STT 200 3pm 2-15-10

Note Title 2/15/2010

RECALL WORK FROM LAST WED.

TASK: COUNT X = # "Y" IN YOUR BLOCK OF n ("Y", "N").

HAD BLOCKS OF $n = 100$, $n = 60$, $n = 40$.

BLOCKS OF $n = 100$ HAVE $p =$ 0.5 0.4 0.6

LIKEWISE $n = 60$ .5 .4 .6

$n = 40$ .5 .4 .6

FLIPPED LAST TIME –

QUESTION: OVER ENTIRE CLASS WE WILL FIND WHAT FRACTION OF 68% CI: $[\hat{p} \pm 1\sqrt{\hat{p}\hat{q}/n}]$ COVER TRUE p.

eg RECITATION 2-16-10 #1.

$\exists$ $\sqrt{}$ Ld $\boxed{n = 60}$, $p = 0.3$; (REVEALED IN 1J.)

FOR THIS DATA, X (COUNT OF "y") = 22

CALC $\hat{p}$ (POINT ESTIMATOR) $= X/n = 22/60$

$z = 1.0$ for 68% CI

YOUR (FAT GUESS) IS 68% CI : $\hat{p} \pm 1 \sqrt{\hat{p}\hat{q}/n}$

$\frac{22}{60} \pm 1 \sqrt{\frac{22}{60} \frac{38}{60} / 60}$

POINT EST OF p

z for 68%

$\underbrace{\sqrt{\frac{22}{60} \frac{38}{60} / 60}}_{\text{OUR EST OF THE SD OF } \hat{p}}$

p = UNKNOWN RATE OF DEFECTIVES IN POP^N

$\sigma_{\hat{p}} = \sqrt{pq/n}$

$\hat{\sigma}_{\hat{p}} = \sqrt{\hat{p}\hat{q}/n}$ PLUG-IN EST OF SD EST OF $\hat{p}$

$\boxed{\hat{p} \pm 1\ \hat{\sigma}_{\hat{p}}}$

FOR THIS DATA THE 68% CI EVALUATES TO

$11/3 = 3.666$

$$.3666 \pm 1\ 0.0822$$

$$.3044 \text{ TO } 0.4288$$

CHECK THIS!

CLAIM IS NOT THAT TRUE p IS FOUND WITHIN THESE LIMITS.

NOT THAT THERE IS A 68% CHANCE IT IS WITHIN THE LIMITS .3044, .4288.

NOTHING RANDOM HERE

CAN SAY THAT AROUND 68% OF SAMPLES OF $n = 60$ PRODUCE A 68% CI COVERING p.

OUR TRUE $p = 0.35$ AND CI [.3044, .4288] DOES COVER $p = 0.35$.

REGARDING THE SIMULATED DATA ASSIGNMENT:

FINISH NOW - BUT NOTE THAT

[100] $p=.5$	[60] $p=.5$	[30] $p=.5$
[100] $p=.4$	[60] $p=.4$	[30] $p=.4$
[100] $p=.6$	[60] $p=.6$	[30] $p=.6$

15 COVER
7 NOT

12 COVER
2 NOT

40 COVER
14 NOT

RATE OF COVERAGE

$15 \quad \frac{40}{54} = .74$

13 COVER
5

WE TRIED 68% CI

WE DID THIS 54 TIMES

FAIRLY GOOD AGREEMENT (68%)

NEWSPAPER MIGHT SAY:

"EST 52.3 % VOTE REPUBLICAN

SAMPLE OF 1000

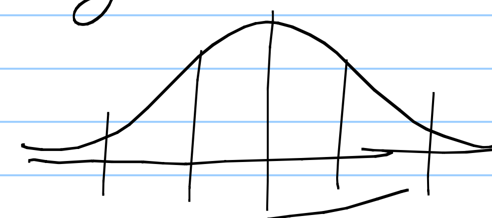
MARGIN OF ERROR $\pm$ .026"

BETTER $z = 1.96$

$\hat{p} = .523$

95% CI $\quad \hat{p} \pm \boxed{2\sqrt{\hat{p}\hat{q}/n}}$

.026



95%

NOT 2 BUT 1.96

[,] $\frac{1}{2}$ WIDTH OF 95% CI IS "MARGIN OF ERROR"

MARGIN OF ERROR IS (THIS CONTEXT) $1.96\sqrt{\hat{p}\hat{q}/n}$

STANDARD ERROR IS (THIS CONTEXT) $\sqrt{\hat{p}\hat{q}/n} = \hat{\sigma}_{\hat{p}}$

I.L. $P\left(p \text{ IS IN } \boxed{\hat{p} \pm 1\sqrt{\hat{p}\hat{q}/n}}\right) \approx .68$

RANDOM 68% CI $\to$ .68 as $n \to$

TROUBLE IS, FOR

$p \sim 0$ or $p \sim 1$ WE REQUIRE n VERY LARGE IN ORDER THAT COVERAGE PROBABILITY $\sim .68$.

THERE IS A QUICK & DIRTY FIX (pg 498) DUE TO AGRESTI – COULL (A-C). 3

(FAKE!)

FIX IS TO $\tilde{p} = \cancel{\frac{X}{n}} \frac{X+2}{n+4}$

"(68% CI)" $\tilde{p} \pm 1\sqrt{\tilde{p}\tilde{q}/(n+4)}$ (A CANDIDATE FOR IMPROVED 68% PERFORMANCE)

BASED ON SIMULATION STUDENTS $P\left(p \text{ IS IN } \boxed{\tilde{p} \pm 1\sqrt{\tilde{p}\tilde{q}/(n+4)}}\right) \sim .68$

AND CLOSER TO .68 IN SOME OVERALL SENSE.

FOR DATA OF #1 WE FOUND $n=60$, $X=22$

SO $\hat{p} = 22/60$ $\tilde{p} = 24/64$

$\tilde{p}$ TILDE BASED 68% CI IS

$$P\left(p \text{ IN } \boxed{\tilde{p} \pm 1\sqrt{\tilde{p}\tilde{q}/(n+4)}}\right) \sim .68$$

AND BETTER AT IT THAN USING $\hat{p}$ HAT (IN OVERALL SENSE)

$$\frac{24}{64} \pm 1\sqrt{\frac{24}{64}\,\frac{40}{64}/64}$$

---

1. m. FPC (FINITE POPULATION CORRECTION)

SIMPLE FACT: SAMPLING WITH REPL (INDEP) $\sigma_{\hat{p}} = \sqrt{\hat{p}\hat{q}/n}$

WITHOUT REPLACEMENT $\sigma_{\hat{p}} = \sqrt{\hat{p}\hat{q}/n}\sqrt{\frac{N-n}{N-1}}$

WHERE N = SIZE OF POPULATION.

IMPLICATION:

WITH-REPL 68% CI $\hat{p} \pm 1\sqrt{\frac{\hat{p}\hat{q}}{n}}$

WITHOUT " 68% CI $\hat{p} \pm 1\sqrt{\frac{\hat{p}\hat{q}}{n}}\sqrt{\frac{N-n}{N-1}}$

FPC

eg $N = 1000$
$n = 80$ (SAY)

$FPC = \sqrt{\frac{1000-80}{1000-1}} \sim .96?$

THIS COMPARISON DEBUNKS THE POPULAR NOTION THAT STATISTICS ACHIEVES ITS RESULTS (EST OF p BY $\hat{p}$) THROUGH EROSION OF THE POPULATION (ie BY SAMPLING A BIG SHARE).