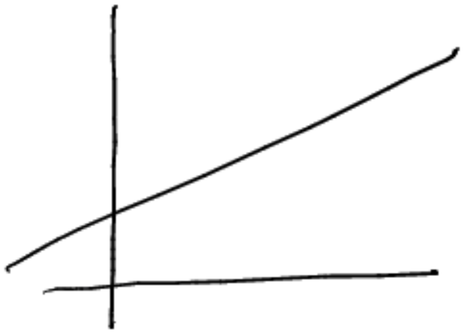
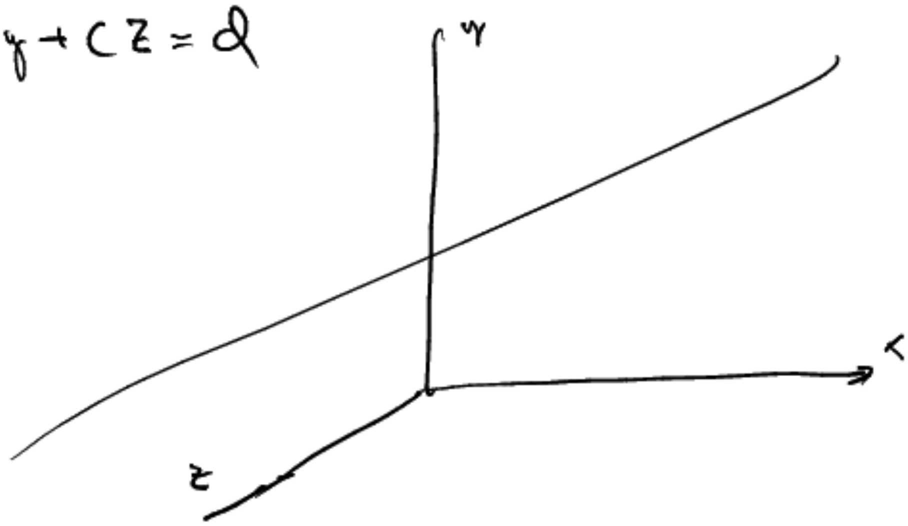
Regression estimate $\hat{\beta}_i$:

Model: $\underbrace{\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \dots + \hat{\beta}_n x_n}_{\text{Multiple linear = more than one x-value}}$

$y = ax + b$:



$ax + by + cz = d$



Goal is always to

1) Find estimates of slope and intercept
2) Assess quality of model:

    a) r and/or r^2

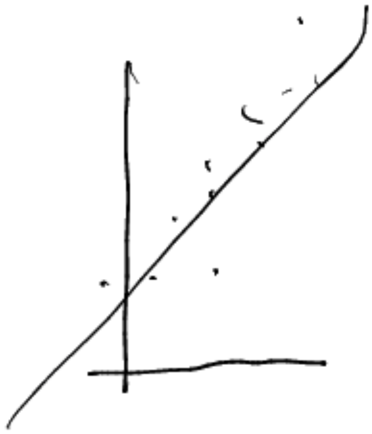
Correlation coefficient

Correlation of determination

r^2: Indicates strength of the linear relation between x and y. And is the proportion of explained $\leq r^2$

r : Measure of the linearity between x and y, so if r=0 no relationship: $-1 \leq r \leq 1$

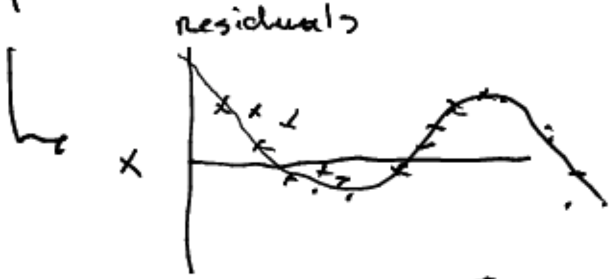
r = 0 → x and y are said to be independent = non-correlated.



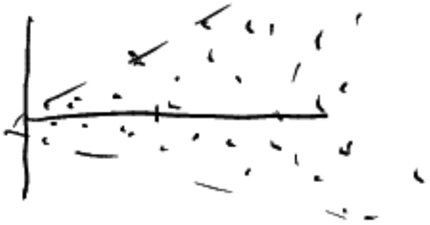
b) Check basic assumptions:

i) Residuals must be normally distributed

ii) Plot residuals against x-values and look for "pattern"



Residuals must have no pattern! →



↓

Pattern means systematic → non random.

c) Setup CI's for

1) Slope

2) Intercept

3) Prediction interval

Estimates:

$$\hat{\beta}_1 = \frac{\sum (y_i - \bar{y}) \cdot (x_i - \bar{x})}{\sum (x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}$$

$$= \boxed{\frac{\sum x \cdot y - n \cdot \bar{x} \cdot \bar{y}}{\sum x^2 - \frac{(\sum x_i)^2}{n}}} = \frac{n \sum xy - \sum x \cdot \sum y}{n \cdot \sum x^2 - (\sum x)^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta} \bar{x}$$

What is the standard error for slope $= S(\hat{\beta}_1)$

$$E_0 = t_{\alpha/2, \underline{n-2}} \cdot S(\hat{\beta}_1)$$

$$= t_{\alpha/2, n-2} \cdot \underbrace{\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}}_{S(\hat{\beta}_1)} \qquad \hat{\sigma}^2 = \frac{SS_E}{\boxed{n-2}}$$

because of $\bar{x}$ and $\bar{y}$