The importance of power and sample size to research conducted

Power:
Power is the additive inverse of the Type 2 error known as Beta (β). As such it is calculated as \( 1 - \beta \).

Here are some examples of this relationship:

If you have a power of 80%, what is your Type 2 error rate? \( .80 = 1 - \beta \). Since 1 is 1.00, you would subtract .20 from it to get .80. Therefore your \( \beta \) is .20, or expressed as a percent 20%. This is a very common level for Type 2 error and is quite a bit larger than the Type 1 error rate widely used, which is .05 or 5%. It means that there is a 20% chance that you will fail to reject a null hypothesis that is false or that you will not be able to support your alternative hypothesis statistically. Why do you think biostatisticians would be willing to accept such a high risk of drawing the wrong conclusion? The answer is that they really are not willing to accept it but to minimize the chance of both Type 1 and Type 2 they would need very large sample sizes (as you will see in the section on sample size) and that is often cost and time prohibitive. In the trade off they decided that it was better to have a Type 2 error than a Type 1 error in most circumstances. There are times, however, when failing to reject a false null hypothesis can be very detrimental to future research.

Here are some other examples of the relationship between \( \beta \) and Power.

- **Power of .85 = 1 - a \( \beta \) of .15**
- **Power of .90 = 1 - a \( \beta \) of .10**
- **Power of .95 = 1 - a \( \beta \) of .05**

Sample size:
Most of the variables used to calculate the Type 1 and Type 2 error rates in statistical calculations are fixed, that is you the statistician cannot control or manipulate them. The only variable you can control is the size of your sample, but even there you are limited by money and time. Typically, the larger the samples size the more money and time it costs the researcher to conduct the research. The exception to this is that large national datasets, such as NHANES, NHIS, and BRFSS (don’t worry here about what the acronyms represent), available through the CDC do not cost any money to use. This makes secondary data analysis of existing data very attractive to researchers. Cost and time are associated.
with the collection of the sample and not the analysis. The costs and time of analysis are fixed regardless of the size of the sample you are examining.

Here is a commonly used sample size calculation (used to calculate the sample size requirements of the $t$ test:

$$n = \left( \frac{Z_{1-\alpha/2} + Z_{1-\beta}}{ES} \right)^2$$

It looks a lot like Greek, doesn’t it? That is because most of the letters used by mathematicians to replace unknown variables are Greek. Do not let that scare you off. You will see what each of the symbols mean and that will help you understand where you plug in your values.

- $n$ is the total sample size you will need to conduct the statistical test
- $Z$ refers to the table in the back of the text beginning on page 272.
- The subscript of the first $Z$ in the equation is $1 - \alpha/2$, where $\alpha$ is the Type 1 error rate you are allowing in this statistical test. It is divided by 2 for a two-tailed test, if your hypothesis is directional and you only need a 1 tailed test it would be $\alpha$. The best part of this is that the table tells you what value you enter here. The text even gives you a simplified table on page 275 you can use for this. The typical value for $Z_{1-\alpha/2}$ is 1.96.
- The subscript of the first $Z$ in the equation is $1 - \beta$, where $\beta$ is the Type 2 error rate you are allowing in this statistical test. This should look familiar to you from the discussion on Power. Unfortunately the book does not have a table to give you the most frequently used values for $\beta$, so this is the table you should use:

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$1-\beta$</th>
<th>$Z_{1-\beta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.20</td>
<td>.80</td>
<td>0.84</td>
</tr>
<tr>
<td>.15</td>
<td>.85</td>
<td>1.04</td>
</tr>
<tr>
<td>.10</td>
<td>.90</td>
<td>1.28</td>
</tr>
<tr>
<td>.05</td>
<td>.95</td>
<td>1.64</td>
</tr>
</tbody>
</table>

The typical value for $Z_{1-\beta}$ is .84.

- Finally $ES$ is the difference you are expecting to observe. This is based on an additional formula:

$$ES = \frac{|\mu_1 - \mu_0|}{\sigma}$$
In this formula the \( \mu \) (1 and 2) are the means you plan to compare. Since you will not have that information prior to collecting and analyzing your data you need to estimate these. Frequently the estimate is based on a review of the literature. The \( \sigma \) is the standard deviation, also unknown and usually based on other research done using similar variables and populations. This is the hardest part of a sample size calculation. There is software available to do the math for you but you need to be able to justify each of the values you input into the software and these are the hardest to justify.

Here is an example of how it is actually used in Public Health:

Here is an example of the formula as used in Public Health:

The mean body mass index (BMI) for boys age 12 is 23.6. An investigator wants to test if the BMI is higher in 12-year-old boys living in New York City. How many boys are needed to ensure that a two-sided test of hypothesis has 80% power to detect a difference in BMI of 2 units? Assume that the standard deviation in BMI is 5.7.

\[
ES = \frac{|\mu_1 - \mu_0|}{\sigma} = \frac{2}{5.7} = 0.35,
\]

\[
n = \left( \frac{Z_{1-\alpha/2} + Z_{1-\beta}}{ES} \right)^2 = \left( \frac{1.96 + 0.84}{0.35} \right)^2 = 64.0.
\]

A sample of size \( n = 64 \) will ensure that the test of hypothesis will have 80% power to detect a difference of 2 units in BMI.

If this was all you needed to do every time, it would be a relatively simple process to calculate sample size. Unfortunately, each of the statistical tests you have learned (or will learn) about this quarter has its own equation. Therefore you need to determine the appropriate formula(s) for the test(s) you plan to use. If you plan on more than one test, you need to calculate the sample size requirement for each one and use the highest as your minimum sample size.

Here are a few other things you need to consider in computing the sample size:
• Potential losses to follow up (or drop outs) in a study conducted over a period of time.
• Potential missing values. These are questions in a survey instrument that the participant may skip for any number of reasons.
• Multiple testing. As you learned in week 7, multiple tests conducted on the same data require adjustments to the Type 1 error rate. To maintain an error rate of .05 as well as a power of .80, you need to increase your sample size.

**Type 1 and Type 2 error and Sample Size:**
If you look again at the equation for sample size you will notice that both the value for $Z_{1-\alpha/2}$ and the value for $Z_{1-\beta}$ are in the numerator of the equation. This means that as these values increase the value for $n$ increases as well (a direct relationship). If you recalculate the example equation with a Power of 90% instead of 80% you get the following:

\[ n = \left( \frac{Z_{1-\alpha/2} + Z_{1-\beta}}{\text{ES}} \right)^2 = \left( \frac{1.96 + 1.28}{0.35} \right)^2 = 86. \]

Just that small increase in Power increases the sample size requirement from 64 to 86.

You should also note that the ES, or expected difference in a $t$ test, is in the denominator. This means that as the ES gets smaller the $n$ becomes larger (an inverse relationship). Therefore, if your ES is smaller, as is often the case, the sample size grows even larger. Use an expected difference of .25 with a Power of 90% and a Type 1 error rate of 5% yields a minimum $n$ of 168.