STT200 10-12-09

Note Title 10/12/2009

BACKGROUND
A POPULATION

SAMPLE $n = 100$ "UNITS"

PERHAPS
PROCESS
IN CONTROL

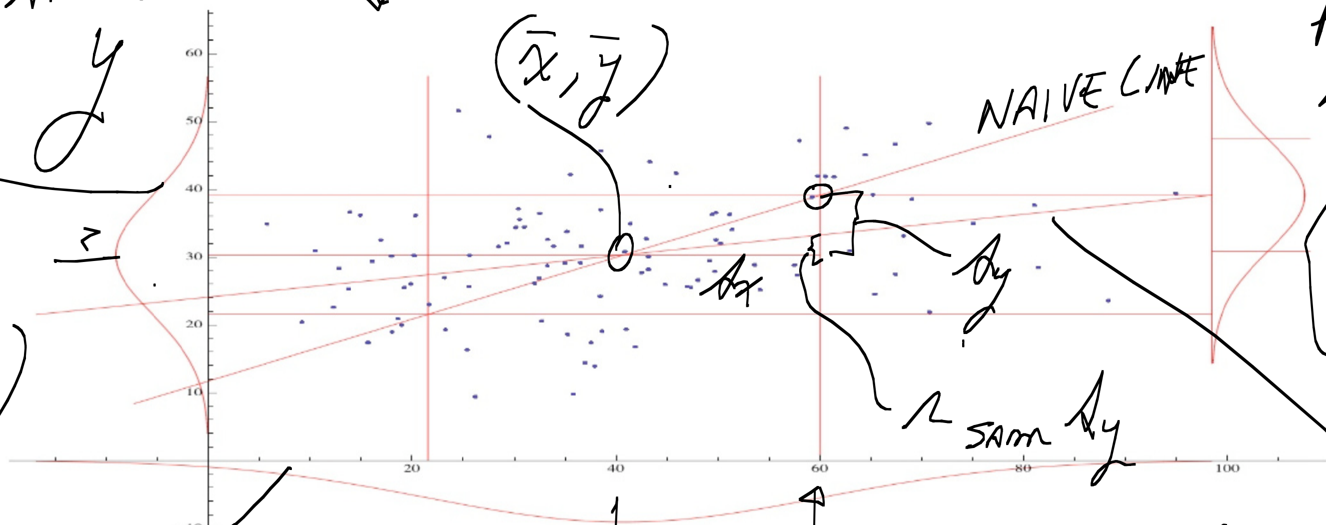
$\bar{y} + \Delta y$

y

$\bar{y}$
(SAMPLE)

$(\bar{x}, \bar{y})$

NAIVE LINE



Δx

Δy

r sam Δy

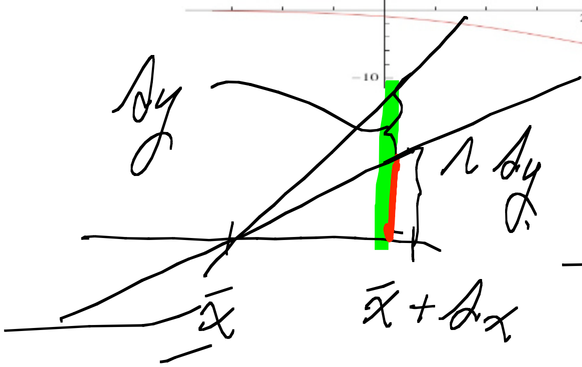
SAMPLE REGRESSION

LEAST SQUARES LINE

$\bar{x}$
(SAMPLE)

$\bar{x} + \Delta x$

x



Δy

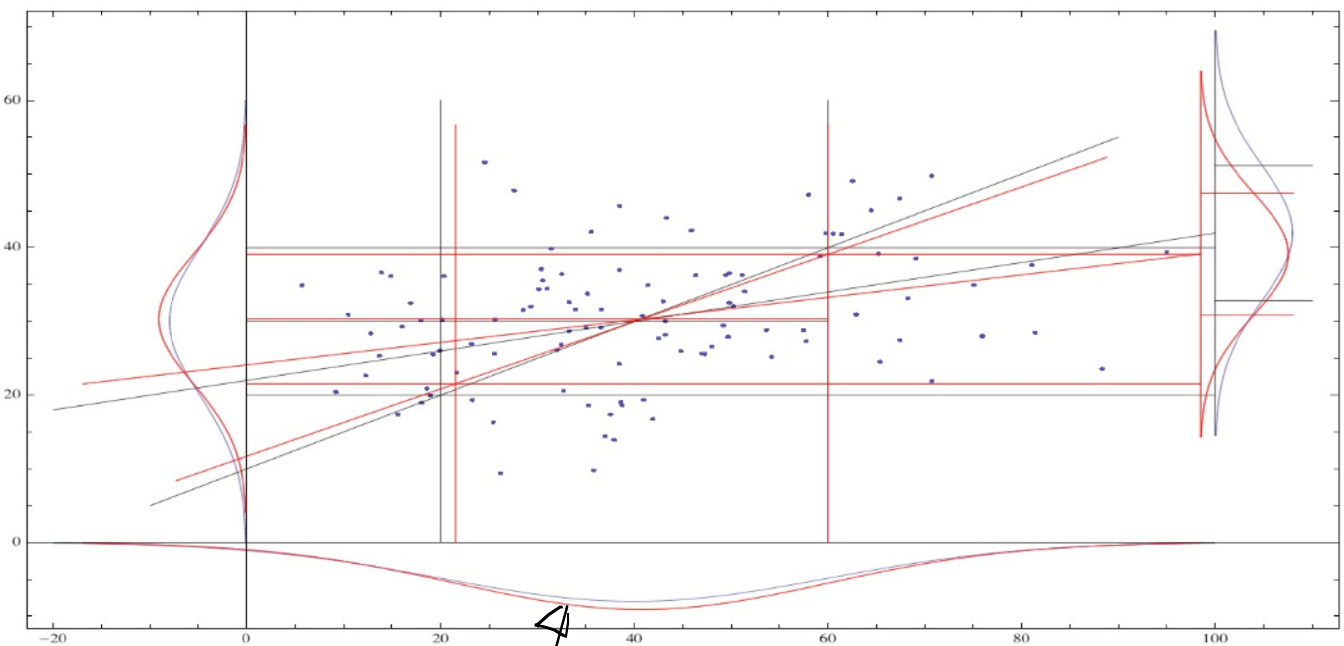
$r\,\Delta y$

$\bar{x}$ $\quad \bar{x} + \Delta x$

$r \quad R \quad \rho$

OBSERVE THAT $r \sim .4$ = RATIO RED/GREEN

Out[1903]=



BLACK: POPULATION REGRESSION

RED: SAMPLE REGRESSION

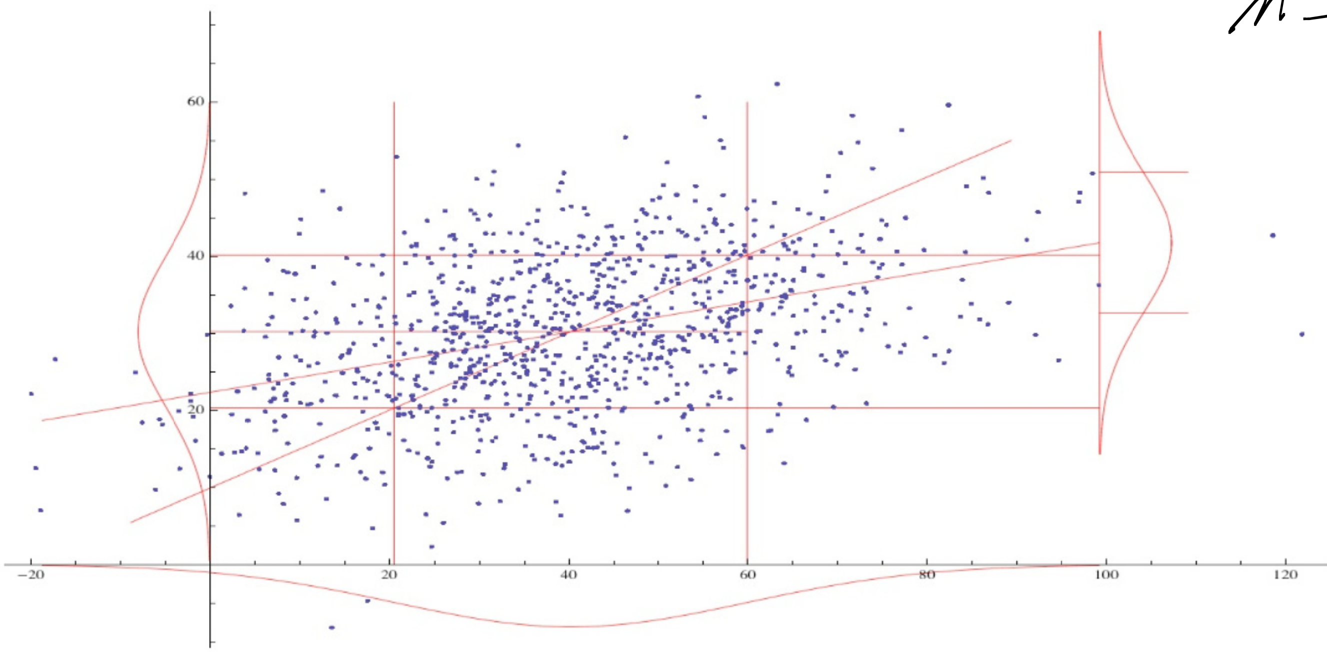
NOTE: LINE OF LEAST SQUARES

x = (SAY) LENGTH OF SARDINE   SAMPLE

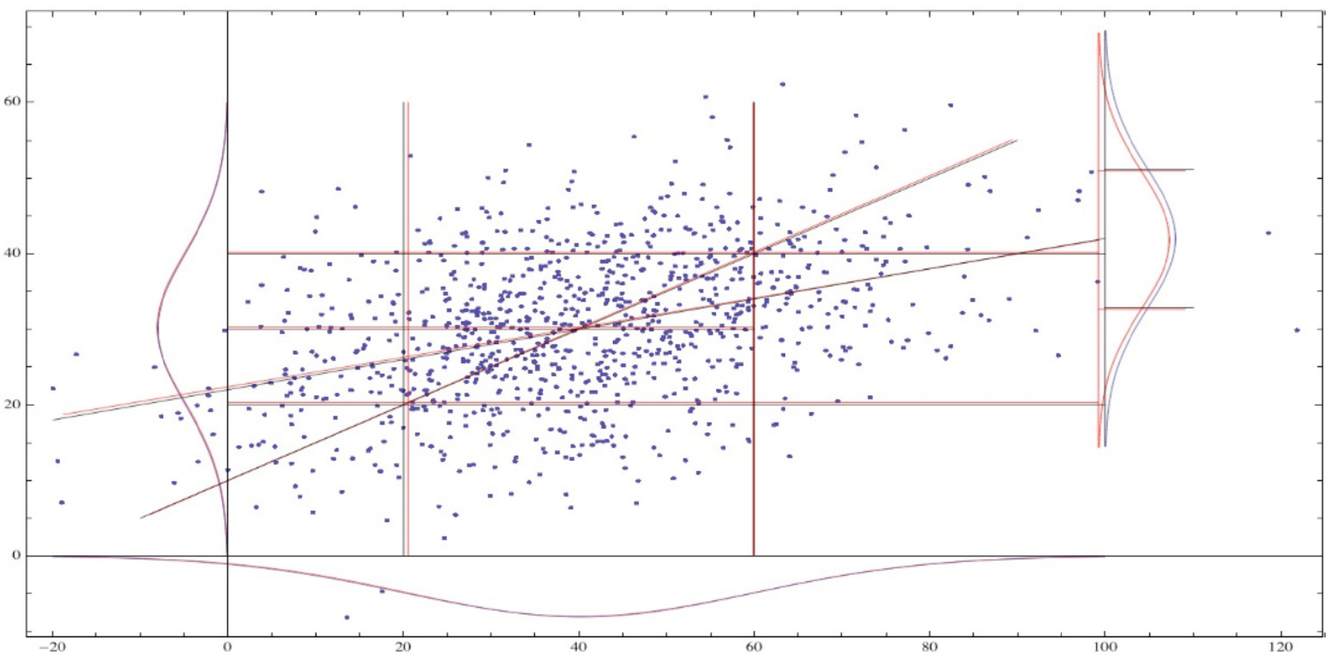
y = " WT " " $n=100$

Same
n = 1000

Out[1974]=



Out[1972]=



TAKE-AWAY: MODEST SIZED SAMPLES CAN DO REMARKABLE JOB OF REVEALING THE POPULATION CHARACTERISTICS.

$z = \frac{116 - 100}{15}$ — IQ, MEAN IQ, SD OF IQ

10. Determine the standard score z of a person having an IQ of ~~124~~ 116.

$\frac{16}{15} = 1.07$

11. Use the table of areas under the standard normal curve to deter
having IQ < ~~124~~ 116. This fraction is the fraction of the standard normal to

$P(IQ < 116) = P(z\ \text{SCORE} < (116-100)/15)$ TABLE $= .8577$

12. What is the IQ you need to have in order to surpass 83% of the p
the 83rd percentile of IQ.

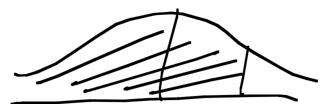
a. To find out, first find the 83rd percentile of z by entering 0.83 (or t
of the table of left-tail z areas, then read off the z-value.

b. After you find the 83rd percentile of z convert this to the 83rd perc
15 z.

So what is the 83rd percentile of IQ?

116

$P(IQ = 116) = 0$



100 116

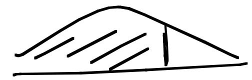
IQ

=

0 $z = \frac{116-100}{15}$

(11) CONT'

$P(z < 1.07)$ (FIX)

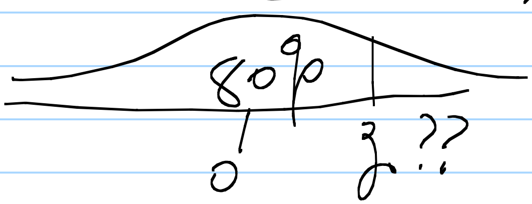


z	.07
1.0	(.8577)

12. a. ASK INSTEAD FOR 80^{TH} PERCENTILE OF IQ.

SOLVE FIRST FOR 80^{TH} PERCENTILE OF z.

1 = SD OF z



z	.04
.8	← .8 (BODY) ↑

.7995 CLOSEST TO .8

SO 80^{TH} PERCENTILE OF z IS $z = .84$

μ_{IQ} σ_{IQ}

12 b. THE 80^{TH} PERCENTILE OF IQ IS $100 + .84(15)$

$\overline{xy} = 9 = \dfrac{\cancel{0(0)} + \cancel{0(9)} + \cancel{9(3)}}{\cancel{3}}$

$n = 3$ POINTS (x, y)

x	y	x^2	y^2	xy
0	0	0	0	0
0	9	0	81	0
9	3	81	9	27
—	—	—	—	—
3.	4.	27.	30.	9.

AVGS →

16. Mean of x = 3

17. Sample standard deviation of x =
Note: Use a calculator program to get it then confirm in a separate ca

$$s_x = \sqrt{\frac{n}{n-1}}\sqrt{\text{mean of } x \text{ squares} - \text{square }}$$

SAMPLE SD

$$s_x = \sqrt{\frac{n}{n-1}}\sqrt{\overline{x^2} - (\bar{x})^2} = \sqrt{\frac{3}{2}}\sqrt{27 - 3^2}$$

18. Correlation R =
Note: Use a calculator program and confirm it is equal to
R =

$$\frac{\text{mean of xy products} - \text{product of } x \text{ mean by } y \text{ mean}}{\sqrt{\text{mean of } x \text{ squares} - \text{square of } x \text{ mean}}\sqrt{\text{mean of } y \text{ squa}}}$$

$r = R$
$((\rho))$

$$(\overline{xy} - \bar{x}\,\bar{y}) / \left(\sqrt{\overline{x^2} - \bar{x}^2}\sqrt{\overline{y^2} - \bar{y}^2}\right) = \frac{9 - 3(4)}{\sqrt{27 - 3^2}\sqrt{30 - 4^2}} =$$

$\bar{x} = 12$, $s_x = 2.4$ $n = 60$

(CROP)

Questions 24 through 28 concern T and normal based confidence
error. Throughout, we suppose a random sample is selected wit
probability from a population (the distribution of the population x sc
unless it is specifically assumed in a problem). Suppose the samp
standard deviation is ~~4.2~~ 2.4, the sample size is ~~50~~ 60, and the population size

$\bar{x} = 12$

23a. Give the *point estimate* of the population mean based on this dat

TYPICALLY → $\bar{x} = 12$

ESTIMATOR = RULE $\bar{x} = 12$
ESTIMATE = VALUE

24. Give the estimated margin of error for your estimate of #23a.

RULE BY ~~WHICH~~ WE GET: EMOE IS $1.96\, s_x/\sqrt{n}$ VALUE IS 1.96

25. Give the 95% confidence interval for the population mean based

$\bar{x} \pm 1.96\, s_x/\sqrt{n}$ CLAIM: $P(\bar{x} \pm 1.96\, s_x/\sqrt{n}$ COVERS $\mu_x) \approx .95$

GOES W/ 1.96

26. Your answer to #25 is an interval calculated from data withou
mean. What is the *approximate probability* that such an interval cove
means out of all possible intervals constructed from random samples.

.95 ANS

27. If instead of sampling with replacement we sampled without rep
answer to #24?

IF INSTEAD SAMPLE WITHOUT REPL: SIMPLE CORRECTION

$\underbrace{\bar{x} \pm 1.96 \frac{s_x}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}}_{\text{ASSOC CI}}$

EMOE

ANS TO #27

NOTE: $\sqrt{\frac{N-n}{N-1}}$ CALLED FINITE POP'N CORRECTION. $\frac{N-n}{N-1} \approx 1$ IF $n << N$

[redacted]

28. If instead of sampling with replacement we sampled *without*
answer to #25?

CIRCLED IN 27.

29. If instead we had this same data but from a sample of only n = 6
known to be close to normal what number would we use in place of th
applicable degrees of freedom?

30. For a sample of n = 6 from a *normal population* give the 95% co
mean. Notice it is widened as compared with the n = 50 case since
larger than s / $\sqrt{6}$ but also the applicable T score exceeds 1.96. THE
EXACT, MEANING IT ACHIEVES EXACTLY THE NOMINAL C
IF THE POPULATION DISTRIBUTION IS INDEED NORMAL,
ACCURATE AND ALL CALCULATIONS ARE INFINITE PRECISI

29. SUPPOSE POPULATION HAS NORMAL DISTRIBUTION.

IF $n = 6$ $\bar{x} = 12$ $s_x = 2.4$

LOOK AT USUAL C.I. $\cancel{\bar{x} \pm 1.96 \frac{s_x}{\sqrt{n}}}$ NOT APPLICABLE

APPROX $n \to \infty$

HOWEVER: USING STUDENT'S T

GET 95% CI FOR μ_x : $\bar{x} \pm T \frac{s_x}{\sqrt{n}}$

REPL 1.96

DEGREES FREEDOM:

SHORT ANS $DF = n - 1$ — PENALTY FOR EST. σ_x BY s_x.

$DF = 6 - 1 = 5$

T

T DF

5 $\boxed{2.571}$

$n \to \infty$ 1.96

CONF LEVELS 95%

A-59

T-CI (95%) IS

$\bar{x} \pm T\ s_x/\sqrt{n}$

$12 \pm \boxed{2.571}\ \frac{2.4}{\sqrt{6}}$

~~1.96~~

HAD n BEEN ONLY 2

DF = 1

$12 \pm \boxed{12.706}\ \frac{2.4}{\sqrt{2}}$