2009 10-7-09 FORMULA REVIEW

Note Title 10/7/2009

POPULATION: POP MEAN μ (mu)

μ_x — AVG AGE OF THOSE PRESENT

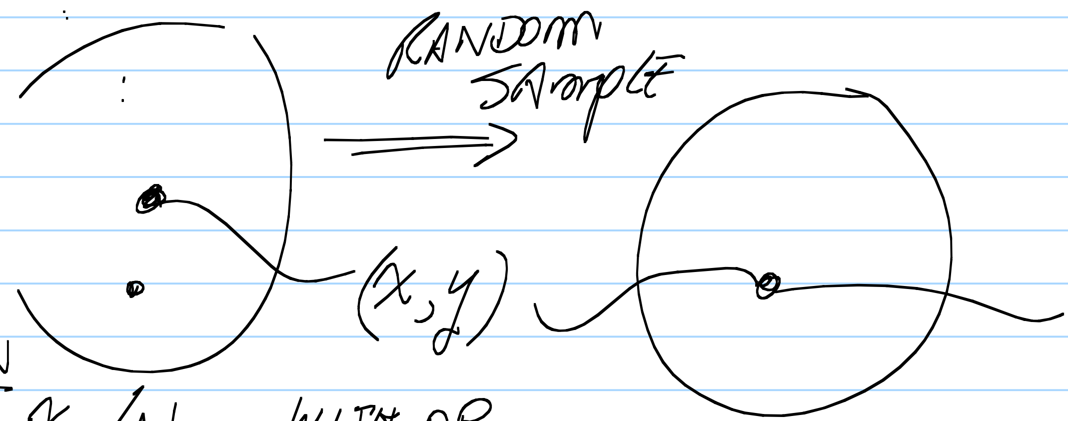
σ_x — SD OF AGE " "

μ_y — AVG CLASS LEVEL

σ_y = SD " "

μ_x

σ_x



$$\mu_x = \sum_{i=1}^{N} x_i / N$$

WITH OR WITHOUT REPLACEMENT

EQUAL PROBABILITY SAMPLING

$$\sigma_x = \sqrt{\sum_{i=1}^{N} (x_i - \mu_x)^2 / N}$$

$$\bar{x} = \sum_{i=1}^{n} x_i / n$$

$$s_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

LOOK AT NICE ALGEBRAIC PROPERTIES OF σ_x CALC$^{\underline{N}}$

$$\sigma_x \underset{\text{pop}^{\underline{n}}}{\overset{\text{ALSO}}{=}} \sqrt{\overline{x^2} - \bar{x}^2}$$

AVE OVER N

RECALL $\underset{\text{JUST}}{=} \sqrt{\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}}$

IF YOU APPLY THIS TO SAMPLE DATA YOU CAN GET s_x OUT OF IT AS

$$s_x = \sqrt{\frac{n}{n-1}}\sqrt{\overline{x^2} - \bar{x}^2}$$

LIKEWISE $\sigma_y = \sqrt{\mu_{y^2} - \mu_y^2}$ $\quad s_y = \sqrt{\dfrac{n}{n-1}}\sqrt{\overline{y^2} - \bar{y}^2}$

SO (EXAMPLE) GIVEN DATA $n = 400$

$\bar{x} = 18.2$ $\quad \overline{x^2} = 500$ $\quad s_x = \sqrt{\dfrac{400}{399}}\sqrt{500 - 18.2^2}$

IF ALSO $\bar{y} = 2.7$ AVG CLASS LEVEL $\overline{y^2} = 11$ (Day)

$\Rightarrow s_y = \sqrt{\frac{400}{399}} \sqrt{11 - 2.7^2}$

CLOSE TO 1

BEWARE PRECISION ISSUES.

IF ALSO $\overline{xy} = 65 \Rightarrow r = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\sigma_x \sigma_y}$ ~ n DIVISOR

= n DIVISOR.

$r = \frac{65 - (18.2)(2.7)}{\sqrt{500 - 18.2^2}\sqrt{11 - 2.7^2}}$

ALL GOOD FOR

$x \quad y \quad x^2 \quad y^2 \quad xy$

n

$\mu_{ax+b} = a\mu_x + b$	$r_{ax+b,\,cy+d}$
$\sigma_{ax+\cancel{b}} = \lvert a\rvert \sigma_x$	$= r_{x,y}$ IF $ac > 0$

AVGS.

STT 200 10-7-09

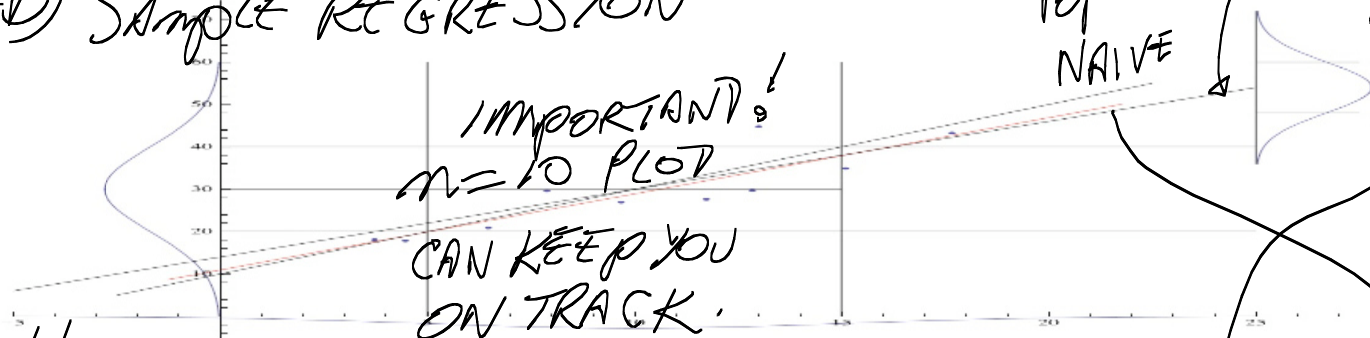
Note Title 10/7/2009

LOOKEE! SAMPLE PLOT $n=10$ POINTS GIVES (RED) SAMPLE REGRESSION CLOSELY COINCIDES WITH THE POPULATION REGRESSION!!

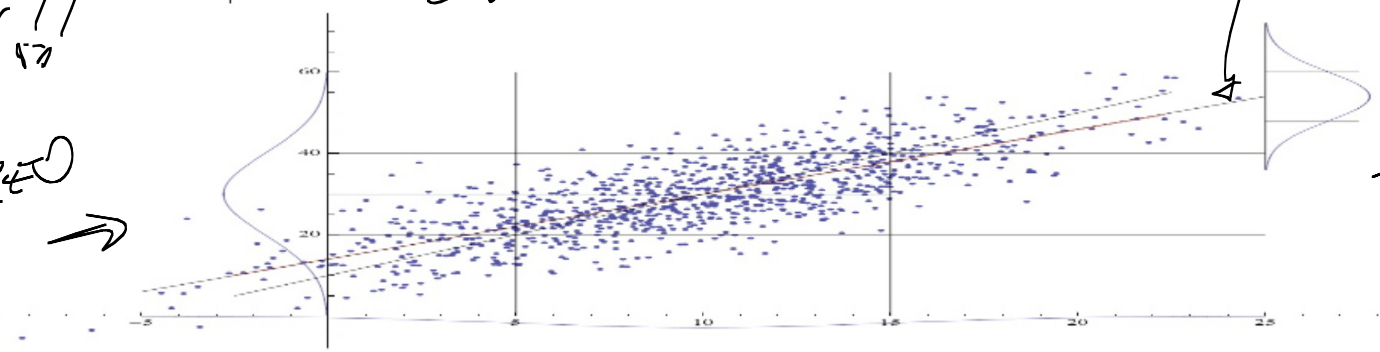
IMPORTANT! $n=10$ PLOT CAN KEEP YOU ON TRACK.

POPN/NAIVE

REGR LINE POPN



DESIRED 2-D NORMAL



RED LINE IS SAMPLE REGR LINE

NOW TO CH 23 READINGS (SYLLABUS)

1 MATTER OF ESTIMATING POPULATION MEAN μ_X.

FROM RANDOM EQ PROBABILITY WITH REPLACEMENT

1 "ONE POSSIBLE ESTIMATOR OF μ_X IS $\bar{X}$."

OTHER POSSIBILITIES INCLUDE

TRIMMED MEAN — eg TOSS OUT $X_{(1)}$ (LEAST) + $X_{(n)}$ (LARGEST) AND AVG $n-2$ REMAINING.

NONPARAMETRIC WEIGHTS $w_1 \cdots w_n$ on $X_{(1)} < \cdots < X_{(n)}$ ETC

METHOD OF ASSIGNING MARGIN OF ERROR FOR $\bar{X}$:

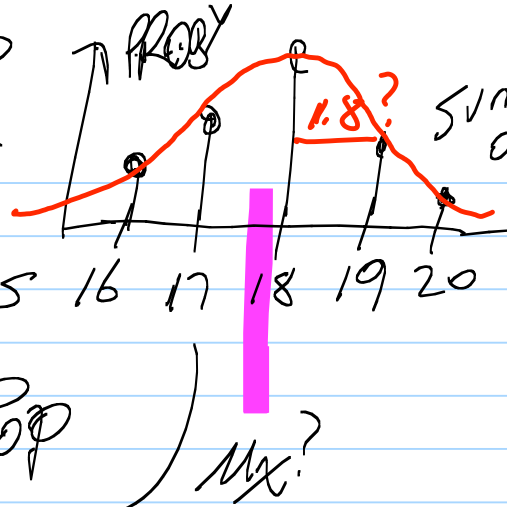
(EST^{d}) MARGIN OF ERROR $1.96\, \sigma_X / \sqrt{n}$

95% CONFIDENCE INTERVAL FOR μ_X: $\bar{X} \pm 1.96 \dfrac{\sigma_X}{\sqrt{n}}$

CLAIM: PROB'Y THAT $\bar{X} \pm 1.96\, \sigma_X/\sqrt{n}$ COVERS μ_X IS $\approx .95$ $n \to \infty$

WHY SO ??

IMAGINE A POP^n OF SCORES



POP

SUPPOSE

$\mu_x = 17.7$ (SAY)

$\sigma_x = 1.8$?? (SAY)

TAKE $n = 100$

ONE SAMPLE OF n

$\Longrightarrow$

RANDOM WITH REPL SAMPLE

$\bar{x}$, s_x, n

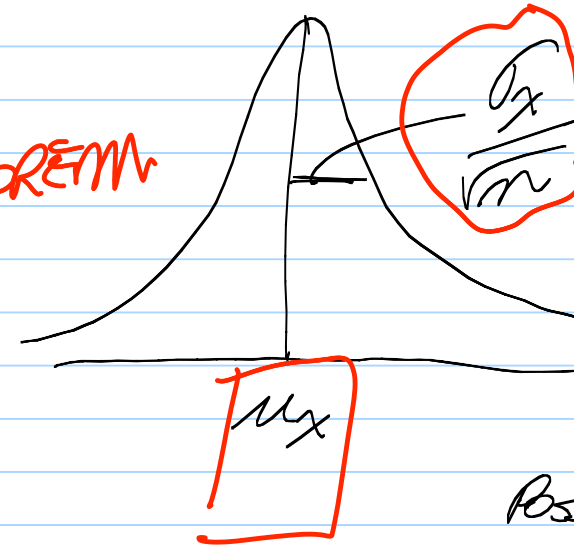
$\bar{x} \pm 1.96 \, s_x / \sqrt{n}$

AN ESD OF

CENTRAL LIMIT THEOREM

$\approx$

APPROX OF DENSITY FOR $\bar{x}$ SCORES !!



POSSIBLE

CENTRAL LIMIT THEOREM

DIST OF $\dfrac{\bar{x}-\mu_x}{\sigma_x/\sqrt{n}} \to$ Z DIST^N $N(0,1)$

ALSO (MORE USEFUL FORM)

$(\bar{x}-\mu_x)/(s_x/\sqrt{n}) \to$ Z DIST^N

APPL. $P\left(\underbrace{\left|\dfrac{x-\mu_{\bar{x}}}{s_x/\sqrt{n}}\right| < 1.00}\right) \approx P(|Z| < 1.00) = .68$

THIS IS EQUIV TO μ_x COVERED BY $\boxed{\bar{x} \pm 1.00 \dfrac{s_x}{\sqrt{n}}}$

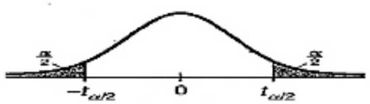
App 5   PROBY M_x IN

for 95% COVERAGE

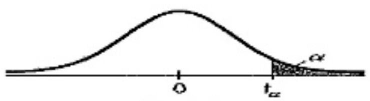
T TABLE

Table T
Values of t_α

	df					
Two-tail probability		0.20	0.10	0.05	0.02	0.01
One-tail probability		0.10	0.05	0.025	0.01	0.005
	1	3.078	6.314	12.706	31.821	63.657
	2	1.886	2.920	4.303	6.965	9.925
	3	1.638	2.353	3.182	4.541	5.841
	4	1.533	2.132	2.776	3.747	4.604
	5	1.476	2.015	2.571	3.365	4.032
	6	1.440	1.943	2.447	3.143	3.707
	7	1.415	1.895	2.365	2.998	3.499
	8	1.397	1.860	2.306	2.896	3.355
	9	1.383	1.833	2.262	2.821	3.250
	10	1.372	1.812	2.228	2.764	3.169
	11	1.363	1.796	2.201	2.718	3.106
	12	1.356	1.782	2.179	2.681	3.055
	13	1.350	1.771	2.160	2.650	3.012
	14	1.345	1.761	2.145	2.624	2.977
	15	1.341	1.753	2.131	2.602	2.947
	16	1.337	1.746	2.120	2.583	2.921
	17	1.333	1.740	2.110	2.567	2.898
	18	1.330	1.734	2.101	2.552	2.878
	19	1.328	1.729	2.093	2.539	2.861
	∞	1.282	1.645	1.960	2.326	2.576



Two tails



One tail

∞ ENTRY

1.96

TO PROB'Y THAT $\bar{x} \pm 2.326 \frac{\sigma_x}{\sqrt{n}}$

COVERS $\mu_x \approx$ (TABLE) 98%

$n \geq 30$

NOTE: IF YOUR SAMPLE IS WITHOUT REPL EQ PROB'Y MODIFY AS FOLLOWS ——

$\bar{x} \pm z \frac{\sigma_x}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$

eg 1.96 for 95%

STUDENTS' HOW TO MAKE DO WITH SMALL n. ?

TYPICALLY MAKES NO IMPACT IF $n \ll N$.

LOOKED AT $\boxed{\dfrac{\bar{x}-\mu_x}{\sigma_x/\sqrt{n}}} \sim z$ $\quad n \to \infty$

REGARDLESS OF FORM OF POP^N !!

IF $x_1 \ldots x_n$ ARE FROM A NORMAL POPULATION

IF POPULATION IS $(\approx)$ NORMAL

THINK $x_i = \sigma_x z_i + \mu_x$ PERSUE THIS THRU

$$\frac{\bar{x}-\mu_x}{\sigma_x/\sqrt{n}} = \bar{z}\Big/\left(\sigma_x/\sqrt{n}\right)$$

TABULATE THIS DIST^N

FOR $n = 2, 3, 4 \ldots$ ad inf $n = \infty$ 3 BACK.

UPSHOT SAY POP^N IS NORMAL eg IQ $\mu_x = 100$

$\sigma_x = 15$

SAMPLE $n = 4$ PERSONS.

$\bar{x} = 106 \quad s_x = 13.8 \quad n = 4$

USE 95% INTERVAL $\bar{x} \pm z \; t \; s_x/\sqrt{n}$

$106 \pm \cancel{1.96}\ 2.776 \dfrac{13.8}{\sqrt{4}}$ for 95%

Deg $= n-1 = 3$

SO T-INTERVAL $106 \pm 2.776 \dfrac{13.8}{\sqrt{4}}$ IS CALC^d BY A RECIPE

$\bar{x} \pm$ (T SCORE) $s_x/\sqrt{n}$ HAVING PROB^Y OF EXACTLY PROB^Y .95 OF COVER μ_x.