$p =$ $y =$ Data

$\text{delta}y\_i = y[i] - \text{Yhat}[i]$

```
for ( j in 1:(p+1)){

    W[j] = W[j] + delta_y_i[j] }
```

02/19 AKA

$p = 1$

$\mathcal{Y} = \mathbb{R}$, p covariants

$\mathcal{H} = \{ \vec{w} \cdot \vec{x} : \vec{w} \in \mathbb{R}^{p+1} \} = \{ w_0 + w_1 x : w_0 \, w_1 \in \mathbb{R} \}$

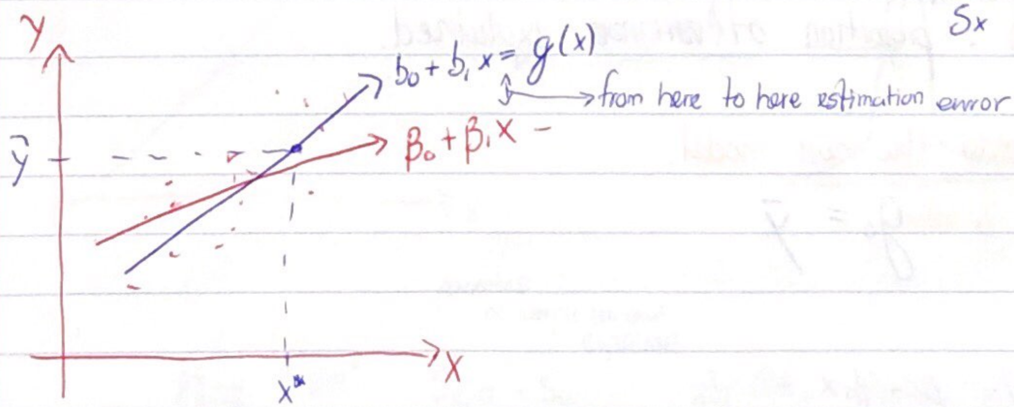
$A = \text{OLS}$ → algorithm minimize SSE

$\xrightarrow{A} g(x) = b_0 + b_1$

$b_0 = \bar{y} - r \frac{S_y}{S_x} \bar{x}$

$h^*(x) = \beta_0 + \beta_1 x$

$b_1 = r \frac{S_y}{S_x}$



4,7 3,1 3 $\frac{6+6}{4}$ $\frac{12}{4} = 3$

Model perform metric

how well does g predict?

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum (y_i - \hat{y})^2 \quad \text{(units y-square)}$$

$$MSE = \frac{1}{n-2} SSE \quad (\text{" " "})$$

mean square error

$$RMSE = \sqrt{MSE} \quad \text{(Units y)}$$

approximate interpretation

g is RMSE of from y on average.

$g(x) \pm 2 \cdot RMSE$ is a 95% confidence set for y. → EMPIRICAL RULE (Rule of thumb)

R^2: sample proportion of response variance explained.

Consider the null model.

$$g_0 = \bar{y}$$

$y = \beta_0 + \beta_1 x$

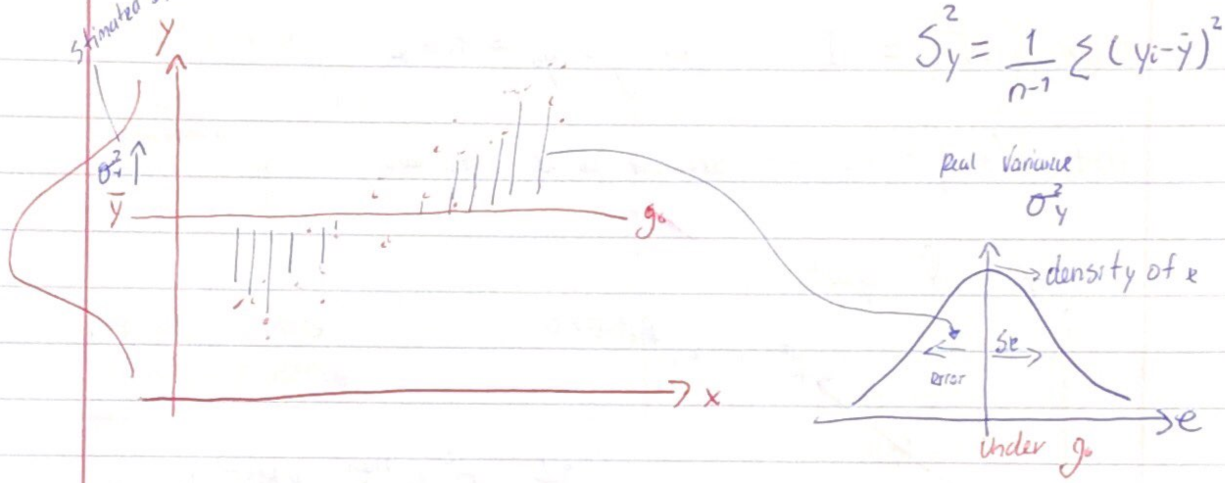
Null model

Sample variance

$$S_y^2 = \frac{1}{n-1} \sum (y_i - \bar{y})^2$$

Real Variance

σ_y^2

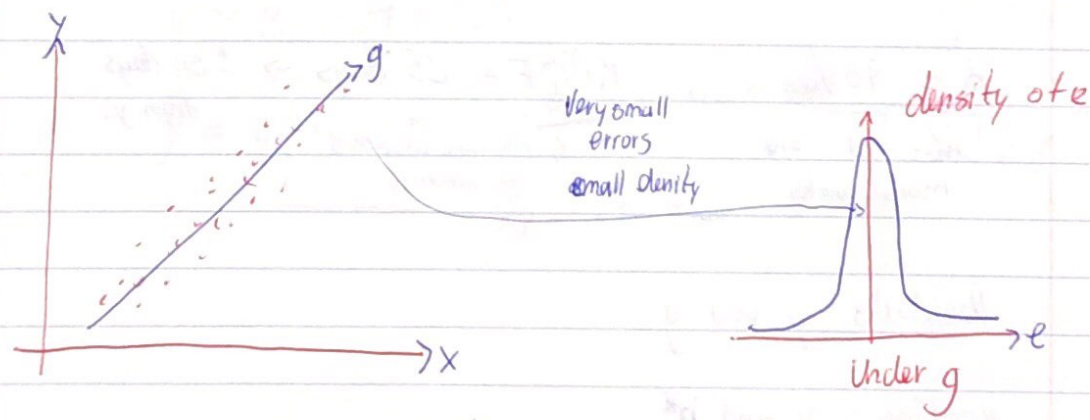


$$S_y^2 = \frac{1}{n-1} \underbrace{\sum (y_i - \bar{y})^2}$$

SSE_0 Sum of squares of g_0

$\shortparallel$

SST = sum squres total



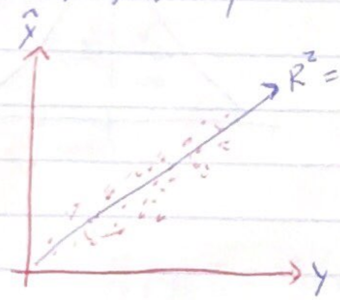
percentace of sample variance explained

$$R^2 := \frac{\Delta se^2}{S_{e,0}^2} = \frac{S_{e,0}^2 - S_e^2}{S_{e,0}^2} = \frac{\frac{1}{n-1} SSE_0 - \frac{1}{n-1} SSE}{\frac{1}{n-1} SSE_0} = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

$\max(R^2) = 1$ If $g = g_0 \Rightarrow R^2 = 0$

$\inf\{R^2\} = -\infty \rightarrow$ because it can be in any plane

y, $\hat{y}$ plot



if $R^2 \uparrow \overset{\text{then}}{\Rightarrow}$ RMSE $\downarrow$

$\rightarrow$ as you approach 100% ERROR go down.

---

y: # payments day late

$\underbrace{R^2 = 98\%}_{\text{this is how the model works}}$

$\underbrace{\text{RMSE} = 25 \text{ days}}$ $\rightarrow \pm 50$ days from y.

$\boxed{\text{we care about this the variance}}$

Residuals y and g

epsilons y and h^*

RMSE $\Rightarrow$ how far off you are from y

Rstudio

$y \sim x$

phenomena ← → explanation

formula

in pack stats function → lm

sigma => RMSE

$p=1$

$X_{raw} \in \{red, green\}$

coded: red → 0, green → 1

→ $x \in \{0, 1\}$ binary

$\mathbb{1}_{green=1}$

$y \sim x$ $\mathcal{A}$: OLS

$$\hat{y} = g(x_{raw}) = \begin{cases} \bar{y}_r & \text{if } x_{raw} = 0 \\ \bar{y}_g & \text{if } x_{raw} = 1 \end{cases} = \underbrace{\bar{y}_r}_{= b_0} + \underbrace{(\bar{y}_g - \bar{y}_r)}_{b_1} x$$

$= b_0 + b_1 x$