Math 584 Exam 1 is on Tuesday, Feb. 23. You are allowed 8 sheets of notes and a calculator. CHECK FORMULAS: YOU ARE RESPONSIBLE FOR ANY ERRORS ON THIS HANDOUT!

Types of problems likely to appear on Exam 1 are numbered:

Let M be a nonempty subset of a vector space V. If i) $\alpha \boldsymbol{x} \in M \ \forall \boldsymbol{x} \in M$ and for any scalar α, and ii) $\boldsymbol{x} + \boldsymbol{y} \in M \ \forall \boldsymbol{x}, \boldsymbol{y} \in M$, then M is a vector space known as a subspace.

The set of all linear combinations of $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ is the vector space known as span$(\boldsymbol{x}_1, ..., \boldsymbol{x}_n)$.

Let $\boldsymbol{x}_1, ..., \boldsymbol{x}_k \in V$. If $\exists$ scalars $\alpha_1, ..., \alpha_k$ not all zero such that $\sum_{i=1}^k \alpha_i \boldsymbol{x}_i = \boldsymbol{0}$, then $\boldsymbol{x}_1, ..., \boldsymbol{x}_k$ are *linearly dependent.* If $\sum_{i=1}^k \alpha_i \boldsymbol{x}_i = \boldsymbol{0}$ only if $\alpha_i = 0 \ \forall \ i = 1, ..., k$, then $\boldsymbol{x}_1, ..., \boldsymbol{x}_k$ are *linearly independent.*

Suppose $\{\boldsymbol{x}_1, ..., \boldsymbol{x}_k\}$ is a linearly independent set and $V = span(\boldsymbol{x}_1, ..., \boldsymbol{x}_k)$. Then $\{\boldsymbol{x}_1, ..., \boldsymbol{x}_k\}$ is a linearly independent spanning set for V, known as a *basis.*

Let $\boldsymbol{A} = [\boldsymbol{a}_1 \ \boldsymbol{a}_2 \ ... \ \boldsymbol{a}_m] = \begin{bmatrix} \boldsymbol{r}_1^T \\ \vdots \\ \boldsymbol{r}_n^T \end{bmatrix}$ be an $n \times m$ matrix. The space spanned by the columns of $\boldsymbol{A}$ = **column space** of $\boldsymbol{A} = C(\boldsymbol{A})$.

Let $\boldsymbol{X} = [\boldsymbol{v}_1 \ \boldsymbol{v}_2 \ ... \ \boldsymbol{v}_p] = \begin{bmatrix} \boldsymbol{x}_1^T \\ \vdots \\ \boldsymbol{x}_n^T \end{bmatrix}$ be an $n \times p$ matrix. Note that

$C(\boldsymbol{X}) = \{\boldsymbol{y} \in \mathbb{R}^n : \boldsymbol{y} = \boldsymbol{X\beta}$ for some $\boldsymbol{\beta} \in \mathbb{R}^p\}$. (If the function $\boldsymbol{X}_f(\boldsymbol{\beta}) = \boldsymbol{X\beta}$ where the f indicates that the operation $\boldsymbol{X}_f : \mathbb{R}^p \to \mathbb{R}^n$ is being treated as a function, then $C(\boldsymbol{X})$ is the range of $\boldsymbol{X}_f$.)

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

Let $\boldsymbol{A}$ be $m \times n$. Then rank$(\boldsymbol{A})$ = rank$(\boldsymbol{A}^T) \leq \min(m, n)$. If rank$(\boldsymbol{A}) = \min(m, n)$, then $\boldsymbol{A}$ has *full rank*, or $\boldsymbol{A}$ is a full rank matrix.

The *row space* of $\boldsymbol{A} = C(\boldsymbol{A}^T)$, the span of the rows of $\boldsymbol{A}$.

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

$V_M^\perp = \{\boldsymbol{y} \in M : \boldsymbol{y} \perp V\}$ is the *orthogonal complement of* V with respect to M where $\boldsymbol{y} \perp V$ means $\boldsymbol{y}^T \boldsymbol{x} = \boldsymbol{0} \ \forall \ \boldsymbol{x} \in V$. If $M = \mathbb{R}^k$, then the subspace $V^\perp = \{\boldsymbol{y} \in \mathbb{R}^k : \boldsymbol{y} \perp V\}$ is the *orthogonal complement of* V.

$N(\boldsymbol{A}^T) = [C(\boldsymbol{A})]^\perp$, so $N(\boldsymbol{A}) = [C(\boldsymbol{A}^T)]^\perp$.

Rank Nullity Theorem: Let $\boldsymbol{A}$ be $m \times n$. Then rank$(\boldsymbol{A})$+ dim$(N(\boldsymbol{A})) = n$.

A **generalized inverse** of an $m \times n$ matrix $\boldsymbol{A}$ is any $n \times m$ matrix $\boldsymbol{A}^-$ satisfying $\boldsymbol{AA}^-\boldsymbol{A} = \boldsymbol{A}$. Other names are conditional inverse, pseudo inverse, g-inverse, and p-inverse. Usually a generalized inverse is not unique, but if $\boldsymbol{A}^{-1}$ exists, then $\boldsymbol{A}^- = \boldsymbol{A}^{-1}$ is unique. Notation: $\boldsymbol{G} := \boldsymbol{A}^-$ means $\boldsymbol{G}$ is a generalized inverse of $\boldsymbol{A}$.

1) **Know:** Be able to show that $\boldsymbol{G} := \boldsymbol{A}^-$.

$\boldsymbol{A}$ is **idempotent** if $\boldsymbol{A}^2 = \boldsymbol{A}$.

2) **Know:** Be able to show whether $\boldsymbol{A}$ is idempotent.

Let V be a subspace of $\mathbb{R}^k$. Then every $\boldsymbol{y} \in \mathbb{R}^k$ can be expressed uniquely as $\boldsymbol{y} = \boldsymbol{w} + \boldsymbol{z}$ where $\boldsymbol{w} \in V$ and $\boldsymbol{z} \in V^\perp$.

Let $\boldsymbol{X} = [\boldsymbol{v}_1\ \boldsymbol{v}_2\ ...\ \boldsymbol{v}_p]$ be $n \times p$, and let $V = C(\boldsymbol{X}) = \text{span}(\boldsymbol{v}_1, ..., \boldsymbol{v}_p)$. Then the $n \times n$ matrix $\boldsymbol{P}_V = \boldsymbol{P}_{\boldsymbol{X}}$ is a **projection matrix** on $C(\boldsymbol{X})$ if $\boldsymbol{P}_{\boldsymbol{X}}\ \boldsymbol{y} = \boldsymbol{w}\ \forall\ \boldsymbol{y} \in \mathbb{R}^n$. (Here $\boldsymbol{y} = \boldsymbol{w} + \boldsymbol{z} = \boldsymbol{w}_{\boldsymbol{y}} + \boldsymbol{z}_{\boldsymbol{y}}$, so $\boldsymbol{w}$ depends on $\boldsymbol{y}$.)

Theorem: a) $\boldsymbol{P}_{\boldsymbol{X}}$ is unique.
b) $\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^-\boldsymbol{X}^T$ where $(\boldsymbol{X}^T\boldsymbol{X})^-$ is any generalized inverse of $\boldsymbol{X}^T\boldsymbol{X}$.
c) $\boldsymbol{A}$ is a projection matrix on $C(\boldsymbol{A})$ iff $\boldsymbol{A}$ is symmetric and idempotent. Hence $\boldsymbol{P}_{\boldsymbol{X}}$ is a projection matrix on $C(\boldsymbol{P}_{\boldsymbol{X}}) = C(\boldsymbol{X})$.
d) $\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}}$ is the projection matrix on $[C(\boldsymbol{X})]^\perp$.
e) $\boldsymbol{A} = \boldsymbol{P}_{\boldsymbol{X}}$ iff i) $\boldsymbol{y} \in C(\boldsymbol{X})$ implies $A\boldsymbol{y} = \boldsymbol{y}$ and ii) $\boldsymbol{y} \perp C(\boldsymbol{X})$ implies $\boldsymbol{A}\boldsymbol{y} = \boldsymbol{0}$.

Theorem: a) $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{X} = \boldsymbol{X}$, and $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{W} = \boldsymbol{W}$ if each column of $\boldsymbol{W} \in C(\boldsymbol{X})$.
b) $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{v}_i = \boldsymbol{v}_i$.
c) If $C(\boldsymbol{X}_R)$ is a subspace of $C(\boldsymbol{X})$, then $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{P}_{\boldsymbol{X}_R} = \boldsymbol{P}_{\boldsymbol{X}_R}\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{P}_{\boldsymbol{X}_R}$.
d) Let $\boldsymbol{X} = [\boldsymbol{Z}\ \boldsymbol{X}_r]$ where $\text{rank}(\boldsymbol{X}) = \text{rank}(\boldsymbol{X}_r) = r$ so the columns of $\boldsymbol{X}_r$ form a basis for $C(\boldsymbol{X})$. Then

$$\begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & (\boldsymbol{X}_r^T\boldsymbol{X}_r)^{-1} \end{bmatrix}$$

is a generalized inverse of $\boldsymbol{X}^T\boldsymbol{X}$, and $\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X}_r(\boldsymbol{X}_r^T\boldsymbol{X}_r)^{-1}\boldsymbol{X}_r^T$.

3) **Know:** Be able to find $\boldsymbol{P}_{\boldsymbol{X}}$ for small $\boldsymbol{X}$, perhaps by finding a basis for $C(\boldsymbol{X})$.

Notation: The matrix $\boldsymbol{A}$ in a quadratic form $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$ is **symmetric.** $\boldsymbol{A}$ is **positive definite** ($\boldsymbol{A} > 0$) if $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} > 0\ \forall\ \boldsymbol{x} \neq \boldsymbol{0}$. $\boldsymbol{A}$ is **positive semidefinite** ($\boldsymbol{A} \geq 0$) if $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \geq 0\ \forall\ \boldsymbol{x}$. If $\boldsymbol{A} \geq 0$ then the eigenvalues of $\boldsymbol{A}$ are real and nonnegative. If $\boldsymbol{A} \geq 0$, let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$. If $\boldsymbol{A} > 0$, then $\lambda_n > 0$.

Theorem: Let $\boldsymbol{A}$ be a $n \times n$ symmetric matrix with eigenvector eigenvalue pairs $(\lambda_1, \boldsymbol{t}_1), (\lambda_2, \boldsymbol{t}_2), ..., (\lambda_n, \boldsymbol{t}_n)$ where $\boldsymbol{t}_i^T\boldsymbol{t}_i = 1$ and $\boldsymbol{t}_i^T\boldsymbol{t}_j = 0$ for $i = 1, ..., n$. Hence $\boldsymbol{A}\boldsymbol{t}_i = \lambda_i\boldsymbol{t}_i$. Then the *spectral decomposition* of $\boldsymbol{A}$ is

$$\boldsymbol{A} = \sum_{i=1}^n \lambda_i\boldsymbol{t}_i\boldsymbol{t}_i^T = \lambda_1\boldsymbol{t}_1\boldsymbol{t}_1^T + \cdots + \lambda_n\boldsymbol{t}_n\boldsymbol{t}_n^T.$$

Let $\boldsymbol{T} = [\boldsymbol{t}_1\ \ \boldsymbol{t}_2\ \ \cdots\ \ \boldsymbol{t}_n]$ be the $n \times n$ orthogonal matrix with ith column $\boldsymbol{t}_i$. Then $\boldsymbol{T}\boldsymbol{T}^T = \boldsymbol{T}^T\boldsymbol{T} = \boldsymbol{I}$. Let $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, ..., \lambda_n)$ and let $\boldsymbol{\Lambda}^{1/2} = \text{diag}(\sqrt{\lambda_1}, ..., \sqrt{\lambda_n})$. If $\boldsymbol{A}$ is a positive definite $n \times n$ symmetric matrix with spectral decomposition $\boldsymbol{A} = \sum_{i=1}^n \lambda_i\boldsymbol{t}_i\boldsymbol{t}_i^T$, then $\boldsymbol{A} = \boldsymbol{T}\boldsymbol{\Lambda}\boldsymbol{T}^T$ and

$$\boldsymbol{A}^{-1} = \boldsymbol{T}\boldsymbol{\Lambda}^{-1}\boldsymbol{T}^T = \sum_{i=1}^n \frac{1}{\lambda_i}\boldsymbol{t}_i\boldsymbol{t}_i^T.$$

The *square root matrix* $\boldsymbol{A}^{1/2} = \boldsymbol{T}\boldsymbol{\Lambda}^{1/2}\boldsymbol{T}^T$ is a positive definite symmetric matrix such that $\boldsymbol{A}^{1/2}\boldsymbol{A}^{1/2} = \boldsymbol{A}$.

ch. 1

The **response variable** Y is the variable you want to predict. The *explanatory variables* $X_1, ..., X_k$ are used to predict Y.

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

For the response variable, conditioning is suppressed. So $E(Y) = E(Y|\boldsymbol{X} = \boldsymbol{x})$ or $E(Y) = E(Y|\boldsymbol{X})$. So $E(Y) = E(Y|X_1 = x_1, ..., X_k = x_k) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$.

Let $\boldsymbol{X} = (X_{ij})$ be a random matrix. Then $E(\boldsymbol{X}) = (E(X_{ij}))$.

Notation: Unless told otherwise, assume expectations exist and that conformable matrices and vectors are used.

The *population mean* of a random $n \times 1$ vector $\boldsymbol{x} = (x_1, ..., x_n)^T$ is $E(\boldsymbol{x}) = \boldsymbol{\mu} = (E(x_1), ..., E(x_n))^T$ and the $n \times n$ *population covariance matrix* $\text{Cov}(\boldsymbol{x}) = \boldsymbol{\Sigma}_{\boldsymbol{x}} = E(\boldsymbol{x} - E(\boldsymbol{x}))(\boldsymbol{x} - E(\boldsymbol{x}))^T = (\sigma_{i,j})$ where $\text{Cov}(x_i, x_j) = \sigma_{i,j}$. The *population covariance matrix* of $\boldsymbol{x}$ with $\boldsymbol{y}$ is

$$\text{Cov}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{\Sigma}_{\boldsymbol{x},\boldsymbol{y}} = E[(\boldsymbol{x} - E(\boldsymbol{x}))(\boldsymbol{y} - E(\boldsymbol{y}))^T].$$

4) **Know:** If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are $n \times 1$ random vectors, $\boldsymbol{a}$ a conformable constant vector, and $\boldsymbol{A}$ and $\boldsymbol{B}$ are conformable constant matrices, then

$$E(\boldsymbol{X} + \boldsymbol{Y}) = E(\boldsymbol{X}) + E(\boldsymbol{Y}), \ E(\boldsymbol{a} + \boldsymbol{Y}) = \boldsymbol{a} + E(\boldsymbol{Y}), \ \& \ E(\boldsymbol{AXB}) = \boldsymbol{A}E(\boldsymbol{X})\boldsymbol{B}.$$

Also

$$\text{Cov}(\boldsymbol{a} + \boldsymbol{AX}) = \text{Cov}(\boldsymbol{AX}) = \boldsymbol{A}\text{Cov}(\boldsymbol{X})\boldsymbol{A}^T.$$

Note that $E(\boldsymbol{AY}) = \boldsymbol{A}E(\boldsymbol{Y})$ and $\text{Cov}(\boldsymbol{AY}) = \boldsymbol{A}\text{Cov}(\boldsymbol{Y})\boldsymbol{A}^T$.

5) If $\boldsymbol{X}$ $(m \times 1)$ and $\boldsymbol{Y}$ $(n \times 1)$ are random vectors, and $\boldsymbol{A}$ and $\boldsymbol{B}$ are conformable constant matrices, then

$$\text{Cov}(\boldsymbol{AX}, \boldsymbol{BY}) = \boldsymbol{A}\text{Cov}(\boldsymbol{X}, \boldsymbol{Y})\boldsymbol{B}^T.$$

6) **Theorem 1.5, expected value of a quadratic form:** Let $\boldsymbol{X}$ be a random vector with $E(\boldsymbol{X}) = \boldsymbol{\mu}$ and $\text{Cov}(\boldsymbol{X}) = \boldsymbol{\Sigma}$. Then

$$E(\boldsymbol{X}^T\boldsymbol{AX}) = tr(\boldsymbol{A\Sigma}) + \boldsymbol{\mu}^T\boldsymbol{A\mu}.$$

Ch. 3

If $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $E(\boldsymbol{X}) = \boldsymbol{\mu}$, $\text{Cov}(\boldsymbol{X}) = \boldsymbol{\Sigma}$, and $m_{\boldsymbol{X}}(\boldsymbol{t}) = \exp(\boldsymbol{t}^T\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t})$.

7) If $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and if $\boldsymbol{A}$ is a $q \times p$ matrix, then $\boldsymbol{AX} \sim N_q(\boldsymbol{A\mu}, \boldsymbol{A\Sigma A}^T)$. If $\boldsymbol{a}$ $(p \times 1)$ and $\boldsymbol{b}$ $(q \times 1)$ are constant vectors, then $\boldsymbol{X} + \boldsymbol{a} \sim N_p(\boldsymbol{\mu} + \boldsymbol{a}, \boldsymbol{\Sigma})$ and $\boldsymbol{AX} + \boldsymbol{b} \sim N_q(\boldsymbol{A\mu} + \boldsymbol{b}, \boldsymbol{A\Sigma A}^T)$.

$$\text{Let} \ \ \boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix}, \ \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \ \text{and} \ \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

8) **All subsets of a MVN are MVN:** $(X_{k_1}, ..., X_{k_q})^T \sim N_q(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}})$ where $\tilde{\boldsymbol{\mu}}_i = E(X_{k_i})$ and $\tilde{\boldsymbol{\Sigma}}_{ij} = \text{Cov}(X_{k_i}, X_{k_j})$. In particular, $\boldsymbol{X}_1 \sim N_q(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $\boldsymbol{X}_2 \sim N_{p-q}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$. If $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent iff $\boldsymbol{\Sigma}_{12} = \boldsymbol{0}$.

$$\text{Let} \ \begin{pmatrix} Y \\ X \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_Y \\ \mu_X \end{pmatrix}, \ \begin{pmatrix} \sigma_Y^2 & \text{Cov}(Y, X) \\ \text{Cov}(X, Y) & \sigma_X^2 \end{pmatrix} \right).$$

Also recall that the *population correlation* between X and Y is given by

$$\rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{VAR}(X)}\sqrt{\text{VAR}(Y)}} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}$$

if $\sigma_X > 0$ and $\sigma_Y > 0$.

9) **Know:** The conditional distribution of a MVN is MVN. If $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the conditional distribution of $\boldsymbol{X}_1$ given that $\boldsymbol{X}_2 = \boldsymbol{x}_2$ is multivariate normal with mean $\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2)$ and covariance matrix $\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$. That is,

$$\boldsymbol{X}_1 | \boldsymbol{X}_2 = \boldsymbol{x}_2 \sim N_q(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}).$$

Notation:

$$\boldsymbol{X}_1 | \boldsymbol{X}_2 \sim N_q(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{X}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}).$$

10) **Know:** Be able to compute the above quantities if X_1 and X_2 are scalars.

11) Theorem 2.5. Let $\boldsymbol{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let $\boldsymbol{U} = \boldsymbol{AY}$ and $\boldsymbol{W} = \boldsymbol{BY}$. Then $\boldsymbol{AY} \perp\!\!\!\perp \boldsymbol{BY}$ iff $\text{Cov}(\boldsymbol{U}, \boldsymbol{W}) = \boldsymbol{A\Sigma B}^T = \boldsymbol{0}$ iff $\boldsymbol{B\Sigma A}^T = \boldsymbol{0}$. Note that if $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_n$, then $\boldsymbol{AY} \perp\!\!\!\perp \boldsymbol{BY}$ iff $\boldsymbol{AB}^T = \boldsymbol{0}$ iff $\boldsymbol{BA}^T = \boldsymbol{0}$.

12) **Theorem 2.7**. Let $\boldsymbol{Y} \sim N_n(\boldsymbol{0}, \boldsymbol{I}_n)$, and let $\boldsymbol{A} = \boldsymbol{A}^T$ be symmetric. Then $\boldsymbol{Y}^T\boldsymbol{AY} \sim \chi^2_r$ iff $\boldsymbol{A}$ is idempotent of rank r.

13) If $\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y}$, then $g(\boldsymbol{X}) \perp\!\!\!\perp h(\boldsymbol{Y})$ where g is a vector valued function of $\boldsymbol{X}$ alone and h is a vector valued function of $\boldsymbol{Y}$ alone.

14) Theorem: Let $\boldsymbol{Y} \sim N_n(\boldsymbol{0}, \boldsymbol{I}_n)$, with $\boldsymbol{A}$ and $\boldsymbol{B}$ symmetric. If $\boldsymbol{Y}^T\boldsymbol{AY} \sim \chi^2_r$ and $\boldsymbol{Y}^T\boldsymbol{BY} \sim \chi^2_d$, then $\boldsymbol{Y}^T\boldsymbol{AY} \perp\!\!\!\perp \boldsymbol{Y}^T\boldsymbol{BY}$ iff $\boldsymbol{AB} = \boldsymbol{0}$.

15) Cor. of Th 2.8. If $\boldsymbol{Y} \sim N_n(\boldsymbol{0}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma} > 0$, and $\boldsymbol{A}$ is symmetric, then $\boldsymbol{Y}^T\boldsymbol{AY} \sim \chi^2_r$ iff $\boldsymbol{A\Sigma}$ is idempotent of rank r.

16) Th. 2.9. If $\boldsymbol{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the population squared Mahalanobis distance $(\boldsymbol{Y} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}) \sim \chi^2_n$.

17) Suppose $Y_1, ..., Y_n$ are independent $N(\mu_i, 1)$ random variables so that $\boldsymbol{Y} = (Y_1, ..., Y_n)^T \sim N_n(\boldsymbol{\mu}, \boldsymbol{I}_n)$. Then $\boldsymbol{Y}^T\boldsymbol{Y} = \sum_{i=1}^n Y_i^2 \sim \chi^2(n, \boldsymbol{\mu}^T\boldsymbol{\mu}/2)$, a *noncentral* χ^2 *distribution*, $(\chi^2(n,\gamma))$, with n degrees of freedom and *noncentrality parameter* $\gamma = \boldsymbol{\mu}^T\boldsymbol{\mu}/2 = \frac{1}{2}\sum_{i=1}^n \mu_i^2 \geq 0$. The noncentrality parameter $\delta = \boldsymbol{\mu}^T\boldsymbol{\mu} = 2\gamma$ is also used. Note that if $Y \sim N(\mu, 1)$ then $Y^2 \sim \chi^2(n = 1, \gamma = \mu^2/2)$, and if $Y \sim N(\sqrt{2\gamma}, 1)$, then $Y^2 \sim \chi^2(n = 1, \gamma)$.

18) If $W \sim \chi^2_n$, then $W \sim \chi^2(n, 0)$ so $\gamma = 0$. The χ^2_n distribution is also called the *central* χ^2 *distribution.*

19) a) If $Y \sim \chi^2(n,\gamma)$, then the mgf of Y is $m_Y(t) = (1-2t)^{-n/2}\exp(-\gamma[1-(1-2t)^{-1}])$ for $t < 0.5$.

b) If $Y_i \sim \chi^2(n_i, \gamma_i)$ are independent for $i = 1, ..., k$, then $\displaystyle\sum_{i=1}^k Y_i \sim \chi^2\left(\sum_{i=1}^k n_i, \sum_{i=1}^k \gamma_i\right)$.

c) If $Y \sim \chi^2(n,\gamma)$, then $E(Y) = n + 2\gamma$ and $V(Y) = 2n + 8\gamma$.

20) If $Y_1, ..., Y_k$ are independent with mgf's $m_{Y_i}(t)$ then the mgf of $\sum_{i=1}^k Y_i$ is

$$m_{\sum_{i=1}^k Y_i}(t) = \prod_{i=1}^k m_{Y_i}(t).$$

21) Theorem: If $\boldsymbol{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} > 0$, then $\boldsymbol{Y}^T\boldsymbol{AY} \sim \chi^2(\text{rank}(\boldsymbol{A}), \boldsymbol{\mu}^{\text{T}}\boldsymbol{A\mu}/2)$ iff $\boldsymbol{A\Sigma}$ is idempotent. If $\boldsymbol{\Sigma} = \boldsymbol{I}_n$, the result holds iff $\boldsymbol{A}$ is idempotent. Note $\boldsymbol{A} = \boldsymbol{A}^T$.

22) **Craig's Theorem:** Let $\boldsymbol{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
a) If $\boldsymbol{\Sigma} > 0$, then $\boldsymbol{Y}^T \boldsymbol{A} \boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{Y}^T \boldsymbol{B} \boldsymbol{Y}$ iff $\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B} = \boldsymbol{0}$ iff $\boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{A} = \boldsymbol{0}$.
b) If $\boldsymbol{\Sigma} \geq 0$, then $\boldsymbol{Y}^T \boldsymbol{A} \boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{Y}^T \boldsymbol{B} \boldsymbol{Y}$ if $\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B} = \boldsymbol{0}$ (or if $\boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{A} = \boldsymbol{0}$).
c) If $\boldsymbol{\Sigma} \geq 0$, then $\boldsymbol{Y}^T \boldsymbol{A} \boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{Y}^T \boldsymbol{B} \boldsymbol{Y}$ iff
$(*)$ $\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}\boldsymbol{\Sigma} = \boldsymbol{0}, \boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}\boldsymbol{\mu} = \boldsymbol{0}, \boldsymbol{\Sigma}\boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\mu} = \boldsymbol{0}$, and $\boldsymbol{\mu}^T \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}\boldsymbol{\mu} = 0$.
Note that if $\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B} = \boldsymbol{0}$, then $(*)$ holds.

23) One way to show $C(\boldsymbol{A}) = C(\boldsymbol{B})$ is to show that i) $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{B}\boldsymbol{y} \in C(\boldsymbol{B})$ and ii) $\boldsymbol{B}\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \in C(\boldsymbol{A})$.

ch. 3

24) Then the (full rank) **multiple linear regression (MLR) model** is $Y_i = x_{i,0}\beta_0 + x_{i,1}\beta_1 + x_{i,2}\beta_2 + \cdots + x_{i,p-1}\beta_{p-1} + \epsilon_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + \epsilon_i$ for $i = 1, \ldots, n$. For the (ordinary) least squares (OLS) model, the ϵ_i are uncorrelated and usually iid with $E(\epsilon_i) = 0$ and $V(\epsilon_i) = \sigma^2$, an unknown positive parameter. Usually $x_{i,0} = 1$ for $i = 1, ..., n$. In matrix form the model is $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where the $n \times p$ design matrix $\boldsymbol{X}$ has full rank $p \leq n$. Also, $\boldsymbol{X}$ is treated as a constant matrix and $\boldsymbol{\beta}$ as an unknown constant vector. If $\boldsymbol{X}$ is a random matrix, condition on $\boldsymbol{X}$.

25) Given an estimate $\boldsymbol{b}$ of $\boldsymbol{\beta}$, the corresponding vector of *predicted* or *fitted values* is $\widehat{\boldsymbol{Y}} \equiv \widehat{\boldsymbol{Y}}(\boldsymbol{b}) = \boldsymbol{X}\boldsymbol{b}$. Thus the ith fitted value

$$\hat{Y}_i \equiv \hat{Y}_i(\boldsymbol{b}) = \boldsymbol{x}_i^T \boldsymbol{b} = x_{i,0} b_0 + \cdots + x_{i,p-1} b_{p-1}.$$

The vector of *residuals* is $\boldsymbol{e}(\boldsymbol{b}) = \boldsymbol{Y} - \widehat{\boldsymbol{Y}}(\boldsymbol{b})$. Thus ith residual $e_i(\boldsymbol{b}) = Y_i - \hat{Y}_i(\boldsymbol{b}) = Y_i - x_{i,0} b_0 - \cdots - x_{i,p-1} b_{p-1}$. Note $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{e}(\boldsymbol{b})$. Let $e_i = e_i(\hat{\boldsymbol{\beta}})$. So $\boldsymbol{Y} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{e}$.

26) The *least squares* (OLS) estimator $\hat{\boldsymbol{\beta}}$ minimizes $Q_{OLS}(\boldsymbol{b}) = \sum_{i=1}^n e_i^2(\boldsymbol{b})$ and $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$. The vector of *predicted* or *fitted values* $\widehat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{H}\boldsymbol{Y}$ where the *hat matrix* $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T = \boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{P}$ since $\boldsymbol{X}$ has full rank. The least squares *regression equation* is $\hat{Y} = \hat{\beta}_0 x_0 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_{p-1} x_{p-1}$ where $x_0 \equiv 1$ if the model contains a constant. The least squares vector of residuals is $\boldsymbol{e} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}}$. Thus ith residual $e_i = Y_i - \hat{Y}_i = Y_i - x_{i,0}\hat{\beta}_0 - \cdots - x_{i,p-1}\hat{\beta}_{p-1}$.

27) **Know:** Let $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where $\boldsymbol{X}$ is full rank, $E(\boldsymbol{\epsilon}) = \boldsymbol{0}$, $\text{Cov}(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}$ and $\boldsymbol{P} = \boldsymbol{P}_{\boldsymbol{X}}$ is the projection matrix on $C(\boldsymbol{X})$. Then $\widehat{\boldsymbol{Y}} = \boldsymbol{P}\boldsymbol{Y}$, $\boldsymbol{e} = (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{Y}$, and $\boldsymbol{P}\boldsymbol{X} = \boldsymbol{X}$ so $\boldsymbol{X}^T\boldsymbol{P} = \boldsymbol{X}^T$. Recall that $\boldsymbol{X}\boldsymbol{\beta}$ is treated as a constant vector. Then
i) $\boldsymbol{X}^T\boldsymbol{e} = \boldsymbol{X}^T(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{Y} = \boldsymbol{0}$. Hence the columns of $\boldsymbol{X}$, which correspond to the predictor variables, and the residual vector are orthogonal. This result is useful for the residual plot of the ith predictor variable versus the residuals.
ii) $E(\boldsymbol{Y}) = \boldsymbol{X}\boldsymbol{\beta}$.
iii) $\text{Cov}(\boldsymbol{Y}) = \text{Cov}(\boldsymbol{\epsilon}) = \sigma^2\boldsymbol{I}$.
iv) $\text{Cov}(\boldsymbol{e}, \widehat{\boldsymbol{Y}}) = \boldsymbol{0}$, an $n \times n$ matrix. Hence the fitted values are uncorrelated with the residuals. This result is useful for the residual plot of the fitted values versus the residuals.

28) **Know:** You should be able to show results such as those in 27).

29) **Know:** The least squares estimator $\hat{\boldsymbol{\beta}}$ satisfies the **normal equations** $\boldsymbol{X}^T\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}^T\boldsymbol{Y}$.

30) **Know:** Suppose $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where $\boldsymbol{X}$ is full rank and $E(\boldsymbol{\epsilon}) = \boldsymbol{0}$. Then a)

$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$. Hence $\hat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}$. b) If $\text{Cov}(\boldsymbol{Y}) = \text{Cov}(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}$, then $\text{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$.

31) The linear estimator $\boldsymbol{a}^T\boldsymbol{Y}$ of $\boldsymbol{c}^T\boldsymbol{\theta}$ is the best linear unbiased estimator (BLUE) of $\boldsymbol{c}^T\boldsymbol{\theta}$ if $E(\boldsymbol{a}^T\boldsymbol{Y}) = \boldsymbol{c}^T\boldsymbol{\theta}$, and if for any other unbiased linear estimator $\boldsymbol{b}^T\boldsymbol{Y}$ of $\boldsymbol{c}^T\boldsymbol{\theta}$, $V(\boldsymbol{a}^T\boldsymbol{Y}) \le V(\boldsymbol{b}^T\boldsymbol{Y})$. Note that $E(\boldsymbol{b}^T\boldsymbol{Y}) = \boldsymbol{c}^T\boldsymbol{\theta}$.

32) Let $\hat{\boldsymbol{\theta}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$ be the least squares estimator of $\boldsymbol{X}\boldsymbol{\beta}$ where $\boldsymbol{X}$ has full rank p. a) $\boldsymbol{c}^T\hat{\boldsymbol{\theta}}$ is the unique BLUE of $\boldsymbol{c}^T\boldsymbol{\theta}$. b) $\boldsymbol{a}^T\hat{\boldsymbol{\beta}}$ is the BLUE of $\boldsymbol{a}^T\boldsymbol{\beta}$ for every vector $\boldsymbol{a}$.

33) Let $Y = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where $\boldsymbol{X}$ has full rank p and the ϵ_i are iid with mean 0 and variance σ^2. Let the residual sum of squares (RSS) be $\text{RSS} = (\boldsymbol{Y} - \hat{\boldsymbol{Y}})^T(\boldsymbol{Y} - \hat{\boldsymbol{Y}}) = \sum_{i=1}^n e_i^2$. Let the $MSE = RSS/(n-p) = S^2$. a) Then MSE is an unbiased estimator of σ^2. b) Let $h_i = \boldsymbol{H}_{ii}$ where $\boldsymbol{H} = \boldsymbol{P}_{\boldsymbol{X}}$ is an $n \times n$ matrix. Then h_i is the ith leverage. If $\max h_i \to 0$ as $n \to \infty$ and if $E(\epsilon_i^4) = \gamma < \infty$, then MSE is a $\sqrt{n}$ consistent estimator of σ^2: $\sqrt{n}(MSE - \sigma^2) = O_P(1)$, implying $n^\delta(MSE - \sigma^2) \xrightarrow{P} 0$ if $0 < \delta < 0.5$.

34) Suppose $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, $\boldsymbol{X}$ full rank, $\boldsymbol{\epsilon} \sim N_n(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n)$, and $\boldsymbol{Y} \sim N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I}_n)$. Then a)

$$\hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}).$$

b)

$$\frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T\boldsymbol{X}^T\boldsymbol{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\sigma^2} \sim \chi_p^2.$$

c) $\hat{\boldsymbol{\beta}} \perp\!\!\!\perp$ MSE.
d)

$$\frac{RSS}{\sigma^2} = \frac{(n-p)MSE}{\sigma^2} \sim \chi^2_{n-p}.$$

35) Consider the MLR model $Y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + \epsilon_i$, and assume that the errors are independent with zero mean and the same variance: $E(\epsilon_i) = 0$ and $\text{V}(\epsilon_i) = \sigma^2$. Also assume that $\max_i(h_1, ..., h_n) \xrightarrow{P} 0$ as $n \to \infty$. Then

a) $\hat{Y}_i = \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}} \xrightarrow{P} E(Y_i|\boldsymbol{x}_i) = \boldsymbol{x}_i^T\boldsymbol{\beta}$ in probability for $i = 1, ..., n$ as $n \to \infty$.

b) All of the least squares estimators $\boldsymbol{a}^T\hat{\boldsymbol{\beta}}$ are asymptotically normal where $\boldsymbol{a}$ is any fixed constant $p \times 1$ vector.

c) (**Least squares CLT**:) Suppose that the ϵ_i are iid and

$$\frac{\boldsymbol{X}^T\boldsymbol{X}}{n} \to \boldsymbol{W}^{-1}.$$

Then the least squares (OLS) estimator $\hat{\boldsymbol{\beta}}$ satisfies

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N_p(\boldsymbol{0}, \sigma^2\ \boldsymbol{W}).$$

Also,

$$(\boldsymbol{X}^T\boldsymbol{X})^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N_p(\boldsymbol{0}, \sigma^2\ \boldsymbol{I}_p).$$

36) **Know:** If $\boldsymbol{Z}_n \xrightarrow{D} N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\boldsymbol{A}\boldsymbol{Z}_n + \boldsymbol{b} \xrightarrow{D} N_m(\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T)$ where $\boldsymbol{A}$ is an $m \times k$ constant matrix and $\boldsymbol{b}$ is an $m \times 1$ constant vector.

Problems from Quiz 1-3 and HW 1-4 are fair game. Appendices A,B, ch. 1, 2, and sections 3.1–3.4.