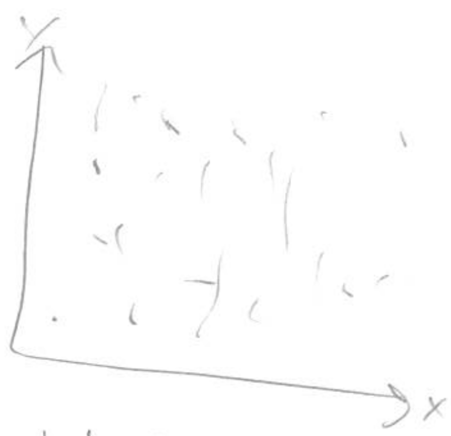
Lecture #20
6/28/11

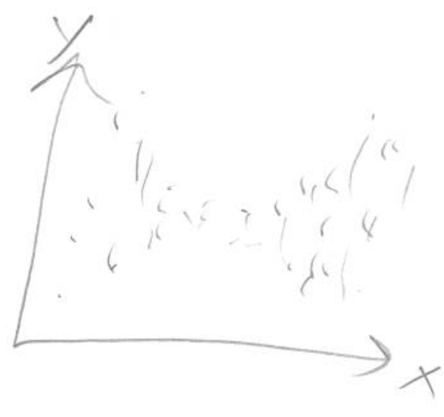
Admin
- off hrs today

Plan
- review LS function
- LS demos
- regression to the mean
- linear model assumptions
- distribution of the prediction
- regression fallacies

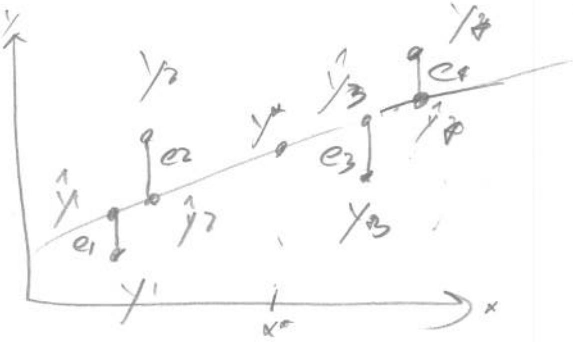
Independent $\Rightarrow$ uncorrelated
Uncorrelated $\not\Rightarrow$ Independent



ind & uncorrelated. Why?



uncorrelated and dependent. Why?



residuals below the line are (−), above the line are (+)

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

By min $\sum_{i=1}^{n} e_i^2$ we found best fit line

$\hat{y} = b_0 + b_1 x$ ← best guess

where

$b_0 = \bar{y} - b_1\bar{x}$, $b_1 = r\frac{s_y}{s_x}$

or equivalently

$z_y = r z_x$ z_y, z_x are not draws from $N(0,1)$ it's just bad notation

Better fit → smaller residuals overall, worse fit, vice-versa

A measure of spread of residuals is

$$s_e = \sqrt{\frac{\sum e_i^2}{n-2}} = \sqrt{MSE} = RMSE$$

RMSE, se interchangeable. This deals with the spread of the residuals.
same $\hat{y}$ formulas, but

Now we'll go one step further.

Right now, we only have algebraic answers for b_0, b_1.

What is the distr of B_0, B_1 when b_0, b_1 are draws?

We don't know ... we haven't imposed a probability model on it.

We now define the simple regression model:

3-param model

$$\text{Assume that } \quad Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \text{where} \quad \varepsilon_1, \ldots, \varepsilon_n \overset{iid}{\sim} N(0, \sigma_e^2)$$

$$\text{We model this via } \quad y_i = \underbrace{b_0 + b_1 x_i}_{\hat{y}_i} + e_i$$

estimation...

Using this model, you will prove the sampling distr's of B_0, B_1 so you can make CI's, hyp tests (first $\frac{1}{2}$ of Stat 102). You will also prove the following:

$$\hat{Y} \sim N\left(b_0 + b_1 x, \, RMSE^2\right)$$

predictions are normally distributed about the center which is the "best guess"

Let's see an example of this in action

LV

More bang for the buck? Is price related to horsepower?

$b_0 = -4797.65, \quad b_1 = 172.17$

$RMSE = \$6687.927$

best guess car with 100 HP 's price?

$\hat{y} = -4797.65 + 172.17 \cdot 100 =$