Lecture 11

11.1 Calculating the Power of a Hypothesis Test

- It is relatively easy to calculate the power of the the kinds of hypothesis tests we have been discussing.

- You'll recall that a test's statistical power is $1-\beta$. It is the probability that the test will falsely reject a positive result, or the probability of the commission of a Type II error.

- Go over handout called "Type I, Type II error."

- So how do we calculate β?

$$
\begin{aligned}
\beta &= \Pr\left(\text{Accept } H_0 | H_A \text{ true}\right) \\
&= \Pr\left(\widehat{\theta} < \theta_0 + z_\alpha \sigma_{\widehat{\theta}} | \theta = \theta_A\right)
\end{aligned}
$$

- Well we know that $\widehat{\theta}$ is distributed Normal with mean θ_A and standard deviation $\sigma_{\widehat{\theta}}$. So we therefore know the probability with which it will take on any value. First standardize $\widehat{\theta}$:

$$
\begin{aligned}
&= \Pr\left(\frac{\widehat{\theta}-\theta_A}{\sigma_{\widehat{\theta}}} < \frac{\theta_0 + z_\alpha \sigma_{\widehat{\theta}} - \theta_A}{\sigma_{\widehat{\theta}}} | \theta = \theta_A\right) \\
&= \Phi\left(\frac{\theta_0 + z_\alpha \sigma_{\widehat{\theta}} - \theta_A}{\sigma_{\widehat{\theta}}}\right). \\
&= \Phi\left(\frac{\theta_0 - \theta_A}{\sigma_{\widehat{\theta}}} + z_\alpha\right)
\end{aligned}
$$

- So a test's power is therefore
$$1 - \Phi\left(\frac{\theta_0 - \theta_A}{\sigma_{\widehat{\theta}}} + z_\alpha\right).$$

- Note that we have all the quantities necessary to calculate β and therefore power:

  - We've specified a θ_0 and θ_A.

– We've also specified a α and therefore a z_α.

– And we use our usual methods to obtain $\sigma_{\widehat{\theta}} = \frac{\sigma}{\sqrt{n}}$.

  * (So we might actually write:

$$Power = 1 - \Phi\left(\frac{\theta_0 - \theta_A}{\frac{\sigma}{\sqrt{n}}} + z_\alpha\right).$$

- Now for something interesting. What is the sign of $\frac{\partial \text{Power}}{\partial \alpha}$? Of $\frac{\partial \text{Power}}{\partial \sigma}$? Of $\frac{\partial \text{Power}}{\partial n}$? Of $\frac{\partial \text{Power}}{\partial (|\theta_0 - \theta_A|)}$?

- NOTE TO SELF: Note that $\theta_0 - \theta_A < 0$!

- Well, since Φ is monotonically increasing in its argument, we know that:

  – $\frac{\partial z_\alpha}{\partial \alpha} < 0$, and so $\frac{\partial \Phi\left(\frac{\theta_0 - \theta_A}{\frac{\sigma}{\sqrt{n}}} + z_\alpha\right)}{\partial \alpha} < 0$, and so $\frac{\partial \text{Power}}{\partial \alpha} > 0$. As we increase α (that is, decrease our confidence coefficient) we increase power.

  – $\frac{\partial \frac{\theta_0 - \theta_A}{\frac{\sigma}{\sqrt{n}}}}{\partial \sigma} = \frac{\partial \frac{(\theta_0 - \theta_A)\sqrt{n}}{\sigma}}{\partial \sigma} > 0$ (since $\theta_0 - \theta_A < 0$), and so $\frac{\partial \Phi\left(\frac{\theta_0 - \theta_A}{\frac{\sigma}{\sqrt{n}}} + z_\alpha\right)}{\partial \sigma} > 0$, and so $\frac{\partial \text{Power}}{\partial \sigma} < 0$. As there is more variance in the population, our estimator becomes less precise, and the power of our statistical tests goes down.

  – Conversely, $\frac{\partial \frac{\theta_0 - \theta_A}{\frac{\sigma}{\sqrt{n}}}}{\partial n} < 0$, and so $\frac{\partial \text{Power}}{\partial n} > 0$.

  – And finally $\frac{\partial \text{Power}}{\partial (|\theta_0 - \theta_A|)} = \frac{\partial \text{Power}}{\partial (\theta_A - \theta_0)}$(since $\theta_0 - \theta_A < 0$). Since $\frac{\partial \Phi\left(\frac{\theta_0 - \theta_A}{\frac{\sigma}{\sqrt{n}}} + z_\alpha\right)}{\partial (\theta_A - \theta_0)} < 0$, $\frac{\partial \text{Power}}{\partial (|\theta_0 - \theta_A|)} > 0$. The farther way you specify an alternative hypothesis away from the null, the less likely you are to falsely reject the null.

- A vice president in charge of sales for a large corporation claims that salespeople are averaging no more than 15 sales contacts per week. (He would like to increase this figure.) As a check on his claim, n = 36 salespeople are selected at random, and the number of contacts made by each is recorded for a single randomly selected week. The mean and variance of the 36 measurements were 17 and 9, respectively. Does the evidence contradict the vice president's claim? Use a test with level α = .05.

  – $H_0 : \mu = 15$

- $H_A : \mu > 15$
- We know that $\bar{Y} \sim \mathcal{N}(\mu_{\bar{Y}}, \sigma_{\bar{Y}})$, meaning our test statistic can be $Z = \frac{\bar{Y}-\mu_0}{\sigma_{\bar{Y}}} = \frac{\bar{Y}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$
- Our rejection region for $\alpha = 0.05$ is given by $z = 1.645$.
- Thus $Z = \frac{\bar{Y}-\mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{17-15}{3/6} = 4$ which is inside the rejection region.

- Now suppose the VP wants to be able to detect a difference equal to one call in the mean number of calls per week. That is, he wishes to test $H_0 : \mu = 15$ against the alternative $H_A : \mu = 16$. What is the β for this test?
  - Given the same set-up as above, the rejection region for $\alpha = .05$ is $\bar{y} > \mu_0 + 1.645\left(\frac{\sigma}{\sqrt{n}}\right)$.
  - Substituting our values, we get $\bar{y} > 15 + 1.645\left(\frac{3}{\sqrt{36}}\right)$ or $\bar{y} > 15.8225$.
  - Thus $\beta = P(\bar{Y} \leq 15.8225 | \mu_A = 16)$ or $P\left(\frac{\bar{Y}-\mu_a}{\sigma/\sqrt{n}} \leq \frac{15.8225-16}{3/\sqrt{36}}\right) = P(Z \leq -.36) = .3594$

11.2 Another way to report results of a statistical test: *p*-values

- You'll recall that α, the probability of a Type I ("false positive") error associated with a statistical test, is often called the test's **significance level.**
- In the hypothesis testing regime we've discussed so far, we:
  - pick a significance level
  - determine the critical value(s) of the test statistic associated with the significance level
  - and then determine whether to accept or reject the null hypothesis by comparing our test statistic with the critical value.
- This is all very black-or-white: the result is either significant or it's not. But if report only whether we reject or accept the null, we're actually failure to report a fair amount of information.
- Another way to report results of a statistical test that provides the reader with this additional information is to report what's called its p-value.

- For any test statistic, the *p-value*, or **attained significance level**, is the smallest level of significance α for which the observed data indicate that the null hypothesis should be rejected.

- The smaller a p-value is, the more compelling is the evidence that the null hypothesis should be rejected. That is because the null should be rejected for any value of α *down to and including* the p-value.

- Reporting the p-value permits the reader to make her own choice about whether the observed data justify a rejection of the null.

- This should be obvious, but:

$$\text{Reject H}_0 \iff p \leq \alpha$$
$$\text{Accept H}_0 \iff \alpha \leq p$$

  - A psychological study was conducted to compare the reaction times of men and women to a stimulus. Independent random samples of 50 men and 50 women were employed in the experiment. The results are shown in the table below. Do the data present sufficient evidence to suggest a difference between true mean reaction times for men and women? Use $\alpha = .05$.

Men	Women
$n_1 = 50$	$n_2 = 50$
$\bar{y}_1 = 3.6$	$\bar{y}_2 = 3.8$
$s_1^2 = .18$	$s_2^2 = .14$

  - $H_0 : (\mu_1 - \mu_2) = 0$
  - $H_A : (\mu_1 - \mu_2) \neq 0$
  - Test statistic: $Z = \frac{(\bar{Y}_1 - \bar{Y}_2) - \theta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$
  - Setting $\theta_0 = 0$ yields $Z = \frac{(3.6-3.8)-0}{\sqrt{\frac{.18}{50} + \frac{.14}{50}}} = -2.5$ which is below the $-z_{\alpha/2} = -1.96$ so we reject the null.
  - What is the $p - value$ associated with this conclusion?

- $P(Z \leq -2.5) = .0062$ which means $\alpha = 0.0062 * 2 = 0.0124$. Note that, if we set $\alpha = .01$, we would NOT be able to reject.