Quiz - open notes -
prepare your notes!

1. What is a least-squares regression line?
2. What is the formula for a least squares regression line? Define each variable.
3. The LSRL always passes through what point?
4. What is a residual?
5. What special property do the residuals have?

6. If a least squares regression line fits the data well, what characteristics should the residual plot exhibit?

(2) $\hat{y} = a + bx$

$\hat{y}$ = predicted value

$b = r \cdot \frac{s_y}{s_x}$ $\quad a = \bar{y} - b\bar{x}$

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

(4) Actual - predicted

Hondas

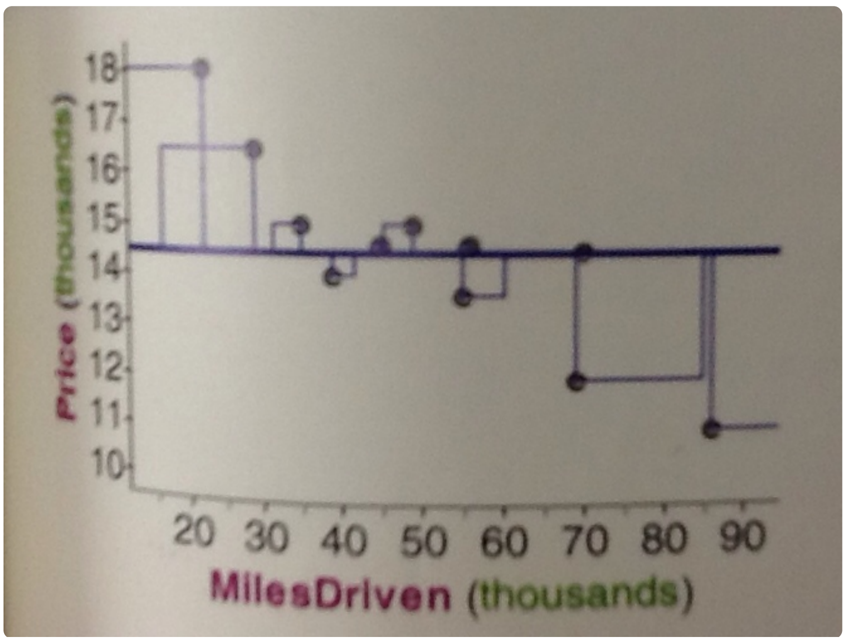
x #miles	
y price	

suppose we want to estimate the advertised price of a used 2002-2006 honda CR-V but don't know the number of miles. The mean price of other used CR-V's would be a reasonable guess.

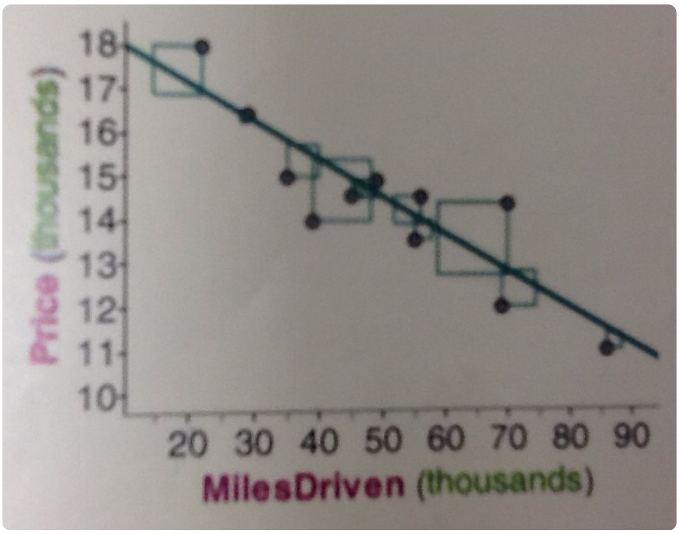
22 17998
29 16450
35 14998

$\bar{y}$

$\bar{y}=14{,}425$



This SCATTERPLOT shows the squared prediction errors when using the mean price $\bar{y}$ as our prediction. The sum of the squared prediction errors is called the total sum of squares, SST, and measures the total variability in the y variable. In this case, SST=36,070,000. However, we could make much better predictions if we know the number of miles driven. How much better? This SCATTERPLOT shows the squared prediction errors when using the LSRL.



The sum of the squared prediction errors when using the LSRL is called the sum of the squared errors, SSE. In this case, SSE=8,497,000. This means that only 8,497,000/36,070,000 or 23.6% of the variation in asking price is UN-accounted for by the LSRL. The remaining variation is due to other factors, such as the condition or age of the car. Therefore, 1 - 8,497,000/36,070,000 or 76.4% of the variability in advertised price is accounted for by the LSRL. This last percent is r^2 and should be interpreted by saying, "76.4% of the variation in advertised price can be accounted for by the linear model relating advertised price to number of miles driven."

$r^2 \rightarrow$ coefficient of determination

$$r^2 = 1 - \frac{SSE}{SST}$$

$SSE = \sum \text{residuals}^2$

$SST = \sum (y_i - \bar{y})^2$

Top p.182 Computer Outputs

$r = -0.874$

$r^2 = (-0.874)^2$

$= 0.7644$

$1 - \dfrac{8{,}497{,}000}{36{,}070{,}000} = 0.7644$

proj due 10/3

Ch3 Test Friday

Wed - comp time