Math 3904
lecture 7 ; 12/18

$p=1$; $y = R$

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

$g(x) = b_0 + b_1 x$

A: OLS $\Rightarrow b_0 = \bar{y} - r\frac{s_y}{s_x}\bar{x}$; $b_1 = r\frac{s_y}{s_x}$

$y = g(x) + e$ ← (e = all error types)

How well does g predict?

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

y minus g squared

$MSE = \frac{1}{n-2} SSE$ — mean squared error

$RMSE = \sqrt{MSE}$ — root mean squared error

e is pulled from a normal distribution

$g(x) \pm RMSE$

95% prediction model

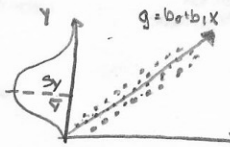
*how accurate your model is

consider the null model

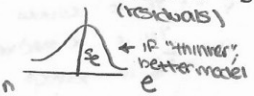
$g_0(x) = \bar{y}$

$SSE_0 = \sum_{i=1}^{n} (y_i - \bar{y})^2 = SST = (n-1)s_y^2$

sum of squares total



← model must beat SST



distr. of e's (residuals)

← IF "thinner", better model

$SSE = \sum_i e_i^2 = (n-1)s_e^2$

$\Delta s^2 = s_y^2 - s_e^2$

← reduction in variance / variance explained

$R^2 = \frac{\Delta s^2}{s_y^2} = \frac{s_y^2 - s_e^2}{s_y^2} = 1 - \frac{s_e^2}{s_y^2} = \frac{(n-1)s_y^2}{(n-1)} : \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST}$

*OLS will always be non-negative

- IF $R^2 \uparrow \Leftrightarrow SSE \downarrow \Leftrightarrow MSE \downarrow \Leftrightarrow RMSE \downarrow$

IF $RMSE = 0 \Leftrightarrow SSE = 0 \Leftrightarrow MSE = 0 \Leftrightarrow R^2 = 1$

← if $R^2 = 1$; RMSE is large

- R^2 vs RMSE

which is more important?

RMSE because it shows how off your model is in units which is more useful than % (usually).

⊖ = RMSE

→ can $R^2 > 1$?

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

since $SSE \& SST \geq 0$

$\frac{SSE}{SST} \geq 0$

$R^2 = 1$ only when $SSE = 0$

→ can $R^2 = 0$

$SSE = SST$

$g = \bar{y}$

→ can $R^2 < 0$

$SSE > SST$

bad model

g_0 predicts worst than $\bar{y}$

$g_0 > \bar{y}$

$p = 1$

↓

x_1 = Binary (0 / 1)

$x \in \mathcal{X} = \{red, green\}$

$x = 1_x = green$ $\Rightarrow$ $g(x) \begin{cases} \bar{y}_{red} \text{ if } x = 0 \\ \bar{y}_{green} \text{ if } x = 1 \end{cases}$

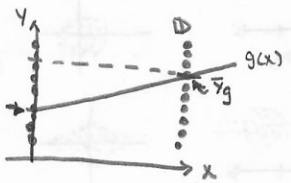
$\mathcal{H} = \{w_0 + w_1 x : w_0, w_1 \in \mathbb{R}\}$

$\hat{y} = g(x) = b_0 + b_1 x$

A: OLS returns ↓

$= \bar{y}_{red} + (\bar{y}_{green} - \bar{y}_{red})x$

b_0 ; b_1



$g(x) = 0.7 + 0.2x$

$\bar{y}_{red} - \bar{y}_{green}$

$\bar{x} = \frac{\sum x_i}{n} = \frac{n_{green}}{n} =$ proportion of green $\rightarrow p$

$\bar{y} = \frac{\sum y_i}{n} = \frac{\sum_{i=g} y_i + \sum_{i=red} y_i}{n} = \sum_{x_i=1} \frac{y_i}{n} \frac{n_g}{n_g} + \sum_{x=0} \frac{y_i}{n} \cdot \frac{n-n_g}{n-n_g} = \frac{\sum y_i}{n_g} \frac{n_g}{n} + \frac{\sum y_i}{n-n_g} \frac{n-n_g}{n} =$

$= p\bar{y}_{green} + (1-p)\bar{y}_{red}$

$b_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{n_g \bar{y}_g - np\bar{y}}{n_g - np^2} \cdot \frac{\frac{1}{n}}{\frac{1}{n}} = \frac{p\bar{y}_g - p\bar{y}}{p - p^2} = \frac{\bar{y}_g - \bar{y}}{1-p} = \frac{\bar{y}_g - p\bar{y}_g - (1-p)\bar{y}_r}{1-p}$

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

$b_0 = \bar{y} - b_1\bar{x} = (p\bar{y}_g - (1-p)\bar{y}_r) - (\bar{y}_g - \bar{y}_r)p + (1-p)\bar{y}_r + p\bar{y}_r = \bar{y}_r$

L = 3

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

$x_1 = \mathbb{1}_{X=\text{green}} \in \{0,1\}$

$x_2 = \mathbb{1}_{X=\text{blue}} \in \{0,1\}$

$\hat{y} = b_0 + b_1 x_1 + b_2 x_2$

$= \bar{y}_{red} + (\bar{y}_g - \bar{y}_r) x_1 + (\bar{y}_g - \bar{y}_b) x_2$

$\quad -.8 \qquad .3 \qquad -.4$

$X \in \{\text{low, med, high}\}$ ordinal

$x_L = \mathbb{1}_{X=m}$

$y_L = \mathbb{1}_{X=H}$

$\hat{y} = \begin{cases} \bar{y}_L & \text{if } X=\text{low} \\ \bar{y}_M & \text{if } X=\text{medium} \\ \bar{y}_H & \text{if } X=\text{high} \end{cases}$

what if you want to constrain

$\hat{y}(\text{low}) < \hat{y}(\text{medium}) < \hat{y}(\text{high})$

A: OLS will not give you ♠

we want monotonicity

consider the r.v's X, Y

they are dependent if (associated if)

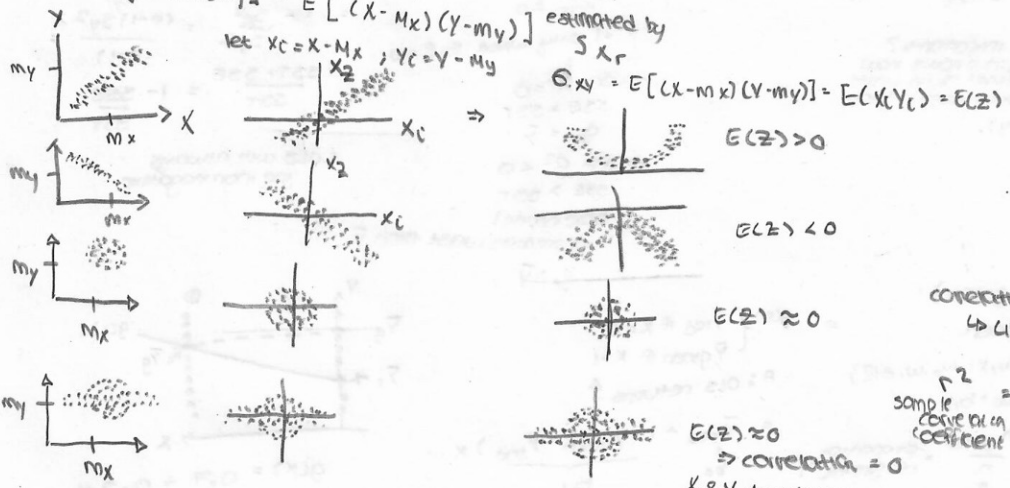
$\exists x_1, x_0$ s.t. $P(Y|X=x_1) \neq P(Y|X=x_0)$

$\rho_{xy} = \text{cor}[X,Y] = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$ estimated by r

$\sigma_{xy} = \text{cov}[X,Y] = E[(X-\mu_x)(Y-\mu_y)]$ estimated by S_{xy}

let $x_c = X - \mu_x$; $y_c = Y - \mu_y$

$\sigma_{xy} = E[(X-\mu_x)(Y-\mu_y)] = E(X_c Y_c) = E(Z)$



correlation ∈ "association"

↳ linear associations

r^2 = sample coefficient of determination $= R^2$ for $p=1$