Lecture 7:

$P = 1$, $y = \mathbb{R}$,

X = binary

$H = \{ \vec{w} \cdot \vec{x} . \vec{w} \in \mathbb{R}^d \}$ linear modeling.

$h^*(x) = \beta_0 + \beta_1 x$, $y = h^*(x) + \varepsilon$ → error due to ignorance and misspecification

$g(x) = b_0 + b_1 x$

A : OLS $\Rightarrow b_0 = \bar{y} - r \frac{s_y}{s_x} \bar{x}$

$b_1 = r \frac{s_y}{s_x}$

$y = g(x) + e$ → residuals (all 3 errors)

How well does g Predict?

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - g(x_i))^2$$

Units?

y metric - Squared

Not so important.

$$MSE = \frac{1}{n-2} SSE$$

Units: Y-metric square
not so interpretable

mean squared error

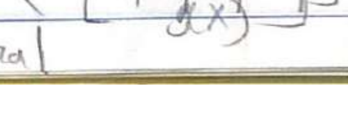
forget this...

$$RMSE = \sqrt{MSE}$$

units: Y, very interepretable.

Root mean squared error.

Imagine if e was a realization from a normal distribution. You can show,

$[g(x) \pm 2RMSE]$ 

$\approx 95\%$ Predictive Integral

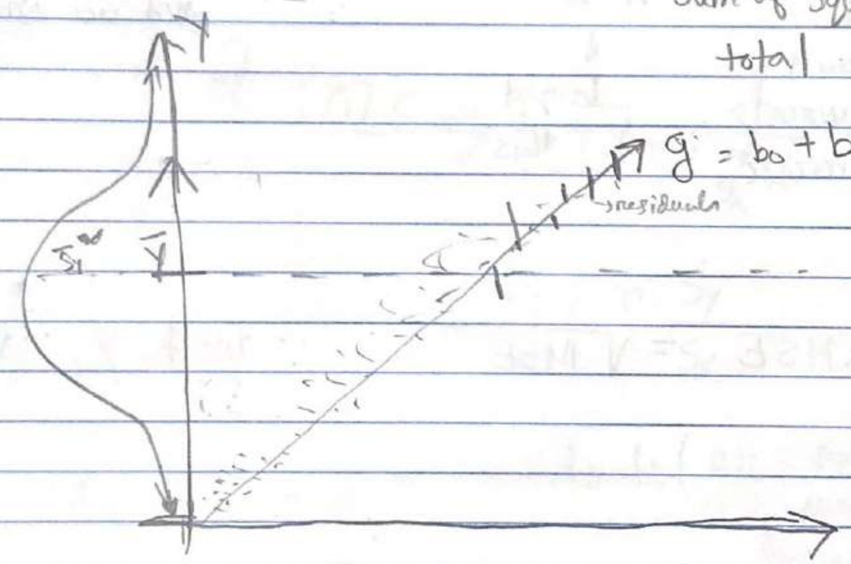
$[\hat{g}(x) \pm 2\,RMSE]$ $\longrightarrow$ How accurate your model is.

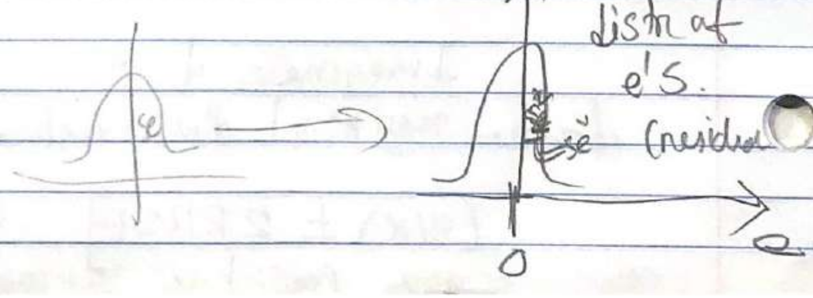
Consider the null model

$$g_0(X) = \bar{Y}$$

$$SSE_0 = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = SST = (n-1)\,s_Y^2$$

SST $\Downarrow$ Sum of Squared total





$$SSE = \sum e_i^2 = (n-1)\, S_e^2$$

$$\Delta S^2 = S_y^2 - S_e^2$$ "reduction in Variance"

"Variance explained"

$$R^2 = \frac{\Delta S^2}{S_y^2} = \frac{\text{amount of explained variance}}{\text{total variance}}$$

$$= \frac{S_y^2 - S_e^2}{S_y^2} \rightarrow \frac{Var[Y] - Var[\overset{\overset{\text{residuals}}{\uparrow}}{E}]}{Var[Y]}$$

$$= 1 - \frac{S_e^2}{S_y^2}$$

$$= \frac{(n-1) S_y^2 - (n-1) S_e^2}{(n-1) S_y^2}$$

$$= \frac{SST - SSE}{SST}$$

$$= 1 - \frac{SSE}{SST}$$

R^2. is the "Proportion of Variance explained"

Can $R^2 > 1$?

$\Downarrow$

$$1 - \frac{SSE}{SST} \leq 1.$$

Since $\Rightarrow$ $SSE \geqslant 0$, $SST \geqslant 0$

$$\Rightarrow \frac{SSE}{SST} \geqslant 0$$

where is, $R^2 = 1 \Rightarrow SSE = 0 \Rightarrow e_i = 0 \ \forall i \Rightarrow Y_i = \hat{Y}_i \ \forall i$

Can $R^2 = 0$?

$SSE = SST = g(x) = \bar{Y}$

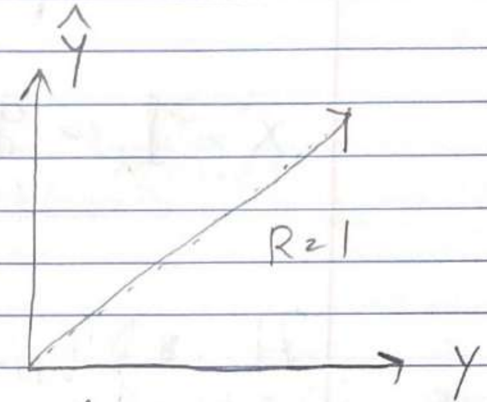
Can $R^2 < 0$?

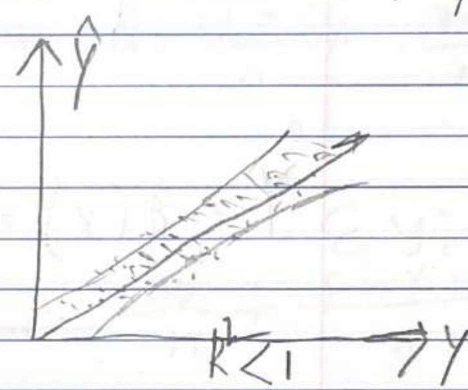
$SSE > SST$

$\Rightarrow g$ Predicts worse then $g_0 = \bar{Y}$

$R^2\uparrow \Leftrightarrow RMSE\downarrow$

SSE$\downarrow$ $\Leftrightarrow$ MSE$\downarrow$ $\Leftrightarrow$





R^2 vs RMSE

Who is more important?

$$X \in \mathcal{X} = \{\underset{0}{\text{red}}, \underset{1}{\text{green}}\}$$

$$X = \mathbb{1}_{X = \text{green}} \in \{0, 1\}$$

$$\mathcal{H} := \{w_0 + w_1 X : w_0, w_1 \in \mathbb{R}\}$$

$$\hat{Y} = g(X) = b_0 + b_1 X$$

$$g(X) = \begin{cases} \bar{Y}_{red} & \text{if } X = 0 \\ \bar{Y}_{green} & \text{if } X = 1 \end{cases} = \overbrace{\bar{Y}_{red}}^{b_0} + \overbrace{(\bar{Y}_{green} - \bar{Y}_{red})}^{b_1} X.$$

$$\Downarrow$$

$$.7 + (\cdot)X$$

Let's Prove

A: OLS
return this

$$\bar{X} = \frac{\sum x_i}{n} = \frac{\overset{\nearrow \text{\# green}}{n_g}}{n} = p \rightarrow \text{Prop of green}$$

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{\sum_{\{i: \text{green}\} = \{i: x_i = 1\}} Y_i + \sum_{\{i: \text{red}\} = \{i: x_i = 0\}} Y_i}{n}$$

$$= \frac{\sum_{\{i: x_i = 1\}} y_i}{n} \cdot \frac{n_g}{n_g} + \frac{\sum_{\{i: x_i = 0\}} y_i}{n} \cdot \frac{n - n_g}{n - n_g}$$

$$= \frac{\sum_{\{i: x_i = 1\}} y_i}{n_g} \frac{n_g}{n} + \frac{\sum_{\{i: x_i = 0\}} y_i}{n - n_g} \frac{n - n_g}{n}$$

$$= p\bar{Y}\text{ green} + (1-p)\,\bar{Y}\text{ red}$$

$$b_1 = \frac{\sum x_i Y_i - n\bar{X}\bar{Y}}{\sum x_i^2 - n\bar{X}^2} = \cancel{n_g \bar{Y}_g - np(p\,\bar{Y}_{\text{green}} + (1-p))}$$

$$= \frac{n_g \bar{Y}_g - np\bar{Y}}{n_g - np^2} \cdot \frac{\frac{1}{n}}{\frac{1}{n}}$$

$$\frac{p\overline{Y_g} - p\overline{Y}}{p - p^2} = \frac{\overline{Y_g} - \overline{Y}}{1-p}$$

$$= \frac{\overline{Y}_g - p\overline{Y}_g - (1-p)\overline{y}_r}{1-p}$$

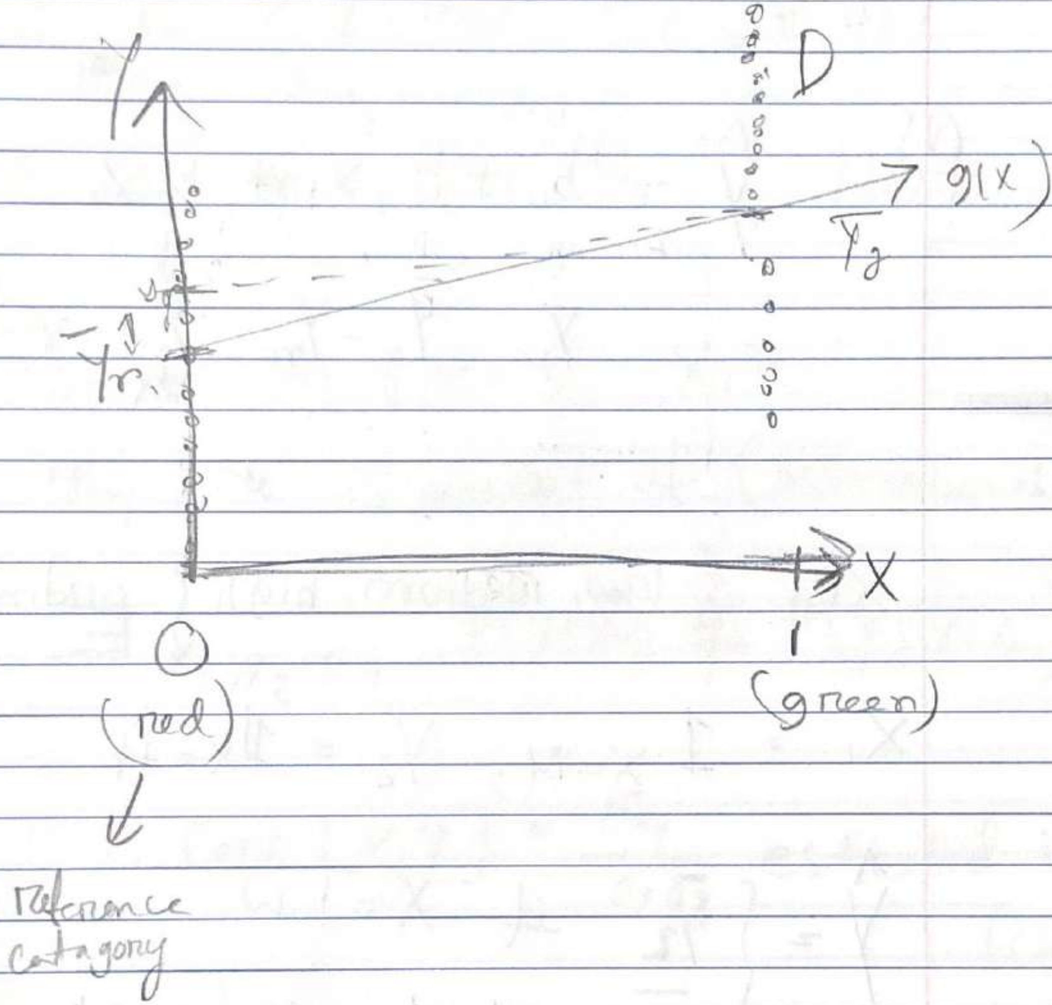
$$= \frac{(1-p)\,\overline{Y}_g - (1-p)\overline{y}_r}{1-p}$$

$$= \overline{Y}_g - \overline{Y}_r$$

$$b_0 = \overline{Y} - b_1\overline{X} = \left(p\overline{Y}_g + (1-p)\overline{Y}_r\right) - \left(\overline{Y}_g - \overline{y}_r\right)p$$

$$= \cancel{p\overline{Y}_g} + (1-p)\overline{Y}_r - \cancel{p\overline{Y}_g} + p\overline{Y}_r$$

$$= \bar{Y}_r - \cancel{p\bar{Y}_r} + \cancel{p\bar{Y}_r} = \bar{Y}_r$$



$l = 3$ e.g

$X \in \{\text{red, green, blue}\}$

$X_1 = \mathbb{1}_{X = \text{green}} \in \{0,1\}$

$X_2 = \mathbb{1}_{X = \text{blue}} \in \{0,1\}$

Q5) $\hat{y} = \overset{0.8}{b_0} + \overset{=0.3}{b_1} X_1 + \overset{=-0.4}{b_2} X_2$

$\Downarrow \quad \Downarrow \quad \Downarrow$

$\bar{Y}_r \quad \bar{Y}_g - \bar{Y}_r \quad \bar{Y}_b - \bar{Y}_r$

$X \in \{\text{low, medium, high}\}$ ordinal categories

$X_1 = \mathbb{1}_{X=M}, \; X_2 = \mathbb{1}_{X=H}$

$$\hat{y} = \begin{cases} \bar{Y}_L & \text{if } X = \text{low} \\ \bar{Y}_M & \text{if } X = \text{Medium} \\ \bar{Y}_H & \text{if } X = \text{High} \end{cases}$$

Can't bear the Pain

However, what if you wish to Constrain

$\hat{Y}(\text{low}) < \hat{Y}(\text{med}) < \hat{Y}(\text{High})$?

Does A: OLS give you this? NO!

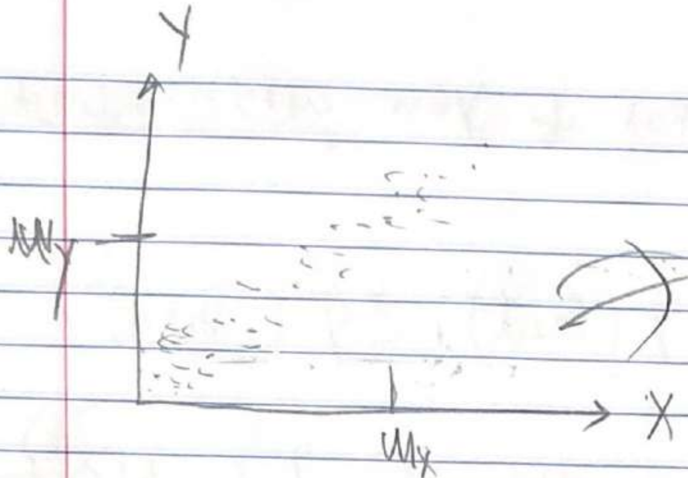
Consider two r.v's X, Y

They are dependent if (association)

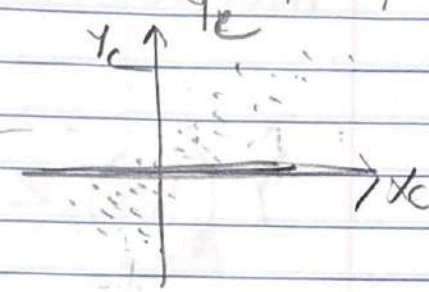
$\exists\, x_1, x_2$ s.t. $P(Y \mid X = x_1) \neq P(Y \mid X = x_2)$

$\rho := \text{Corr}[X, Y] = \dfrac{\sigma_{XY}}{\sigma_X \sigma_Y}$ estimated by r. (Correlation

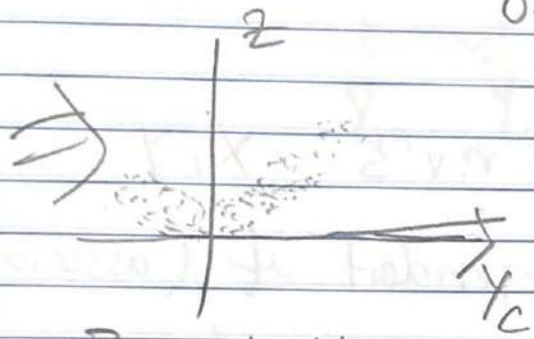
$\sigma_{XY} := \text{Cov}[X, Y] = E[(X - \mu_X)(Y - \mu_Y)]$ estimated by S_{XY}



let $X_c = X - \mu_X$
$Y_c = Y - \mu_Y$



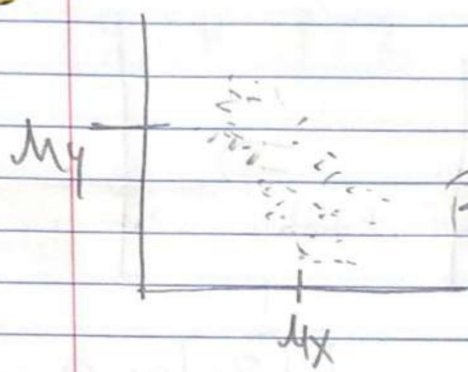
$$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)]$$
$$= E[X_c Y_c] = E[Z]$$
$$E[Z] > 0$$

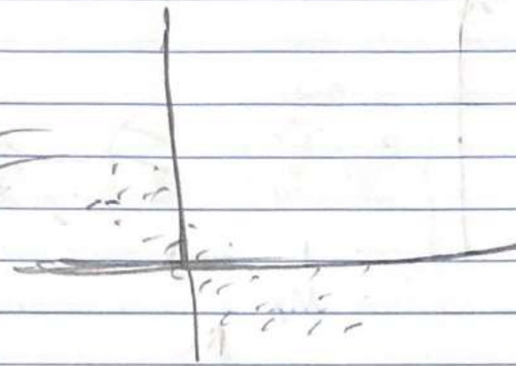


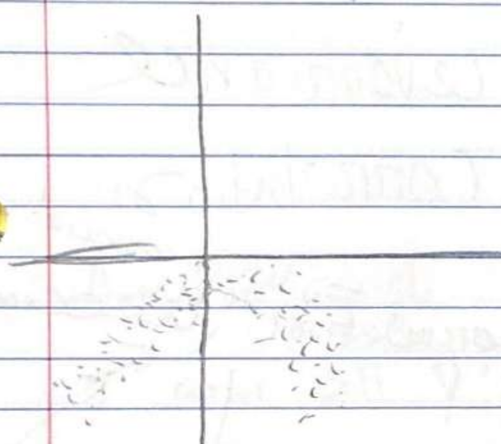
$$Z = X_c Y_c$$

C — centered

expectation
0

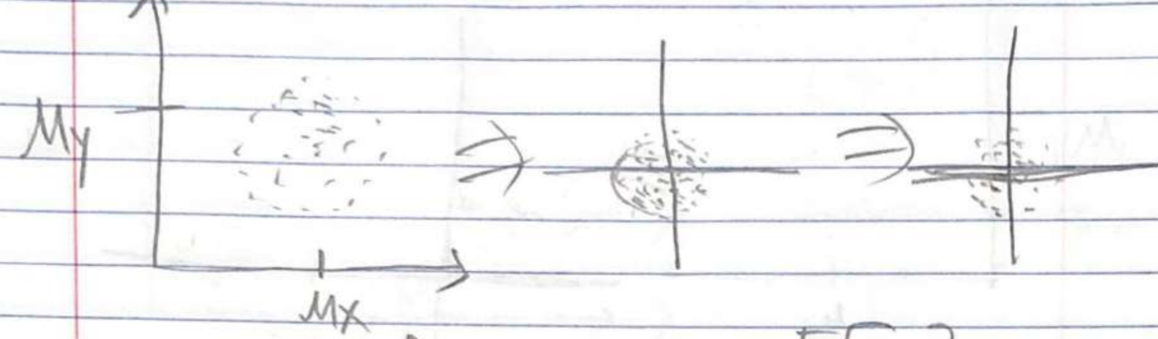






$E(z) < 0$

HW: $r^2 = R^2$ Pon $P = 1$



$E[Z] \approx 0$

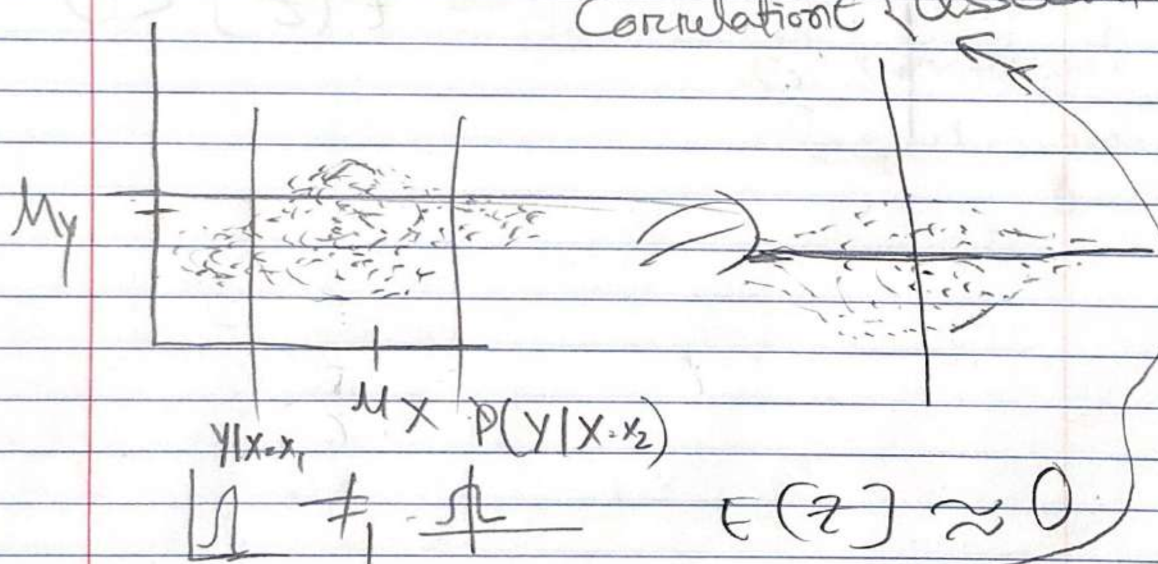
No Covarriance

about 0 Covarriance

" 0 Correlation

⇒ "linear association"

Correlation $\in$ {association}



$E[Z] \approx 0$

⇒ Correlation ≈ 0

but X, Y

dependent or "associated"!