Math 1710 Class 25

Sampling, Confidence Intervals

Dr. Back

Oct. 26, 2009

Choose an SRS of size 5 from 30 people

Step 1: Assign ID numbers 01,02,. . .,29,30.
all 2-digit!

Choose an SRS of size 5 from 30 people

Step 1: Assign ID numbers 01,02,. . .,29,30.
all 2-digit!

Step 2: Pick a random place to start in the table.
(e.g. row 12, col 3) You might use a wristwatch.

Choose an SRS of size 5 from 30 people

Step 1: Assign ID numbers 01,02,. . .,29,30.
all 2-digit!

Step 2: Pick a random place to start in the table. (e.g. row 12, col 3) You might use a wristwatch.

69167 38894 00172 02999 97377 33305

Choose an SRS of size 5 from 30 people

Step 1: Assign ID numbers 01,02,. . .,29,30.
all 2-digit!

Step 2: Pick a random place to start in the table.
(e.g. row 12, col 3) You might use a wristwatch.

69167 38894 00172 02999 97377 33305

Step 3: Read off pairs of ID digits, selecting those that correspond to real units, rejecting others.

Choose an SRS of size 5 from 30 people

Step 1: Assign ID numbers 01,02,. . .,29,30.
all 2-digit!

Step 2: Pick a random place to start in the table.
(e.g. row 12, col 3) You might use a wristwatch.

69167 38894 00172 02999 97377 33305

Step 3: Read off pairs of ID digits, selecting those that correspond to real units, rejecting others.

69 16 73 88 94 00 17 20 29 99 97 37 73 33 05

Choose an SRS of size 5 from 30 people

Step 1: Assign ID numbers 01,02,. . .,29,30.
all 2-digit!

Step 2: Pick a random place to start in the table.
(e.g. row 12, col 3) You might use a wristwatch.

69167 38894 00172 02999 97377 33305

Step 3: Read off pairs of ID digits, selecting those that correspond to real units, rejecting others.

69 16 73 88 94 00 17 20 29 99 97 37 73 33 05

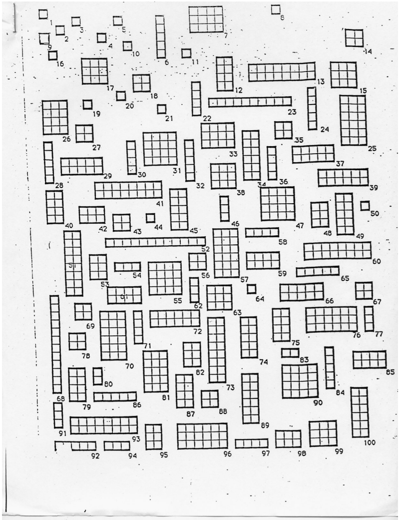
Our SRS is units 16 17 20 29 and 05.

Average Area?

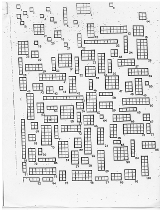
Eyeball Estimate: Take a 5 second look and make a guess:

Average Area?

Eyeball Estimate: Take a 5 second look and make a guess:

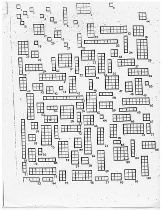


Average Area?



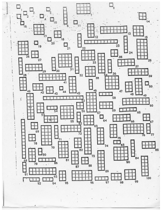
Judgement Sample: Each person picks 5 representative rectangles, and computes their average area.

Average Area?



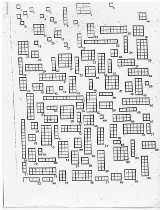
SRS: Each person uses a table of random numbers to choose 5 random rectangles, and computes their average area.

Average Area?



Stratified Random Sample: Call a rectangle big if its area is above 5; small otherwise.

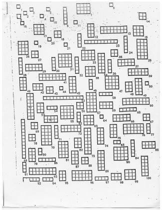
Average Area?



Stratified Random Sample: Call a rectangle big if its area is above 5; small otherwise.

Have 70% of people find averages of big rectangles and 30% study small ones.

Average Area?



Stratified Random Sample: Call a rectangle big if its area is above 5; small otherwise.

Have 70% of people find averages of big rectangles and 30% study small ones.

Our overall estimate will be

$$.48\bar{x}_{small} + .52\bar{x}_{large}$$

Average Area?

Stratified Random Sample: Call a rectangle big if its area is above 5; small otherwise.
Have 70% of people find averages of big rectangles and 30% study small ones.
Our overall estimate will be

$$.48\bar{x}_{small} + .52\bar{x}_{large}$$

The numbers .48 and .52 come from the fact that there are 48 small and 52 big rectangles.

Average Area?

Stratified Random Sample: Call a rectangle big if its area is above 5; small otherwise.

Have 70% of people find averages of big rectangles and 30% study small ones.

The fractions 70% and 30% come from the following calculation:

Type	n_i	σ_i	$n_i\sigma_i$
small rectangles	48	1.557	74.74
big rectangles	52	3.847	200.04

$$\frac{200.04}{200.04 + 74.74} = .73 \sim 70\%$$

$$\frac{74.74}{200.04 + 74.74} = .27 \sim 30\%$$

Average Area?

Stratified Random Sample: Call a rectangle big if its area is above 5; small otherwise.

Have 70% of people find averages of big rectangles and 30% study small ones.

The fractions 70% and 30% come from the following calculation:

Type	n_i	σ_i	$n_i\sigma_i$
small rectangles	48	1.557	74.74
big rectangles	52	3.847	200.04

$$\frac{200.04}{200.04+74.74} = .73 \sim 70\%$$

$$\frac{74.74}{200.04+74.74} = .27 \sim 30\%$$

Average Area?

Stratified Random Sample: Call a rectangle big if its area is above 5; small otherwise.

Have 70% of people find averages of big rectangles and 30% study small ones.

The fractions 70% and 30% come from the following calculation:

Type	n_i	σ_i	$n_i\sigma_i$
small rectangles	48	1.557	74.74
big rectangles	52	3.847	200.04

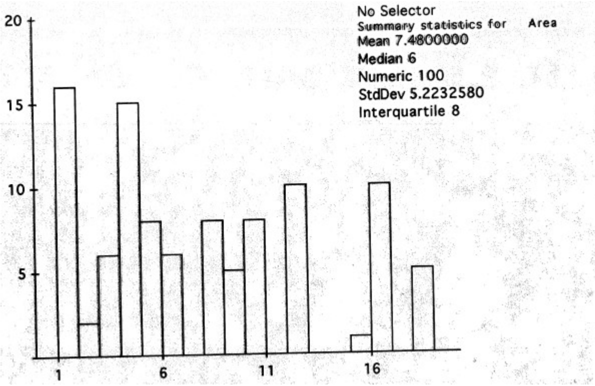
$$\frac{200.04}{200.04 + 74.74} = .73 \sim 70\%$$

$$\frac{74.74}{200.04 + 74.74} = .27 \sim 30\%$$

These proportions give the stratified random sample with smallest variance for a given sample size.

Average Area?

Histogram of Actual Areas



Average Area?

Frequency Breakdown of Actual Areas

Area

Frequency breakdown of Area
No Selector

Group	Count	Cumulative Count	%	Cumulative %
1	16	16	16	16
2	2	18	2	18
3	6	24	6	24
4	15	39	15.0	39
5	8	47	8	47
6	6	53	6	53.0
8	8	61	8	61
9	5	66	5	66
10	8	74	8	74
12	10	84	10	84
15	1	85	1	85
16	10	95	10	95
18	5	100	5	100
total	100			

Newspaper Poll

Poll size $n = 400$.
144 say yes, so

$$\hat{p} = \frac{144}{400} = .36.$$

Newspaper Poll

Poll size $n = 400$.
144 say yes, so

$$\hat{p} = \frac{144}{400} = .36.$$

What does this say about the true proportion p?

Newspaper Poll

Poll size $n = 400$.
144 say yes, so

$$\hat{p} = \frac{144}{400} = .36.$$

What does this say about the true proportion p?
The sampling distribution of $\hat{p}$ is $N(p, SD(\hat{p}))$.

Newspaper Poll

$$\hat{p} = \frac{144}{400} = .36.$$

What does this say about the true proportion p?

The sampling distribution of $\hat{p}$ is $N(p, SD(\hat{p}))$.

Recall $SD(\hat{p}) = \sqrt{\frac{pq}{n}}$, $SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}$

Newspaper Poll

$$\hat{p} = \frac{144}{400} = .36.$$

What does this say about the true proportion p?

The sampling distribution of $\hat{p}$ is $N(p, SD(\hat{p}))$.

p	**SD($\hat{p}$)**
.36	.0240
.35	.0238
.3	.0229
.4	.0245
.45	.0249
.5	.0250
.25	.0217

Newspaper Poll

p	SD($\hat{p}$)
.36	.0240
.35	.0238
.3	.0229
.4	.0245
.45	.0249
.5	.0250
.25	.0217

All these SD($\hat{p}$) are quite close, so let's approximate them all by $\text{SE}(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}} = .024$.

Newspaper Poll

p	**SD($\hat{p}$)**
.36	.0240
.35	.0238
.3	.0229
.4	.0245
.45	.0249
.5	.0250
.25	.0217

All these SD($\hat{p}$) are quite close, so let's approximate them all by $SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}} = .024$.

Our sampling distribution of $\hat{p}$ is $N(p, .024)$ approximately. For which p is $\hat{p} = .36$ reasonably consistent with this?

CI Idea:

Answer will be an interval (a, b) of possible values of p.

CI Idea:

Answer will be an interval (a, b) of possible values of p.
Include all values of the parameter p which are reasonably consistent with the observed $\hat{p}$

CI Idea:

Answer will be an interval (a, b) of possible values of p.
Include all values of the parameter p which are reasonably consistent with the observed $\hat{p}$
Use the sampling distribution of $\hat{p}$ ($N(p, .24)$ in our example) (together with the confidence level) to decide which p are reasonably consistent.

CI Idea:

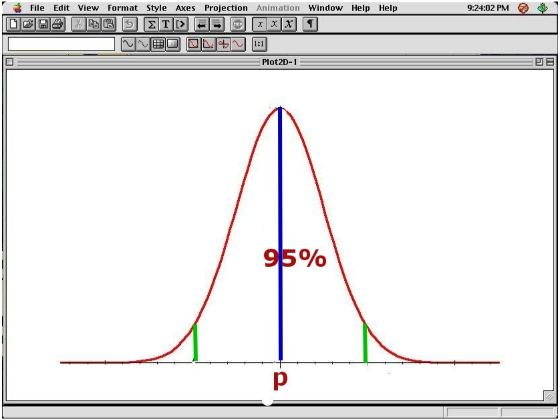
Include all values of the parameter p which are reasonably consistent with the observed $\hat{p}$

Use the sampling distribution of $\hat{p}$ ($N(p, .24)$ in our example) (together with the confidence level) to decide which p are reasonably consistent.

e.g. for a 95% CI, a $\hat{p}$ centered in the middle 95% zone of the sampling distribution of $\hat{p}$ centered at p is reasonably consistent.

CI Idea:

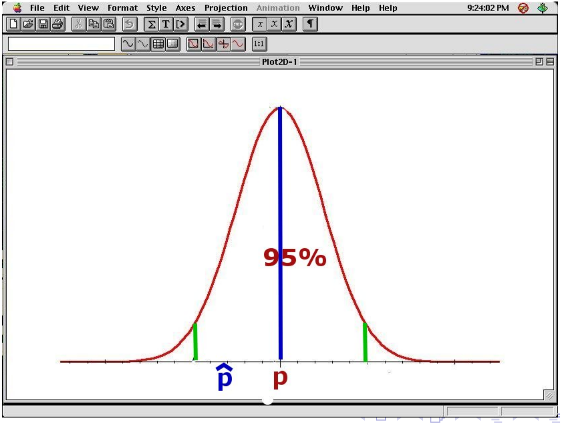
e.g. for a 95% CI, a $\hat{p}$ centered in the middle 95% zone of the sampling distribution of $\hat{p}$ centered at p is reasonably consistent.



CI Idea:

e.g. for a 95% CI, a $\hat{p}$ centered in the middle 95% zone of the sampling distribution of $\hat{p}$ centered at p is reasonably consistent.

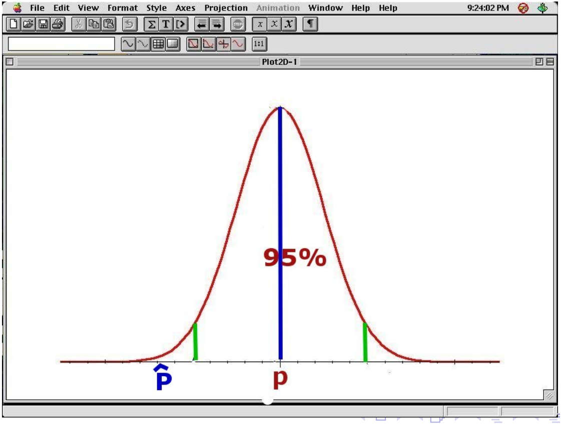
$\hat{p}$ here means p is in the CI



CI Idea:

e.g. for a 95% CI, a $\hat{p}$ centered in the middle 95% zone of the sampling distribution of $\hat{p}$ centered at p is reasonably consistent.

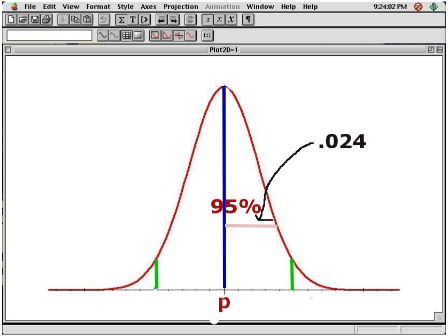
$\hat{p}$ here means p is NOT in the CI



CI Idea:

e.g. for a 95% CI, a $\hat{p}$ centered in the middle 95% zone of the sampling distribution of $\hat{p}$ centered at p is reasonably consistent.

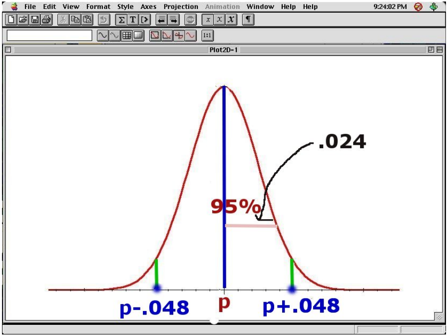
At the level of the 68-95-99.7 rule, middle 95% means within 2 standard deviations.



CI Idea:

e.g. for a 95% CI, a $\hat{p}$ centered in the middle 95% zone of the sampling distribution of $\hat{p}$ centered at p is reasonably consistent.

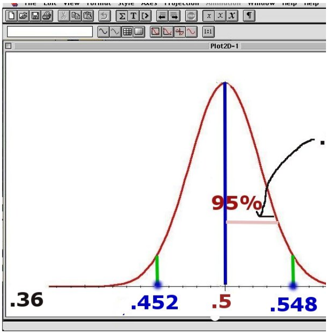
At the level of the 68-95-99.7 rule, middle 95% means within 2 standard deviations.



CI Idea:

e.g. for a 95% CI, a $\hat{p}$ centered in the middle 95% zone of the sampling distribution of $\hat{p}$ centered at p is reasonably consistent.

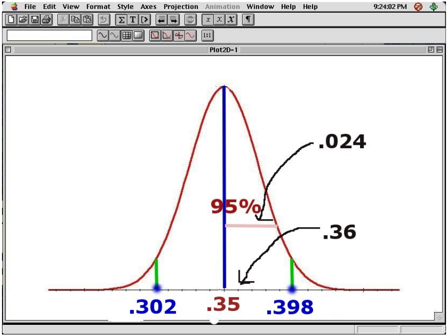
This shows $p = .5$ should not be in the CI:



CI Idea:

e.g. for a 95% CI, a $\hat{p}$ centered in the middle 95% zone of the sampling distribution of $\hat{p}$ centered at p is reasonably consistent.

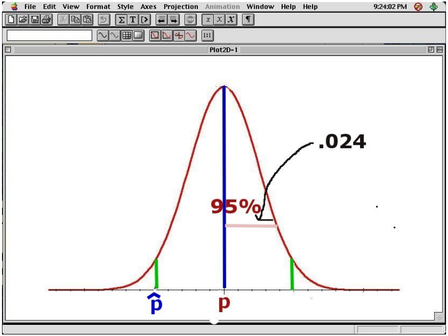
This shows $p = .35$ should be in the CI:



CI Idea:

e.g. for a 95% CI, a $\hat{p}$ centered in the middle 95% zone of the sampling distribution of $\hat{p}$ centered at p is reasonably consistent.

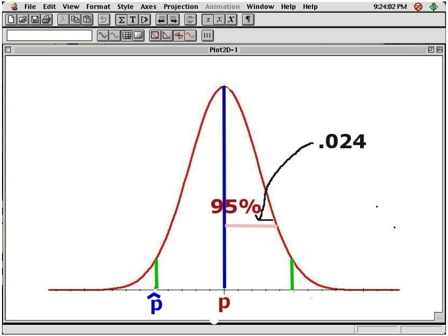
Picture when p is at the upper limit of the CI:



CI Idea:

e.g. for a 95% CI, a $\hat{p}$ centered in the middle 95% zone of the sampling distribution of $\hat{p}$ centered at p is reasonably consistent.

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

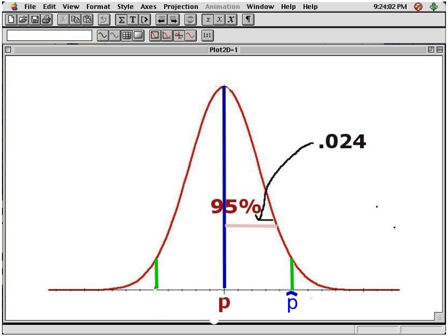


So the upper limit is $\hat{p} + .048 = .408$.

CI Idea:

e.g. for a 95% CI, a $\hat{p}$ centered in the middle 95% zone of the sampling distribution of $\hat{p}$ centered at p is reasonably consistent.

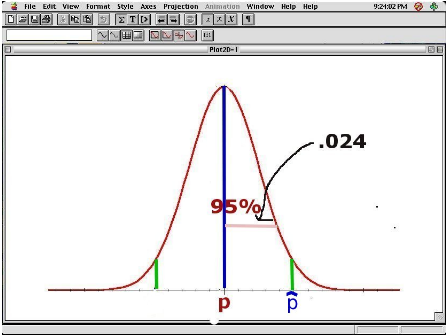
Picture when p is at the lower limit of the CI:



CI Idea:

e.g. for a 95% CI, a $\hat{p}$ centered in the middle 95% zone of the sampling distribution of $\hat{p}$ centered at p is reasonably consistent.

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



So the lower limit is $\hat{p} - .048 = .312$.

CI Idea:

e.g. for a 95% CI, a $\hat{p}$ centered in the middle 95% zone of the sampling distribution of $\hat{p}$ centered at p is reasonably consistent.

A more accurate reading of Table Z tells us $z = 1.96$ is really the cutoff for the middle 95%.

CI Idea:

e.g. for a 95% CI, a $\hat{p}$ centered in the middle 95% zone of the sampling distribution of $\hat{p}$ centered at p is reasonably consistent.

A more accurate reading of Table Z tells us $z = 1.96$ is really the cutoff for the middle 95%.

The value $z^* = 1.96$ is called a critical value.

CI Idea:

e.g. for a 95% CI, a $\hat{p}$ centered in the middle 95% zone of the sampling distribution of $\hat{p}$ centered at p is reasonably consistent.

A more accurate reading of Table Z tells us $z = 1.96$ is really the cutoff for the middle 95%.

The value $z^* = 1.96$ is called a critical value.

The same argument shows that our CI runs from $\hat{p} - z^* SE(\hat{p})$ to $\hat{p} + z^* SE(\hat{p})$

CI Idea:

e.g. for a 95% CI, a $\hat{p}$ centered in the middle 95% zone of the sampling distribution of $\hat{p}$ centered at p is reasonably consistent.

A more accurate reading of Table Z tells us $z = 1.96$ is really the cutoff for the middle 95%.

The value $z^* = 1.96$ is called a critical value.

The same argument shows that our CI runs from $\hat{p} - z^* SE(\hat{p})$ to $\hat{p} + z^* SE(\hat{p})$

Here

$$\hat{p} \pm z^* SE(\hat{p}) = .36 \pm (1.96)(.024) = .36 \pm .047 = (.313, .407)$$

CI Idea:

e.g. for a 95% CI, a $\hat{p}$ centered in the middle 95% zone of the sampling distribution of $\hat{p}$ centered at p is reasonably consistent.

We are 95% confident that the true value of p is between .313 and .407.