The Power of a Statistical Test

- Definition
- Ways to Increase the Power
- Two Types of Error Revisited

The power of a statistical test measures its ability to detect an alternative hypothesis.

The power against a specific alternative is calculated as the probability that the test will reject $H_0$ when that specific alternative is true.

This calculation requires knowledge of the sampling distribution of the test statistic under the alternative hypothesis.

Example: Does exercise make strong bones?

Can a 6-month exercise program increase the total body bone mineral content (TBBMC) of young women? A team of researchers is planning a study to examine this question. Based on the results of a previous study, they are willing to assume that $\sigma = 2$ for the percent change in TBBMC over the 6-month period. A change in TBBMC of 1% change would be considered important, and the researcher would like to have a reasonable chance of detecting a change this large or larger. Are 25 subjects a large enough sample for this project?

Ways to Increase the Power

- Increase $\alpha$. A 5% test of significance will have a greater chance of rejecting the alternative than a 1% test because the strength of evidence required for rejection is less.
- Consider a particular alternative that is farther away from $\mu_0$. Values of $\mu$ that are in $H_a$ but lie close to the hypothesized value $\mu_0$ are harder to detect than values of $\mu$ that are far from $\mu_0$.
- Increase the sample size. More data will provide more information about $\bar{x}$ so we have a better chance of distinguishing values of $\mu$.
- Decrease $\sigma$. This has the same effect as increasing the sample size: it provides more information about $\mu$. Improving the measurement process and restricting attention to a subpopulation are two common ways to decrease $\sigma$. 

1. State the hypotheses: let $\mu$ denote the mean percent change.

   $H_0 : \mu = 0$
   $H_a : \mu > 0$

2. Calculate the rejection region: The $z$ test rejects $H_0$ at the $\alpha = 0.05$ level whenever

   $$ z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \geq \frac{\bar{x} - 0}{2/\sqrt{25}} \geq 1.645 $$

   that is, we reject $H_0$ when

   $\bar{x} \geq 1.645 - \frac{2}{\sqrt{25}} = 0.658$

3. Compute the power at a specific alternative:

   The power of the test at alternative $\mu = 1$ is

   $$ P(\bar{x} \geq 0.658|\mu = 1) = P(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \geq \frac{0.658 - 1}{2/\sqrt{25}}|\mu = 1) $$

   $$ = P(Z \geq -0.855) $$

   $$ = 0.80 $$
Example: Does exercise make strong bones? (cont.)

- Change the significance level to $\alpha = 0.01$.

The $z$ test rejects $H_0$ at the $\alpha = 0.01$ level whenever

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{x} - p}{2/\sqrt{25}} \geq 2.32$$

that is, we reject $H_0$ when $\bar{x} \geq 2.32 \frac{1}{\sqrt{25}} = 0.928$

The power of the test at alternative $\mu = 1$ is

$$P(\bar{x} \geq 0.928|\mu = 1) = P\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \geq \frac{0.928 - 1}{2/\sqrt{25}} | \mu = 1\right)$$

$$= P(Z \geq -0.18)$$

$$= 0.57$$

- Change the alternative to $\mu = 2$.

The power of the test at alternative $\mu = 2$ at $\alpha = 0.01$ level is

$$P(\bar{x} \geq 0.928|\mu = 2) = P\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \geq \frac{0.928 - 2}{2/\sqrt{25}} | \mu = 2\right)$$

$$= P(Z \geq -2.68)$$

$$= 0.996$$

- Increase the sample size $n = 100$.

The $z$ test rejects $H_0$ at the $\alpha = 0.01$ level whenever

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{x} - p}{2/\sqrt{100}} \geq 2.32$$

that is, we reject $H_0$ when $\bar{x} \geq 2.32 \frac{1}{\sqrt{100}} = 0.464$

The power of the test at alternative $\mu = 1$ is

$$P(\bar{x} \geq 0.464|\mu = 1) = P\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \geq \frac{0.464 - 1}{2/\sqrt{100}} | \mu = 1\right)$$

$$= P(Z \geq -2.68)$$

$$= 0.996$$

- Decrease $\sigma = 1$

The $z$ test rejects $H_0$ at the $\alpha = 0.01$ level whenever

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{x} - p}{1/\sqrt{100}} \geq 2.32$$

that is, we reject $H_0$ when $\bar{x} \geq 2.32 \frac{1}{\sqrt{100}} = 0.232$

The power of the test at alternative $\mu = 1$ is

$$P(\bar{x} \geq 0.232|\mu = 1) = P\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \geq \frac{0.232 - 1}{1/\sqrt{100}} | \mu = 1\right)$$

$$= P(Z \geq -7.68) \approx 1$$

### Two Types of Errors Revisited

Recall that the following four outcomes are possible when conducting a test:

<table>
<thead>
<tr>
<th>Reality</th>
<th>Our Decision</th>
<th>$H_0$</th>
<th>$H_a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$</td>
<td>$\checkmark$</td>
<td>Type I Error</td>
<td>(Prob = $1 - \alpha$)</td>
</tr>
<tr>
<td>$H_a$</td>
<td>Type II Error</td>
<td>$\checkmark$</td>
<td>(Prob = $\beta$)</td>
</tr>
</tbody>
</table>

The significance level $\alpha$ of any fixed level test is the probability of a Type I error.

The power of a fixed level test against a particular alternative is $1 - \beta$ for that alternative.

In practice, we first choose an $\alpha$ and consider only tests with probability of Type I error no greater than $\alpha$.

Among all level $\alpha$, we select one that makes the probability of Type II error as small as possible (i.e. the most powerful possible test).