

# Lecture 7

26/4/22

## Computing LS solutions using SVD

Why?

- SVD handles rank-deficient  $Z$  and can identify covariates causing the deficiency.
- SVD is numerically stable, as opposed to solving  $Z^T Z \beta = Z^T y$  using Gauss elimination
- SVD is commonly used in multivariate statistics and ML

### The SVD

$$Z \in \mathbb{R}^{n \times p}$$

$$Z = U \Sigma V^T$$

where

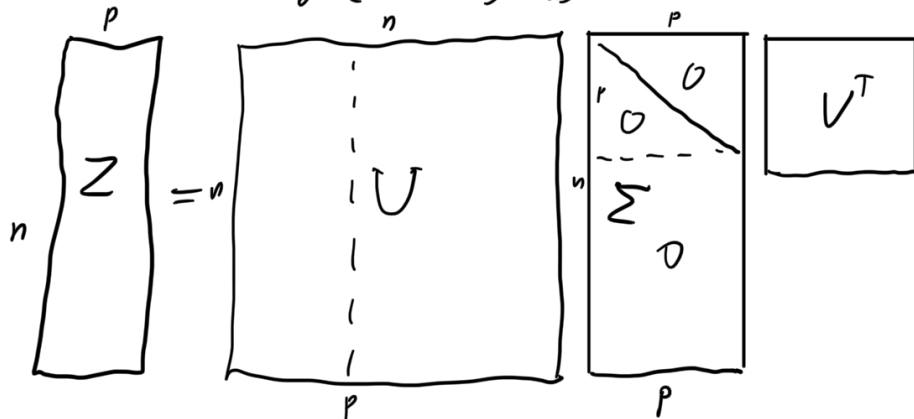
$$U \in \mathbb{R}^{n \times n}$$

$$\Sigma \in \mathbb{R}^{n \times p}$$

$$V \in \mathbb{R}^{p \times p}$$

$$U^T U = U U^T = I_n \quad V^T V = V V^T = I_p$$

$$\Sigma = \text{diag}(\sigma_1, \dots, \sigma_k), \quad \sigma_i \geq 0, \quad k = \min\{n, p\}$$

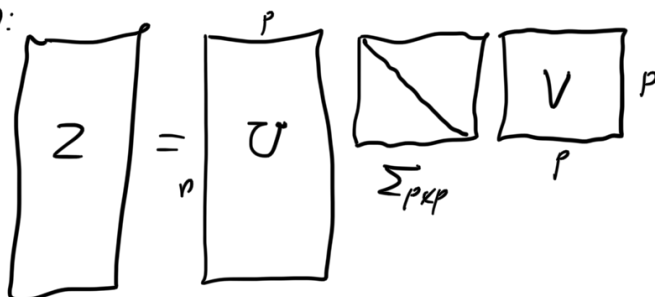


Standard case  $n > p$

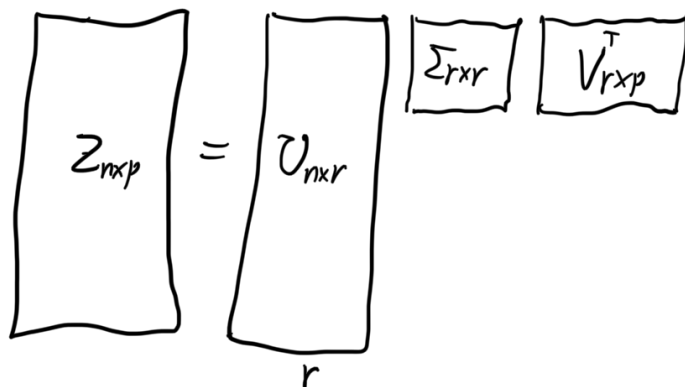
$$\tilde{y} = Z\tilde{\beta} = \underbrace{U}_{\text{III}} \underbrace{\Sigma}_{\text{II}} \underbrace{V^T \beta}_{\text{I}}$$

- (I) Rotating  $\beta$  by multiplying it by  $V^T$
- (II) Stretching each coordinate of  $V^T \beta$  by  $\Sigma$
- (III) Rotating  $\Sigma V^T \beta$  by  $U$

Skinny SVD:



In the general case where  $\text{rank}(Z) \leq p$ , we can also eliminate coordinates corresponding to  $\sigma_i = 0$ . Assume  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ , we let



$r = \text{rank}(Z)$  is the number of  $\sigma_i > 0$ .

Also note that eigenvalues of

$$Z^T Z = V \underbrace{\Sigma^T \Sigma}_{\text{I}} V^T \text{ are } \sigma_1^2, \dots, \sigma_r^2.$$

$$\tilde{y} = Z\tilde{\beta}$$

$$\|y - \tilde{y}\|^2 = \|y - Z\tilde{\beta}\|^2 = \|y - U \Sigma V^T \tilde{\beta}\|^2$$

$$\begin{aligned}
&= \underbrace{\| \underbrace{\sigma^T y}_{y^*} \underbrace{U^T U}_{I} \underbrace{\Sigma}_{\Sigma} \underbrace{V^T \tilde{\beta}}_{\beta^*} \|}_{\beta^*}^2 & \sigma_i = 0 \text{ if } i \geq r \\
& & i=1, \dots, n \\
&= \| y^* - \Sigma \beta^* \|^2 = \sum_{i=1}^n (y_i^* - \sigma_i \beta_i^*)^2 \\
&= \sum_{i=1}^r (y_i^* - \sigma_i \beta_i^*)^2 + \sum_{i=r+1}^p (y_i^* - 0 \cdot \beta_i^*)^2 + \sum_{i=p+1}^n (y_i^* - 0 \cdot \beta_i^*)^2 \\
&= \underbrace{\sum_{i=1}^r (y_i^* - \sigma_i \beta_i^*)^2} + \underbrace{\sum_{i=r+1}^n (y_i^*)^2}
\end{aligned}$$

The LS solution satisfies:

$$\beta_i^* = \begin{cases} y_i^* / \sigma_i & i=1, \dots, r \\ \text{anything} & i=r+1, \dots, p \end{cases}$$

usually we take  $\beta_i^* = 0$ ,  $i=r+1, \dots, p$   
to get the shortest solution

overall:

$$(I) \quad Z = U \Sigma V^T$$

$$(II) \quad y^* := \sigma^T y$$

$$(III) \quad \beta_j^* = \begin{cases} y_j^* / \sigma_j & \sigma_j \neq 0 \\ 0 & \sigma_j = 0 \end{cases}$$

$$(IV) \quad \hat{\beta} = V \beta^*$$

Computing SVD is done in  $\mathcal{O}(np^2)$

The same SVD can be used for multiple  $y$ 's

## ANOVA

Recap:

- The main idea: test if there exists a significant difference in the mean values between columns

comparing means between subgroups

- The cell means model:

$$Y_{ij} = \mu_i + \varepsilon_{ij} \quad \varepsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$$

$$i=1, \dots, k \quad j=1, \dots, n_i$$

- When  $k=2 \Rightarrow$  Use  $t$ -test

- The effect means model,

$$\sum d_i n_i = 0 \quad Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

The main statistical problem is

testing:  $H_0: \mu_1 = \dots = \mu_k$

or  $H_0: \alpha_1 = \dots = \alpha_k = 0$

$$Y \sim N(\beta Z, \sigma^2 I)$$

$$\beta = (\mu_1, \dots, \mu_k)^T$$

$$y = (Y_{11}, Y_{12}, \dots, Y_{1n_1}, Y_{21}, \dots, Y_{2n_2}, \dots, Y_{kn_1}, \dots, Y_{kn_{n_k}})^T$$

$$Z = \begin{bmatrix} 1 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \dots & 0 \\ 0 & 1 & & \vdots \\ \vdots & \vdots & & \vdots \\ 0 & 1 & & \vdots \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & 0 \\ \vdots & \vdots & & 0 \\ \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & 1 \end{bmatrix}$$

$$\hat{\beta} = (Z^T Z)^{-1} Z^T y = \begin{bmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_k \end{bmatrix}$$

$$\bar{y}_{i.} := \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \quad \left[ \begin{array}{c} \vdots \\ \bar{y}_{k.} \end{array} \right]$$

- For the effect model:

$$\hat{y}_i = \bar{y}_{i.} = \frac{1}{n} \sum_{i=1}^k \bar{y}_{i.} \cdot n_i = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij}$$

- The SS for the cell means:

$$SS_{within} := SS_{res} = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2$$

- The total SS:

$$SS_{tot} = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2$$

- We can test  $H_0$  using the extra SS principle:

$$F = \frac{\frac{1}{k-1} (SS_{tot} - SS_{within})}{\frac{1}{n-k} SS_{within}}$$

reject  $H_0$  at level  $\alpha$  if

$$F \geq F_{k-1, n-k}^{1-\alpha}$$

- we also define

$$\begin{aligned} SS_{between} &:= SS_{fit} = SS_{tot} - SS_{within} \\ &= \sum_{i=1}^k n_i (\bar{y}_{i.} - \bar{y}_{..})^2 \end{aligned}$$

$$SS_{tot} = SS_{between} + SS_{within}$$

- We usually summarize values in so-called ANOVA table:

ANOVA TABLE

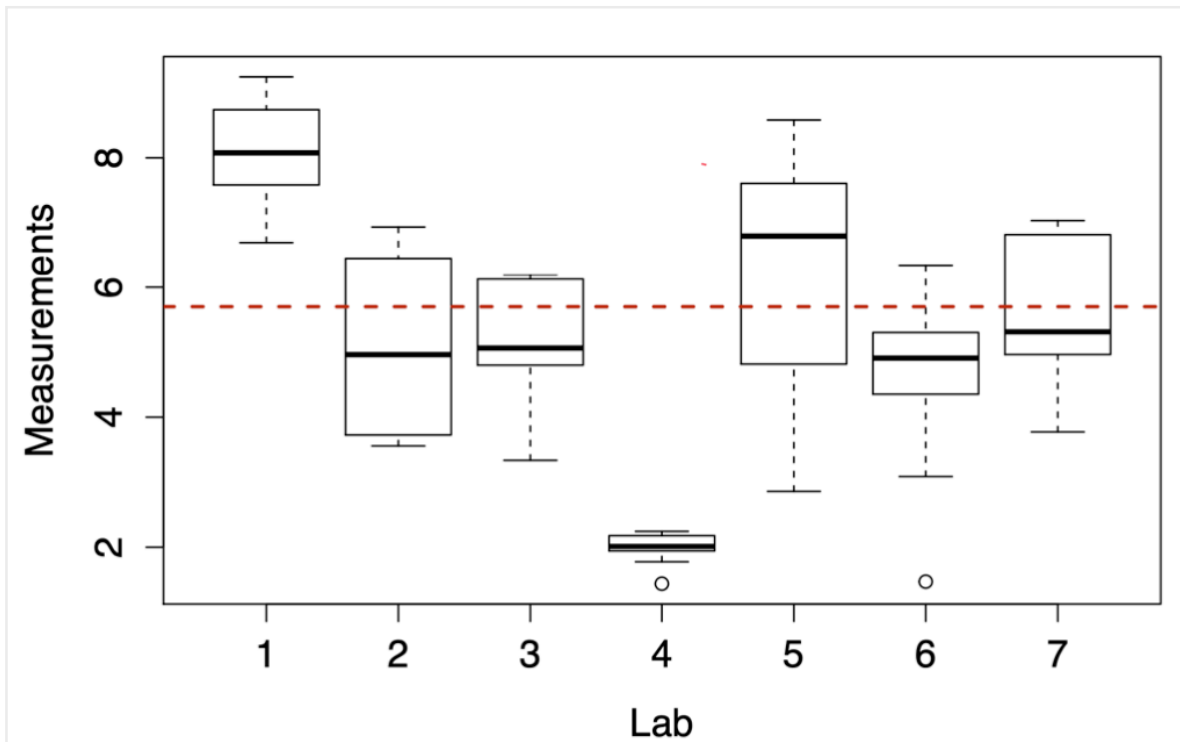
Source	DoF	SS	MS	F
Groups	$k-1$	$SS_{\text{between}}$	$SS_{\text{between}}/k-1$	$MS_{\text{between}}$
Error	$n-k$	$SS_{\text{within}}$	$SS_{\text{within}}/n-k$	$MS_{\text{within}}$
Total	$n-1$	$SS_{\text{tot}}$		

$$MS_{\text{within}} = S^2$$

Example: we send 10 pills to each of 7 different labs:



$H_0$ : there is no significant difference in measurements across all the labs



we have the ANOVA table:  
 $n = 7 \times 10 = 70$

	DoF	SS	MS	F	P
Labs	6	1125	187.5	5.66	0.0015

	0	0.125	0.021	0.000	$7.75 \times 10^{-4}$
Error	63	0.231	0.0037		
Total	64	0.356			

we reject  $H_0$

## Contrasts

- suppose that we detected some effect, which one the groups causing it?
- we can check differences between individual groups:

$$\frac{\bar{y}_i - \bar{y}_j}{s \sqrt{\frac{1}{n_j} + \frac{1}{n_i}}} \sim t_{n-k} \quad s^2 = \frac{SS_{within}}{n-k}$$

there are  $\binom{k}{2} = \frac{k(k-1)}{2}$  such comparisons

- Is it possible that we fail to reject  $H_0: \mu_1 = \dots = \mu_k$  but find some of the group differences significant? Yes
- Is it possible that we reject  $H_0$  but do not find any of the differences significant? Yes
- we can also look at contrast:  
suppose that we want to test the effectiveness of three detergents with phosphates against four detergents without.  
 $H_0$ : phosphates detergents are no different from non-phosphates

ur ur:

$$t = \frac{\frac{\bar{y}_1 + \bar{y}_2 + \bar{y}_3}{3} - \frac{\bar{y}_4 + \bar{y}_5 + \bar{y}_6 + \bar{y}_7}{4}}{S \sqrt{\frac{1}{9} \left( \frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3} \right) + \frac{1}{16} \left( \frac{1}{n_4} + \dots + \frac{1}{n_7} \right)}}$$

$S^2 = \text{MSE}_{\text{within}}$

$$t \sim t_{n-7}$$

- Another option: a single product with some "treatment" against four other product without:

$$t = \frac{\bar{y}_1 - \frac{1}{4}(\bar{y}_2 + \dots + \bar{y}_5)}{S \sqrt{\frac{1}{n_1} + \frac{1}{16} \left( \frac{1}{n_2} + \dots + \frac{1}{n_5} \right)}}$$

- In general:

$$\sum_{i=1}^k \lambda_i = 0 \quad \sum_{i=1}^k \lambda_i^2 > 0 \quad t = \frac{\sum_{i=1}^k \lambda_i \bar{y}_i}{S \sqrt{\sum_{i=1}^k \frac{\lambda_i^2}{n_i}}} \sim t_{n-k}$$

Example:

	Potassium	No. Potassium
Sulfur	$\bar{y}_1$	$\bar{y}_2$
No Sulfur	$\bar{y}_3$	$\bar{y}_4$

We can examine several effects:

- Sulfur  $\bar{y}_1 + \bar{y}_2 - \bar{y}_3 - \bar{y}_4$



- Potassium:  $\bar{y}_1 + \bar{y}_3 - \bar{y}_2 - \bar{y}_4$ .
- combined effect:  $\bar{y}_1 - \frac{\bar{y}_4 + \bar{y}_2 + \bar{y}_3}{3}$ .

## Multiple Comparisons

- How can we find out which groups are motivating the rejection of the null?
- We can do all pairs using  $\binom{k}{2}$  t-tests. However, these additional tests inflate the prob. of making Type I error (falsely rejecting the null).

if  $k=10$ , then we do 45 tests and reject each one under the null w.p.  $\alpha$ , so  $\Pr(\text{reject}) > \alpha$ .

## Multiple Hypothesis Testing

- Multiple comparisons is a special case of multiple hypothesis testing.
- We formulate a family of  $m$  null hypotheses  $\{H_{0,i}\}_{i=1}^m$  and make up a test with a family-wise type I error rate (FWER) of at most  $\alpha$ .

$$H_0 = \bigcap_{i=1}^m H_{0,i} \quad \Pr(\text{reject } H_0 | H_0 \text{ is true}) \leq \alpha$$

## Bonferroni's Union Bound

- we have  $m$  tests (e.g.  $m = \binom{k}{2}$ )
- we conduct each test at level  $\alpha/m$  (e.g. for t-tests, we reject

based on  $t_{n-k}^{1-\frac{\alpha}{m}}$

- Under this procedure:

$\Pr(\text{reject something} | H_0 \text{ is true})$

$$\leq \sum_{l=1}^m \Pr(\text{reject } l\text{-th test} | H_0 \text{ is true})$$

$$\leq \sum_{l=1}^m \frac{\alpha}{m} = m \cdot \frac{\alpha}{m} = \alpha$$