3-2 Day 3 Linear Regression, Residuals

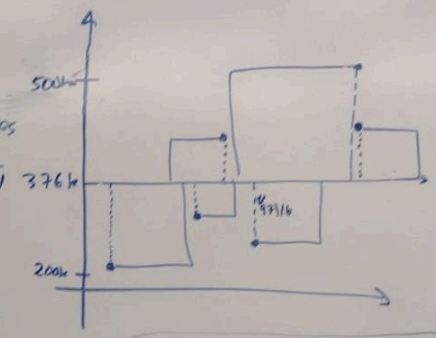
i	L_1 X Size of our home (ft^2)	L_2 Y Cost($)
1	1920	279000
2	2682	529900
3	1064	210000
4	1616	308000
5	2696	489900
6	1728	439900

$\hat{y} = \frac{174.65}{1}x + 35363$

SSTo : Total sum of squared differences from the mean

$\sum(y-\bar{y})^2 = 8.233 \times 10^{10}$

This represents all of the variability in home price about the mean \$376116.67

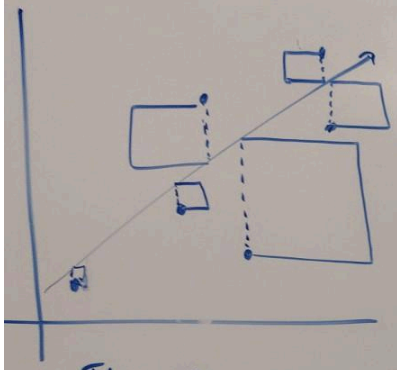


$$r^2 = 1 - \frac{SSResid}{SSTo}$$

$\bar{y} = \$376116.67$ $\bar{x} = 1951\ ft^2$

SSResid : sum of squared residuals

$\hat{y} = \ldots$



$$\sum (y - \hat{y})^2 = 2.013 \times 10^{10}$$

This # represents the variability in Y (price) that can't be explained by LSRL w/ x = square footage.

condition
age ...

$\frac{SSResid}{SSTo}$ = prop. of variability left unexplained by the line

r^2 "Coefficient of Determination."

prop. of variability the the LSRL does explain w/ x = square footage

r = "Correlation Coefficient"

$r^2 = 0.486$ $\quad r = \sqrt{0.486} = 0.697$

$$r = \sqrt{r^2}$$

$$r^2 = 1 - \frac{2.013 \times 10^{10}}{8.233 \times 10^{10}} = 1 - \frac{2.013}{8.233}$$

$r^2 = 0.756$

About 75.6% of variability in price of a home in Fairbanks (y in context) can be accounted for with the least-squares regression line with x = size of home (ft^2)

Facts about LSRL

1. residuals have y units
2. $(\bar{x}, \bar{y})$ is always on LSRL

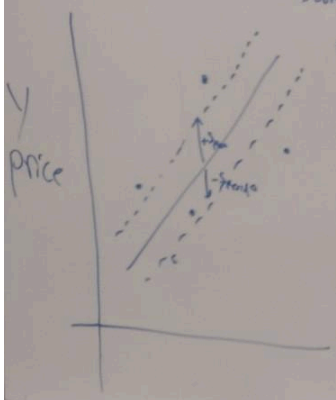
Standard Deviation of Residuals

$$S_e = S = S_{Resid} = \sqrt{\frac{SS\ Resid}{n-2}} = \sqrt{\frac{2.013 \times 10^{10}}{n-2}}$$

book

Calc.

$$= \sqrt{\frac{20130000000}{6-2}}$$



Interpretation:

The actual price of a home in FBX typically varies about \$70940.12 from the price predicted by least-squares-regression line w/ x being size of home in ft²

$= \$70940.12$