The probability of *not* committing a Type II Error is called the **Power** of a hypothesis test.

Effect Size

To compute the power of the test, one offers an alternative view about the "true" value of the population parameter, assuming that the null hypothesis is false. The **effect size** is the difference between the true value and the value specified in the null hypothesis.

Effect size = True value - Hypothesized value

For example, suppose the null hypothesis states that a population mean is equal to 100. A researcher might ask: What is the probability of rejecting the null hypothesis if the true population mean is equal to 90? In this example, the effect size would be 90 - 100, which equals -10.

Factors That Affect Power

The power of a hypothesis test is affected by three factors.

- Sample size (*n*). Other things being equal, the greater the sample size, the greater the power of the test.
- Significance level (α). The higher the significance level, the higher the power of the test. If you increase the significance level, you reduce the region of acceptance. As a result, you are more likely to reject the null hypothesis. This means you are less likely to accept the null hypothesis when it is false; i.e., less likely to make a Type II error. Hence, the power of the test is increased.
- The "true" value of the parameter being tested. The greater the difference between the "true" value of a parameter and the value specified in the null hypothesis, the greater the power of the test. That is, the greater the effect size, the greater the power of the test.

Problem 1: Other things being equal, which of the following actions will reduce the power of a hypothesis test?

I. Increasing sample size.
II. Increasing significance level.
III. Increasing beta, the probability of a Type II error.

(A) I only (B) II only (C) III only (D) All of the above (E) None of the above

Solution

The correct answer is (C). Increasing sample size makes the hypothesis test more sensitive - more likely to reject the null hypothesis when it is, in fact, false. Increasing the significance level reduces the region of acceptance, which makes the hypothesis test more likely to reject the null hypothesis, thus increasing the power of the test. Since, by definition, power is equal to one minus beta, the power of a test will get smaller as beta gets bigger.

Problem 2: Suppose a researcher conducts an experiment to test a hypothesis. If she doubles her sample size, which of the following will increase?

I. The power of the hypothesis test.
II. The effect size of the hypothesis test.
III. The probability of making a Type II error.

(A) I only (B) II only (C) III only (D) All of the above (E) None of the above

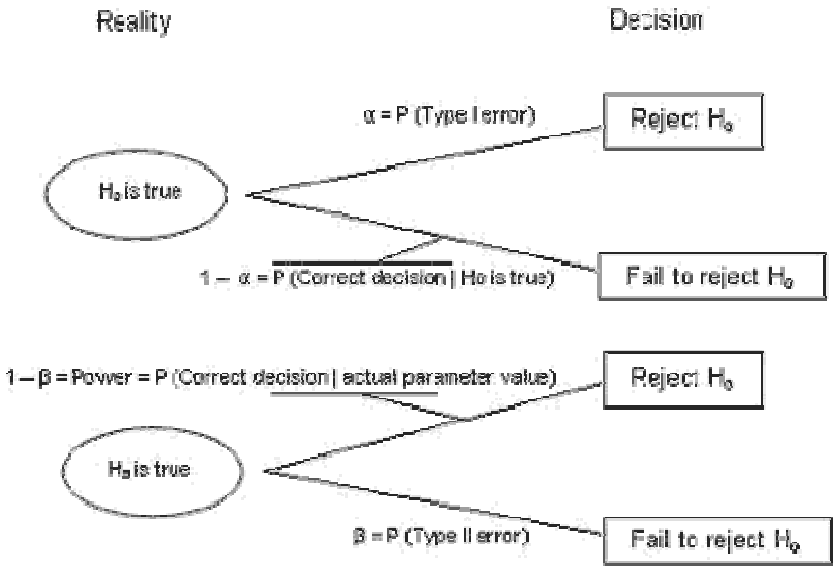
Solution

The correct answer is (A). Increasing sample size makes the hypothesis test more sensitive - more likely to reject the null hypothesis when it is, in fact, false. Thus, it increases the power of the test. The effect size is not affected by sample size. And the probability of making a Type II Error gets smaller, not bigger, as sample size increases.

What Does Power Mean?
Quite simply, the power of a hypothesis test is the probability that it will lead to a rejection of the null hypothesis. In other words, power is the probability of **correctly** rejecting the null hypothesis.

Let's make that clearer before continuing, since it is quite important in a discussion of power. A hypothesis test begins with a null hypothesis, which usually proposes a very **particular** value for a parameter or the difference between two parameters (for example, "$\mu = \mu_0$ " or "$p_1 - p_2 = 0$ “).[1] Then it includes "an" alternate hypothesis, which is usually in fact a **collection** of possible parameter values competing with the one proposed in the null hypothesis (for example, "$\mu \neq \mu_0$ " which is really a collection of possible values of μ , and $p_1 - p_2 \neq 0$," which allows for many possible values of). The **power** of a hypothesis test is the probability of rejecting the null, but this implicitly depends upon what the value of the parameter or the difference in parameter values **really is**.

The following tree diagram may help appreciate the fact that α, β, and power are all conditional probabilities.



Power may be expressed in several different ways. Here are a few different ways to describe what power is:

- Power is the probability of rejecting the null hypothesis when in fact it is false.
- Power is the probability of making a correct decision (to reject the null hypothesis) when the null hypothesis is false.
- Power is the probability that a test of significance will pick up on an effect that is present.
- Power is the probability that a test of significance will detect a deviation from the null hypothesis, should such a deviation exist.
- Power is the probability of avoiding a Type II error.

I have found it helpful as we discuss power to continually restate what power means throughout discussions, using different language each time. For example, if we do a test of significance at level α = 0.1, I might say, "That's a pretty big alpha level. This test is ready to reject the null at the drop of a hat. Is this a very powerful test?" (Yes, it is. Or at least, it's more powerful than it would be with a smaller alpha value.) If a student answers a question about Type II errors and says that the consequences of a Type II error are very severe, then I may follow up with the question, "So you really want to avoid Type II errors, huh? What does that say about what we require of our test of significance?" (We want a very powerful test.)

The computation of statistical power depends on specific a model (or test) and is not part of the AP Stats currilculum but it may clarify things for you. The easiest model is the T-test with a relatively simple formula.

Imagine a random variable like the number of deaths per 100 thousand people from lung cancer. Suppose that the variable is known to be normally distributed with a mean of 20 and a standard deviation of 4: a null hypothesis that population mean is 20. See the probability distribution A in Figure 2.

Now, we believe that the mean is not 20, but 22 with the same standard deviation. See the probability distribution B in Figure 2. We also think that test size .05 will be fine. So, we are going to conduct a two-tailed T-test at the .05 significance level with the alternative hypothesis that the population mean is 22. How can we test our conjecture (alternative hypothesis)? Suppose we took a random sample with 44 observations from the population.

Think about the ordinary T-test first. We need to know how far the 22 is deviated from the baseline 20. Of course, the distance (effect size) is 2 (=22-20). But we do not know how big the effect size 2 is. Put differently, we do not know exactly how likely such a sample mean 22 can be observed if the true mean is 20. This is why people try to take advantage of using standardized probability distributions (e.g., T, F, and Chi-squared). By looking through these distribution tables, we are able to know the likelihood of observing a sample statistic in fairly easy manner. The A' in Figure 2 is a standardized probability distribution of A. The 20 in A corresponds to 0 in A' and 21.2161 in A is equivalent to 2.0167 in A'.

The t statistic here is 3.3166248=(22-20)/4*SQRT(44) . This value is located all the way to the right in A', indicating the p-value is extremely small. If the null hypothesis of population mean 20 is true, it is quite unlikely to observe the sample mean 22 (p<=.01). Obviously, the conjecture of population mean 20 is not likely. Thus, the null hypothesis is undoubtedly rejected in favor of the alternative hypothesis.

However, this test does not tell anything about the power of the test. We may want to know the extent that a test can detect the effect, if any. So, the question here is, "How powerful is this T-test?" There are four steps to compute the statistical power.

Figure 2. Statistical Power of a T-test