Math 1710 Class 26

Hypothesis Testing
Dr. Back

Oct. 28, 2009
Critical Value $z^*$. ($P(Z < z^*) = .975$)
CI for $p$

Critical Value $z^*$ for a level C CI

$z^*$

$C$

$(1-C)/2$

$-z^*$

$z^*$

Three ways to Carry Out an HT
CI for $p$

\[ \hat{p} \pm z^* SE(\hat{p}) \]

where \( SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}} \).
CI for $p$

$$\hat{p} \pm z^* SE(\hat{p})$$

where $SE(\hat{p}) = \sqrt{\frac{\hat{p} \hat{q}}{n}}$.

Poll size $n = 400$, 144 say yes, $\hat{p} = .36$
CI for $p$

$$\hat{p} \pm z^* SE(\hat{p})$$

where $SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}$.

Poll size $n = 400$, 144 say yes, $\hat{p} = .36$

$$SE(\hat{p}) = \sqrt{\frac{.36 \cdot .64}{400}} = .024.$$
CI for $\hat{p}$

$$\hat{p} \pm z^* SE(\hat{p})$$

where $SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}$

Poll size $n = 400$, 144 say yes, $\hat{p} = .36$

$$SE(\hat{p}) = \sqrt{\frac{.36 \cdot .64}{400}} = .024.$$  

Sampling distribution of $\hat{p}$ is $N(p, .024)$ approximately.
CI for $p$

\[ \hat{p} \pm z^* SE(\hat{p}) \]

where \( SE(\hat{p}) = \sqrt{\frac{\hat{p} \hat{q}}{n}} \).

Poll size \( n = 400 \), 144 say yes, \( \hat{p} = .36 \)

This shows \( p = .5 \) should not be in the CI:
Cl for $p$

\[ \hat{p} \pm z^* SE(\hat{p}) \]

where $SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}$.

Poll size $n = 400$, 144 say yes, $\hat{p} = .36$

This shows $p = .35$ should be in the CI:
CI for \( p \)

\[
\hat{p} \pm z^* SE(\hat{p})
\]

where \( SE(\hat{p}) = \sqrt{\frac{\hat{p} \hat{q}}{n}} \).

Poll size \( n = 400 \), 144 say yes, \( \hat{p} = .36 \)

Picture when \( p \) is at the upper limit of the CI:
CI for $p$

$$\hat{p} \pm z^{*} SE(\hat{p})$$

where $SE(\hat{p}) = \sqrt{\frac{\hat{p}q}{n}}$.

Poll size $n = 400$, 144 say yes, $\hat{p} = .36$

Picture when $p$ is at the upper limit of the CI:

The upper limit is $\hat{p} + z^{*} SE(\hat{p}) = .36 + 1.96 \cdot .024 = .407$. 
CI for $p$

$\hat{p} \pm z^* SE(\hat{p})$

where

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}.$$  

Poll size $n = 400$, 144 say yes, $\hat{p} = .36$

Picture when $p$ is at the lower limit of the CI:
CI for $p$

$$\hat{p} \pm z^* SE(\hat{p})$$

where $SE(\hat{p}) = \sqrt{\hat{p}\hat{q}/n}$.

Poll size $n = 400$, 144 say yes, $\hat{p} = .36$

Picture when $p$ is at the lower limit of the CI:

The lower limit is $\hat{p} - z^* SE(\hat{p}) = .36 - 1.96 \cdot .024 = .313$. 
Suppose we use a sample of size $n$ to determine a 95% CI $(a,b)$ for a parameter $p$. 
If repeated with a large number of different samples of size n, (maybe different p’s in each case as well) the defining property of CI’s is that approximately 95% of CI’s using samples of size n will contain the true parameter(s) p.
But the convention in elementary statistics is to view the statement “There is a 95% chance the CI contains $p$” to be false. That’s because by the time the CI is computed, $p$ already has a value and the statement $p$ is in the CI is simply either true or false. No probability between 0 and 1.
But it is correct to say that 95% of the time, a 95% CI contains the true parameter.
Colloquially: A 95% CI uses a method which works 95% of the time.
CI Interpretation

works $\iff$ covers the true parameter
Conditions for Prop Tests

- plausible independence
- random sampling
- 10% condition
- success/failure

What Happens if Not Satisfied:
Conditions for Prop Tests

- plausible independence
- random sampling
- 10% condition
- success/failure

What Happens if Not Satisfied:

random sampling -
could be critical;
might be ok if ”representative”
representative hard/impossible to define
Conditions for Prop Tests

- plausible independence
- random sampling
- 10% condition
- success/failure

What Happens if Not Satisfied:

plausible independence -
could be critical
sometimes just a working hypothesis
Conditions for Prop Tests

- plausible independence
- random sampling
- 10% condition
- success/failure

What Happens if Not Satisfied:

**10% condition** -
results in overestimation of samp. dist. st dev
gradual breakdown in formulas, not method
Conditions for Prop Tests

- plausible independence
- random sampling
- 10% condition
- success/failure

What Happens if Not Satisfied:

success/failure -
progressive reduction of accuracy
accuracy varies regardless for smaller values of n
Width of CI’s

The **margin of error** in a CI is half its width:
Width of CI’s

\[ CI(\text{red}) : \hat{p} \pm z^* SE(\hat{p}) \]

\[ MOE(\text{pink}) : z^* SE(\hat{p}) \]

where \( SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}} \).
Width of CI’s

Previous 95% CI Newspaper Poll Example

\[ CI(\text{red}) : \hat{p} \pm z^* SE(\hat{p}) \quad .36 \pm .047 \]

\[ MOE(\text{pink}) : z^* SE(\hat{p}) \quad 1.96 \cdot .024 = .047 \]

where \( SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}} = .024. \)
How big should a CI you are 100% certain of be?
How big should a CI you are 100% certain of be?

**Answer:** A 100% CI would be (0, 1).
Margins of error increase as the level of confidence increases.

\[
\text{Margin of Error} = \frac{1-C}{2} \times Z^* 
\]
Width of CI’s

But not linearly.

<table>
<thead>
<tr>
<th>Level of Confidence C</th>
<th>$z^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>80%</td>
<td>1.282</td>
</tr>
<tr>
<td>90%</td>
<td>1.645</td>
</tr>
<tr>
<td>95%</td>
<td>1.96</td>
</tr>
<tr>
<td>98%</td>
<td>2.326</td>
</tr>
<tr>
<td>99%</td>
<td>2.576</td>
</tr>
</tbody>
</table>
Width of CI’s

These numbers could come from table Z

\[ z^* = 1.28 \quad C = 80\% \]

<table>
<thead>
<tr>
<th>z</th>
<th>0.00</th>
<th>0.01</th>
<th>0.02</th>
<th>0.03</th>
<th>0.04</th>
<th>0.05</th>
<th>0.06</th>
<th>0.07</th>
<th>0.08</th>
<th>0.09</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.5000</td>
<td>0.5040</td>
<td>0.5080</td>
<td>0.5120</td>
<td>0.5160</td>
<td>0.5199</td>
<td>0.5239</td>
<td>0.5279</td>
<td>0.5319</td>
<td>0.5359</td>
</tr>
<tr>
<td>0.1</td>
<td>0.5398</td>
<td>0.5438</td>
<td>0.5478</td>
<td>0.5517</td>
<td>0.5557</td>
<td>0.5596</td>
<td>0.5636</td>
<td>0.5675</td>
<td>0.5714</td>
<td>0.5753</td>
</tr>
<tr>
<td>0.2</td>
<td>0.5793</td>
<td>0.5832</td>
<td>0.5871</td>
<td>0.5910</td>
<td>0.5948</td>
<td>0.5987</td>
<td>0.6026</td>
<td>0.6064</td>
<td>0.6103</td>
<td>0.6141</td>
</tr>
<tr>
<td>0.3</td>
<td>0.6179</td>
<td>0.6217</td>
<td>0.6255</td>
<td>0.6293</td>
<td>0.6331</td>
<td>0.6368</td>
<td>0.6406</td>
<td>0.6443</td>
<td>0.6480</td>
<td>0.6517</td>
</tr>
<tr>
<td>0.4</td>
<td>0.6554</td>
<td>0.6591</td>
<td>0.6628</td>
<td>0.6664</td>
<td>0.6700</td>
<td>0.6736</td>
<td>0.6772</td>
<td>0.6808</td>
<td>0.6844</td>
<td>0.6879</td>
</tr>
<tr>
<td>0.5</td>
<td>0.6915</td>
<td>0.6950</td>
<td>0.6985</td>
<td>0.7019</td>
<td>0.7054</td>
<td>0.7088</td>
<td>0.7123</td>
<td>0.7157</td>
<td>0.7190</td>
<td>0.7224</td>
</tr>
<tr>
<td>0.6</td>
<td>0.7257</td>
<td>0.7291</td>
<td>0.7324</td>
<td>0.7357</td>
<td>0.7389</td>
<td>0.7422</td>
<td>0.7454</td>
<td>0.7486</td>
<td>0.7517</td>
<td>0.7549</td>
</tr>
<tr>
<td>0.7</td>
<td>0.7580</td>
<td>0.7611</td>
<td>0.7642</td>
<td>0.7673</td>
<td>0.7704</td>
<td>0.7734</td>
<td>0.7764</td>
<td>0.7794</td>
<td>0.7823</td>
<td>0.7852</td>
</tr>
<tr>
<td>0.8</td>
<td>0.7881</td>
<td>0.7910</td>
<td>0.7939</td>
<td>0.7967</td>
<td>0.7995</td>
<td>0.8023</td>
<td>0.8051</td>
<td>0.8078</td>
<td>0.8106</td>
<td>0.8133</td>
</tr>
<tr>
<td>0.9</td>
<td>0.8159</td>
<td>0.8186</td>
<td>0.8212</td>
<td>0.8238</td>
<td>0.8264</td>
<td>0.8289</td>
<td>0.8315</td>
<td>0.8340</td>
<td>0.8365</td>
<td>0.8389</td>
</tr>
<tr>
<td>1.0</td>
<td>0.8413</td>
<td>0.8438</td>
<td>0.8461</td>
<td>0.8485</td>
<td>0.8508</td>
<td>0.8531</td>
<td>0.8554</td>
<td>0.8577</td>
<td>0.8599</td>
<td>0.8621</td>
</tr>
<tr>
<td>1.1</td>
<td>0.8643</td>
<td>0.8665</td>
<td>0.8686</td>
<td>0.8708</td>
<td>0.8729</td>
<td>0.8749</td>
<td>0.8770</td>
<td>0.8790</td>
<td>0.8810</td>
<td>0.8830</td>
</tr>
<tr>
<td>1.2</td>
<td>0.8849</td>
<td>0.8869</td>
<td>0.8888</td>
<td>0.8907</td>
<td>0.8925</td>
<td>0.8944</td>
<td>0.8962</td>
<td>0.8980</td>
<td>0.8997</td>
<td>0.9015</td>
</tr>
<tr>
<td>1.3</td>
<td>0.9032</td>
<td>0.9049</td>
<td>0.9066</td>
<td>0.9082</td>
<td>0.9099</td>
<td>0.9115</td>
<td>0.9131</td>
<td>0.9147</td>
<td>0.9162</td>
<td>0.9177</td>
</tr>
<tr>
<td>1.4</td>
<td>0.9192</td>
<td>0.9207</td>
<td>0.9222</td>
<td>0.9236</td>
<td>0.9251</td>
<td>0.9265</td>
<td>0.9279</td>
<td>0.9292</td>
<td>0.9306</td>
<td>0.9319</td>
</tr>
</tbody>
</table>
These numbers could come from table Z

e.g. \( z^* = 1.28 \) \( C = 80\% \)
Width of CI’s

<table>
<thead>
<tr>
<th></th>
<th>80%</th>
<th>90%</th>
<th>95%</th>
<th>98%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>1.285</td>
<td>1.651</td>
<td>1.969</td>
<td>2.341</td>
<td>2.596</td>
</tr>
<tr>
<td>400</td>
<td>1.284</td>
<td>1.649</td>
<td>1.966</td>
<td>2.336</td>
<td>2.588</td>
</tr>
<tr>
<td>1000</td>
<td>1.282</td>
<td>1.646</td>
<td>1.962</td>
<td>2.330</td>
<td>2.581</td>
</tr>
<tr>
<td>∞</td>
<td>1.282</td>
<td>1.645</td>
<td>1.960</td>
<td>2.326</td>
<td>2.576</td>
</tr>
</tbody>
</table>

These numbers actually come from the bottom of table T.
Smoke Detectors

**CPSC ’96:** 90% of American homes have at least one smoke detector. After a public safety campaign, a city observes that 376 out of 400 randomly selected homes have a detector.
Smoke Detectors

CPSC ’96: 90% of American homes have at least one smoke detector. After a public safety campaign, a city observes that 376 out of 400 randomly selected homes have a detector. Is this strong evidence the local rate is greater than the national rate?
**Smoke Detectors**

**CPSC ’96:** 90% of American homes have at least one smoke detector. After a public safety campaign, a city observes that 376 out of 400 randomly selected homes have a detector. Is this strong evidence the local rate is greater than the national rate?

So \( \hat{p} = \frac{376}{400} = .94. \)
Null Hypothesis $H_0$ - retained unless disproven.
Logic of Hypothesis Testing

Null Hypothesis $H_0$ - retained unless disproven.
Alternative Hypothesis $H_A$ - Only thing which possibly can be proven in the HT.
Logic of Hypothesis Testing

Null Hypothesis $H_0$ - retained unless disproven.
Alternative Hypothesis $H_A$ - Only thing which possibly can be proven in the HT.

Procedure:

1. . .
2. . .
3. Consider the sampling distribution of $\hat{p}$ which would hold if $H_0$ were true.
4. Retain $H_0$ if $\hat{p}$ is reasonably consistent with this sampling distribution. Otherwise reject $H_0$.
5. . .
Logic of Hypothesis Testing

Procedure:

1 ... 

2 ... 

3 Consider the sampling distribution of $\hat{p}$ which would hold if $H_0$ were true. 

4 Retain $H_0$ if $\hat{p}$ is reasonably consistent with this sampling distribution. Otherwise reject $H_0$. 

5 ... 

Three ways to carry out step 4:

1 Calculate a $Z$-statistic and determine a p-value. 

2 Calculate a $Z$-statistic and compare with a critical value $z^*$. 

3 Use a confidence interval.
Hypothesis Testing Vocabulary

- Null Hypothesis $H_0$
- Alternative Hypothesis $H_1$ (One Sided vs Two Sided)
- $SD(\hat{p})$ vs. $SE(\hat{p})$
- Z-statistic
- Tail Probability
- P-value
- Significance Level
- Retain Null H vs. Accept Alt H (equiv. reject Null H)
Let $p$ be the true proportion of smoke detectors homes in our city have.
CPSC Example

Let $p$ be the true proportion of smoke detectors homes in our city have.

$H_0 : p = 0.9, \quad H_A : p > 0.9$
Let $p$ be the true proportion of smoke detectors homes in our city have.

$H_0 : p = .9, \quad H_A : p > .9$

$$SD(\hat{p}) = \sqrt{\frac{.9 \cdot .1}{400}} = .015.$$
Let $p$ be the true proportion of smoke detectors homes in our city have.

$H_0 : p = .9, \quad H_A : p > .9$

$$SD(\hat{p}) = \sqrt{\frac{.9 \cdot .1}{400}} = .015.$$ 

z-statistic

$$z = \frac{.94 - .9}{.015} = 2.67.$$
z-statistic

\[ z = \frac{.94 - .9}{.015} = 2.67. \]

z-statistic is the z-score of \( \hat{p} \) on the samp dist centered at the hypothesized value of \( p \)
CPSC Example

z-statistic

\[ z = \frac{.94 - .9}{.015} = 2.67. \]

z-statistic is the z-score of \( \hat{p} \) on the samp dist centered at the hypothesized value \( p_0 \) of \( p \).
CPSC Example

\[ z = \frac{.94 - .9}{.015} = 2.67. \]

Two primary ways to continue:
CPSC Example

z-statistic

\[ z = \frac{.94 - .9}{.015} = 2.67. \]

Two primary ways to continue:
Method 1: Tail Prob = \( P(\hat{p} > .94) = P(Z > 2.67) = .0038 \)
Since the test is 1-sided, P-value=tail probability=.0038.
z-statistic

\[ z = \frac{.94 - .9}{.015} = 2.67. \]

Two primary ways to continue:
Method 1: Tail Prob = \( P(\hat{p} > .94) = P(Z > 2.67) = .0038 \)
Since the test is 1-sided, P-value=tail probability=.0038.
This is small, so conclusion is we reject \( H_0 \).
CPSC Example

z-statistic

\[ z = \frac{.94 - .9}{.015} = 2.67. \]

Two primary ways to continue:
Method 2: Say significance level is \( \alpha = .05 \).
Since our z-statistic of 2.67 is more extreme than \( z^* = 1.645 \) (and supports \( H_a \)), we reject \( H_a \) at \( \alpha = .05 \).
z-statistic

\[
z = \frac{.94 - .9}{.015} = 2.67.
\]

Two primary ways to continue:

**N(0,1) Picture behind**  
\[z^* = 1.645\]