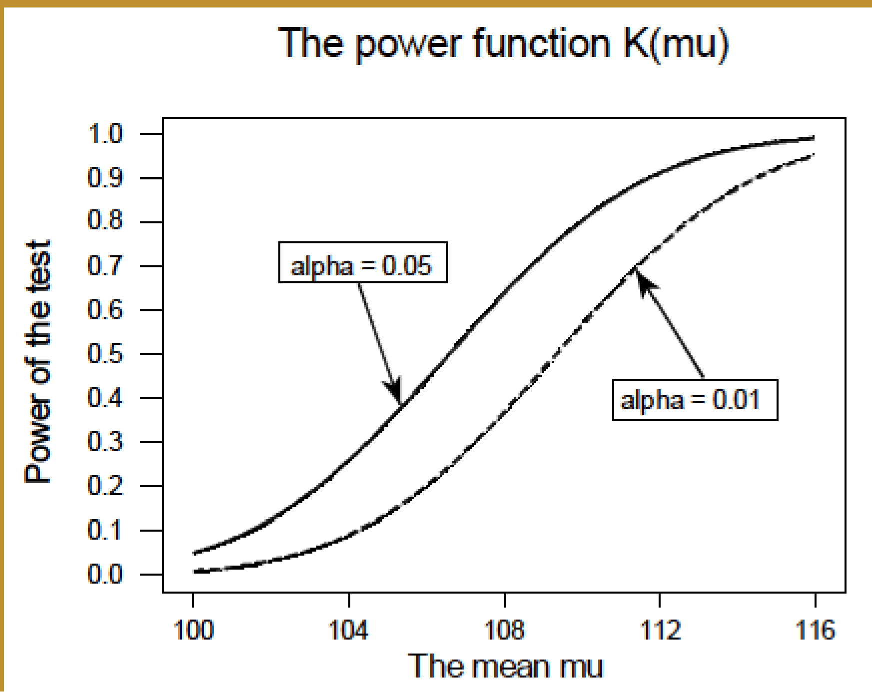
$$= 1 - \Phi(0.326) = 1 - 0.6278 = 0.3722$$

So, the power when $\mu = 108$ and $\alpha = 0.01$ is smaller (0.3722) than the power when $\mu = 108$ and $\alpha = 0.05$ (0.6406)! Perhaps we can see this graphically:



By the way, we could again alternatively look at the glass as being half-empty. In that case, the probability of a Type II error when μ = 108 and α = 0.01 is 1−0.3722 = 0.6278. In this case, the probability of a Type II error is *greater* than the probability of a Type II error when μ = 108 and α = 0.05.

All of this can be seen graphically by plotting the two power functions, one where α = 0.01 and the other where α = 0.05, simultaneously. Doing so, we get a plot that looks like this:

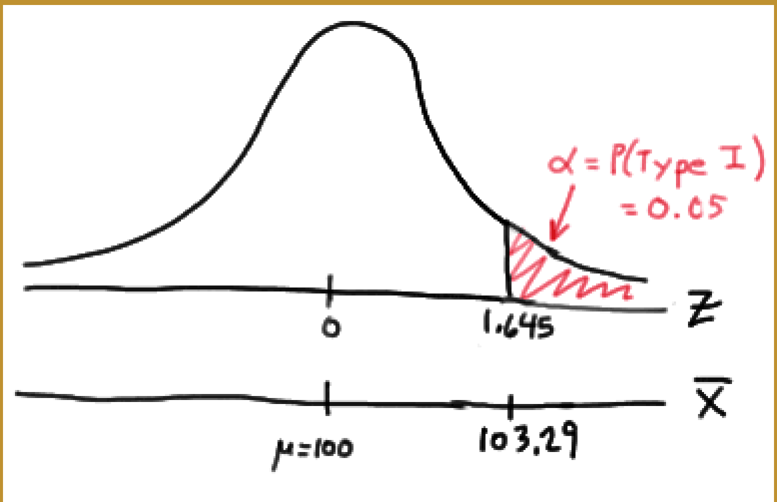


Example

Let X denote the IQ of a randomly selected adult American. Assume, a bit unrealistically again, that X is normally distributed with unknown mean μ and (a strangely known) standard deviation of 16. This time, instead of taking a random sample of n = 16 students, let's increase the sample size to $\boldsymbol{n}$ **= 64**. And, while setting the probability of committing a Type I error to α = 0.05, test the null hypothesis H_0: μ = 100 against the alternative hypothesis that H_A: μ > 100.

What is the power of the hypothesis test when μ = 108, μ = 112, and μ = 116?

Solution. Setting α, the probability of committing a Type I error, to 0.05, implies that we should reject the null hypothesis when the test statistic $Z \geq 1.645$, or equivalently, when the observed sample mean is 103.29 or greater:

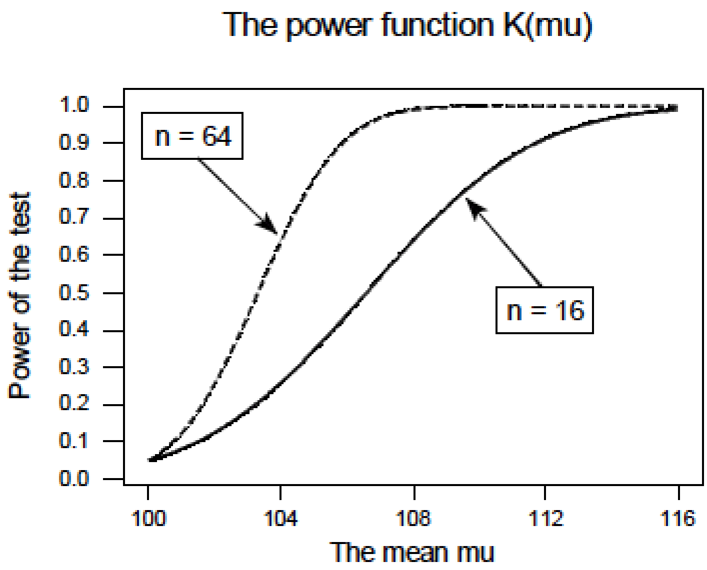


r calculations:

POWER	K(108)	K(112)	K(116)
n = 16	0.6406	0.9131	0.9909
n = 64	0.9907	0.9999...	0.999999...

As you can see, our work suggests that **for a given value of the mean μ under the alternative hypothesis, the larger the sample size n, the greater the power $K(\mu)$**.

Perhaps there is no better way to see this than graphically by plotting the two power functions simultaneously, one when n = 16 and the other when n = 64:

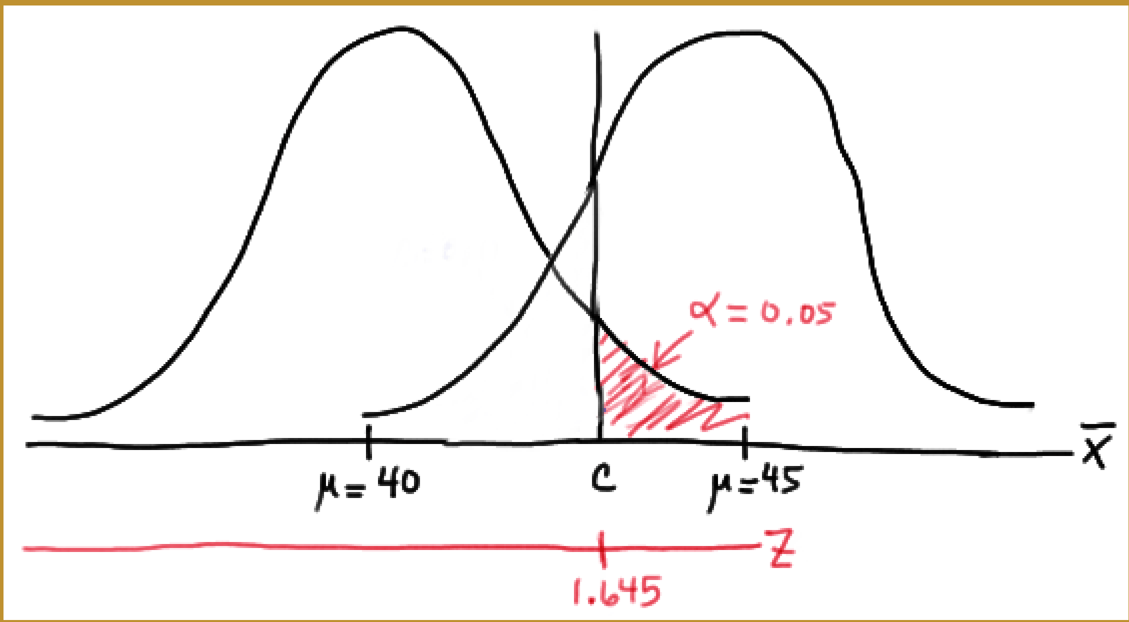


As this plot suggests, if we are interested in increasing our chance of rejecting the null hypothesis when the alternative hypothesis is true, we can do so by increasing our sample size n. This benefit is perhaps even greatest for values of the mean that are close to the value of the mean assumed under the null hypothesis. Let's take a look at two examples that illustrate the kind of **sample size calculation** we can make to ensure our hypothesis test has sufficient power.

Example

Let X denote the crop yield of corn measured in the number of bushels per acre. Assume (unrealistically) that X is normally distributed with unknown mean μ and standard deviation σ = 6. An agricultural researcher is working to increase the current average yield from 40 bushels per acre. Therefore, he is interested in testing, at the α = 0.05 level, the null hypothesis H_0: μ = 40 against the alternative hypothesis that H_A: μ > 40. Find the sample size n that is necessary to achieve 0.90 power at the alternative μ = 45.

Solution. As is always the case, we need to start by finding a threshold value *c*, such that if the sample mean is larger than *c*, we'll reject the null hypothesis:



the equations, and solve for *n*. Doing so, we get:

$$40 + 1.645\left(\frac{6}{\sqrt{n}}\right) = 45 - 1.28\left(\frac{6}{\sqrt{n}}\right)$$

$$(1.645 + 1.28)\frac{6}{\sqrt{n}} = 45 - 40$$

$$\frac{17.55}{\sqrt{n}} = 5$$

$$\sqrt{n} = \frac{17.55}{5} = 3.51$$

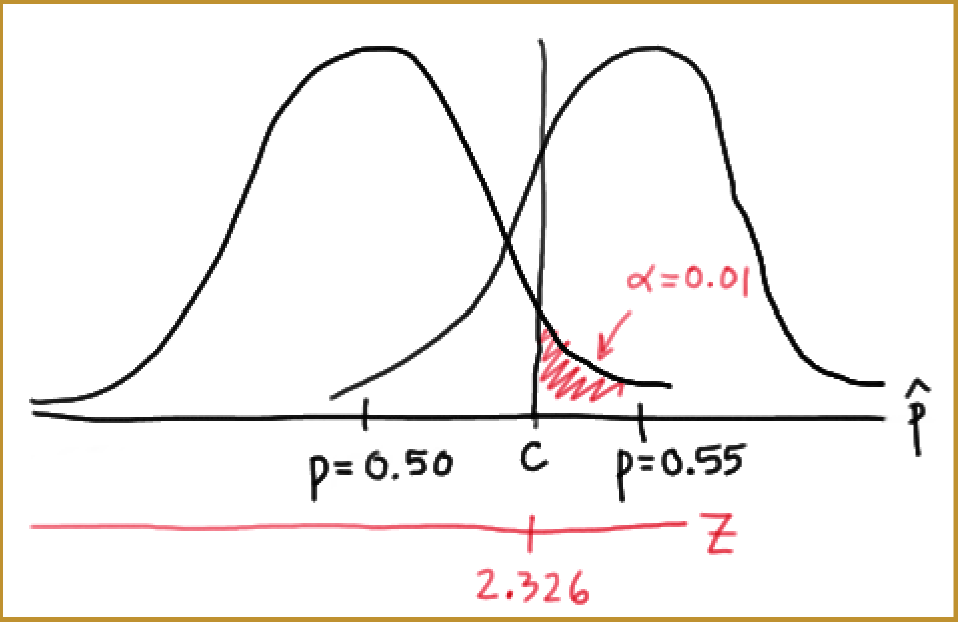
$$n = 3.51^2 = 12.3 \approx 13$$

So, in summary, if the agricultural researcher collects data on n = 13 corn plots, and rejects his null hypothesis H_0: μ = 40 if the average crop yield of the 13 plots is greater than 42.737 bushels per acre, he will have a 5% chance of committing a Type I error and a 10% chance of committing a Type II error if the population mean μ were actually 45 bushels per acre.

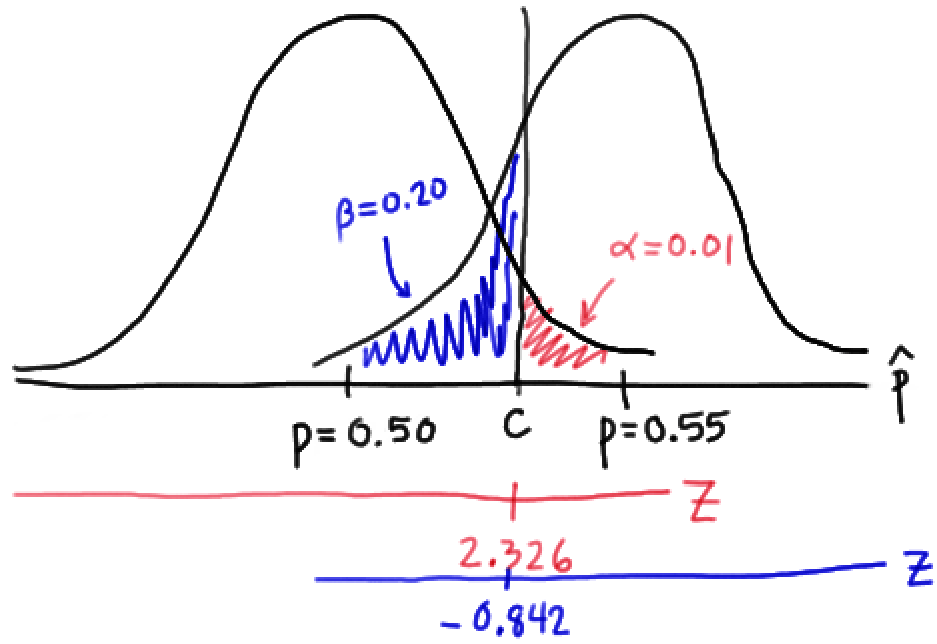
Example

Consider p, the true proportion of voters who favor a particular political candidate. A pollster is interested in testing at the $\alpha = 0.01$ level, the null hypothesis H_0: $p = 0.50$ against the alternative hypothesis that H_A: $p > 0.50$. Find the sample size n that is necessary to achieve 0.80 power at the alternative $p = 0.55$.

Solution. In this case, because we are interested in performing a hypothesis test about a population proportion p, we use the Z-statistic:



But, again, that's not the only condition that c must meet, because c also needs to be defined to ensure that our power is 0.80 or, alternatively, that the probability of a Type II error is 0.20. That would happen if there was a 20% chance that our test statistic fell short of c when $p = 0.55$, as the following drawing illustrates in blue:



$$0.5 + 2.326\sqrt{\frac{(0.5)(0.5)}{n}} = 0.55 - 0.842\sqrt{\frac{(0.55)(0.45)}{n}}$$

$$2.326\,\frac{\sqrt{(0.25)}}{\sqrt{n}} + 0.842\,\frac{\sqrt{0.2475}}{\sqrt{n}} = 0.55 - 0.5$$

$$\frac{1}{\sqrt{n}}\left(1.5818897\right) = 0.05$$

$$n \approx \left[\frac{1.5818897}{0.05}\right]^2 = 1000.95 \approx 1001$$