"Section 9.6: The Power of a Test"

Some of the concepts and terminology in this handout are on pp. 303-304 of your text. I would suggest reading (or rereading) those pages. Actually, I recommend that you read Section 9.1 thoroughly. It is only a few pages. Let us proceed.

The following is an event:

$Event:$ Rejecting $\text{H}_0 \mid \text{H}_0$ is true

This event is called a Type I error. It is the equivalent of "convicting the innocent." Let us now determine the probability of this event.

$P(\text{Rejecting H}_0 \mid \text{H}_0 \text{ is true}) = P(\text{Type I error}) = \alpha$

Here is an event that automatically follows from the event of a Type I error:

$Event:$ Failing to reject $\text{H}_0 \mid \text{H}_0$ is true

This is what you are aiming to do. It is the equivalent of "acquitting the innocent." Its probability is the following:

$P(\text{Failing to reject H}_0 \mid \text{H}_0 \text{ is true}) = 1-\alpha$

Pay attention to this terminology:

α is called "significance." As α *decreases*, significance *increases*, a concept that you should keep in mind.

$1-\alpha$ is called "confidence." As α decreases, confidence increases.

* * *

Now, let's move on to the main topic of Section 9.6. Consider this event:

$Event:$ Failing to reject $\text{H}_0 \mid \text{H}_0$ is false

This event is called a Type II error. It is the equivalent of "acquitting the guilty." Let us now determine the probability of this event.

$P(\text{Failing to reject H}_0 \mid \text{H}_0 \text{ is false}) = P(\text{Type II error}) = \beta$

Here is an event that automatically follows from the event of a Type II error:

$Event:$ Rejecting $\text{H}_0 \mid \text{H}_0$ is false

You are also aiming to achieve this event. It is the equivalent of "convicting the guilty." Its probability is the following:

$P(\text{Rejecting H}_0 \mid \text{H}_0 \text{ is false}) = 1-\beta$

Note that you first determine what you consider to be the more serious error. You call that error α and fix its value at whatever you think to be appropriate (for example, 1%, 2%, 5%, or 10% are typical values). Before you take action — in other words, before you reject the null hypothesis — you accept the probability (α) that you are erroneously rejecting the null hypothesis. After you have set α, you can go on to set β if you wish. If you do not want to get involved with worrying about making a Type II error, then forget about β. In that case, either you reject the null hypothesis (hopefully, you did not make a Type I error), or you fail to reject the null hypothesis and accept the null hypothesis by default.

So far, so good. But what if you fail to reject the null hypothesis, and you do want to go on to prove that the null is true? In other words, you do not wish to accept the null hypothesis by default, but you want to try to *prove* that the null is true. It is the equivalent of proving that the defendant who has just been acquitted is actually innocent.

Now pay attention to this terminology:

β has no special word attached to it.

$1-\beta$ is called "power." As β decreases, power increases.

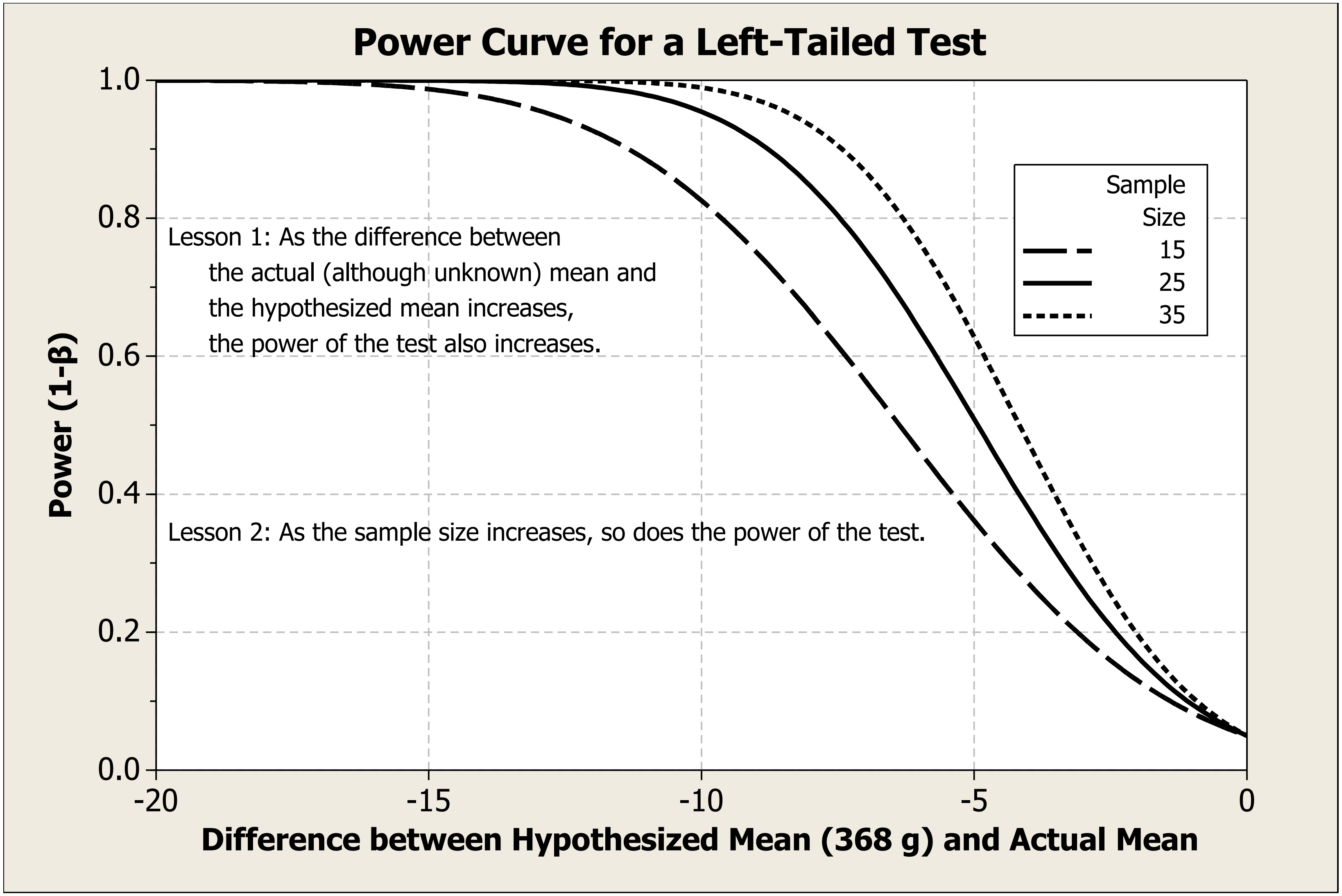
Several factors control how powerful a test is.

- $\alpha \downarrow \Rightarrow \beta \uparrow \Rightarrow (1-\beta) \downarrow$. In words, we are saying that "as significance increases, power decreases." I am not proving this somewhat advanced topic here, but I am stating it as a fact. It comes down to this: "There is no free lunch. If you want to minimize a Type I error, it will cost you power."
- $n \uparrow \Rightarrow \left[\alpha \downarrow \text{ and } (1-\beta) \uparrow\right]$. Again, I am not proving this here, although I demonstrate it in the accompanying graph. Intuitively, however, it makes sense: If you acquire more data (a bigger sample size), then you obtain more significance and more power.
- Finally, one of the main lessons to take away is this: As the actual (but unknown) mean gets farther away from the hypothesized mean, power increases. *Do* try to follow this one. If the actual value of the mean is much less than the hypothesized mean, then probably your sample mean will be substantially less than the hypothesized mean. So you will have less chance of falsely failing to reject H_0 (you will have a smaller β), and therefore you have a greater chance of correctly failing to reject a true H_0 (more power).

Consider the hypotheses of the Oxford Cereal Company:

$$\mathrm{H}_0 : \mu \geq 368 \text{ g}$$
$$\mathrm{H}_1 : \mu < 368 \text{ g}$$

So let's say that the true (but unknown) mean were 367 g. Then the result of the sample could easily be near or even above the hypothesized mean of 368 g, and you would incorrectly fail to reject H_0. But now let's say that the true (but unknown) mean were 352 g. Then the result of the sample would probably much lower than the hypothesized mean of 368 g, so you would have a much better chance of correctly rejecting H_0. That is to say, β would be small, so power would be high.



Power Curve for a Left-Tailed Test
Power (1-β)
1.0
0.8
0.6
0.4
0.2
0.0
-20
-15
-10
-5
0
Difference between Hypothesized Mean (368 g) and Actual Mean
Sample Size
15
25
35
Lesson 1: As the difference between the actual (although unknown) mean and the hypothesized mean increases, the power of the test also increases.
Lesson 2: As the sample size increases, so does the power of the test.