A, B, C LM

1) A linear model is $\underset{n\times 1}{\underset{\sim}{Y}} = \underset{n\times p}{X}\underset{p\times 1}{\underset{\sim}{\beta}} + \underset{n\times 1}{\underset{\sim}{\varepsilon}}$

$$Y_i = \underset{\sim}{x}_i^T \underset{\sim}{\beta} + \varepsilon_i.$$

For multiple linear regression (MLR), at least one of the predictors in $\underset{\sim}{x}_i^T = (x_1, x_2, \ldots, x_p)$ is quantitative. For design of experiments (DOE), ~~For~~ ANOVA, or experimental design models, all of the predictors are qualitative, often coded as indicator variables.

2) Linear models are often fit by least squares and $\hat{\underset{\sim}{Y}} = P_X \underset{\sim}{Y}$ is the projection of $\underset{\sim}{Y}$ onto the column space of X.

13) A set $V \subseteq \mathbb{R}^n$ is a vector space if for any $\underset{\sim}{x}, \underset{\sim}{y}, \underset{\sim}{z} \in V$, and scalars α and β, the operations of vector addition and scalar multiplication are defined such

1) $(\underline{x}+\underline{y})+\underline{z} = \underset{\sim}{x}+(\underline{y}+\underline{z})$ associative

2) $\underline{x}+\underline{y} = \underline{y}+\underline{x}$ commutative

3) $\exists\ \underset{\sim}{0} \in V \ni \underline{x}+\underline{0} = \underset{\sim}{x} = \underline{0}+\underline{x}$ additive identity

4) $\forall\ \underline{x} \in V\ \exists\ \underline{y} = -\underline{x} \ni \underline{x}+\underline{y} = \underline{0} = \underline{y}+\underline{x}$ additive inverse

5) $\alpha(\underline{x}+\underline{y}) = \alpha\underline{x}+\alpha\underline{y}$ distributive

6) $(\alpha+\beta)\underline{x} = \alpha\underline{x}+\beta\underline{x}$ scalar distributive

7) $(\alpha\beta)\underline{x} = \alpha(\beta\underline{x})$ associative

8) $1\underline{x} = \underline{x}$ scalar identity

4) *Let M be a nonempty subset of vector space V. If

i) $\alpha\underline{x} \in M \quad \forall\ \underline{x} \in M$ and for any scalar α

ii) $\underline{x}+\underline{y} \in M \quad \forall\ \underline{x},\underline{y} \in M$

then M is a vector space called a <u>subspace</u> of V.

Note that M is closed under addition and scalar multiplication.

ex) $\mathbb{R}^n$ is a vector space

5) span $(\underline{x}_1, \ldots, \underline{x}_r)$ = set of all linear combinations of $\underline{x}_1, \ldots, \underline{x}_r$ is a vector space

16) Let $\underline{x}_1, \ldots, \underline{x}_K$ be vectors in V.

If $\exists$ scalars $\alpha_1, \ldots, \alpha_K$ not all $0 \ni$ $\sum_{i=1}^{K} \alpha_i \underline{x}_i = \underline{0}$, then $\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_K$ are <u>linearly dependent</u>. If $\sum_{i=1}^{K} \alpha_i \underline{x}_i = \underline{0}$ only if $\alpha_i = 0 \ \forall i = 1, \ldots, K$, then $\underline{x}_1, \ldots, \underline{x}_K$ are <u>linearly independent</u>.

17) Suppose $\{\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_K\}$ is a linearly independent set and $V = \text{span}(\underline{x}_1, \ldots, \underline{x}_K)$.

Then $\{\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_K\}$ is a linearly independent & spanning set for V, called a <u>basis</u> for V.

18) p459 *Let $A = [\underline{a}_1 \ \underline{a}_2 \cdots \underline{a}_m] = \begin{bmatrix} \underline{r}_1^T \\ \vdots \\ \underset{\sim}{r}_n^T \end{bmatrix}$ be an $n \times m$ matrix. The space spanned by the columns of A = <u>column space</u> of $A = C(A)$.

19) Notation $\underset{n \times p}{X} = \begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix} = \underbrace{[\underline{v}_1 \ \cdots \ \underline{v}_p]}$

rows correspond to measurements on ith object = ith case

columns correspond to variables $\underline{v}_1, \ldots, \underset{\sim}{v}_p$

10} The dimension of a vector space V (25)

= $\dim(V)$ = number of vectors in a basis of V. The rank of a matrix A = $\text{rank}(A) = \dim(C(A))$, the dimension of $C(A)$.

11} The row space of A = $C(\overset{\text{transpose}}{A^T})$ = span of the rows of A.

ex} $$\underset{n\times p}{\underline{X}}\ \underset{p\times 1}{\underline{\beta}} = \begin{bmatrix} \underline{x}_1^T \\ \vdots \\ \underline{x}_n^T \end{bmatrix} \underline{\beta} = \begin{bmatrix} \underline{x}_1^T \underline{\beta} \\ \vdots \\ \underline{x}_n \underline{\beta} \end{bmatrix})$$

$$= [\underline{v}_1 \cdots \underline{v}_p] \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} = \beta_1 \underline{v}_1 + \cdots + \beta_p \underline{v}_p$$

$\in C(\underline{X}) = \{\underline{y} \in \mathbb{R}^n \ni \underline{y} = \underline{X}\underline{\beta}$ for some $\underline{\beta} \in \mathbb{R}^p\}$.

So $\underline{X}\underline{\beta} \in C(\underline{X})$ if $\underline{\beta} \in \mathbb{R}^p$.

11) Let A be $m \times n$. Then

$rank(A) = rank(A^T) \leq \min(m,n)$.

If $rank(A) = \min(m,n)$ then A has full rank, or A is a full rank matrix.

13) p458 $N(A)$ = null space of A = Kernel of A

$\{x : Ax = 0\}$. The nullity of $A = \dim[N(A)]$.

14) p458-9 a) If AB exists $rank(AB) \leq \min(rank(A), rank(B))$,

b) $rank(A) = rank(A^T) = rank(A^TA) = rank(AA^T)$

c) If A is symmetric ($A = A^T$ and $m = n$)

$rank(A)$ = number of nonzero eigenvalues of A. (If λ has multiplicity k, think of λ being an eigenvalue k times.)

d) $rank(A+B) \leq rank(A) + rank(B)$

e) For conformable matrices with P and Q nonsingular

$rank(PAQ) = rank(A)$.

15) p473 The subspace $V_M^{\perp} = \{y \in M : y \perp V\}$ is the orthogonal complement of V wrt M.

$\mathbb{R}^k$ the subspace $V^\perp = \{x \ldots\}$ (5.5)

is called the orthogonal complement of V.

$y \perp V \Rightarrow y^T x = 0 \ \forall x \in V$

16) p477 $\underbrace{N(A^T)}_{\text{null space of } A^T} = \underbrace{[C(A)]^\perp}_{\text{orthognal complement of } C(A)}$ so $N(A) = [C(A^T)]^\perp$

17) p458 Rank Nullity Theorem: Let A be $m \times n$. Then, rank(A) + dim$[N(A)] = n$.

18) If V is a subspace of $\mathbb{R}^n$

then $\dim(V) + \dim(V^\perp) = n$,

and if $\{x_1, \ldots, x_r\}$ is a basis for V

and $\{x_{r+1}, \ldots, x_n\}$ is a basis for $V^\perp$,

then $\{x_1, \ldots, x_n\}$ is a basis for $\mathbb{R}^n$

(Leon P. 167).

19) * p469 A generalized inverse

of an $m \times n$ matrix A is any

$n \times m$ matrix A^- satisfying $AA^-A = A$.

Notation $G = A^-$ means G is a generalized inverse of A.

20) Usually a generalized inverse is not unique. If A^{-1} exists then $A^- = A^{-1}$ is unique.

21) p475 Let V be a subspace of $\mathbb{R}^k$. Every $\underline{y} \in \mathbb{R}^k$ can be expessed uniquely as $\underline{y} = \underline{w} + \underline{z}$ where $\underline{w} \in V$ and $\underline{z} \in V^\perp$.

22) Let $\underline{X} = [\underline{v}_1, \ldots, \underline{v}_p]$ be $n \times p$ and let $V = C(\underline{X}) = \text{span}(\underline{v}_1, \ldots, \underline{v}_p)$. The $n \times n$ matrix $P_V = P_X$ is a projection matrix on $C(\underline{X})$ if $P_X \underline{y} = \underline{w}$ $\forall \underline{y} \in \mathbb{R}^n$

orthogonal is bad since P is not an orthogonal matrix

$(\underline{y} = \underline{w} + \underline{z} = \underline{w}_y + \underline{z}_y$ so $\underline{w}$ depends on $\underline{y})$

(Note: if $\underline{Y} \perp C(X)$ then $\underline{Y} = \underline{0} + \underline{z}$ so $P_X \underline{Y} = \underline{0}$.)

23) p464 A is idempotent if $A^2 = AA = A$.

Theorem a) P_X is unique.

p475-6 b) $P_X = \underset{n \times p}{\underline{X}} (\underset{p \times n}{\underline{X}^T} \underset{n \times p}{\underline{X}})^- \underset{p \times n}{\underline{X}^T}$ where inverse (4.5)

of $\underline{X}^T\underline{X}$.

c) A is a projection matrix on $C(A)$ iff A is symmetric and idempotent.

Hence P_X is a projection matrix of $C(P_X) = C(\underline{X})$.

d) $I_n - P_X$ is the projection matrix on $[C(\underline{X})]^\perp$.

e) $A = P_X$ iff i) $\underline{y} \in C(X) \Rightarrow A\underline{y} = \underline{y}$, ii) $\overbrace{\underline{y} \perp C(\underline{X})}^{\underline{y} \in [C(X)]^\perp} \Rightarrow A\underline{y} = \underline{0}$.

Theorem a) $P_X \underline{X} = \underline{X}$ and $P_X \underline{W} = \underline{W}$ if each column of $\underline{W} \in C(\underline{X})$.

b) $P_X \underline{v}_i = \underline{v}_i$

c) If $C(\underline{X}_R)$ is a subspace of $C(\underline{X})$,

then $P_X P_{X_R} = P_{X_R} P_X = P_{X_R}$

d) Let $\underline{X} = [\underline{Z} \ \underline{X}_r]$ where $\text{rank}(\underline{X}) = r = \text{rank}(\underline{X}_r)$ so the columns of $\underline{X}_r$ form a basis for $C(\underline{X})$.

Then $\begin{bmatrix} 0 & 0 \\ 0 & (\underline{X}_r^T\underline{X}_r)^{-1} \end{bmatrix} = (\underline{X}^T\underline{X})^-$ is a generalized inverse of X^TX

and $P_X = \underline{X}_r(\underline{X}_r^T\underline{X}_r)^{-1}\underline{X}_r^T$ $(= P_{X_r}$ since $C(\underline{X}) = C(\underline{X}_r))$.

$$X'X = \begin{pmatrix} Z' \\ X_r^T \end{pmatrix} [Z \ \ X_r] = \begin{bmatrix} Z^T Z & Z^T X_r \\ X_r^T Z & X_r^T X_r \end{bmatrix}$$

So $X^T X \begin{pmatrix} 0 & 0 \\ 0 & (X_r^T X_r)^{-1} \end{pmatrix} X^T X =$

$$\begin{pmatrix} 0 & Z^T X_r (X_r^T X_r)^{-1} \\ 0 & I \end{pmatrix} X^T X =$$

$$\begin{pmatrix} Z^T X_r (X_r^T X_r)^{-1} X_r^T Z & Z^T X_r \\ X_r^T Z & X_r^T X_r \end{pmatrix} = X^T X$$

Since $\underbrace{X_r (X_r^T X_r)^{-1} X_r^T}_{P_X} \underbrace{Z}_{\in C(X)} = Z$

Note that $P_X = X (X^T X)^{-} X^T = [0 \ \ X_r (X_r^T X_r)^{-1}] \begin{pmatrix} Z^T \\ X_r^T \end{pmatrix}$

$$= X_r (X_r^T X_r)^{-1} X_r^T$$

24] Be able to show whether A is idempotent: $AA = A$. 5.5

25] Know for E1 Be able to show $G \overset{"}{:=} A^-$: show $AGA = A$.

ex] i) Suppose $G \overset{"}{:=} A^-$ show $\underbrace{G^T \overset{"}{:=} (A^T)^-}_{G^T \text{ is a generalized inverse of } A^T}$.

soln] $AGA = A$. Transpose both sides so $A^T G^T A^T = A^T$ So $G^T \overset{"}{:=} (A^T)^-$.

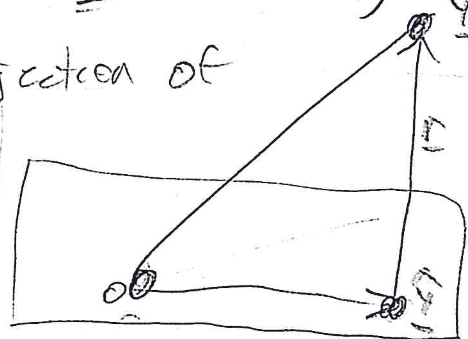
ii) If A is symmetric and G is a generalized inverse of A, show G^T is a generalized inverse of A

soln] by i) $AG^TA = A$ so $G^T \overset{"}{:=} A^-$

iii) If A is symmetric show that $\exists$ a symmetric generalized inverse of A.

Soln $\frac{G + G^T}{2}$ works

26) The least squares estimator $\underline{\hat{Y}} \in C(X)$ of $\underline{Y}$ is $\underline{\hat{Y}} = P_X \underline{Y} \in C(X)$, $\underline{Y} \in \mathbb{R}^n$. Then $\underline{\hat{Y}}$ is the projection of $\underline{Y}$ on $C(X)$.



$\underline{r} = \underline{Y} - \underline{\hat{Y}} = (I - P_X)\underline{Y}$
$= (I - H)\underline{Y}$

$\underline{\hat{Y}} = P_X \underline{Y} = H\underline{Y} = X\underline{\hat{\beta}} =$
$\underline{v}_1 \hat{\beta}_1 + \cdots + \underline{v}_p \hat{\beta}_p$

ex) Find a matrix $A \ni \mathbb{R}^k = C(A)$.

Soln Find a basis for $\mathbb{R}^k$ and use the basis vectors as columns of A.

So $A = I_k$. Note $A = P_A = P_{\mathbb{R}^k}$.

27) p471 If $A\underline{x} = \underline{0}$ $\forall \underline{x}$ then $A = 0$

28) p464 For an orthogonal projection matrix P_X, $\operatorname{rank}(P_X) = \underbrace{tr(P_X)}_{\text{trace}} = \sum_{i=1}^{n} P_{X_{ii}}$

29) p460-1 *Let A be symmetric. A is <u>positive definite</u> if $\underline{x}'A\underline{x} > 0$ $\forall \underline{x} \neq \underline{0}$. $(A > 0)$

A is <u>positive semidefinite</u> if $\underline{x}'A\underline{x} \geq 0$ $\forall \underline{x}$. $(A \geq 0)$

30] If A is positive definite then all eigenvalues of $A > 0$; $\lambda_i > 0$. If A is positive semidefinite then all eigenvalues of A satisfy $\lambda_i \geq 0$ $\quad A \geq 0 \Rightarrow \lambda_i \geq 0$

31] p461 If A is positive definite so is A^{-1}. $\quad A > 0$ then $A^{-1} > 0$

32] p461 If X is $\overbrace{n \times p}^{\text{so } n \geq p}$ of rank p then $X^T X$ is positive definite.

33] p464 If A is positive definite $\exists$ positive definite square root matrix $A^{\frac{1}{2}}$ $\ni$ $A^{\frac{1}{2}} A^{\frac{1}{2}} = A$, $\quad A = (A^{\frac{1}{2}})^2$ $\quad A^{\frac{1}{2}} = T \Lambda^{\frac{1}{2}} T^T$ see 34].

✓34] p458 For a symmetric matrix A, the <u>spectral decomposition</u> of A is

$A = T \Lambda T^T$ where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$,

λ_i are the eigenvalues of A, $T = [\underline{t}_1, \ldots, \underline{t}_n]$,

$A \underline{t}_i = \lambda_i \underline{t}_i$. So the $\underline{t}_i$ are orthonormal eigenvectors of A: $\underline{t}_i^T \underline{t}_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$

and T is an orthogonal matrix: $T^T T = I = T T^T$.

A has a set of n orthonormal eigenvectors and $C(A)$ is the space spanned by those eigenvectors corresponding to the nonzero eigenvalues.

36] p461 If A is positive semidefinite then $X^T A X = 0 \Rightarrow AX = 0$

matrix

Note! May refer back to appendices.

Ch 1] p 1-2 know $tr(A)$, $A' = A^T$, $det(A)$, A^{-1}, $C(A)$, $N(A)$.

2] p3 The response variable Y is the variable you want to predict. The explanatory variables $x_1, \ldots, x_p$ are used to predict Y.

3] Use regression models for description, prediction, and hypothesis testing.

4] p4 Conditioning is usually suppressed when finding $E(Y|\underset{\sim}{x})$ or $E\{Y|\underset{\sim}{X}=\underset{\sim}{x}\}$ 7.5

So $\underbrace{E\{Y\}}_{\text{notation}} = E(Y|X_1=x_1, \ldots, X_k=x_k) = E\{Y|\underset{\sim}{X}=\underset{\sim}{x}\}$

response

$= \beta_0+\beta_1 x_1 + \cdots + \beta_k x_k.$

§1.4 5] p4 Now Let $\underset{\sim}{X} = [X_{ij}] = \begin{bmatrix} X_{11} & \cdots & X_{1n} \\ \vdots & & \\ X_{m1} & & X_{mn} \end{bmatrix}$

be a random matrix.

Then $E(\underset{\sim}{X}) = [E\,X_{ij}]$, if all expectations exist. A random vector $\underset{\sim}{X}$ is a special case.

6] Unless told otherwise assume expectations exist and that conformable matrices are used.

7] Let A, B, and C be constant matrices.

Then $E[A Z B + C] = A\,E[Z]B + C$

8.5' If second moments exist,

the population mean $E(\underset{\sim}{X}) = \underset{\sim}{\mu} = E\begin{pmatrix} X_1 \\ X_n \end{pmatrix}$

$= \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{pmatrix}$ and the $n \times n$ population

covariance matrix $\text{COV}(\underset{\sim}{X}) = \text{Var}(\underset{\sim}{X}) = \Sigma_{\underset{\sim}{X}}$

$= E[(\underset{\sim}{X} - E(\underset{\sim}{X}))(\underset{\sim}{X} - E\underset{\sim}{X})^T]$

← like short cut rule: $V(X) = EX^2 - (EX)^2$

$= E(\underset{\sim}{X}\underset{\sim}{X}^T) - E(\underset{\sim}{X})E(\underset{\sim}{X}^T) = (\sigma_{ij})$,

so the ij entry of $\text{Cov}(\underset{\sim}{X})$ is

$\text{COV}(X_i, X_j) = \sigma_{ij}$.

9] P6 Let $\underset{\sim}{X}$ be $p \times 1$ and $\underset{\sim}{Y}$ $q \times 1$.

Assuming expected values exist, the

$p \times q$ pop covariance matrix of $\underset{\sim}{X}$ and $\underset{\sim}{Y}$

is $\text{Cov}(\underset{\sim}{X}, \underset{\sim}{Y}) = E[(\underset{\sim}{X} - E(\underset{\sim}{X}))(\underset{\sim}{Y} - E(\underset{\sim}{Y}))^T]$

$= E\underset{\sim}{X}\underset{\sim}{Y}^T - E(\underset{\sim}{X})E(\underset{\sim}{Y}^T) = \Sigma_{\underset{\sim}{X}\underset{\sim}{Y}}$. $\text{Cov}(\underset{\sim}{X}, \underset{\sim}{X}) = \text{Cov}(\underset{\sim}{X})$

Notation: sometimes use $\underline{x}$ for a random vector, but when conditioning $Y|\underset{\sim}{X} = \underline{x}$ makes more sense. So boldface x could be a matrix or vector. On board a $\underset{\sim}{x}$ indicates a vector. 8.5

10) Let $\underline{a}$ be a constant vector and A, B constant matrices (conformable). Then

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

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

$$E[A \underset{\sim}{X} B] = A\, E(\underset{\sim}{X})\, B$$

$$\mathrm{Cov}(\underset{\sim}{a} + A\underset{\sim}{X}) = \mathrm{Cov}(A\underset{\sim}{X}) = A\, \mathrm{Cov}(\underset{\sim}{X})\, A^T$$

11) p7 If $\underset{\sim}{X}$ is $m \times 1$ and Y $n \times 1$ are random vectors and A, B are constant matrices, then

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

12) know p9 Let $\underset{\sim}{X}$ be a RV with $E\underset{\sim}{X} = \underset{\sim}{\mu}$ and $\mathrm{Cov}(\underset{\sim}{X}) = \Sigma$. Then