Skip §36 outliers Skip notes 39-42

end exam 2 material

ch 4] 1] A random vector (matrix) is a vector (matrix) with elements that are random variables.

$$\underset{\sim}{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$$

2] p163 The population mean of $\underset{\sim}{Y}$ is $\underset{n \times 1}{E(\underset{\sim}{Y})} = \begin{pmatrix} E(Y_1) \\ \vdots \\ E(Y_n) \end{pmatrix}$ provided each EY_i exists, otherwise $E(\underset{\sim}{Y})$ does not exist.

The population covariance matrix of $\underset{\sim}{Y}$ is $\underset{n \times n}{\text{cov}(\underset{\sim}{Y})} = E\left[(\underset{\sim}{Y} - E(\underset{\sim}{Y}))(\underset{\sim}{Y} - E(\underset{\sim}{Y}))^T\right] = (\sigma_{ij})$

$$= \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & & & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix} = \begin{pmatrix} V(Y_1) & \text{cov}(Y_1, Y_2) & \cdots & \text{cov}(Y_1, Y_n) \\ \text{cov}(Y_2, Y_1) & V(Y_2) & \cdots & \text{cov}(Y_2, Y_n) \\ \vdots & & & \vdots \\ \text{cov}(Y_n, Y_1) & \text{cov}(Y_n, Y_2) & \cdots & V(Y_n) \end{pmatrix}$$

where $\sigma_{ij} = \text{cov}(Y_i, Y_j)$, provided each σ_{ij} exists. otherwise $\text{cov}(\underset{\sim}{Y})$ does not exist.

Note $\sigma_{ii} = V(Y_i)$, $\sigma_{ij} = \sigma_{ji}$

so $\text{cov}(\underset{\sim}{Y})$ is symmetric.

3] Know p163-4 Suppose $\underset{\sim}{Z}$ and $\underset{\sim}{Y}$ are $n \times 1$ random vectors, $\underset{\sim}{a}$ a conformable constant vector and A and B are conformable constant matrices. 43.5

Then i) $E(A) = A$, $E(\underset{\sim}{a}) = \underset{\sim}{a}$

ii) $E(\underset{\sim}{a} + \underset{\sim}{Y}) = \underset{\sim}{a} + E(\underset{\sim}{Y})$

iii) $E(\underset{\sim}{Y} + \underset{\sim}{Z}) = E(\underset{\sim}{Y}) + E(\underset{\sim}{Z})$

iv) $E(A\underset{\sim}{Y}) = A\,E(\underset{\sim}{Y})$, $E(A\underset{\sim}{Y}B) = A\,E(\underset{\sim}{Y})\,B$

v) $\mathrm{cov}(\underset{\sim}{a} + A\underset{\sim}{Y}) = \mathrm{cov}(A\underset{\sim}{Y}) = A\,\mathrm{cov}(\underset{\sim}{Y})\,A^T$

4] Analogy with random variables:
$E(a) = a$, $E[aY] = a\,E(Y)$,
$\mathrm{Var}(aY) = a^2\,V(Y) = a\,V(Y)\,a^T$

5] If $Y_1, \dots, Y_n$ are uncorrelated or independent, $\mathrm{cov}(\underset{\sim}{Y}) = \begin{pmatrix} V(Y_1) & & 0 \\ & \ddots & \\ 0 & & V(Y_n) \end{pmatrix}$

If also $V(Y_i) = \sigma^2$, then $\mathrm{cov}(\underset{\sim}{Y}) = \sigma^2 I_n$
where $I_n = \mathrm{diag}(1, \dots, 1)$ is the $n \times n$ identity matrix.

6] Let $A\underset{\sim}{Y} = \underset{\sim}{W}$. Then
$\mathrm{cov}(A\underset{\sim}{Y}) = \mathrm{cov}(\underset{\sim}{W}) = E[(\underset{\sim}{W} - E(\underset{\sim}{W}))(\underset{\sim}{W} - E(\underset{\sim}{W}))^T]$

$= E[A(\underset{\sim}{Y} - E\underset{\sim}{Y})(\underset{\sim}{Y} - E\underset{\sim}{Y})^T A^T] = A\,E[(\underset{\sim}{Y} - E\underset{\sim}{Y})(\underset{\sim}{Y} - E\underset{\sim}{Y})^T]A^T$

$= A\,\mathrm{cov}(\underset{\sim}{Y})\,A^T.$

ex] $\underset{\sim}{Y} = \underbrace{X\beta}_{\text{constant}} + \underset{\sim}{e}$, $E(\underset{\sim}{e}) = \underset{\sim}{0}$, $\operatorname{cov}(\underset{\sim}{e}) = \sigma^2 I_n$

$E(\underset{\sim}{Y}) = X\beta + \underset{\sim}{0} = X\underset{\sim}{\beta}$

$\operatorname{Cov}(\underset{\sim}{Y}) = \operatorname{Cov}(\underset{\sim}{e}) = \sigma^2 I_n$

ex] $\hat{\underset{\sim}{\beta}} = (X^T X)^{-1} X^T Y$

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

$\operatorname{Cov}(\hat{\underset{\sim}{\beta}}) = (X^T X)^{-1} X^T \operatorname{Cov}(\underset{\sim}{Y}) \left[(X^T X)^{-1} X^T\right]^T$

$= (X^T X)^{-1} X^T \sigma^2 I \, X (X^T X)^{-1}$

$= \sigma^2 (X^T X)^{-1}$

ex] $\hat{\underset{\sim}{Y}} = H \underset{\sim}{Y}$

So $E(\hat{Y}) = H \, E Y = H X \underset{\sim}{\beta} = X (X^T X)^{-1} X^T X \underset{\sim}{\beta}$

$= X\underset{\sim}{\beta} = E(Y)$

$\operatorname{cov}(\hat{Y}) = H \operatorname{cov}(Y) H^T = H \sigma^2 I H$

$= \sigma^2 H H = \sigma^2 H$ since $H = H^T$, $H^2 = H$

7] Generalized Least Squares GLS

has $Y = X\beta + e$ with

$E(\underset{\sim}{e}) = \underset{\sim}{0}$ and $\operatorname{Cov}(\underset{\sim}{e}) = \sigma^2 \underset{n \times n}{V}$

where V is known.

8] p165 $\hat{\underset{\sim}{\beta}}_{GLS} = (\mathbf{X}^T V^{-1} \mathbf{X})^{-1} \mathbf{X}^T V^{-1} \underset{\sim}{Y}$

$\hat{\underset{\sim}{Y}}_{GLS} = \mathbf{X} \hat{\underset{\sim}{\beta}}_{GLS}$.

9] p165 Weighted Least Squares (WLS) is the special case where $V = diag(v_1, \ldots, v_n)$.

The weights $w_i = \frac{1}{v_i}$.
Then $E(\underset{\sim}{e}) = \underset{\sim}{0}$, $cov(\underset{\sim}{e}) = \sigma^2 diag(v_1, \ldots, v_n)$

$= \sigma^2 diag\left(\frac{1}{w_1}, \ldots, \frac{1}{w_n}\right)$.
$Y_i = \underset{\sim}{x}_i^T \underset{\sim}{\beta} + e_i$ where the

e_i are independent, $E\, e_i = 0$ $V(e_i) = \frac{\sigma^2}{w_i}$.

10] $\hat{\underset{\sim}{\beta}}_{WLS} = (\mathbf{X}^T V^{-1} \mathbf{X})^{-1} \mathbf{X}^T V^{-1} \underset{\sim}{Y}$,

$\hat{\underset{\sim}{Y}}_{WLS} = \mathbf{X} \hat{\underset{\sim}{\beta}}_{WLS}$.

11] p165 Feasible Generalized Least Squares is like GLS except
$V = V(\underset{\sim}{\theta})$ where $\underset{q \times 1}{\underset{\sim}{\theta}}$ is unknown.

Let $\hat{V} = V(\hat{\underset{\sim}{\theta}})$. Then
$\hat{\underset{\sim}{\beta}}_{FGLS} = (\mathbf{X}^T \hat{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \hat{V}^{-1} \underset{\sim}{Y}$,

$\hat{\underset{\sim}{Y}}_{FGLS} = \mathbf{X} \hat{\underset{\sim}{\beta}}_{FGLS}$

12] Feasible weighted least squares FWLS is the special case where $V(\underset{\sim}{\theta})$ is diagonal

$$\hat{w}_i = \frac{1}{\hat{v}_i} = \frac{1}{v_i(\hat{\underset{\sim}{\theta}})}.$$

13] FGLS and FWLS have $p+q+1$ unknown parameters $\underset{p\times 1}{\underset{\sim}{\beta}}$, $\underset{q\times 1}{\underset{\sim}{\theta}}$ and σ^2.

14] P168 GLS and WLS can be found from the OLS regression (without intercept) of a transformed model.

From (numerical) linear algebra, there is a matrix $\underset{n\times n}{K}$ such that $V = K K^T$.

Let $\underset{\sim}{Z} = K^{-1} \underset{\sim}{Y}$, $U = K^{-1} X$ and $\underset{\sim}{\varepsilon} = K^{-1} \underset{\sim}{e}$.

Then a) $\underset{\sim}{Z} = U \underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ has $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$

$\text{cov}(\underset{\sim}{\varepsilon}) = \sigma^2 I_n$ and

b) $\hat{\underset{\sim}{\beta}}_{GLS}$ can be obtained from the OLS regression (without intercept) of $\underset{\sim}{Z}$ on U.

Proof] a) $E(\underset{\sim}{\varepsilon}) = K^{-1} E(\underset{\sim}{e}) = \underset{\sim}{0}$

$\text{cov}(\underset{\sim}{\varepsilon}) = \text{cov}(K^{-1} \underset{\sim}{e}) = K^{-1} \text{cov}(\underset{\sim}{e}) (K^{-1})^T$

45.5

$= \sigma^2 K^{-1} V (K^{-1})^T = \sigma^2 K^{-1} K K^T (K^{-1})^T = \sigma^2 I$

OLS without intercept needs to be used since the 1st column of U is $K^{-1}\underset{\sim}{1} \neq \underset{\sim}{1}$.

b) Let $\hat{\beta}_{ZU}$ be the OLS estimator without intercept obtained from regressing Z on U.

Then $\hat{\underset{\sim}{\beta}}_{ZU} = (U^T U)^{-1} U^T \underset{\sim}{Z} =$

$(X^T (K^{-1})^T K^{-1} X)^{-1} X^T (K^{-1})^T K^{-1} \underset{\sim}{Y}$

$= (X^T V^{-1} X)^{-1} X^T V \underset{\sim}{Y} = \hat{\beta}_{GLS}$

since $V^{-1} = (K K^T)^{-1} = (K^T)^{-1} K^{-1} = (K^{-1})^T K^{-1}$.

15] For FGLS and FWLS, use $\hat{V} = \hat{K} \hat{K}^T$ and sub $\hat{K}$ for K in 14).

16] If Y_i = average or median of n_i obs's, use $w_i = n_i$ since $V(Y_i) = V(e_i) \propto \frac{1}{n_i}$.

$\left(V(\tilde{Y}_{n_i}) \approx \frac{\sigma^2}{n_i} \right)$

17] Under regularity conditions $\hat{\underset{\sim}{\beta}}_{OLS}$ is a consistent estimator of β when GLS assumptions hold.

So make the OLS response and residual plots. Plotted points should scatter about the identity and $r=0$ line but not in evenly populated bands since the constant variance assumption is violated. Plots based on $\hat{\beta}_{OLS}$ instead of

if $\hat{\beta}_{GLS}$ is good

$\hat{\beta}_{GLS} \rightarrow \hat{\beta}_{OLS}$ should be similar (since $\hat{\beta}_{OLS} \approx \hat{\beta}_{GLS} \approx \underline{\beta}$). To check the GLS model V or FGLS model $\hat{V}$, do the OLS regression of $\underset{\sim}{Z}$ on U and make the response and residual plots based on $\underset{\sim}{Z}$ and U. These should look like OLS plots and scatter about the identity and $r=0$ lines in evenly populated band if $\underset{\sim}{\varepsilon} = K^{-1}\underset{\sim}{e}$ is not from a highly skewed distribution.

FF and RR plots $\underline{x}^T\hat{\beta}_{OLS} = \hat{Y}_{OLS}$ vs $\hat{Y}_{OLS}$

18] A major problem with GLS, WLS, FGLS, and FWLS is often the response and residual plots from regressing $\underset{\sim}{Z}$ on U look worse than the OLS plots from regressing Y on X. This suggests that the GLS or FGLS model (V or $V(\hat{\underline{\theta}})$) is wrong. Often the plots from regressing $\underset{\sim}{Z}$ on U have outliers or high leverage points.

19] p170 pvalues for FGLS and FWLS are generally incorrect, but can be used for exploratory purposes.

20] §4.3 Inference for GLS $\underset{\sim}{Y} = X\beta + \underset{\sim}{e}$

can be performed using the partial F test for the equivalent no intercept OLS model $\underset{\sim}{Z} = U\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$

covered in §2.10. Knowing U is too strong

21] p170 For WLS supply weights to the MLR software. The output needs correct weights to be valid (prediction intervals are likely wrong even with valid weights).

too strong

ch5] 1] p175 Models where Y is quantitative but all of the predictor variables are qualitative are called analysis of variance (ANOVA) models, experimental design models or design of experiments (DOE) models.

2] A unit is experimental material assigned treatments, the conditions the experimenter wants to study. The unit is