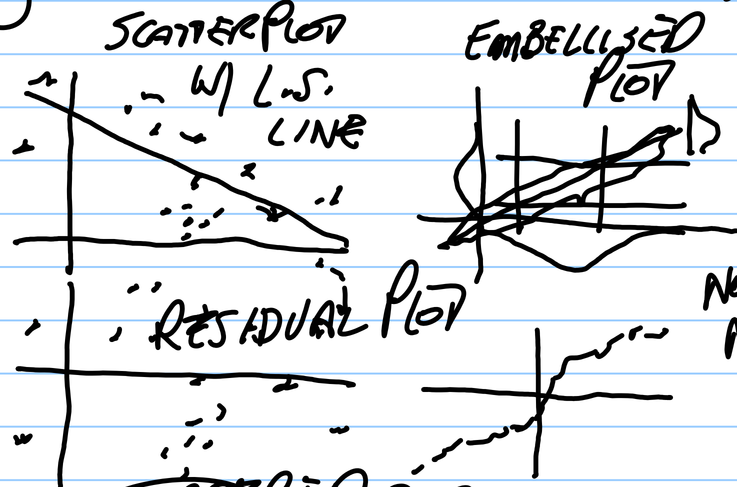
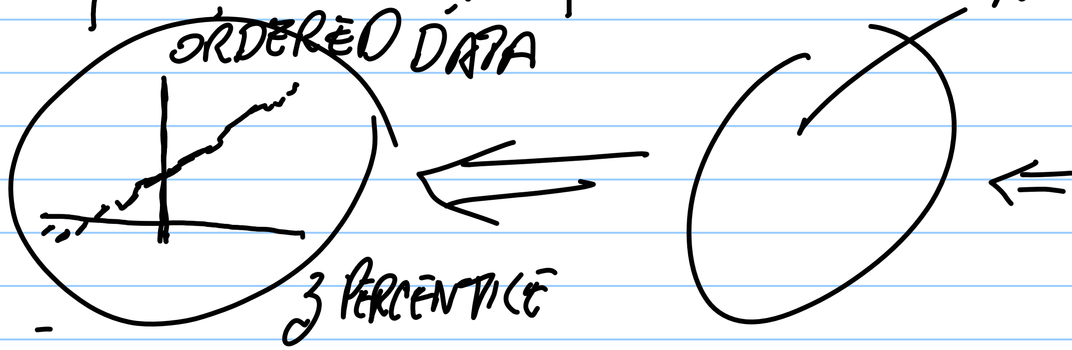
STT 200 5:30pm 4-12-10a

Note Title 4/12/2010

#18 NORMAL PROBABILITY PLOT TAKES 1 DIM NORMAL SAMPLES INTO ≈ STRAIGHT LINE.



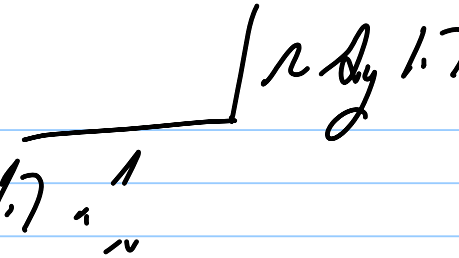
NORMAL PR PLOT NOT SO GOOD.



NORMAL PROB'Y PLOT DEVICE

SOME DATA $x_1 \dots x_n$

#1. $z_x = 1.7$, STD X SCR. = L.S. LINE

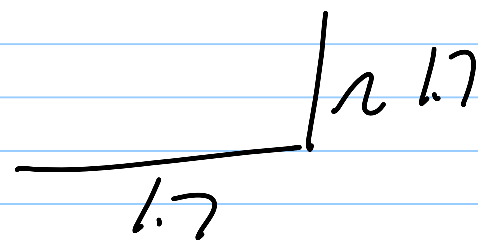


L.S. PRED FOR $z_y = r\,1.7$

ANS: IF JOHN IS 1.7 SD_x ABOVE MEAN x

L.S. PREDICTS ONLY $\underbrace{0.9}_{r}(1.7) = 1.53$ SD_y ABOVE MEAN y,

IN STD SCORES



($\sigma_x = \sigma_y = 1$ IN STD SCORES)

#2. IF IN #1. DIST OF x IS NORMAL

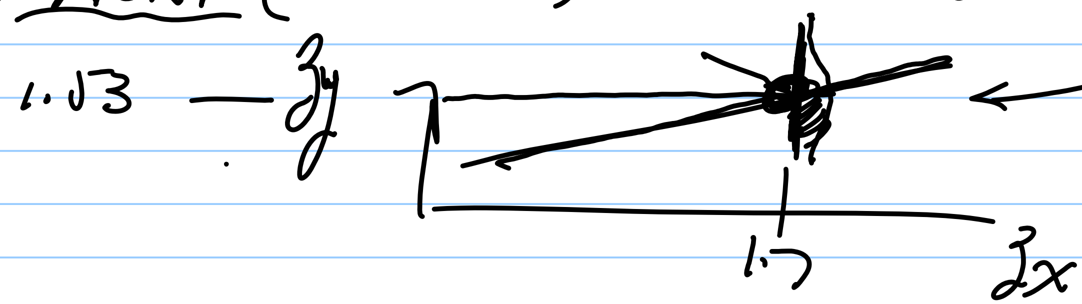
JOHN'S PERCENTILE IS $P(Z < 1.7) = .9554$

.00
1.7 .9554

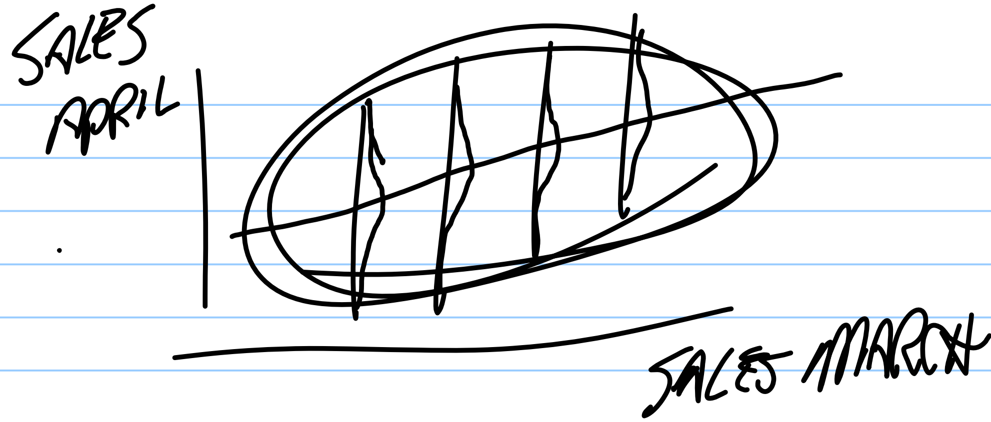
3. IF JOHN EARNS STD SCORE ~ 1.7 = 1.53 (0.9)
AS PREDICTED BY L.S., 1.53 = PRED z_y FOR JOHN.
IF JOHN EARNS THAT THEN JOHN'S
EXAM 2 PERCENTILE WOULD BE 0.937

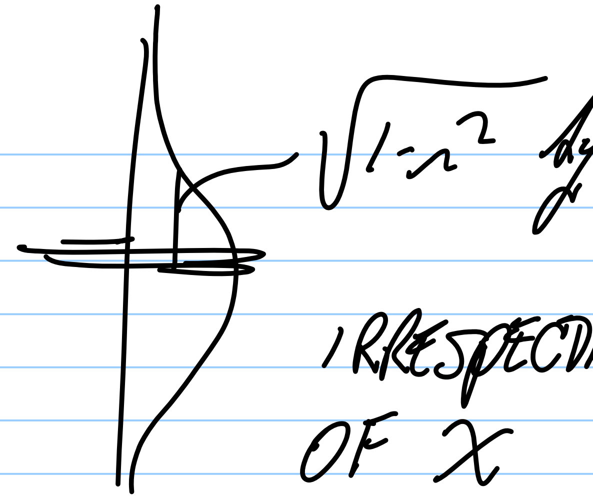
1.5 .03 $\boxed{0.937}$

4. IF JOHN'S EXAM 2 STD SCORE IS 1.53 AS
PREDICTED WHAT IS JOHN'S PERCENTILE IN HIS
COHORT (LIKE HE, EARNING STD SCORE 1.7 ON EXAM 1)



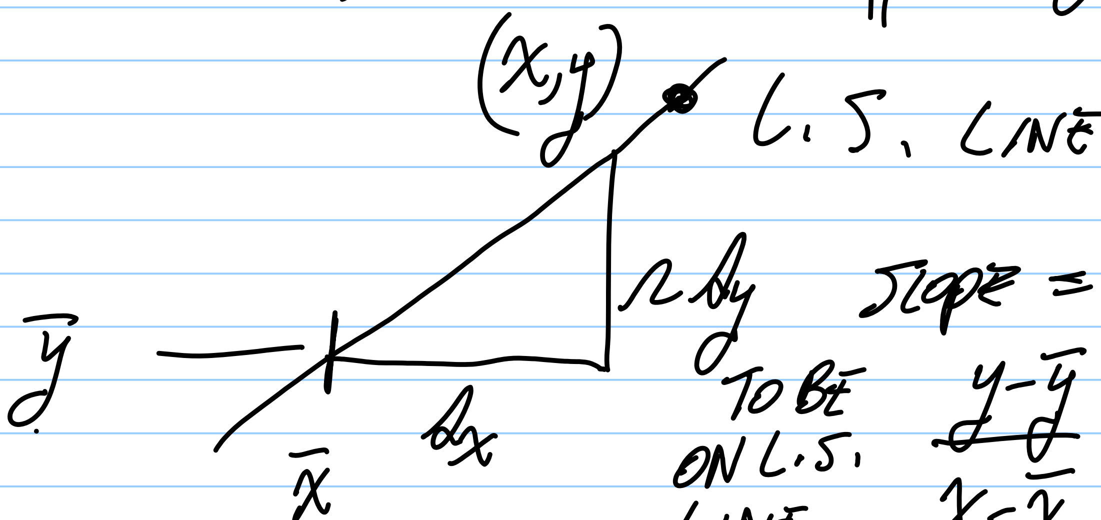
FOR 2D NORMAL PLOTS.





IRRESPECTIVE OF X

#5.



SLOPE $= r\,\sigma_y/\sigma_x$

TO BE ON L.S. LINE $\quad \dfrac{y-\bar{y}}{x-\bar{x}} = r\,\sigma_y/\sigma_x$

\#6. EQ^N OF L.S. (SAMPLE LINE) $y = \bar{y} + (x - \bar{x})\, r\, s_y / s_x$

LIKEWISE FOR POPULATION

$$y = \mu_y + (x - \mu_x)\, \rho\, \sigma_y / \sigma_x$$

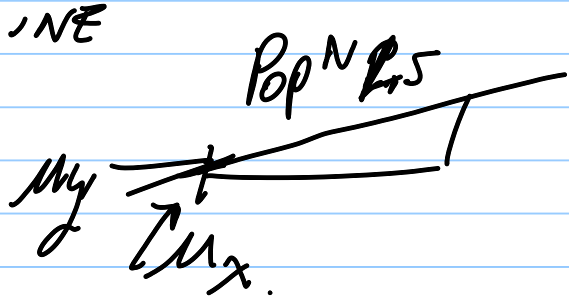
ρ POP CORRELATION

SO SAMPLE L.S. LINE WANTS TO BE CLOSE TO THE POPULATION L.S. LINE.

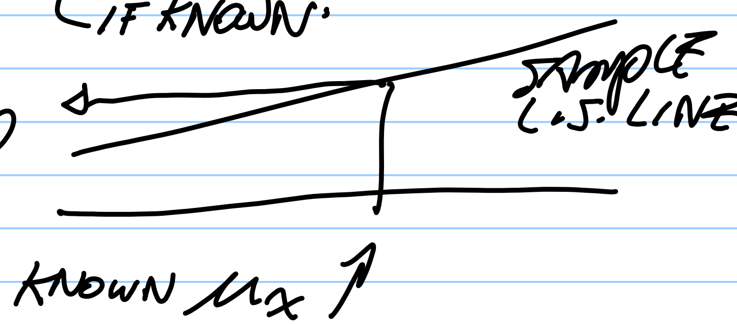
\#7. GIVING MOTIVATION FOR REGR-BASED ESTIMATOR OF μ_y. WHICH IS "INSERT $x = \mu_x$ INTO SAMPLE L.S. LINE"

μ_x IF KNOWN.

PICS



INSTEAD



SAMPLE L.S. LINE $y = \bar{y} + (x - \bar{x}) r \, s_y / s_x$

μ_x

REGR-BASED

ESTIMATOR OF μ_y IS NOT $\bar{y}$ BUT IS

$$y^*_{REGR} = \bar{y} + (\mu_x - \bar{x}) r \, s_y / s_x$$

INSERT 42

FURTHERMORE 95% z-BASED CI FOR μ_y IS

$$y^*_{REGR} \pm \sqrt{1-r^2} \, 1.96 \, s_y / \sqrt{n} \quad \Big\| \quad \bar{y} \pm 1.96 \, s_y / \sqrt{n}$$

NARROWED CI

IF $r = .7$

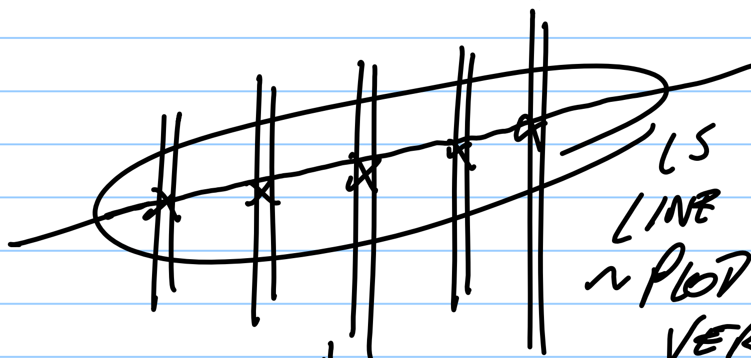
$\sqrt{1-r^2} \approx \sqrt{.5} \sim .7$

EQUIV TO FAIRLY LARGE INCR IN M.

SETUP y = 2010 TAX

x = 2009 TAX. (FREE INFO) (KNOW μ_x)

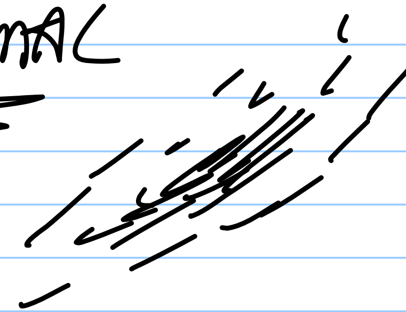
#8.



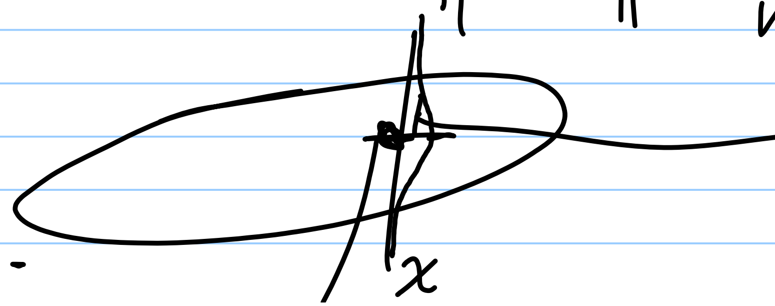
LS LINE

~ PLOT OF VERT STRIP AVGS.

2D **NORMAL**



#9.



$\sqrt{1-r^2}\,\sigma_y$

SAME FOR EVERY X !!!

#10. $\bar{y} + (x - \bar{x})\, r\, \sigma_y/\sigma_x$ REGR PREDICTION FOR y.

#11. 2D NORMAL (AGAIN)

AVG HT OF SONS WHOSE FATHER'S HEIGHT ~62"

$\mu_x = 65 \quad \sigma_x = 4 \quad \mu_y = 67 \quad \sigma_y = 5 \quad \rho = 0.76$

$\mu_y + (x - \mu_x)\, r\, \sigma_y/\sigma_x$

$67 + (\underset{\text{"pop"}}{62} - 65)\, 0.76 \left(5/4\right)$ PRED

THE SD IS $\sqrt{1-r^2}\,\sigma_y = \sqrt{1-0.76^2}\,(5)$

#12. 2D NORMAL SAME AS #11

REVERSE ROLES OF FATHER & SON.

IF A SON IS 62" PREDICT FATHER'S HT

$$= \mu_x + (y - \mu_y)\, r\, \sigma_x / \sigma_y$$

$$= 65 + (\underbrace{62}_{\text{SON}} - 67)\, 0.76\, (4/5)$$

& SD OF HTS OF FATHERS WHOSE SON'S HAVE HT 62"

IS $\sqrt{1 - r^2}\, \sigma_x = \sqrt{1 - .76^2}\, \cancel{(4)}\, (5)$

INTERCHANGE x, y

I CHANGED $\sigma_x = 5$ (NOT 4)

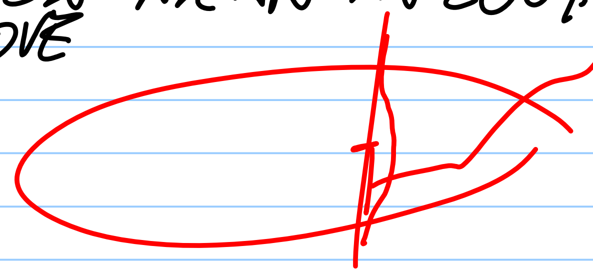
#13. $r = 0.4$ COHORT IS 1 SD_x BELOW MEAN X

PREDICT $0.4(1)$ SD_y " " y.

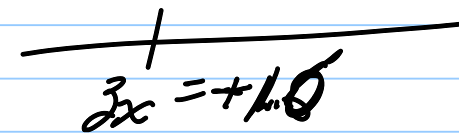
REGRESSION TO MEDIOCRITY.

#14. SD OF SERVERS 2010 INCOME WITHIN ANY FIXED COHORT

1.0 SD ~~BELOW~~ ABOVE MEAN IN 2009.



SD IN COHORT IS $\sqrt{1-r^2}\ \sigma_y$



#15. MORE VARIABLE?

(a) 2010 EARNINGS FOR COHORT
2009 " = 27,000

(b) 2010 " " "
" " = 38,000

(c) SAME

16.