Matrix Derivatives: Week 1

$f: \mathbb{R}^p \to \mathbb{R} \qquad f(x) = a^T x$

$\nabla f(x) = a$

$g: \mathbb{R}^p \to \mathbb{R} \qquad g(x) = x^T A x$, where $A_{p \times p}$ is symmetric

$\nabla g(x) = 2Ax$

$g''(x) = 2A$

Ex:

$$z = \| y_{n\times 1} - X_{n\times p}\, \beta_{p\times 1} \|^2 = y^T y - 2y^T X\beta + \beta^T X^T X \beta$$

$$\frac{\partial z}{\partial \beta} = -2X^T y + 2X^T X \beta = 0$$

$$\Rightarrow \hat{\beta} = (X^T X)^{-1} X^T y$$

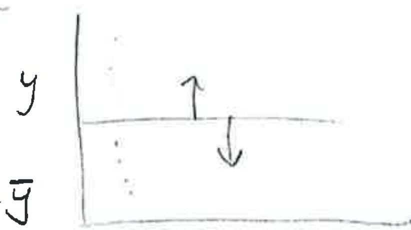
$$\frac{\partial^2 z}{\partial \beta \partial \beta^T} = 2X^T X$$

$$a^T X^T X a = \| Xa \|^2 \geq 0 \quad (\text{positive definite})$$

$$\Rightarrow \hat{\beta} = (X^T X)^{-1} X^T y \text{ is a minimum}$$

Centering by matrix multiplication

$$\underset{n\times 1}{J_n} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$



$$\min \| y - J_n \beta \|^2 \Rightarrow \hat{\beta} = (J_n^T J_n)^{-1} J_n^T y = \bar{y}$$

$$y - J_n \bar{y} = y - J_n (J_n^T J_n)^{-1} J_n^T y = (I - J_n (J_n^T J_n)^{-1} J_n^T) y$$

$$\frac{1}{n} J_n^T (I - J_n (J_n^T J_n)^{-1} J_n^T) y = \frac{1}{n} (J_n^T - \cancel{J_n^T J_n} (J_n^T J_n)^{-1} J_n^T) y$$

gives average

$$= \frac{1}{n} (J_n^T - J_n^T) y = 0$$

premultiply any vector by $(I - J_n(J_n^T J_n)^{-1} J_n^T)$ mean centers it.

$(I - J_n(J_n^T J_n)^{-1} J_n^T)\, x_{n\times p}$ mean centers the columns of x

$x(I - J_p(J_p^T J_p)^{-1} J_p^T)$ mean centers the rows of x

$$I - J_n(\underbrace{J_n^T J_n}_{n})^{-1} J_n^T = I - \frac{1}{n} J_{n\times n}$$

Variance via matrix multiplication

$$s^2 = \frac{\Sigma(y_i - \bar{y})^2}{n-1} = \frac{\|y - \bar{y} J_n\|^2}{n-1} = \frac{1}{n-1} \tilde{y}^T \tilde{y}$$

$$\tilde{y} = (I - J_n(J_n^T J_n)^{-1} J_n^T) y$$

$$\tilde{y}^T \tilde{y} = y^T \underbrace{(I - J_n(J_n^T J_n)^{-1} J_n^T)}_{\text{Symmetric}} (I - J_n(J_n^T J_n)^{-1} J_n^T) y$$

$$= y^T \underbrace{(I - J_n(J_n^T J_n)^{-1} J_n^T)}_{\text{idempotent}} y \quad = \text{sum of squared deviations around the mean}$$

$X_{n\times p}$

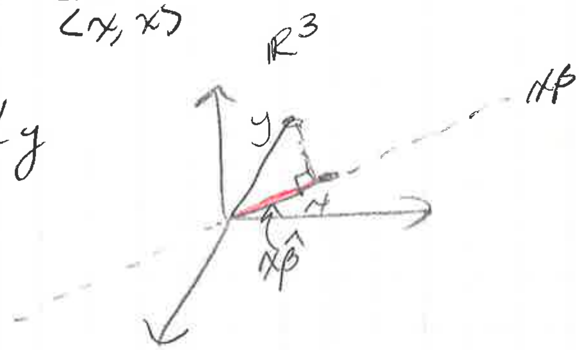
$$X^T(I - \underbrace{J_n(J_n^T J_n)^{-1} J_n^T}_{H})X = \underbrace{X^T(I-H)^T}_{\tilde{X}^T} \underbrace{(I-H)X}_{\tilde{X}} = X^T(I-H)(I-H)X$$

$$= \tilde{X}^T \tilde{X} \Rightarrow \frac{1}{n-1} \tilde{X}^T \tilde{X} = Cov(X)$$

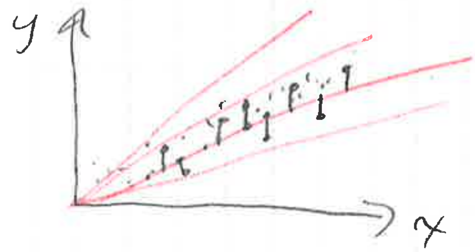
Regression through the origin

$$\min \| \underset{n\times 1}{y} - \underset{n\times 1}{x}\beta \|^2 \Rightarrow \hat{\beta} = \frac{\langle y, x\rangle}{\langle x, x\rangle}$$

$x\hat{\beta}$ is the projection of y onto $x\beta$

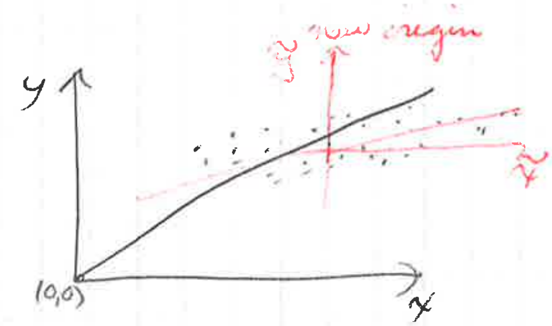


When $X = J_n \Rightarrow \langle y, x\rangle = n\bar{y}$ and $\langle x, x\rangle = n$
thus $\hat{\beta} = \bar{y}$



Centering First

$$\min_{\beta} \| \underset{n\times 1}{y} - \underset{n\times 1}{x}\beta \|^2 \Leftrightarrow \hat{\beta} = \frac{\langle y, x\rangle}{\langle x, x\rangle}$$



$$\left.\begin{aligned} \tilde{y} &= (I - \overbrace{J_n(J_n^T J_n)^{-1} J_n^T}^{H})y \\ \tilde{x} &= (I - J_n(J_n^T J_n)^{-1} J_n^T)x \end{aligned}\right\} \text{centered}$$

$$\min_{\gamma} \|\tilde{y} - \tilde{x}\gamma\|^2 \Leftrightarrow \hat{\gamma} = \frac{\langle \tilde{y}, \tilde{x}\rangle}{\langle \tilde{x}, \tilde{x}\rangle} = \frac{y^T(I-H)^T(I-H)x}{x^T(I-H)^T(I-H)x} = \frac{(n-1)\,\mathrm{cov}(x,y)}{(n-1)\,\mathrm{var}(x)}$$

$$\hat{\gamma} = \frac{(n-1)\hat{\rho}_{xy}\hat{\sigma}_x\hat{\sigma}_y}{(n-1)\hat{\sigma}_x^2} = \boxed{\hat{\rho}_{xy}\frac{\hat{\sigma}_y}{\hat{\sigma}_x}}$$

$\hat{\rho}_{xy}\frac{\hat{\sigma}_y}{\hat{\sigma}_x}$ has units $\frac{y}{x}$

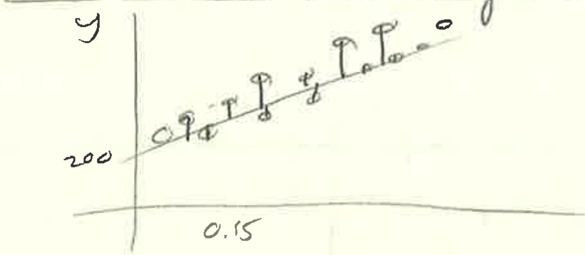
If we switch x and y:

$$\min_{\gamma} \|\tilde{x} - \tilde{y}\gamma\|^2 \Leftrightarrow \hat{\gamma} = \hat{\rho}_{xy}\frac{\hat{\sigma}_x}{\hat{\sigma}_y}$$

If we also set $\hat\sigma_y = \hat\sigma_x = 1$ by scaling (and centering)

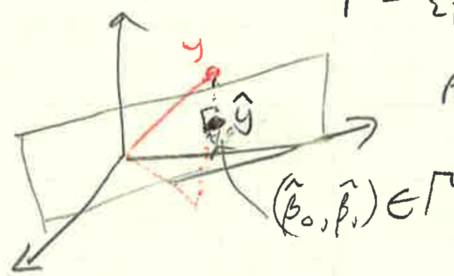
$\Rightarrow \hat\gamma = \hat\rho_{xy}$

Connection with linear regression



$$\text{minimize} \| y - \{\beta_0 J_n - \beta_1 x\} \|^2 \qquad J_n = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$\Gamma = \left\{ \beta_0 J_n + \beta_1 x \,\middle|\, \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \in \mathbb{R}^2 \right\}$$



projection in n-dim space

$$\|\underbrace{y - \beta_1 x}_{\text{fix } \beta_1} - \beta_0 J_n\|^2 \Rightarrow \min_{\beta_0} \|y - \beta_1 x - \beta_0 J_n\|^2 := Z$$

$$\hat\beta_0(\beta_1) = \frac{1}{n} J_n^T (y - \beta_1 x) = \bar y - \beta_1 \bar x$$

$$\Rightarrow Z \geq \|y - \beta_1 x - (\bar y - \beta_1 \bar x) J_n\|^2$$

$$= \|y - \bar y J_n - \beta_1 (x - \bar x J_n)\|^2$$

$$= \|\tilde y - \beta_1 \tilde x\|^2 \quad , \tilde y = y - \bar y J_n \ , \ \tilde x = x - \bar x J_n$$

$$\boxed{\hat\beta_1 = \frac{\langle \tilde y, \tilde x \rangle}{\langle \tilde x, \tilde x \rangle} = \hat\rho_{xy} \frac{\hat\sigma_y}{\hat\sigma_x}}$$

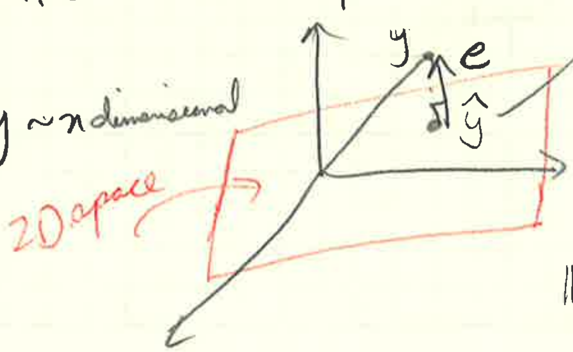
$$\Rightarrow \boxed{\hat\beta_0 = \bar y - \hat\beta_1 \bar x}$$

Fitted values and residuals

$\|y - \{\beta_0 J_n + \beta_1 x\}\|^2$ is to be minimized

projection of y on 2D space spanned by (β_0, β_1)

y ~ n dimensional



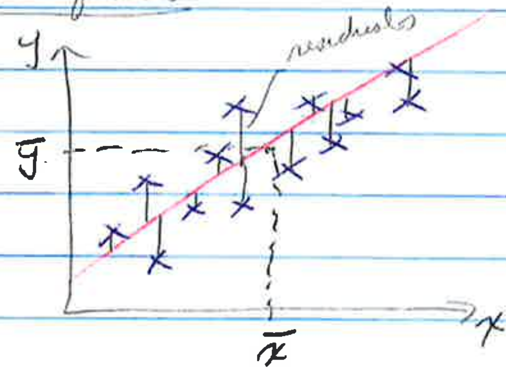
$$\hat y_0 = \hat\beta_0 J_n + \hat\beta_1 x$$

$$e = y - \hat y \quad \text{(residuals)}$$

$$\|y - \hat y\|^2 = \|e\|^2$$

TOPS 35500

Least Squares



$y_{n\times1} + x_{n\times1}$

$Z := \|y - \beta_1 x - \beta_0 J_n\|^2$

Find $\min_\beta Z$

$\Rightarrow$ Fix β_1 $\Rightarrow$ treat $y - \beta_1 x$ as a single vector

$\Rightarrow$ $\beta_0(\beta_1) = \frac{1}{n}(y - \beta_1 x)^T J_n = \bar{y} - \beta_1 \bar{x}$

$\Rightarrow Z \geq \|y - \bar{y} J_n - \beta_1(x - \bar{x} J_n)\|^2$

$= \|\tilde{y} - \beta \tilde{x}\|^2$

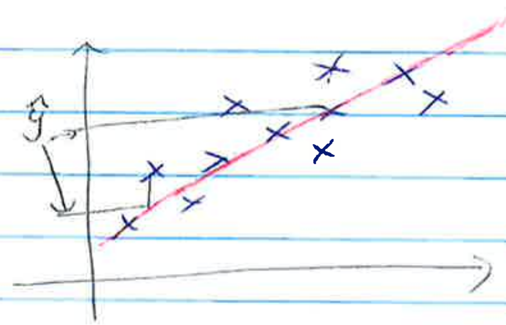
$\Rightarrow \hat{\beta}_1 = \frac{\langle \tilde{y}, \tilde{x} \rangle}{\langle \tilde{x}, \tilde{x} \rangle} = \hat{\rho}_{xy} \frac{\hat{\sigma}_y}{\hat{\sigma}_x}$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ Thus a linear fit line always passes through $(\bar{x}, \bar{y})$

Prediction

$Z := \|y - \beta_0 J - \beta_1 x\|^2$

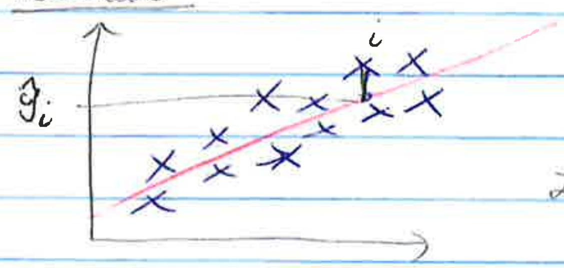
$\Rightarrow \hat{y}_{pred} = \hat{\beta}_0 + \hat{\beta}_1 x_{new}$



$\hat{\beta}_0 J_n + \hat{\beta}_1 x = \hat{y}$

Least squares minimizes $\|y - \hat{y}\|^2$

Residuals

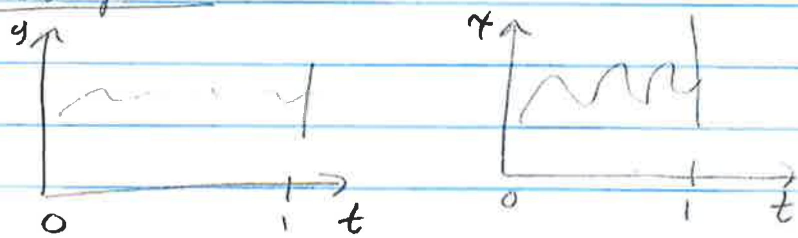


$e := y - \hat{y} = y - \hat\beta_0 J_n - \hat\beta_1 x$ (residuals)

Least squares $\Rightarrow \min \|e\|^2$

plot residuals vs. x with horizontal line at $e=0$

Generalizations



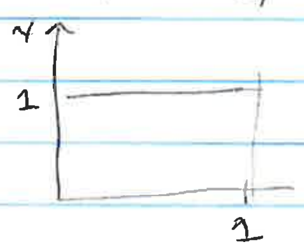
$\mathcal{L}^2[0,1]$

$$\langle f,g\rangle = \int_0^1 f(t)g(t)\,dt$$

$z := \|y-\beta x\|^2$, show $\hat\beta = \frac{\langle y,x\rangle}{\langle x,x\rangle}$

$$\begin{aligned} z &= \|y-\hat\beta x+\hat\beta x-\beta x\|^2 \\ &= \|y-\hat\beta x\|^2 + \underbrace{\cancel{2\langle y-\hat\beta x, \hat\beta x-\beta x\rangle}}_{0} + \overbrace{\|\hat\beta x-\beta x\|^2}^{\text{positive}} \\ &\geq \|y-\hat\beta x\|^2 \end{aligned}$$

Let $x=1$ $\Rightarrow$ $\hat\beta = \frac{\langle y,x\rangle}{\langle x,x\rangle} = \int_0^1 y(t)\,dt = \bar{y}$



$\tilde{y}(t) := y(t) - \bar{y}J(t)$ $\quad J(t)=1,\ t\in[0,1]$

$\tilde{x}(t) := x(t) - \bar{x}J(t)$

$$Cov(y,x) = \int_0^1 (y(t) - \bar{y}J(t))(x(t) - \bar{x}J(t))\,dt$$

$var(y) := cov(y,y)$

$\min z := \| y - \beta_0 J - \beta_1 x \|^2$

$\beta_0(\beta_1) = \bar{y} - \beta_1 \bar{x}$

$\Rightarrow \hat{\beta}_1 = Cor(y,x) \dfrac{sd(y)}{sd(x)}$ and $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

Example: $y(t) = t + 2t^2$, $x(t) = t$

$\min \| y - \beta x \|^2 \Rightarrow \hat{\beta} = \dfrac{\langle y,x \rangle}{\langle x,x \rangle}$

$$\hat{\beta} = \frac{\int_0^1 (t+2t^2)t\,dt}{\int_0^1 t^2\,dt} = \frac{\frac{1}{3} + \frac{1}{2}}{\frac{1}{3}} = \frac{\frac{2}{6} + \frac{3}{6}}{\frac{2}{6}} = \frac{5}{2} = 2.5$$

In R, call lm(y~x-1) to get 2.5

Least Squares

$$y_{n\times 1} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \qquad X_{n\times p} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} = [x_1 \cdots x_p]$$

X is of full column rank, $p \le n$

$$z := \|y - X\beta\|^2 = y^T y - 2y^T X\beta + \beta^T X^T X \beta$$

$$\frac{\partial z}{\partial \beta} = -2X^T y + 2X^T X\beta = 0$$

$$\Rightarrow \underbrace{X^T X}_{p\times p}\beta = X^T y \qquad \text{Normal equations}$$

$$\hat\beta = (X^T X)^{-1} X^T y$$

$$\frac{\partial^2 z}{\partial\beta\,\partial\beta^T} = 2X^T X \ge 0 \quad \text{since } X^T X \text{ is positive definite}$$

$\Rightarrow \hat\beta$ minimizes z

In R beta_hat = solve(t(x) %*% x) %*% t(x) %*% y

or solve(t(x) %*% x, t(x) %*% y)

Second Derivation of Least Squares

$(X^T X)^{-1} X^T y = \hat\beta$, $y_{n\times 1}$ $X_{n\times p}$ (full rank)

$$z := \|y - X\beta\|^2$$

Let $H_X = X(X^T X)^{-1} X^T$

$H_X^T = H_X$, $H_X H_X = H_X$

$(I - H_X)$ is symmetric and idempotent

$(I - H_X) Xa = 0$

$$[I - X(X^T X)^{-1} X^T] Xa = Xa - XIa = 0$$

$$z = \|y - X\hat\beta + X\hat\beta - X\beta\|^2$$

$$= \|y - X\hat\beta\|^2 + 2(y - X\hat\beta)^T (X\hat\beta - X\beta) + \|X\hat\beta - X\beta\|^2$$

$$\ge \|y - X\hat\beta\|^2 + 2(y - X\hat\beta)^T (X\hat\beta - X\beta)$$

$$= \|y - X\hat\beta\|^2 + 2(y - X(X^T X)^{-1} X^T y)^T (X\hat\beta - X\beta)$$

$$= \|y - X\hat\beta\|^2 + 2y(I - H_X)^T X(\hat\beta - \beta)$$

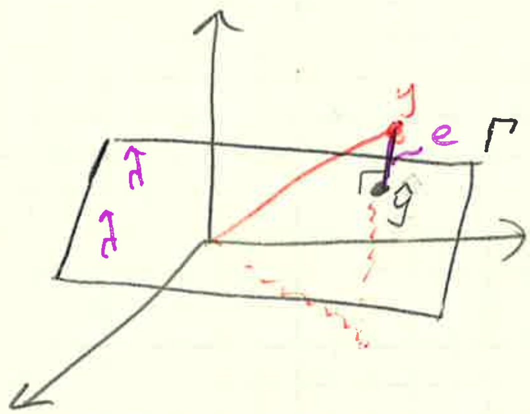
$$= \|y - X\hat\beta\|^2 + 2y\underbrace{(I - H_X)X}_{0}(\hat\beta - \beta)$$

$$= \|y - X\hat\beta\|^2 \quad \therefore \ \hat\beta \text{ is the minimizer of } z$$

Projections

$\underset{3\times 1}{y}$, $\underset{3\times 2}{X}$ $\quad * := \|y - X\beta\|^2$, $\Gamma = \{X\beta \mid \beta \in \mathbb{R}^2\}$

$$\min_{z \in \Gamma} \|y - z\|^2$$



$\hat y \in \Gamma$ minimizes $*$

Suppose $\tilde X = [X_1 \ X_2 \ X_1 + X_2]$

$\tilde\Gamma = \{\tilde X\beta \mid \beta \in \mathbb{R}^3\} = \Gamma$

Since $\tilde X$ has rank 2

Just need LI subset to do regression

$\hat y$ is the projection of y onto Γ

$$\hat y = X(X^TX)^{-1}X^Ty = H_X y$$

$z \longmapsto H_X z$ $\quad H_X$ is a projection operator from $\mathbb{R}^N$ to Γ

$$e = y - \hat y = y - X(X^TX)^{-1}X^Ty$$
$$= (I - H_X)y$$

$e^Tz = 0$ for $z = X\gamma$ for some γ

$$e^Tz = y^T\underbrace{(I - H_X)X}_{0}\gamma = 0$$

Suppose $X = [J_n \cdots] \Rightarrow e^TJ_n = 0 = \sum e_i$

and $e^TX_{*k} = 0$ for $X_{*k} = k^{th}$ column of X

Third Derivation of Least Squares

$$* := \|y - X_1\beta_1 - \cdots - X_p\beta_p\| \qquad e(a,b) = a - b\frac{\langle a,b\rangle}{\langle b,b\rangle}$$

fix $\beta_2, \ldots, \beta_p$

$$* = \|\underbrace{y - X_2\beta_2 \cdots X_p\beta_p}_{\text{Single outcome}} - X_1\beta_1\|^2$$

$$\beta_1(\beta_2, \ldots, \beta_p) = \frac{\langle y - x_2\beta_2 - \cdots - x_p\beta_p, x_1\rangle}{\langle x_1, x_1\rangle}$$

$$= \frac{\langle y, x_1\rangle}{\langle x_1, x_1\rangle} - \frac{\langle x_2, x_1\rangle}{\langle x_1, x_1\rangle}\beta_2 - \cdots - \frac{\langle x_p, x_1\rangle}{\langle x_1, x_1\rangle}\beta_p$$

$$* \geq \|e(y, x_1) - e(x_2, x_1)\beta_2 - \cdots - e(x_p, x_1)\beta_p\|^2$$

fix $\beta_3, \ldots, \beta_p$

$$\Rightarrow * \geq \Big\| e(e(y,x_1), e(x_2,x_1)) - e(e(x_3,x_1), e(x_2,x_1))\beta_3 - \cdots - e(e(x_p,x_1), e(x_2,x_1))\beta_p \Big\|^2$$

$$* \ \|y - \underset{n\times p_1}{X_1}\beta_1 - \underset{n\times p_2}{X_2}\beta_2\|^2 \qquad , \ X = [X_1 \ \ X_2]$$

hold β_1 fixed $\Rightarrow \hat\beta_2(\beta_1) = (X_2^TX_2)^{-1}X_2^T(y - X_1\beta_1)$

$$= (X_2^TX_2)^{-1}X_2^Ty - (X_2^TX_2)^{-1}X_2^TX_1\beta_1$$

$$* \geq \|(I - X_2^T(X_2^TX_2)^{-1}X_2^T)y - (I - X_2^T(X_2^TX_2)^{-1}X_2^T)X_1\beta_1\|^2$$

In R

```
ey = y - x1 %*% solve(t(x1) %*% x1, t(x1) %*% y)
ex2 = x2 - x1 %*% solve(t(x1) %*% x1) %*% t(x1) %*% x2
solve(t(ex2) %*% ex2, t(ex2) %*% ey)
```

TOPS 35500

Basic Examples of Design Matrices and Fits

$$\min_{\beta_0} \|y - J_n \beta_0\|^2 \Rightarrow \hat\beta_0 = \bar{y}$$

$$\hat\beta_0 = (J_n^T J_n)^T J_n^T y = \frac{1}{n} \sum y_i = \bar{y}$$

$$\min_{\beta} \|y - X\beta\|^2 \Rightarrow \hat\beta = \frac{\langle x, y\rangle}{\langle x, x\rangle}$$

$$\hat\beta = (X^T X)^{-1} X^T y = \frac{\langle x, y\rangle}{\langle x, x\rangle}$$

$$\min_{\beta} \|y - \beta_0 J_n - \beta_1 X\|^2 \Leftrightarrow \min_{\beta} \|y - W\beta\|^2 \quad \text{where } W = \begin{bmatrix} 1 & \\ \vdots & X \\ 1 & \end{bmatrix} = [J_n\ X]$$

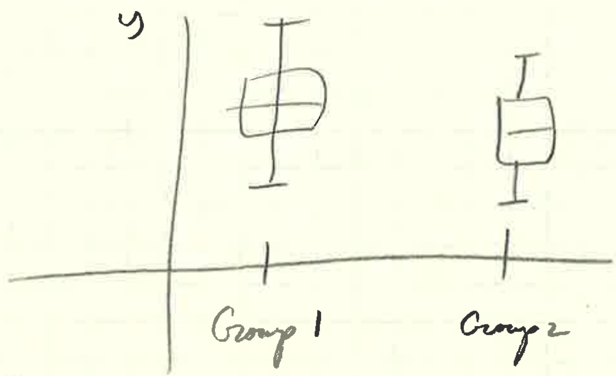
$$\text{and } \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

$$\hat\beta = (W^T W)^{-1} W^T y$$

ANOVA

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \text{ where } y_1 = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1,n/2} \end{bmatrix} \text{ and } y_2 = \begin{bmatrix} y_{21} \\ \vdots \\ y_{2,n/2} \end{bmatrix}$$

$$\text{Then } X = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix} \begin{matrix} \\ \\ \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} n/2 \text{ ones} \end{matrix}$$



$$\|y - X\beta\|^2$$

$$\hat\beta = (X^T X)^{-1} X^T y$$

$$= \left(\begin{bmatrix} 1 \cdots 1 & 0 \cdots 0 \\ 0 \cdots 0 & 1 \cdots 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix} \right)^{-1} \left(\begin{bmatrix} 1 \cdots 1 & 0 \cdots 0 \\ 0 \cdots 0 & 1 \cdots 1 \end{bmatrix} y \right)$$

$$= \begin{pmatrix} n/2 & 0 \\ 0 & n/2 \end{pmatrix}^{-1} \begin{pmatrix} J_{n/2}^T \cdot y_1 \\ J_{n/2}^T \cdot y_2 \end{pmatrix} = \begin{pmatrix} 1/n/2 & 0 \\ 0 & 1/n/2 \end{pmatrix} \begin{pmatrix} J_{n/2}^T y_1 \\ J_{n/2}^T y_2 \end{pmatrix} = \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix}$$

Fitted values $= J_{n/2} \bar{y}_1$ if in group 1

$= J_{n/2} \bar{y}_2$ if in group 2

Change of Parametrization

$\min \|y - X_1\beta\|^2$, $X_1 = \begin{bmatrix} J_{n_1} & 0 \\ 0 & J_{n_2} \end{bmatrix}$, $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$

$\Rightarrow \hat{\beta} = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \end{bmatrix}$

$\min \|y - X_2\gamma\|^2$, $X_2 = \begin{bmatrix} J_{n_1+n_2} & \begin{matrix} J_{n_1} \\ 0_{n_2} \end{matrix} \end{bmatrix}$, $\gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}$

Fitted values

Group 1 : $\hat{\gamma}_1 + \hat{\gamma}_2$ from $X_2\gamma$

Group 2 : $\hat{\gamma}_1$

$\Rightarrow \bar{y}_1 = \hat{\gamma}_1 + \hat{\gamma}_2$
$\bar{y}_2 = \hat{\gamma}_1$

thus $\hat{\gamma}_1 = \hat{\beta}_2$
$\hat{\gamma}_2 = \hat{\beta}_1 - \hat{\beta}_2$

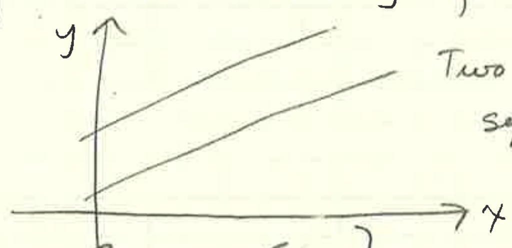
$J_{n_1+n_2} = [X_1]_{*1} + [X_1]_{*2} = [X_2]_{*1}$

$[X_2]_{*1} - [X_2]_{*2} = [X_1]_{*2}$

$\therefore$ X_1 and X_2 have the same column space

ANCOVA

$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$, $W = [\, Z_{n\times 2} \;\; X_{n\times 1} \,]$, $Z = \begin{bmatrix} J_{n_1} & 0_{n_1} \\ 0_{n_2} & J_{n_2} \end{bmatrix}$



Two different groups
seperate intercepts, common slope

$\min \|y - W\gamma\|^2$, $\gamma = \begin{bmatrix} \mu \\ \beta \end{bmatrix}$, $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$

$\|y - W\gamma\|^2 = \|y - X\beta - Z\mu\|^2$

$\mu_1(\beta) = \bar{y}_1 - \bar{x}_1\beta$
$\mu_2(\beta) = \bar{y}_2 - \bar{x}_2\beta$

$\Rightarrow \|y - X\beta - Z\mu\|^2 \geq \left\| \begin{pmatrix} y_1 - \bar{y}_1 J_{n_1} \\ y_2 - \bar{y}_2 J_{n_2} \end{pmatrix} - \begin{pmatrix} x_1 - \bar{x}_1 \\ x_2 - \bar{x}_2 \end{pmatrix} \beta \right\|^2$

$\Rightarrow \hat{\beta} = \dfrac{\langle \tilde{y}, \tilde{x} \rangle}{\langle \tilde{x}, \tilde{x} \rangle}$, $\tilde{x} = \begin{pmatrix} x_1 - \bar{x}_1 \\ x_2 - \bar{x}_2 \end{pmatrix}$, $\tilde{y} = \begin{pmatrix} y_1 - \bar{y}_1 \\ y_2 - \bar{y}_2 \end{pmatrix}$

$$= \frac{\sum_{i=1}^{2} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)(x_{ij} - \bar{x}_i)}{\sum\sum (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)} = \rho\hat{\beta}_1 + (1-\rho)\hat{\beta}_2$$

$\rho = \dfrac{\sum (x_{1j} - \bar{x}_1)^2}{\sum\sum (x_{ij} - \bar{x}_i)^2}$ = % of total variation in Xs from group 1

$$\hat{\mu}_1 = \bar{y}_1 - \bar{x}_1 \hat{\beta}_1 \qquad \Rightarrow \hat{\mu}_1 - \hat{\mu}_2 = \bar{y}_1 - \bar{y}_2 - (\bar{x}_1 - \bar{x}_2)\hat{\beta}$$
$$\hat{\mu}_2 = \bar{y}_2 - \bar{x}_2 \hat{\beta}_2$$

Week 6 1

Bases

Consider $\underset{n \times n}{X^T X} = I$, X is full rank, $X = [x_1 \cdots x_n]$

$$\min \|y - X\beta\|^2 \Rightarrow \hat{\beta} = (X^T X)^{-1} X^T y = X^T y$$
$$= \begin{pmatrix} \langle x_1, y \rangle \\ \langle x_2, y \rangle \\ \vdots \\ \langle x_n, y \rangle \end{pmatrix}$$

$$\hat{y} = X X^T y = \sum_{i=1}^{n} x_i \langle x_i, y \rangle$$

Let X_{S_p} contain a subset of colums of X

$$\Rightarrow X_S^T X_S = I$$
$$\Rightarrow \hat{y}_S = \sum_{i \in S} x_i \langle x_i, y \rangle$$

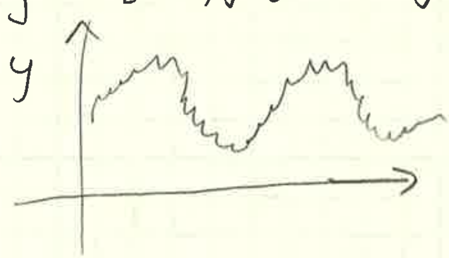
Fourier Bases

$X^T X = I$, Columns of X are trigonometric terms of different periods with increasing frequency

Y is a time series

$y \mapsto \{\langle x_i, y \rangle\} \Leftrightarrow y \mapsto X^T y$ is the discrete Fourier transform

Ex:



$\sum x_i \langle x_i, y \rangle$ to reconstruct the signal or choose $\subseteq i$ to filter, for example

SVDs

$X_{n\times p}$ $p \le n$ full column rank

SVD: $\underset{n\times p}{X} = \underset{n\times p}{U}\underset{p\times p}{D}\underset{p\times p}{V^T}$, $U^TU = V^TV = I$

X is centered $\Rightarrow X^TX$ is essentially $Cov(X)$

$X^TX = VDU^TUDV = VD^2V^T$ V eigenvectors of X^TX

$trace(X^TX) = trace(VD^2V^T)$

$= trace(D^2V^TV) = trace(D^2)$

$XVD^{-1} = U$ D^{-1} is like dividing by a standard deviation

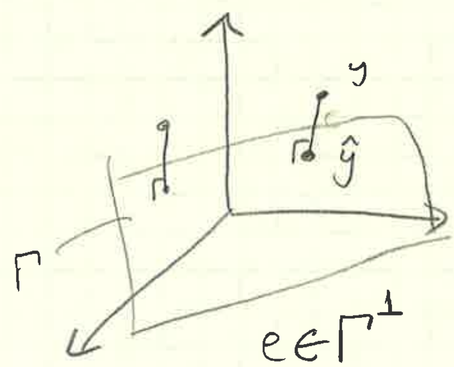
$\min \|y - U_k\gamma\|^2$ U_k is first k columns of U

$\hat{\gamma} = U_k^Ty$ since $U_k^TU_k = I$

Be careful of the units of X

Introduction to Residuals

$e = y - \hat{y} = y - H_xy = (I - H_x)y$ $H_x = X(X^TX)^{-1}X^T$



$e \in \Gamma^\perp$

$y \in \mathbb{R}^3$, $X \in \mathbb{R}^{3\times 2}$

$\Gamma = \{X\beta \mid \beta \in \mathbb{R}^2\}$

$e^TX\gamma = y^T(I-H_x)^TX\gamma$

$= y^T(I-H_x)X\gamma$

$= y^T(I - X(X^TX)^{-1}X^T)X\gamma$

$= y^T(X - X(X^TX)^{-1}X^TX)\gamma$

$= y^T(X - X)\gamma = 0$

$e^TJ_n = 0$ since $J_n \in Col(X)$

$e^Te = y^T(I-H_x)^2y$

$= y^T(I-H_x)y$ since $(I-H_x)$ is symmetric and idempotent

e^Te is a quadratic form

$X = [J \sim]$ $\quad H_J := J(J^TJ)^{-1}J^T$, $H_X := X(X^TX)^{-1}X^T$

$(I-H_X)J = 0 = J - H_XJ = 0 \quad \Rightarrow \quad J = H_XJ$

$SS_{Tot} = \|y - \bar{y}J_n\|^2 = \|y - H_Jy\|^2 = y^T(I-H_J)y$

$SS_{Res} = \|e\|^2 = \|y-\hat{y}\|^2 = y^T(I-H_X)y$

$$\begin{aligned} SS_{Reg} &= \|\bar{y}J - \hat{y}\|^2 = \|H_Jy - H_Xy\|^2 \\ &= y^T(H_J-H_X)(H_J-H_X)y \qquad (H_J-H_X) \text{ is symmetric} \\ &= y^T(H_J - H_XH_J - H_JH_X + H_X)y \end{aligned}$$

$J(J^TJ)^{-1}J^T = H_XJ(J^TJ)^{-1}J^T$

$\Rightarrow H_J = H_XH_J$

$\Rightarrow H_J = H_JH_X$ symmetry

$\Rightarrow SS_{Reg} = y^T(H_J - H_X - H_X + H_X)y = y^T(H_J - H_X)y$

$$\begin{aligned} SS_{Tot} = \|y - \bar{y}J_n\|^2 &= y^T(I-H_J)y \\ &= y^T(I - H_X + H_X - H_J)y \\ &= y^T(I-H_X)y + y^T(H_X-H_J)y \\ &= SS_{Res} + SS_{Reg} \end{aligned}$$

$$\boxed{SS_{Tot} = SS_{Reg} + SS_{Res}}$$

$R^2 := \dfrac{SS_{Reg}}{SS_{Tot}}$ = Percentage of total variability explained by the linear association with the added regressors