STT 200 5:30 pm 4-19-10

Assigned exercises beginning pg. 746: $\boxed{14,\ 16}$, 17, 36, 37, 38.

Exam 4 Prep 21-22. Sampling Distribution slope b_1 of sample regression line. This is keyed to #14 and #16 of page 748.

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

$s_e = \sqrt{\sum (y - \hat{y})^2 / (n-2)} = \frac{\sqrt{n}}{\sqrt{n-2}} \sqrt{1 - r^2} \sqrt{\overline{y^2} - (\bar{y})^2}$

$SE(b_1) = \frac{s_e}{\sqrt{n-1}\ s_x}$ estimate of SD of b_1

$= \frac{\sqrt{1-r^2}\ s_y}{\sqrt{n-2}\ s_x} = SE(b_1)$

This estimate of the standard deviation of b_1 is useful for CI and tests about the population slope β_1.

CH 27. SLOPE OF L.S. LINE ON SAMPLE

$\underbrace{SE(b_1)}$ EST OF SD OF b_1

RECALL CI $\bar{y} \pm z\, s_y/\sqrt{n}$

EST OF SD OF $\bar{y}$ $SE(\bar{y})$

SCRIPT IS: EST $\pm$ t_{crit} SE(EST)

TASKED WITH GIVING CI FOR SLOPE β_1 OF THE POPULATION L.S. LINE

$b_1 \pm (t_{crit})\, SE(b_1)$

SLOPE OF SAMPLE L.S. LINE

$r\, s_y/s_x \pm t_{DF=n-2} \frac{\sqrt{1-r^2}}{\sqrt{n-2}}\, s_y/s_x$

APPROPRIATE IF POPULATION DISTRIBUTION IS ~NORMAL

$SE(b_1)$ = EST SD OF b_1

NORMAL Pop

If the population is 2 D normal then for n > 1

CI for β_1:

* $b_1 \pm t_{\mathrm{df}=n-2}\, \mathrm{SE}(b_1)$ (exact) ✓ $P(\text{CI covers } \beta_1) = \text{e.g. } 95\%$

$n>2$

* test statistic $\dfrac{b_1 - \beta_1}{\mathrm{SE}(b_1)}$ has exact $t_{\mathrm{df}=n-2}$ distribution.

Any Pop $n \to$ LARGE

Even if the population is not normal, for large n:

CI for β_1:

* $b_1 \pm t_{\mathrm{df}=n-2}\, \mathrm{SE}(b_1)$ (approx)

z

* test statistic $\dfrac{b_1 - \beta_1}{\mathrm{SE}(b_1)}$ has approx z distribution.

Note Title 4/19/2010

As applied to #14 and #16 pg. 748

CH. 27

x = AGE USED TOYOTA COROLLA

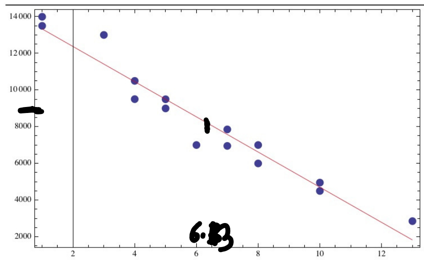
y = ASKING PRICE $.

x	y	x^2	y^2	xy
1	13 990	1	195 720 100	13 990
1	13 495	1	182 115 025	13 495
3	12 999	9	168 974 001	38 997
4	9500	16	90 250 000	38 000
4	10 495	16	110 145 025	41 980
5	8995	25	80 910 025	44 975
5	9494	25	90 136 036	47 470
6	6999	36	48 986 001	41 994
7	6950	49	48 302 500	48 650
7	7850	49	61 622 500	54 950
8	6999	64	48 986 001	55 992
8	5995	64	35 940 025	47 960
10	4950	100	24 502 500	49 500
10	4495	100	20 205 025	44 950
13	2850	169	8 122 500	37 050
—	—	—	—	—
6.13333	8403.73	48.2667	8.09945×10^7	41 330.2

$\bar{x}$ → $\bar{y}$ →

Note Title 4/19/2010

8400



x = age (yr), y = advertised price (dollars)

$r \sim -0.97$

n = 15 pairs (x, y)
means (6.1333, 8403.73) ← $(\bar{x}, \bar{y})$
s_x = 3.3778 s_y = 3333.56
r = -0.971767 t_{13} = 2.16 for 95%
b_1 = -959.039
SE(b_1) = 64.5816 (applicable df = 15-2 = 13)
95%CI = -959.039 + {-1, 1} (2.16) (64.5816)
= {-1098.54, -819.543}

IF THIS LOOKS LIKE ~ NORMAL DATA WE MIGHT BE OK WITH CI FOR β_1.

$$- \; b_1 \pm t_{DF = 15-2 = 13} \; SE(b_1)$$

FOR 95% CONF
FIND $t_{13} = 2.16$

$$r \frac{\Delta y}{\Delta x} \pm 2.16 \frac{\sqrt{1-r^2}}{\sqrt{n-2}} \Delta y / \Delta x$$

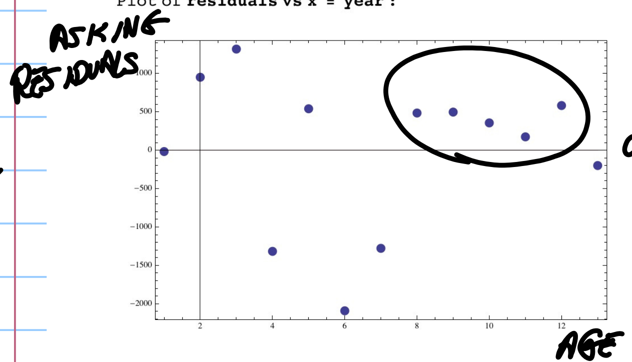
$$\underbrace{-.97 \frac{3333}{3.38}}_{\substack{EST \\ -959.039}} \pm \underbrace{2.16}_{t_{DF\,13}} \underbrace{\frac{\sqrt{1-(-.97)^2}}{\sqrt{15-2}} \frac{3333}{3.38}}_{SE(b_1).}$$

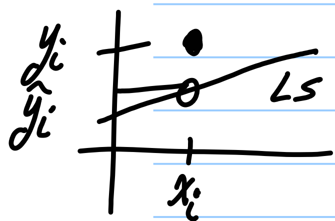
$\{-1098.54, -819.543\}$

LOOKS LIKE ~ INTERVAL FOR AVG LOSS IN VALUE PER YEAR.

Plot of **residuals** vs **x** = **year** :

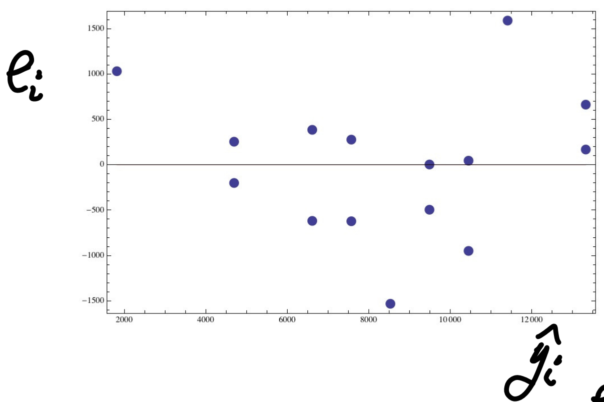
e_i





$e_i = y_i - \hat{y}_i$

Plot of **residuals vs. predicted** values as favored by the textbook:

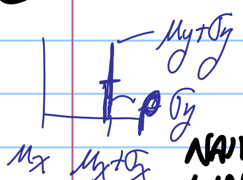


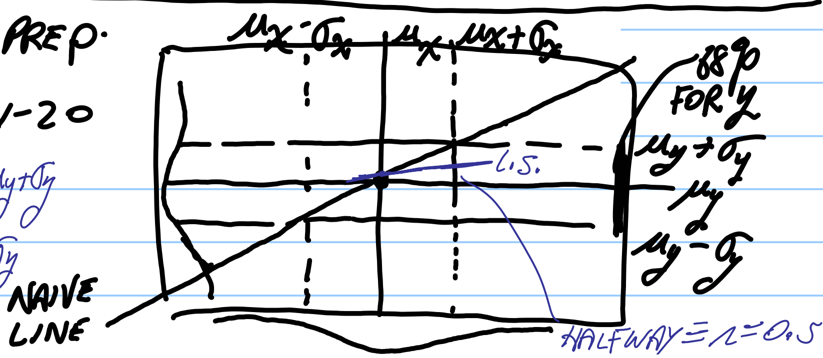
CASE IS MADE THIS TYPE OF RESID PLOT IS LESS PRONE TO ARTIFACTS LIKE ASYMMETRIC CLUMPS SUCH AS SEEN IN PREVIOUS PLOT.

OTHERS OF 4-20-10 REC. EXERCISES GIVE YOU $SE(b_1)$ $\boxed{SE_{-}.}$

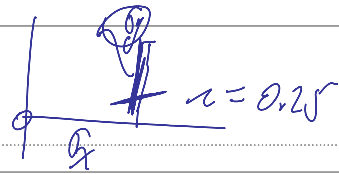
EXAM 4 PREP.

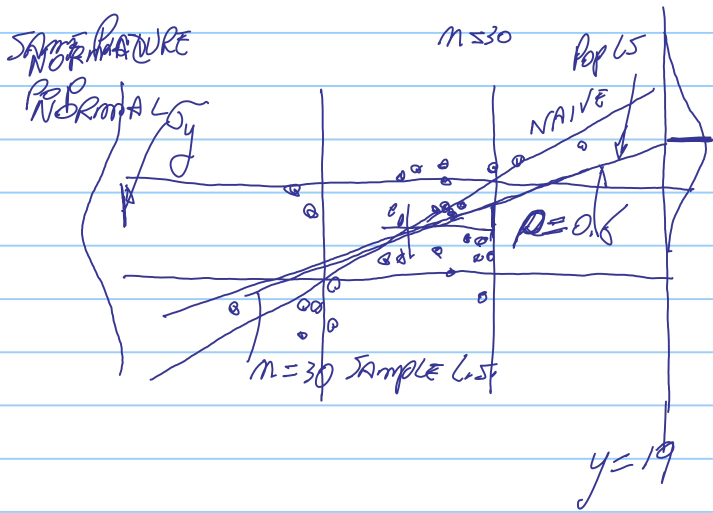
① ~~1-5~~ 1-20





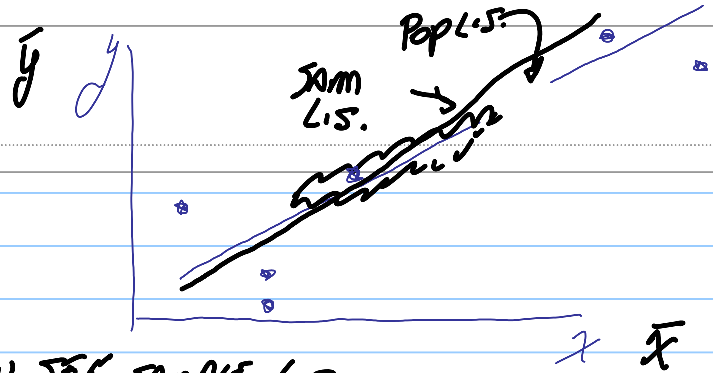
Note Title 4/19/2010





Note Title 4/19/2010

NOT NORMAL POPULATION



WHEN YOU SEE SAMPLE L.S.

① ~ Pop L.S. $(\bar{x}, \bar{y})$

② ~ L.S. ON "ALL" POSSIBLE $(\bar{x}, \bar{y})$, $n = 30$

② ⇒ REGRESSION BASED ESTIMATOR OF μ_y (GIVEN BY $\bar{y} + (\mu_x - \bar{x}) r \, s_y/s_x$) IS WHAT YOU GET WHEN YOU INSERT μ_x TO SAMPLE L.S. LINE (WHICH IS BY ② APPROX L.S. LINE ON THE ~NORMAL PLOT OF ALL $(\bar{x}, \bar{y})$).

SO REGR-BASED ESTIMATOR IS GOOD.

Note Title 4/19/2010

Note Title 4/19/2010

Note Title 4/19/2010