**Vanderbilt University Political Science - Stats I**
**Fall 2024 - Prof. Jim Bisbee**

# Lecture 10

## 10.1 Finishing up last example

- Consider this example where we left-off from last class:

  - * In a Zogby Poll conducted with 1,203 likely voters nationwide between Sep 24-28, 2023, Donald Trump led Joe Biden, 52.5 percent to 47.5 percent, among those expressing a preference.
    * This is a tracking poll. In the previous three-day window of the poll (Sep 21-23), Trump led Biden 55.6 to 44.4 percent (N=1,203).
    * According to the poll, Trump's lead shrunk by about six points in three days. How confident are we that this change is not due to sampling error?
    * Set it up:
    * The parameter we seek is now $p_1 - p_2$, where $p_1$ = Trump's true support in the first poll (Sep 21-23) and $p_2$ = Trump's true support in the second poll.
    * The polls may be considered two binomial experiments in which $Y_1$ is the number of "successes" (here, the # favoring Trump) in the first poll, (no ideological agenda) and $Y_2$ is the number of of such "successes" in the second poll.
    * An intuitive estimator for this quantity would be $\widehat{p}_1 - \widehat{p}_2$, where the p-hats are the proportions of respondents favoring Trump in the two polls. Is it an unbiased estimator for $p_1 - p_2$?

$$
\begin{aligned}
E(\widehat{p}_1 - \widehat{p}_2) &= E(\widehat{p}_1) - E(\widehat{p}_2) \\
&= E\left(\frac{Y_1}{n_1}\right) - E\left(\frac{Y_2}{n_2}\right) \text{ [}\widehat{p}_1 \text{ and } \widehat{p}_2 \text{ are functions of the RVs } Y_1, Y_2\text{]} \\
& \frac{1}{n_1} E\left(Y_1\right) - \frac{1}{n_2} E\left(Y_2\right) \\
&= \frac{1}{n_1} n_1 p_1 - \frac{1}{n_2} n_2 p_2 \text{ [}E\left(Y\right) = np \text{ if } Y \text{ is distributed binomial]} \\
&= p_1 - p_2.
\end{aligned}
$$

* Our next step is to say how precise $\widehat{p}_1 - \widehat{p}_2$ tends to be as an estimator of $p_1 - p_2$.

* We do this by figuring out what the estimator's standard error is. It's

$$\begin{aligned}\sqrt{VAR(\widehat{p}_1 - \widehat{p}_2)} &= \sqrt{VAR(\widehat{p}_1) + VAR(\widehat{p}_2)} \text{ [assume samples drawn independently]} \\ &= \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\end{aligned}$$

* We make the substitution

$$(\widehat{p}_1 - \widehat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}$$

* Plugging in, we have

$$\begin{aligned}&(55.6 - 52.5) \pm z_{\alpha/2}\sqrt{\frac{(55.6)\,(100-55.6)}{1,203} + \frac{(52.5)(100-52.5)}{1,203}} \\ &3.1 \pm z_{\alpha/2}(2.031).\end{aligned}$$

* Do you recall how we find $z_{\alpha/2}$? We type `qnorm`$(\frac{\alpha}{2})$, substituting our chosen $\alpha$. You'll remember that $z_{\alpha/2}$ associated with an $\alpha = .05$ is $z_{.025} = -1.96$. So our 95% CI is:

$$3.1 \pm 1.96(2.031) = 3.1 \pm 3.98, \text{ or } [-.9, 7.1].$$

* We are 95% confident that the true change between the two polls was between -.9 and 7.1 percentage points.

- Note that this CI includes zero. So another interpretation of this CI is that we are **not** 95% confident that there was zero change between the two polls. And this, of course, is what we really wanted to know: was there truly any movement between Oct 21-23 and Oct 24-26?

- Now, does the 90% confidence interval about our point estimate include zero?

  – Let's see: our alpha is .10.

  – typing `qnorm`(.05) gives us -1.64. So our 90% CI is:

$$3.1 \pm 1.64(2.031) = 3.1 \pm 3.33, \text{ or } [-.23,\ 6.43].$$

- Still no cigar. At what level of confidence would we be satisfied that there was movement between the two surveys?

- Think: we wish to find some $\alpha^*$ such that the lower bound of the $100 * (1-\alpha)$ CI is greater than zero. That is, find some $\alpha^*$ meeting this criterion:

$$\alpha^* : 3.1 - z_{\alpha^*/2}(2.031) > 0.$$

- To do this, manipulate the expression

$$\begin{aligned} -z_{\alpha^*/2}(2.031) &> -3.1 \\ z_{\alpha^*/2} &< \frac{3.1}{2.031} \\ z_{\alpha^*/2} &< 1.5263 \end{aligned}$$

- So for any alpha such that $z_{\alpha/2} < 1.5263$, we will be $100 * (1-\alpha)$ percent confident that the true change was greater than zero. How do we find this $\alpha$? Well, if

$$\begin{aligned} z_{\frac{\alpha}{2}} &= -\Phi^{-1}\left(\frac{\alpha}{2}\right), \text{ then} \\ \Phi\left(-z_{\frac{\alpha}{2}}\right) &= \frac{\alpha}{2}, \text{ and} \\ \alpha &= 2\Phi\left(-z_{\frac{\alpha}{2}}\right). \end{aligned}$$

- So in this particular case, $\alpha = 2\Phi(-1.5263)$.

  - To find this alpha, we now type `pnorm`(-1.5263) in `R`, which is the CDF of the standard Normal evaluated at its argument. This returns **.063**.

  - Thus $\alpha/2 = .063$ and alpha is thus .126.

  - And thus if we are working with confidence intervals of $100 * (1 - .126) = 87.4\%$ or smaller, we will conclude that there was true movement between the two polls.

## 10.2 Hypothesis Testing

- This way of framing the question motivates a process known as **hypothesis testing**. A hypothesis test consists of four elements:

  1. A **null hypothesis about a parameter**, which we write as $H_0$.
     - This is typically either what the "conventional wisdom" says is the the value of the parameter–or that the parameter is equal to zero.
  2. An **alternative hypothesis about the parameter**, $H_A$.
     - This is typically that the parameter is equal to *something different* than the null hypothesis. It may be more specific: that the parameter is either greater than or less than the null hypothesis.
  3. A **test statistic derived from an estimator of the parameter.**
  4. A **rejection region**.
     - The RR specifies the range of values of the test statistic for which the null $H_0$ is to be *rejected* in favor of the alternative $H_A$.

- Choosing the rejection region:
  - RR's are associated with two kinds of error:
    * Type I error (a.k.a. a "false positive") is made if $H_0$ is rejected when $H_O$ is actually true.
      · $Pr(\text{Type I error}) = \alpha$. (Yes, the very same $\alpha$ we've been working with.)
    * Type II error (a.k.a. a "false negative") is made if $H_0$ is accepted when $H_A$ is actually true.
      · $Pr(\text{Type II error}) = \beta$.
  - $\alpha$ and $\beta$ are two very practical ways to measure the goodness of a statistical test. We call $\alpha$ the test's **level of significance**. We call the quantity $1 - \beta$ the test's **statistical power**. In the best of all worlds, we want a test's level of significance to be low and its power to be high. In reality, we always face a tradeoff between these two goals.

– To illustrate this tradeoff, consider the data from which we constructed the earlier CI about Trump and Biden. Let's re-pose this question in terms of a hypothesis test, where

$$
\begin{aligned}
H_0 &: p_1 - p_2 = 0 \\
H_A &: p_1 - p_2 \neq 0
\end{aligned}
$$

– Here our *test statistic* is the difference between our two sample proportions, $\widehat{p}_1 - \widehat{p}_2$. And our rejection region includes the values of the statistic for which we reject the null for our chosen $\alpha$.

  * Here, the rejection region are those values of $\widehat{p}_1 - \widehat{p}_2$ for which the constructed CI does not include zero. This would lead us to say (with $100 * (1-\alpha)\%$ confidence) that the change between the two polls was greater than zero.
  * What is this region? Let's look at our CI again:

$$
(\widehat{p}_1 - \widehat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}.
$$

  * Had $(\widehat{p}_1 - \widehat{p}_2)$ been big enough that $(\widehat{p}_1 - \widehat{p}_2) - z_{\alpha/2}\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}} > 0$, then our CI would not have incorporated zero. That is, if

$$
\begin{aligned}
(\widehat{p}_1 - \widehat{p}_2) &> z_{\alpha/2}\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}} \\
\frac{(\widehat{p}_1 - \widehat{p}_2)}{\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}} &> z_{\alpha/2},
\end{aligned}
$$

  we should reject the null and accept $H_A$.

  * So, a few questions:
  * what sample sizes would we have needed for our difference in sample proportions $(\widehat{p}_1 - \widehat{p}_2)$ to have been found statistically different from zero with 95% confidence?

(Assume $n_1$=$n_2 = n$.)

$$
\begin{aligned}
\frac{3.1}{\sqrt{\frac{(55.6)(100-55.6)}{n}+\frac{(52.5)(100-52.5)}{n}}} &> 1.96 \\
\frac{3.1}{\sqrt{\frac{4962.4}{n}}} &> 1.96 \\
\frac{3.1}{1.96} &> \sqrt{\frac{4962.4}{n}} \\
\left(\frac{3.1}{1.96}\right)^2 &> \frac{4962.4}{n} \\
n &> 1,983.7
\end{aligned}
$$

* We would have needed two samples of at least 1,984 in size.

## 10.3 Hypothesis tests vis-a-vis confidence intervals

- Now let's pose the following question:
  - The conventional wisdom is that our parameter $\theta$ is equal to a certain value. Specifically the null hypothesis says $\theta = \theta_0$.
  - I have theoretical reasons to believe that $\theta \neq \theta_0$. This is my alternative hypothesis.
  - I've obtained a point estimate $\widehat{\theta}$ that is not equal to $\theta_0$.
  - This, of course, does not in itself destroy the null hypothesis. Why? Because it's possible that the null is true and I happened to get this very different estimate from the null simply by chance.
  - This leads to a key question:
- Having obtained a point estimate $\widehat{\theta} \neq \theta_0$, how sure am I now that $\theta$ is not equal to $\theta_0$?
  - This is the question we ask when we conduct *hypothesis tests*.
  - Often this value is $\theta_0 = 0$, but in practice it can be any value.
  - Regardless of the value of $\theta_0$, we can fully define the distribution of our estimator, $\widehat{\theta}$ if $\theta_0$ is true.

- We know that CLT tells us that the standardized version of *any* estimator $\widehat{\theta}$ that is a linear combination of i.i.d. random variables is distributed Normal in large samples: $\frac{\widehat{\theta}-\mu_{\widehat{\theta}}}{\sigma_{\widehat{\theta}}} \tilde{} N(0,1)$.
- of course $\mu_{\widehat{\theta}} = E\left(\widehat{\theta}\right)$ and if $\widehat{\theta}$ unbiased and if $\theta_0$ is true then by definition $E\left(\widehat{\theta}\right) = \theta_0$.
- So now rewrite as $\frac{\widehat{\theta}-\theta_0}{\sigma_{\widehat{\theta}}} \tilde{} N(0,1)$
- And this is the distribution of the standardized version of our estimator $\widehat{\theta}$ "under the null." It is centered around 0, but due to *chance variation* in the sampling process individual estimates are closer or further away from 0 in a pattern described by the standard Normal density.
- [Draw another density curve.] Sometimes it's easier to envision this as the unstandardized version of $\widehat{\theta}$. It is centered about $\theta_0$, but due to *chance variation* in the sampling process individual estimates are closer or further away from $\theta_0$ in a pattern described by the Normal density with variance $\sigma^2_{\widehat{\theta}}$.

### 10.4 A Two-Tailed Hypothesis Test

- With these tools in place, we can conduct hypothesis tests.
- Recall that such tests consist of:

1. A **null hypothesis about a parameter**, which we write as $H_0$.
   - This is typically either what the "conventional wisdom" says is the the value of the parameter–or that the parameter is equal to zero.
2. An **alternative hypothesis about the parameter**, $H_A$.
   - This is typically that the parameter is equal to *something different* than the null hypothesis. It may be more specific: that the parameter is either greater than or less than the null hypothesis.
3. A **test statistic derived from an estimator of the parameter.**
4. A **rejection region**.
   - The RR specifies the range of values of the test statistic for which the null $H_0$ is to be *rejected* in favor of the alternative $H_A$.

- We begin by picking a level of confidence, $\alpha$. Recall that this is the probability of a Type I error, that is $\Pr(\text{reject } H_0 | H_0 \text{ true})$. Typically in our discipline this number is .05, or 5 percent.
- We then look at the standard Normal density curve, and identify the range of extreme values of $\widehat{\theta}$ that we will observe $\alpha$ percent of the time in repeated sampling.
- These extreme values are those greater than $z_{\frac{\alpha}{2}}$ and those less than $-z_{\frac{\alpha}{2}}$, where $\Phi\left(-z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$.
- Values in these ranges are the **rejection region** for our test. If $\widehat{\theta}$ falls in the rejection region, we reject $H_0 : \theta = \theta_0$ in favor of $H_A : \theta \neq \theta_0$.
- If $\widehat{\theta}$ does not fall in the rejection region, we fail to reject $H_0 : \theta = \theta_0$.
- In practice, we are usually working with the unstandardized version of our estimator, and so:
  - We reject $H_0$ if $\widehat{\theta} < \theta_0 - z_{\frac{\alpha}{2}}\sigma_{\widehat{\theta}}$ or if $\widehat{\theta} > \theta_0 + z_{\frac{\alpha}{2}}\sigma_{\widehat{\theta}}$.
  - Otherwise, we fail to reject $H_0$ .

## 10.5 A One-Tailed Hypothesis Test

- Now consider a case where we have a stronger alternative hypothesis.
- Specifically, our hypothesis is **signed**. Rather than $H_A : \theta \neq \theta_0$, I have theoretical reason to claim say, $H_A : \theta > \theta_0$
- How does this change things?
- We again pick a level of confidence, $\alpha$, again typically 5 percent.
- We then look at the standard Normal density curve, and identify the range of values of $\widehat{\theta}$ greater than $\theta_0$ that we will observe $\alpha$ percent of the time in repeated sampling.
- Draw ONE-TAILED HYPOTHESIS TEST diagram on board

## 10.6 Cooking the books

- So here's the thing. Let's say you have a test statistic (some realization of $\widehat{\theta}$) whose value is greater than $\theta_0$. You make a *ex post* ("based on actual results") hypothesis that $H_A : p_1 - p_2 \geq 0$ and conduct a one-tailed hypothesis test. This hypothesis is not based on theory. Are you cooking the books?
  - Yes. Knowing that $\widehat{\theta} > \theta_0$, to reject $H_0$ with a one-tailed test, you need

$$\widehat{\theta} > z_\alpha.$$

  - But to reject $H_0$ with a two-tailed test, you need

$$\widehat{\theta} > z_{\frac{\alpha}{2}}.$$

  - Typically political scientists are skeptical of one-tailed tests because they can look awfully post hoc. Most of the hypothesis tests you'll see in journals are two-tailed.