Introduction to Inference

Hypothesis Tests

Last Class: Confidence Intervals

- We used the sample mean $\bar{X}$ as our best estimate of the population mean $\mu$, but we realized that our sample mean will vary between different samples
- Our solution was to use our sample mean as the center of an entire confidence interval of likely values for our population mean $\mu$
- 95% confidence intervals are most common, but we can calculate interval for any confidence level
- Also did confidence interval for population proportion $p$
- Formulas for confidence intervals are based on results about sampling distribution of sample mean and sample proportion (chapter 5)

This Class: Hypothesis Testing

- Today, we will again use our sampling distribution results for a different type of inference: testing a specific hypothesis
- In some problems, we are not interested in calculating a confidence interval, but rather we want to see whether our data confirm a specific hypothesis
- This type of inference is sometimes called statistical decision making, but the more common term is hypothesis testing

Example: Blackout Baby Boom

- New York City experienced a major blackout on November 9, 1965
- Many people were trapped for hours in the dark and on subways, in elevators, etc.
- Nine months afterwards (August 10, 1966), the NY Times claimed that the number of births were way up
- They attributed the increased births to the blackout, and this has since become urban legend!
- Does the data actually support the claim of the NY Times?
- Using data, we will test the hypothesis that the birth rate in August 1966 was different than the usual birth rate

Number of Births in NYC, August 1966

<table>
<thead>
<tr>
<th>Sun</th>
<th>Mon</th>
<th>Tue</th>
<th>Wed</th>
<th>Thu</th>
<th>Fri</th>
<th>Sat</th>
</tr>
</thead>
<tbody>
<tr>
<td>452</td>
<td>470</td>
<td>431</td>
<td>448</td>
<td>467</td>
<td>377</td>
<td></td>
</tr>
<tr>
<td>344</td>
<td>449</td>
<td>440</td>
<td>457</td>
<td>483</td>
<td>405</td>
<td></td>
</tr>
<tr>
<td>377</td>
<td>453</td>
<td>499</td>
<td>461</td>
<td>442</td>
<td>444</td>
<td>419</td>
</tr>
<tr>
<td>356</td>
<td>470</td>
<td>519</td>
<td>443</td>
<td>449</td>
<td>418</td>
<td>394</td>
</tr>
<tr>
<td>399</td>
<td>451</td>
<td>468</td>
<td>432</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$\bar{X} = 433.6$  
$s = 39.4$  
n = 14

- We want to test this data against the usual birth rate in NYC, which is 430 births/day
Steps for Hypothesis Testing

1. Formulate your hypotheses:
   • Need a Null Hypothesis and an Alternative Hypothesis
2. Calculate the test statistic:
   • Test statistic summarizes the difference between data and your null hypothesis
3. Find the p-value for the test statistic:
   • How probable is your data if the null hypothesis is true?

Null and Alternative Hypotheses

• Null Hypothesis ($H_0$) is (usually) an assumption that there is no effect or no change in the population
• Alternative hypothesis ($H_a$) states that there is a real difference or real change in the population
  • If the null hypothesis is true, there should be little discrepancy between the observed data and the null hypothesis
    • If we find there is a large discrepancy, then we will reject the null hypothesis
  • Both hypotheses are expressed in terms of different values for population parameters

Example: NYC blackout and birth rates

• Let $\mu$ be the mean birth rate in August 1966
• Null Hypothesis:
  • Blackout has no effect on birth rate, so August 1966 should be the same as any other month
  • $H_0: \mu = 430$ (usual birth rate)
• Alternative Hypothesis:
  • Blackout did have an effect on the birth rate
  • $H_a: \mu \neq 430$
  • This is a two-sided alternative, which means that we are considering a change in either direction
  • We could instead use a one-sided alternative that only considers changes in one direction
    • E.g. only alternative is an increase in birth rate $H_a: \mu > 430$

Test Statistic

• Now that we have a null hypothesis, we can calculate a test statistic
• The test statistic measures the difference between the observed data and the null hypothesis
• Specifically, the test statistic answers the question: “How many standard deviations is our observed sample value from the hypothesized value?”
• For our birth rate dataset, the observed sample mean is 433.6 and our hypothesized mean is 430
• To calculate the test statistic, we need the standard deviation of our sample mean

Test Statistic for Sample Mean

• Sample mean has a standard deviation of $\sigma/\sqrt{n}$ so our test statistic $Z$ is:
  $$ Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} $$
  • $Z$ is the number of standard deviations between our sample mean and the hypothesized mean
  • $\mu_0$, is the notation we use for our hypothesized mean
  • To calculate our test statistic $Z$, we need to know the population standard deviation $\sigma$
  • For now we will make the assumption that $\sigma$ is the same as our sample standard deviation $s$
  • Later, we will correct this assumption!

Test Statistic for Birth Rate Example

• For our NYC births/day example, we have a sample mean of 433.6, a hypothesized mean of 430 and a sample standard deviation of 39.4
• Our test statistic is:
  $$ Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{433.6 - 430}{39.4/\sqrt{14}} = 0.342 $$
  • So, our sample mean is 0.342 standard deviations different from what it should be if there was no blackout effect
  • Is this difference statistically significant?
Probability values (p-values)
- Assuming the null hypothesis is true, the p-value is the probability we get a value as far from the hypothesized value as our observed sample value.
- The smaller the p-value is, the more unrealistic our null hypothesis appears.
- For our NYC birth-rate example, $Z = 0.342$.
  - Assuming our population mean really is 430, what is the probability that we get a test statistic of 0.342 or greater?

p-value for NYC dataset
- To calculate the p-value, we use the fact that the sample mean has a normal distribution.
- If our alternative hypothesis was one-sided ($H_a: \mu > 430$), then our p-value would be 0.367.
- Since our alternative hypothesis was two-sided, our p-value is the sum of both tail probabilities.
  - p-value = 0.367 + 0.367 = 0.734.

Statistical Significance
- If the p-value is smaller than $\alpha$, we say the data are statistically significant at level $\alpha$.
- The most common $\alpha$-level to use is $\alpha = 0.05$.
- Later, we will see this relates to 95% confidence intervals.
- The $\alpha$-level is used as a threshold for rejecting the null hypothesis.
- If the p-value < $\alpha$, we reject the null hypothesis that there is no change or difference.

Conclusions for NYC birth-rate data
- The p-value = 0.734 for the NYC birth-rate data, so we cannot reject the null hypothesis at $\alpha$-level of 0.05.
- Another way of saying this is that the difference between null hypothesis and our data is not statistically significant.
- So, we conclude that the data do not support the idea that there was a different birth rate than usual for the first two weeks of August, 1966. No blackout baby boom effect!

Tests and Intervals
- There is a close connection between confidence intervals and two-sided hypothesis tests.
- 100 C% confidence interval is contains likely values for a population parameter, like the pop. mean $\mu$.
  - Interval is centered around sample mean $\bar{X}$.
  - Width of interval is a multiple of $\sigma \sqrt{n}$.
- A $\alpha$-level hypothesis test rejects the null hypothesis that $\mu = \mu_0$ if the test statistic $Z$ has a p-value less than $\alpha$.
  $$ Z = \frac{\bar{X} - \mu_0}{\sigma \sqrt{n}} $$

Tests and Intervals
- If our confidence level C is equal to 1 - $\alpha$ where $\alpha$ is the level of the hypothesis test, then we have the following connection between tests and intervals:
  - A two-sided hypothesis test rejects the null hypothesis ($\mu = \mu_0$) if our hypothesized value $\mu_0$ falls outside the confidence interval for $\mu$.
  - So, if we have already calculated a confidence interval for $\mu$, then we can test any hypothesized value $\mu_0$ just by seeing whether or not $\mu_0$ is in the interval!
Example: NYC blackout baby boom

- Births per day from two weeks in August 1966
  \[ \bar{X} = 433.6 \quad s = 39.4 \quad n = 14 \]
- Difference between our sample mean and the population mean \( \mu_0 = 430 \) had a p-value of 0.734, so we did not reject the null hypothesis at a level of 0.05
- Could have calculated 100 \( (1-\alpha) \% \) = 95% confidence interval:
  \[ \left[ \bar{X} - Z \cdot \frac{\sigma}{\sqrt{n}} , \bar{X} + Z \cdot \frac{\sigma}{\sqrt{n}} \right] = \left[ 433.6 - 1.96 \cdot \frac{39.4}{\sqrt{14}} , 433.6 + 1.96 \cdot \frac{39.4}{\sqrt{14}} \right] = (413.0 , 454.2) \]
- Since our hypothesized \( \mu_0 = 430 \) is within our interval of likely values, we do not reject the null hypothesis.
- If hypothesis was \( \mu_0 = 410 \), then we would reject it!


Hypothesis Test for Calcium

- Let \( \mu \) be the mean calcium intake for people below the poverty line
- Null hypothesis is that calcium intake for people below poverty line is not different from RDA: \( \mu_0 = 850 \) mg/day
- Two-sided alternative hypothesis: \( \mu \neq 850 \) mg/day
- To calculate test statistic, we need to know the population standard deviation of daily calcium intake.
  - From previous study, we know \( \sigma = 188 \) mg
  \[ Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{747 - 850}{188/\sqrt{18}} = -2.32 \]
- Need p-value: if \( \mu_0 = 850 \), what is the probability we get a sample mean as extreme (or more) than 747?


Confidence Interval for Calcium

- Alternatively, we calculate a confidence interval for the calcium intake of people below poverty line
- Use confidence level 100 C = 100 \( (1-\alpha) \% \) = 95%
- 95% confidence level means critical value \( Z^* = 1.96 \)
  \[ \left( \bar{X} - Z^* \cdot \frac{\sigma}{\sqrt{n}} , \bar{X} + Z^* \cdot \frac{\sigma}{\sqrt{n}} \right) = \left( 747 - 1.96 \cdot \frac{188}{\sqrt{18}} , 747 + 1.96 \cdot \frac{188}{\sqrt{18}} \right) = (660.1 , 833.9) \]
- Since our hypothesized value \( \mu_0 = 850 \) mg is not in the 95% confidence interval, we can reject that hypothesis right away!


Another Example: Calcium in the Diet

- Calcium is a crucial element in body. Recommended daily allowance (RDA) for adults is 850 mg/day
- Random sample of 18 people below poverty level:
  \[ \bar{X} = 747.4 \text{ mg} \quad n = 18 \]
- Does the data support claim that people below the poverty level have a different calcium intake than the recommended daily allowance?


p-value for Calcium

- We have two-sided alternative, so p-value includes standard normal probabilities on both sides:
  \[ \text{prob} = 0.010 \]
- Looking up probability in table, we see that the two-sided p-value is 0.010
- Since the p-value is less than 0.05, we can reject the null hypothesis
  - Conclusion: people below the poverty line have significantly (at a \( \alpha = 0.05 \) level) lower calcium intake than the RDA


Cautions about Hypothesis Tests

- Statistical significance does not necessarily mean real significance
  - If sample size is large, even very small differences can have a low p-value
- Lack of significance does not necessarily mean that the null hypothesis is true
  - If sample size is small, there could be a real difference, but we are not able to detect it
- Many assumptions went into our hypothesis tests
  - Presence of outliers, low sample sizes, etc. make our assumptions less realistic
  - We will try to address some of these problems next class

Next Class - Lecture 15

• Practice Problems in Chapter 6

• Enjoy Spring Break!