13.1: Power

Power is a concept that applies to all statistical testing. Here we will look at power quantitatively for the z-test for means (t-test with large n). We will see explicitly in that case some principles that apply to other tests. These principles are: the bigger your sample size (n), the higher the power; the larger α is, the more power there is[1]; the larger the "effect size" is the more power there is. A final principle, that we can't show by restricting ourselves to a z-test, is that the simpler the statistical test, the more power it has — being clever doesn't get you anywhere in statistics.

Let's being by recalling the "confusion matrix" (here labelled a little differently than the one shown in Chapter 9 to emphasize the decision making). Note: The α, β, etc. quantities are the probabilities that each conclusion will happen.

		Reality	
		H_0	H_1
Conclusion of Test	H_1	Type I error α	Correct decision $1-\beta$
	H_0	Correct decision $1-\alpha$	Type II error β

Recall that $1-\beta$ is the power, the probability of correctly rejecting H_0. With the definition of H_1 as not H_0, we cannot actually compute a power because this definition is too vague. The confusion matrix with H_0 and H_1 as given here is purely a conceptual device. To actually compute a power number we need to nail down a specific *alternate hypothesis* H_a and compute β for the more specific confusion matrix:

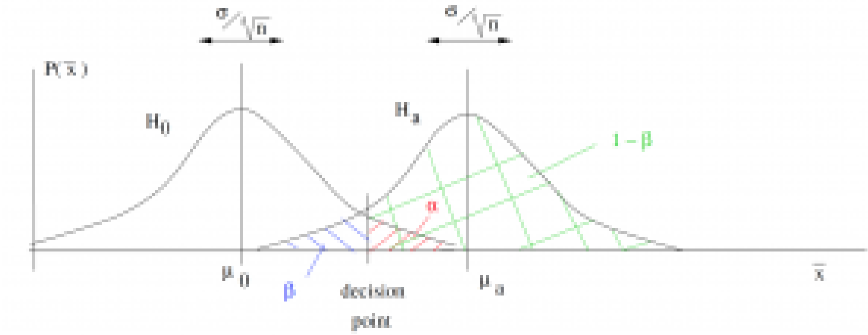
		Reality	
		H_0	H_a
Conclusion of Test	H_a	Type I error α	Correct decision $1-\beta$ **(power)**
	H_0	Correct decision $1-\alpha$	Type II error β

We will define H_0 and H_a be three parameters. The first is that we assume that the populations associated with H_0 and H_a both have the same standard deviation σ. Then, assuming that both populations are normal, H_0 is defined by its population mean μ_0 (we used k in Chapter 9) and H_a is defined by its population mean μ_a.

We can define two flavors of power :

1. **Predicted power.** Based on a *pre-defined* alternate mean μ_a of interest and an estimate of $\sigma/\sqrt{n}$. The population standard deviation σ is frequently estimated from the sample standard deviation s of a small pilot study.
2. **Observed power.** Based on the *observed* sample mean $\overline{x}$ which is then used as the alternate mean μ_a and sample standard deviation s which is used for σ.

The type II error rate β (and power $1-\beta$) is calculated by considering the populations associated with H_0 and H_1 :



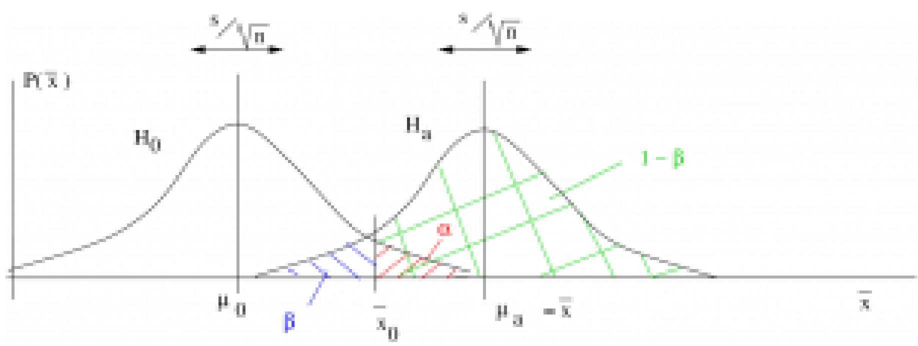
This picture follows directly from the Central Limit Theorem. Hypothesis testing is a decision process. In the picture above, which shows a one-tailed z-test for means, you reject H_0 if $\overline{x}$ falls to the right of the decision point. The decision point is set by the value of α. Note that the alternate mean μ_a needs to be in the rejection region of H_0 for the picture to make sense. The value of β (and hence the power $1-\beta$) depends on the magnitude of the effect size[2] $\mu_a - \mu_0$. We can see that power will increase if the effect size that we are looking for in our experiment increases. This makes sense because larger differences should be easier to measure. Also note that if n increases, as it would by replicating an experiment with a larger sample size, then the two distributions of sample means will get skinner and, for a given effect size, the power will increase. Again, this makes intuitive sense because more data is always better. We will illustrate these features in the numerical examples that follow.

For the purpose of learning the mechanics of statistical power we focus on **observed power**. With observed power we use the sample data for the power calculations; set $\mu_a = \overline{x}$ and $\sigma = s$. Since μ_a needs to be in the rejection region of H_0, observed power can only be computed when the conclusion of the hypothesis test is to reject H_0. In real life if you reject H_0 you don't care about what power the experiment had to reject H_0. It's a bit like calculating if you have enough gas to drive to Regina after you've arrived at Regina. In real life you will care about power only if you fail to reject H_0 because you will want to know the problem was that you tried to measure too small of an effect size or if a larger sample might lead to a decision to reject H_0. In that case you will need to decide what effect size, or sample size, to use in computing a predicted power. You will use predicted power in your experiment design. If your experiment design has a predicted power of about 0.80 then you have a reasonable chance of rejecting the null hypothesis. If your research involves invasive intervention with people (needles, surgery, etc.) then you may need to present a power calculation to prove to an ethics committee that your experiment has a reasonable chance of finding what you think it will find.

In addition to $\mu_a = \overline{x}$ and $\sigma = s$ we need the value of the decision point $\overline{x}_0$ which is the inverse z-transform of $z_\alpha = z_{\text{crit}}$. We'll consider three cases :

***Case* 1. Right tailed test:**

$H_0 : \mu \leq \mu_0$ $H_1 : \mu > \mu_0$ \mu_{0} \]" title="Rendered by QuickLaTeX.com">

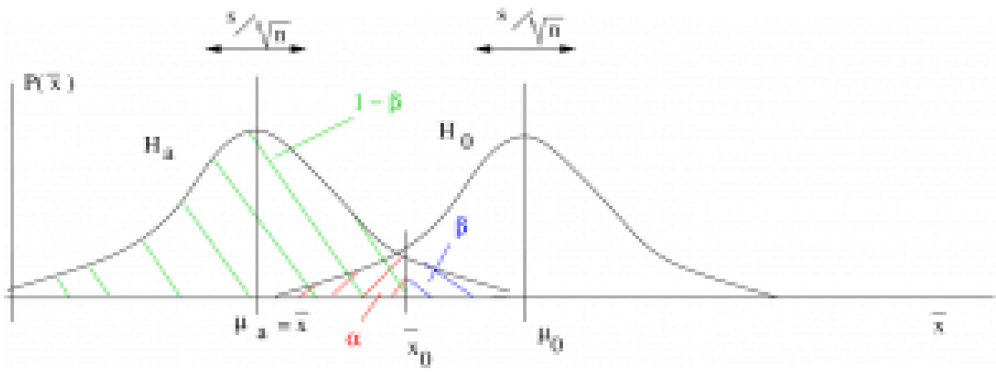


where, In this case

$$\overline{x}_0 = \mu_0 + z_\alpha \left(\frac{s}{\sqrt{n}} \right) \tag{13.1.1}$$

***Case* 2. Left tailed test:**

$$H_0 : \mu \geq \mu_0 \qquad H_1 : \mu < \mu_0 \tag{13.1.2}$$



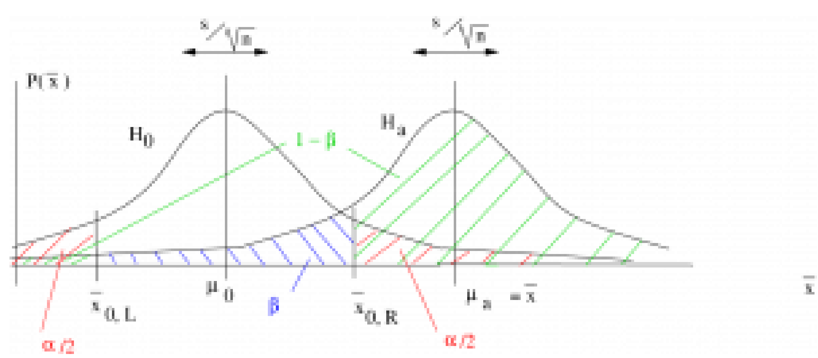
where, In this case

$$\bar{x}_0 = \mu_0 - z_\alpha \left(\frac{s}{\sqrt{n}} \right) \tag{13.1.3}$$

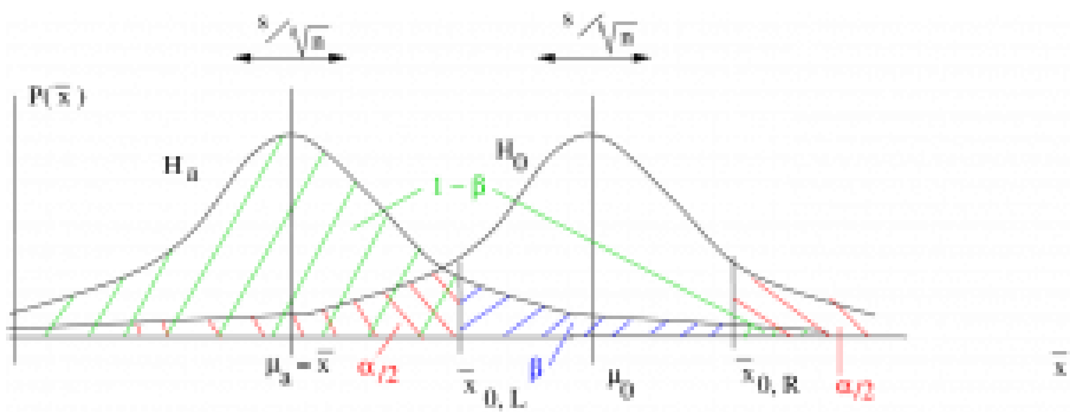
***Case* 3. Two-tailed test:**

$$H_0 : \mu = \mu_0 \qquad H_1 : \mu \neq \mu_0 \tag{13.1.4}$$

(a) $\bar{x}$ in the right tail :



(b) $\bar{x}$ in the left tail:



where, in both cases:

Rendered by QuickLaTeX.com

In both two-tailed cases, notice the small piece of $1-\beta$ area on the side of the H_a distribution on the opposite side from $\bar{x}$. It turns out that the area of that small part is so incredibly small that we can take it to be zero. This will be obvious she we work through the examples. So the upshot is that going from a one-tailed test to a two-tailed test effectively decreases α to $\alpha/2$ which increases β and decreases the power $1-\beta$. One-tailed tests have more power than two-tailed tests for the same α.

Example 13.1 Right tailed test.

Given :

$image$ 150" title="Rendered by QuickLaTeX.com" height="17" width="179" style="vertical-align: -4px;">

$n=50$, $\alpha=0.05$, $s=15$, $\overline{x}=\mu_a=155$

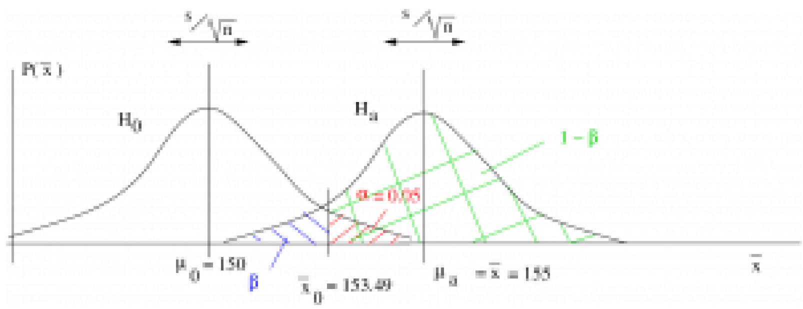
Find the observed power.

Step 1 : Look up $z_\alpha = z_{0.05}$ in the ***t* Distribution Table** for a one-tailed test: $z_\alpha = 1.645$.

Step 2 : Compute :

$$\begin{aligned}\overline{x}_0 &= \mu_0 + z_\alpha \left(\frac{s}{\sqrt{n}}\right)\\ &= 150 + (1.645)\left(\frac{15}{\sqrt{50}}\right)\\ &= 153.49\end{aligned}$$

Step 3 : Draw picture :



Step 4 : Compute the z-transform of $\overline{x}_0$ relative to H_a :

Rendered by QuickLaTeX.com

Step 5 : Look up the area $A(-z_a)$ in the **Standard Normal Distribution Table**. That area will be $0.5-\beta$: $0.5-\beta=0.2611$, so $\beta=0.5-0.2611=0.2389$ and **power** = $1-\beta=1-0.2389=$ **0.7611**.

□

Example 13.2 : Another right tailed test with the data the same as in Example 13.1 but with a smaller α. This example shows how reducing α will reduce the power. With reduced power, it is harder to reject H_0.

Given :

$image$ 150" title="Rendered by QuickLaTeX.com" height="17" width="179" style="vertical-align: -4px;">

$n=50$, $\alpha=0.01$, $s=15$, $\overline{x}=\mu_a=155$

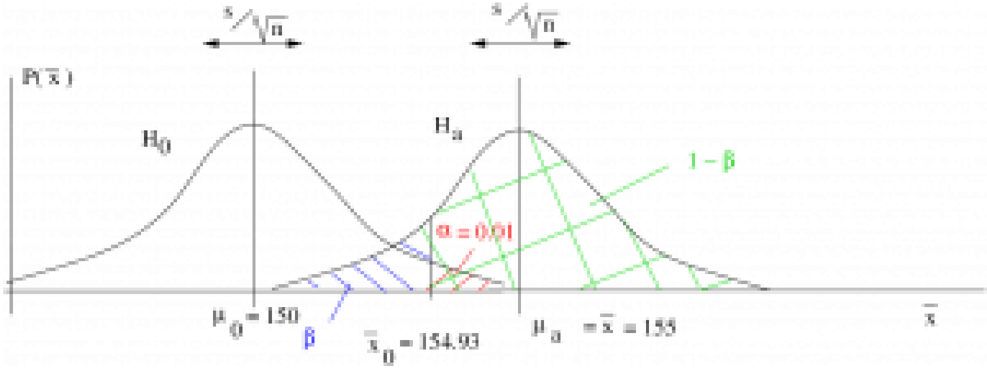
Find the observed power.

Step 1 : Look up $z_\alpha = z_{0.01}$ in ***t* Distribution Table** for a one-tailed test: $z_\alpha = 2.326$.

Step 2 : Compute :

$$\begin{aligned}\overline{x}_0 &= \mu_0 + z_\alpha \left(\frac{s}{\sqrt{n}}\right)\\ &= 150 + (2.326)\left(\frac{15}{\sqrt{50}}\right)\\ &= 154.93\end{aligned}$$

Step 3 : Draw picture :



Step 4 : Compute the z-transform of $\overline{x}_0$ relative to H_a :

Rendered by QuickLaTeX.com

Step 5 : Look up the area $A(0.03)$ in the **Standard Normal Distribution Table**. That area will be $A(0.03) = 0.0120$. So $\beta = 0.5 - 0.0120 = 0.4880$ and **power** = $1 - \beta =$ **0.5120** which is smaller than the power found in Example 13.1.

▢

Example 13.3 : Another right tailed test with the data the same as in Example 13.2 but with larger n. This example shows how increasing the sample size increases the power. This makes sense because more data is always better.

Given :

$image$ 150" title="Rendered by QuickLaTeX.com" height="17" width="179" style="vertical-align: -4px;">

$n = 150$, $\alpha = 0.01$, $s = 15$, $\overline{x} = \mu_a = 155$

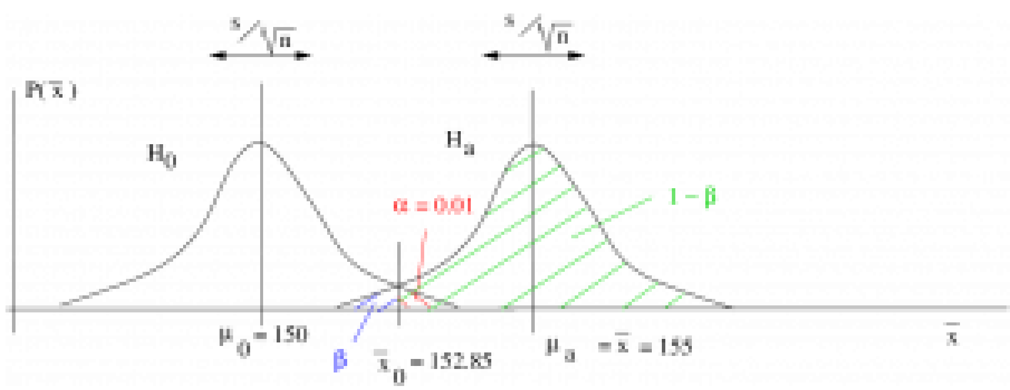
Find the observed power.

Step 1 : Look up $z_\alpha = z_{0.01}$ in ***t* Distribution Table** for a one-tailed test: $z_\alpha = 2.326$.

Step 2 : Compute :

$$\begin{aligned}\overline{x}_0 &= \mu_0 + z_\alpha\left(\frac{s}{\sqrt{n}}\right)\\ &= 150 + (2.326)\left(\frac{15}{\sqrt{150}}\right)\\ &= 152.85\end{aligned}$$

Step 3 : Draw picture :



Step 4 : Compute the z-transform of $\overline{x}_0$ relative to H_a :

Rendered by QuickLaTeX.com

Step 5 : Look up the area $A(0.03)$ in the **Standard Normal Distribution Table**. That area will be $A(0.03) = 0.4808$. So $\beta = 0.5 - 0.4808 = 0.0392$ and **power** = $1 - \beta =$ **0.9608** which is larger than the power found in Example 13.2.

Example 13.4 : Another right tailed test with the data the same as in Example 13.3 but with a smaller value for $\overline{x} = \mu_a$ which leads to a smaller effect size. This example shows how decreasing the effect size decreases the power. This makes sense because it is harder to detect a smaller signal.

Given :

$image$ 150" title="Rendered by QuickLaTeX.com" height="17" width="179" style="vertical-align: -4px;">

$n = 150$, $\alpha = 0.01$, $s = 15$, $\overline{x} = \mu_a = 153$

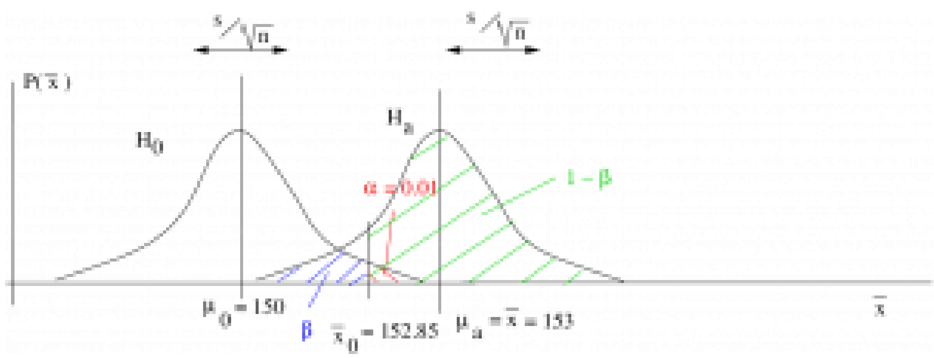
Find the observed power.

Step 1 : Look up $z_\alpha = z_{0.01}$ in the ***t* Distribution Table** for a one-tailed test: $z_\alpha = 2.326$.

Step 2 : Compute :

$$\begin{aligned}\overline{x}_0 &= \mu_0 + z_\alpha \left(\frac{s}{\sqrt{n}}\right)\\ &= 150 + (2.326)\left(\frac{15}{\sqrt{150}}\right)\\ &= 152.85\end{aligned}$$

Step 3 : Draw picture :



Step 4 : Compute the z-transform of $\overline{x}_0$ relative to H_a :

Rendered by QuickLaTeX.com

Step 5 : Look up the area $A(0.12)$ in the **Standard Normal Distribution Table**. That area will be $A(0.12) = 0.0478$. So $\beta = 0.5 - 0.0478 = 0.4522$ and **power** = $1 - \beta =$ **0.5478** which is smaller than the power found in Example 13.3.

Example 13.5 : Left tailed test.

Given :

$H_0 : \mu \geq 150$, $H_1 < 150$

$n = 50$, $\alpha = 0.05$, $s = 15$, $\overline{x} = \mu_a = 144$

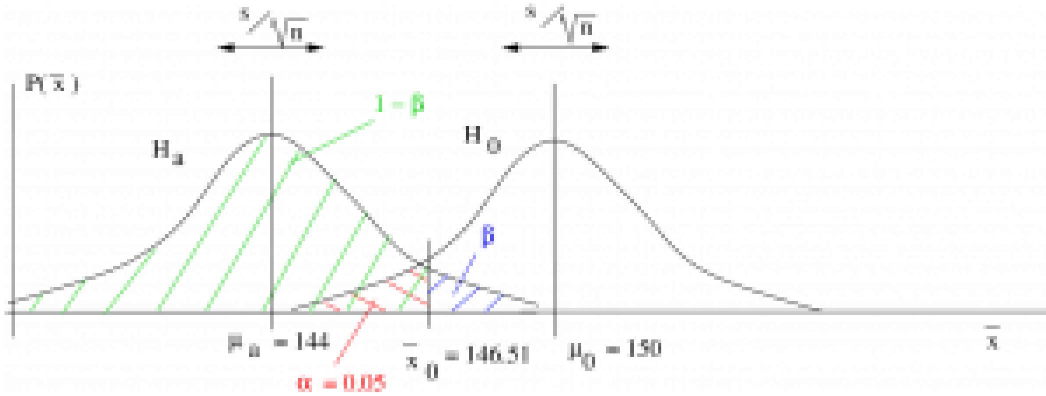
Find the observed power.

Step 1 : Look up $z_\alpha = z_{0.05}$ in the ***t* Distribution Table** for a one-tailed test: $z_{0.05} = 1.645$.

Step 2 : Compute :

$$\begin{aligned}\overline{x}_0 &= \mu_0 - z_\alpha \left(\frac{s}{\sqrt{n}}\right)\\ &= 150 - (1.645)\left(\frac{15}{\sqrt{50}}\right)\\ &= 146.51\end{aligned}$$

Step 3 : Draw picture :



Step 4 : Compute the z-transform of $\overline{x}_0$ relative to H_a :

Rendered by QuickLaTeX.com

Step 5 : Look up the area $A(1.18)$ in the **Standard Normal Distribution Table**. That area will be $A(1.18) = 0.3810$. So $\beta = 0.5 - 0.3810 = 0.1190$ and **power** = $1 - \beta =$ **0.8810**.

▢

Example 13.6 : Two tailed z-test with data the same as Example 13.5.

Given :

$H_0 : \mu = 150$, $H_1 \neq 150$

$n = 50$, $\alpha = 0.05$, $s = 15$, $\overline{x} = \mu_a = 144$

Find the observed power.

Step 1 : Look up $z_{\alpha/2} = z_{0.025}$ in the ***t* Distribution Table** for a one-tailed test: $z_{0.025} = 1.960$.

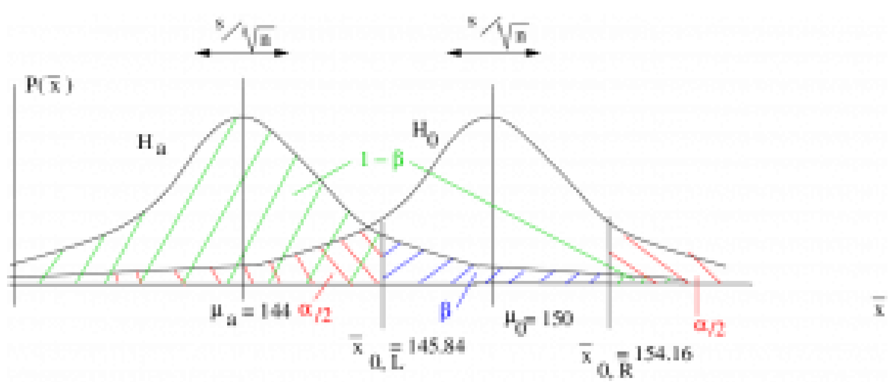
Step 2 : Compute:

Rendered by QuickLaTeX.com

and

Rendered by QuickLaTeX.com

Step 3 : Draw picture :



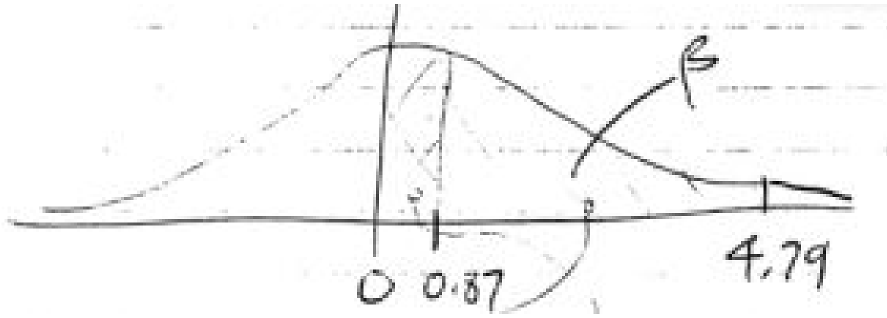
Step 4 : Compute the z-transform of $\overline{x}_{0,L}$ and $\overline{x}_{0,R}$ relative to H_a:

Rendered by QuickLaTeX.com

and

Rendered by QuickLaTeX.com

Step 5 : The two values, $z_{a,L}$ and $z_{a,R}$ appear on the z-distribution as :



So using the areas $A(z)$ from the **Standard Normal Distribution Table** we find

$$\beta = A(4.79) - A(0.87) = 0.5 - 0.3078 = 0.1922 \tag{13.1.5}$$

Notice that $z = 4.79$ is way the heck out there, it is higher than any z given in the **Standard Normal Distribution Table**. So $A(7.79)$ is essentially 0.5; the tail area past $z = 4.79$ is essentially zero. So the effect of going to from a one-tail to a two-tail test is only felt by the size of the H_0 critical region on the side where the test statistic ($\overline{x}$ here) is, which is half the size of the critical region in a one-tail test for a fixed α. In this case, then, the **power** = $1 - \beta =$ **0.8078** which is smaller than the value found in Example 13.5.

▢

Using observed power

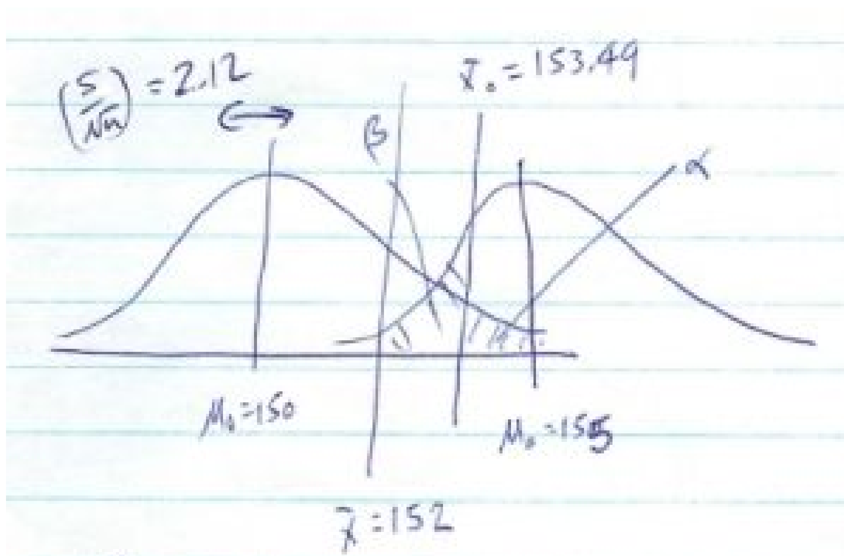
As mentioned earlier, almost no one is interested in observed power because we must reject H_0 to compute it. People are interested in β and power only when you report a failure to reject H_0.

Suppose in the situation of Example 13.1 we wanted to find evidence that $\mu_a = 155$ but measured $\overline{x} = 152$ (fail to reject H_0). Then, with our given information of

$H_0 : \mu \leq 150, image$ 150" title="Rendered by QuickLaTeX.com" height="16" width="73" style="vertical-align: -3px;">

$n = 50$, $\alpha = 0.05$, $s = 15$, $\overline{x} = 152$ and $\mu_a = 155$

we have



Based on the calculation we did in Example 13.1 we would report that we had a power of 0.7611 to detect an effect of $\mu_a = 155$ but with $\overline{x} = 152$ we were unable to detect μ_a.

---

1. And a corollary of this will be that one-tailed tests are more powerful than two-tailed tests. ↩

2. Effect size as defined in the Green and Salkind SPSS book would be $\mu_a - \mu_0/\sigma$. But that quantity is not useful here, so we define effect size as the difference of the means for the purpose of this discussion on power. Reference: Green SB, Salkind NJ. *Using SPSS for Windows and Macintosh: Analyzing and Understanding Data*, new edition pretty much every year, Pearson, Toronto, circa 2005. ↵
