# Biostatistics/Statistics 533 <br> Classical Theory of Linear Models <br> Spring 2008 <br> Midterm Solutions 

Name: KEY
Problems do not have equal value and some problems will take more time than others. Spend your time wisely.

| Problem | $-1-$ | $-2-$ | $-3-$ | $-4-$ | $-5-$ | $-6-$ | Total |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Value | 20 | 19 | 20 | 13 | 8 | 10 | 90 |
| Median Score | 19 | 14 | 12 | 7 | 7 | 10 | 65.5 |

1. (20 points) Consider the linear model $\mathbf{Y}_{n \times 1}=\mathbf{X}_{n \times p} \boldsymbol{\beta}_{p \times 1}+$ $\boldsymbol{\varepsilon}_{n \times 1}$ with $E[\boldsymbol{\varepsilon}]=\mathbf{0}, \operatorname{cov}(\boldsymbol{\varepsilon})=\sigma^{2} \mathbf{I}$. Let $\mathbf{P}$ be the projection operator onto $\mathcal{R}(\mathbf{X})$. For least-squares estimation, recall that $\hat{\mathbf{Y}}=\mathbf{P Y}$. Derive the following. Simplify your answer as much as possible.
(a) $E(\hat{\mathbf{Y}})$

$$
E(\hat{\mathbf{Y}})=E(\mathbf{P Y})=\mathbf{P} E(\mathbf{Y})=\mathbf{P X} \boldsymbol{\beta}=\mathbf{X} \boldsymbol{\beta}
$$

(b) $\operatorname{cov}(\hat{\mathbf{Y}})$

$$
\operatorname{cov}(\hat{\mathbf{Y}})=\operatorname{cov}(\mathbf{P Y})=\mathbf{P} \operatorname{cov}(\mathbf{Y}) \mathbf{P}^{\prime}=\mathbf{P} \sigma^{2} \mathbf{I} \mathbf{P}^{\prime}=\sigma^{2} \mathbf{P}
$$

(c) $\operatorname{cov}(\hat{\mathbf{Y}}, \mathbf{Y})$

$$
\operatorname{cov}(\hat{\mathbf{Y}}, \mathbf{Y})=\operatorname{cov}(\mathbf{P Y}, \mathbf{Y})=\mathbf{P} \operatorname{cov}(\mathbf{Y}, \mathbf{Y})=\mathbf{P} \sigma^{2} \mathbf{I}=\sigma^{2} \mathbf{P}
$$

(d) $E\left[\hat{\mathbf{Y}}^{\prime} \hat{\mathbf{Y}}\right]$

$$
E\left[\hat{\mathbf{Y}}^{\prime} \hat{\mathbf{Y}}\right]=E\left[(\mathbf{P Y})^{\prime}(\mathbf{P} \mathbf{Y})\right]=E\left[\mathbf{Y}^{\prime} \mathbf{P Y}\right]=\operatorname{tr}(\mathbf{P} \operatorname{cov}(\mathbf{Y}))+E[\mathbf{Y}]^{\prime} \mathbf{P} E[\mathbf{Y}]
$$

$$
=\sigma^{2} \operatorname{tr}(\mathbf{P})+[\mathbf{X} \boldsymbol{\beta}]^{\prime} \mathbf{P}[\mathbf{X} \boldsymbol{\beta}]=\sigma^{2} \operatorname{tr}(\mathbf{P})+\boldsymbol{\beta}^{\prime} \mathbf{X}^{\prime} \mathbf{P} \mathbf{X} \boldsymbol{\beta}=\sigma^{2} r+\boldsymbol{\beta}^{\prime} \mathbf{X}^{\prime} \mathbf{X} \boldsymbol{\beta}
$$

where $r$ is the rank of $\mathbf{X}$ (and also the rank of $\mathbf{P}$ ).
2. (19 points) Let $\mathbf{X}$ be an $n \times p$ matrix with linearly independent columns. Let $\mathbf{Y}$ be an $n \times 1$ vector. Suppose $\mathbf{Y}=\mathbf{a}+\mathbf{b}$ where $\mathbf{a} \in \mathcal{R}(\mathbf{X})$ and $\mathbf{b} \in \mathcal{R}(\mathbf{X})^{\perp}$.
(a) (4 points)

Is a unique? Circle one: YES
Is $\mathbf{b}$ unique? Circle one: YES
(b) (2 points) Compute $\mathbf{a}^{\prime} \mathbf{b}$.

$$
\mathbf{a}^{\prime} \mathbf{b}=0
$$

(c) (3 points) Write a as a function of $\mathbf{X}$ and $\mathbf{Y}$. In the context of linear models, what do we typically call a (in words)?

$$
\mathbf{a}=\mathbf{X}\left(\mathbf{X}^{\prime} \mathbf{X}\right)^{-1} \mathbf{X}^{\prime} \mathbf{Y}
$$

"vector of fitted values"
(d) (3 points) Write $\mathbf{b}$ as a function of $\mathbf{X}$ and $\mathbf{Y}$. In the context of linear models, what do we typically call $\mathbf{b}$ (in words)?

$$
\mathbf{b}=\left(\mathbf{I}-\mathbf{X}\left(\mathbf{X}^{\prime} \mathbf{X}\right)^{-1} \mathbf{X}^{\prime}\right) \mathbf{Y}=\mathbf{Y}-\mathbf{X}\left(\mathbf{X}^{\prime} \mathbf{X}\right)^{-1} \mathbf{X}^{\prime} \mathbf{Y}
$$

"residual vector" or "vector of fitted errors"
(e) (3 points) Compute $\mathbf{X}^{\prime} \mathbf{b}$.

$$
\text { since } \mathbf{b} \in R(X)^{\perp}, \mathbf{X}^{\prime} \mathbf{b}=0
$$

(f) (4 points) Consider a linear model with intercept:

$$
y=\beta_{0}+\beta_{1} x_{1}+\beta_{2} x_{2}+\ldots+\beta_{p} x_{p}+\epsilon
$$

Let $\hat{\epsilon}_{i}$ be the residuals from the fitted model. Prove $\sum \hat{\epsilon}_{i}=0$. Since $\hat{\boldsymbol{\varepsilon}} \in R(X)^{\perp}, \mathbf{X}^{\prime} \hat{\boldsymbol{\varepsilon}}=0$. In particular, $\mathbf{X}$ has a column of 1's corresponding to the intercept in the model. So $\mathbf{1}^{\prime} \hat{\boldsymbol{\varepsilon}}=$ $\sum \hat{\epsilon}_{\mathbf{i}}=\mathbf{0}$.
3. (a) (10 points) We know a least-squares estimate of $\boldsymbol{\beta}$ must satisfy the normal equations $\mathbf{X}^{\prime} \mathbf{X} \boldsymbol{\beta}=\mathbf{X}^{\prime} \mathbf{Y}$.
Prove that $\hat{\boldsymbol{\beta}}=\mathbf{G X}^{\prime} \mathbf{Y}$ satisfies the normal equations, where $\mathbf{G}$ is a generalized inverse of $\mathbf{X}^{\prime} \mathbf{X}$.

Solution 1 (intended solution); In the left-hand side, replace $\mathbf{X}^{\prime} \mathbf{X}$ with $\mathbf{X}^{\prime} \mathbf{X G} \mathbf{X}^{\prime} \mathbf{X}$. The normal equations become:

$$
\begin{aligned}
\mathbf{X}^{\prime} \mathbf{X G X} \mathbf{X} \mathbf{X} & =\mathbf{X}^{\prime} \mathbf{Y} \\
\mathbf{X}^{\prime} \mathbf{X G}\left(\mathbf{X}^{\prime} \mathbf{X} \boldsymbol{\beta}\right) & =\mathbf{X}^{\prime} \mathbf{Y} \\
\mathbf{X}^{\prime} \mathbf{X G}\left(\mathbf{X}^{\prime} \mathbf{Y}\right) & =\mathbf{X}^{\prime} \mathbf{Y} \\
\left(\mathbf{X}^{\prime} \mathbf{X}\right)\left(\mathbf{G X}^{\prime} \mathbf{Y}\right) & =\mathbf{X}^{\prime} \mathbf{Y}
\end{aligned}
$$

Therefore, $\mathbf{G X}^{\prime} \mathbf{Y}$ is a solution to the normal equations.
Solution 2: If you use the lemma given in part (b), the problem is almost trivial:

$$
\begin{aligned}
\mathbf{X}^{\prime} \mathbf{X}\left(\mathbf{G X}^{\prime} \mathbf{Y}\right) & =\left(\mathbf{X}^{\prime} \mathbf{X} \mathbf{G} \mathbf{X}^{\prime}\right) \mathbf{Y} \\
& =\left(\mathbf{X}^{\prime}\right) \mathbf{Y}
\end{aligned}
$$

(b) (10 points) For a linear model with $E(\boldsymbol{\varepsilon})=0$ and $\operatorname{cov}(\boldsymbol{\varepsilon})=$ $\sigma^{2} \mathbf{I}$, let $\mathbf{c}^{\prime} \boldsymbol{\beta}$ be estimable. Show $\operatorname{cov}\left(\mathbf{c}^{\prime} \hat{\boldsymbol{\beta}}\right)=\sigma^{2} \mathbf{c}^{\prime} \mathbf{G c}$ where $\mathbf{G}$ is a generalized inverse of $\mathbf{X}^{\prime} \mathbf{X}$. (You might find some of the following facts that we proved in a lemma useful: $\mathbf{X G X} \mathbf{X}^{\prime} \mathbf{X}=$ $\left.\mathbf{X}, \mathbf{X}^{\prime} \mathbf{X G X} \mathbf{X}^{\prime}=\mathbf{X}^{\prime}, \mathbf{X G X} \mathbf{X}^{\prime}=\mathbf{X G}^{\prime} \mathbf{X}^{\prime}\right)$
(Hint: use what we know about $\mathbf{c}$ when $\mathbf{c}^{\prime} \boldsymbol{\beta}$ is estimable)

$$
\begin{aligned}
\operatorname{cov}\left(\mathbf{c}^{\prime} \hat{\boldsymbol{\beta}}\right) & =\operatorname{cov}\left(\mathbf{c}^{\prime} \mathbf{G} \mathbf{X}^{\prime} \mathbf{Y}\right) \\
& =\mathbf{c}^{\prime} \mathbf{G} \mathbf{X}^{\prime} \sigma^{2} I \mathbf{X} \mathbf{G}^{\prime} \mathbf{c} \\
& =\sigma^{2} \mathbf{c}^{\prime} \mathbf{G} \mathbf{X}^{\prime} \mathbf{X} \mathbf{G}^{\prime} \mathbf{c}
\end{aligned}
$$

Since $\mathbf{c}^{\prime} \boldsymbol{\beta}$ is estimable, $\mathbf{c} \in R\left(\mathbf{X}^{\prime}\right)$ and $\mathbf{c}^{\prime}=\mathbf{T}_{1 \times n} \mathbf{X}_{n \times p}$ for some T

$$
\begin{aligned}
\operatorname{cov}(\mathbf{c} \hat{\boldsymbol{\beta}}) & =\sigma^{2} \mathbf{T} \underbrace{\mathbf{X G} \mathbf{X}^{\prime} \mathbf{X}}_{\mathbf{X}} \mathbf{G}^{\prime} \mathbf{X}^{\prime} \mathbf{T}^{\prime} \\
& =\sigma^{2} \mathbf{T} \mathbf{X G} \mathbf{X}^{\prime} \mathbf{T}^{\prime} \\
& =\sigma^{2} \mathbf{T} \mathbf{X G} \mathbf{X}^{\prime} \mathbf{T}^{\prime} \\
& =\sigma^{2} \mathbf{c}^{\prime} \mathbf{G} \mathbf{c}
\end{aligned}
$$

4. This linear model with intercept has $p+1$ parameters:
$y=\alpha+\beta_{1} x_{1}+\beta_{2} x_{2}+\ldots+\beta_{p} x_{p}+\epsilon$
Let $\mathbf{X}=\left[\mathbf{x}_{1}, \mathbf{x}_{2}, \ldots, \mathbf{x}_{p}\right]$ be the design matrix for the parameters $\beta_{1}, \beta_{2}, \ldots, \beta_{p}$. The least squares parameter estimates minimize $\|\mathbf{Y}-\alpha \mathbf{1}-\mathbf{X} \boldsymbol{\beta}\|$ or equivalently, $\|\mathbf{Y}-\alpha \mathbf{1}-\mathbf{X} \boldsymbol{\beta}\|^{2}$, where $\boldsymbol{\beta}$ is the vector $\left(\beta_{1}, \beta_{2}, \ldots, \beta_{p}\right)^{\prime}$.
(a) (8 points) Let $\overline{\mathbf{Y}}=\bar{Y} \mathbf{1}$ (the vector where every entry is the average of the Y 's) and let $\overline{\mathbf{X}}$ be the matrix with all rows equal to $\overline{\mathbf{x}}^{\prime}=\left[\overline{\mathbf{x}}_{1}, \overline{\mathbf{x}}_{2}, \ldots, \overline{\mathbf{x}}_{p}\right]$. Prove that $\|\mathbf{Y}-\alpha \mathbf{1}-\mathbf{X} \boldsymbol{\beta}\|^{2}$ can be written

$$
\|\overline{\mathbf{Y}}-\alpha \mathbf{1}-\overline{\mathbf{X}} \boldsymbol{\beta}\|^{2}+\|\mathbf{Y}-\overline{\mathbf{Y}}-(\mathbf{X}-\overline{\mathbf{X}}) \boldsymbol{\beta}\|^{2}
$$

$$
\|Y-\alpha 1-X \beta\|^{2}=\|\mathbf{Y}-\alpha \mathbf{1}-\mathbf{X} \boldsymbol{\beta}-\overline{\mathbf{Y}}+\overline{\mathbf{Y}}-\overline{\mathbf{X}} \boldsymbol{\beta}+\overline{\mathbf{X}} \boldsymbol{\beta}\|^{2}=
$$

$$
\|\overline{\mathbf{Y}}-\alpha \mathbf{1}-\overline{\mathbf{X}} \boldsymbol{\beta}\|^{2}+\|\mathbf{Y}-\overline{\mathbf{Y}}-(\mathbf{X}-\overline{\mathbf{X}}) \beta\|^{\mathbf{2}} \text { as long as we can }
$$ show $\bar{Y}-\alpha \mathbf{1}-\bar{X} \boldsymbol{\beta}$ and $\mathbf{Y}-\overline{\mathbf{Y}}-(\mathbf{X}-\overline{\mathbf{X}}) \boldsymbol{\beta}$ are orthogonal. Notice that $\overline{\mathbf{Y}}-\alpha \mathbf{1}-\overline{\mathbf{X}} \boldsymbol{\beta}$ is an $\mathrm{n} \times 1$ vector with constant entries. So we can write $\overline{\mathbf{Y}}-\alpha \mathbf{1}-\overline{\mathbf{X}} \boldsymbol{\beta}=c \mathbf{1}$ for some constant c. But $\mathbf{1}^{\prime}(\mathbf{Y}-\overline{\mathbf{Y}})=0$ and $\mathbf{1}^{\prime}(\mathbf{X}-\overline{\mathbf{X}})=0$ so we have established orthogonality.

(b) ( 5 points) Let $\hat{\boldsymbol{\beta}}$ be the least squares estimate of $\boldsymbol{\beta}$. Prove that the least-squares estimate of $\alpha$ is $\hat{\alpha}=\bar{y}-\overline{\mathbf{x}}^{\prime} \boldsymbol{\beta}$.

From (a), minimizing RSS is equivalent to minimizing $\| \overline{\mathbf{Y}}-$ $\alpha \mathbf{1}-\overline{\mathbf{X}} \boldsymbol{\beta}\left\|^{2}+\right\| \mathbf{Y}-\overline{\mathbf{Y}}-(\mathbf{X}-\overline{\mathbf{X}}) \boldsymbol{\beta} \|^{2}$. If $\hat{\boldsymbol{\beta}}$ is the LSE of $\boldsymbol{\beta}$, then we can make the first term equal to 0 by setting $\hat{\alpha}=\bar{y}-\bar{x}^{\prime} \hat{\boldsymbol{\beta}}$. Since $\alpha$ only appears in the first term and since the first term is $\geq 0$, this establishes that $\hat{\alpha}=\bar{y}-\bar{x}^{\prime} \hat{\boldsymbol{\beta}}$ is the LSE of $\alpha$.
5. (8 points) Let $\mathbf{Y} \sim M V N\left(\mathbf{X} \boldsymbol{\beta}, \sigma^{2} \mathbf{I}\right)$, where $\mathbf{X}$ is an $n \times p$ matrix with linearly independent columns.
(a) What is the distribution of $(\mathbf{I}-\mathbf{P}) \mathbf{Y}$, where $\mathbf{P}$ is the projection onto the column space of $\mathbf{X}$ ?
$E((\mathbf{I}-\mathbf{P}) \mathbf{Y})=(\mathbf{I}-\mathbf{P}) E(\mathbf{Y})=(\mathbf{I}-\mathbf{P}) \mathbf{X} \boldsymbol{\beta}=\mathbf{X} \boldsymbol{\beta}-\mathbf{P X} \boldsymbol{\beta}=$ $\mathbf{X} \boldsymbol{\beta}-\mathbf{X} \boldsymbol{\beta}=\mathbf{0}$
$\operatorname{cov}((\mathbf{I}-\mathbf{P}) \mathbf{Y}))=(\mathbf{I}-\mathbf{P}) \operatorname{cov}(\mathbf{Y})(\mathbf{I}-\mathbf{P})^{\prime}=\sigma^{2}(\mathbf{I}-\mathbf{P})$
So $(\mathbf{I}-\mathbf{P}) \mathbf{Y} \sim M V N\left(\mathbf{0}, \sigma^{2}(\mathbf{I}-\mathbf{P})\right)$.
(b) What is the distribution of $\mathbf{Y}^{\prime}(\mathbf{I}-\mathbf{P}) \mathbf{Y} / \sigma^{2}$ ?

$$
\frac{\mathbf{Y}^{\prime}(\mathbf{I}-\mathbf{P}) \mathbf{Y}}{\sigma^{2}}=\frac{R S S}{\sigma^{2}} \sim \chi_{n-p}^{2}
$$

6. (10 points) Argue directly from the definition of estimability and the facts we know about estimable quantities that if the design matrix $\mathbf{X}$ has full rank, every $\mathbf{a}^{\prime} \boldsymbol{\beta}$ is estimable and in particular every individual parameter $\beta_{i}$ is estimable.
Solution 1: Since $\mathbf{X}$ has full rank, the least squares estimate of $\beta$ is

$$
\hat{\beta}=\left(\mathbf{X}^{\prime} \mathbf{X}\right)^{-1} \mathbf{X}^{\prime} Y
$$

which satisfies $E[\hat{\beta}]=\beta$. Hence

$$
\mathbf{a}^{\prime} E[\hat{\beta}]=E\left[\mathbf{a}^{\prime}\left(\mathbf{X}^{\prime} \mathbf{X}\right)^{-1} \mathbf{X}^{\prime} Y\right]=\mathbf{a}^{\prime} \beta
$$

So take $\mathbf{b}^{\prime}=\mathbf{a}^{\prime}\left(\mathbf{X}^{\prime} \mathbf{X}\right)^{-1} \mathbf{X}^{\prime}$. To see that every component $\beta_{i}$ is estimable, take $\mathbf{a}^{\prime}=(0, \ldots, 0,1,0, \ldots, 0)$, where the 1 is in the $i$ th column.

Solution 2: Since $\mathbf{X}$ has full rank, the rows of $\mathbf{X}$ span $\mathbb{R}^{p}$. Therefore, every $p$-dimensional vector $\mathbf{a}$ is in the row space of $\mathbf{X}$ so every $\mathbf{a}^{\prime} \boldsymbol{\beta}$ is estimable. In particular, each $\mathbf{e}_{\mathbf{i}}$ is in the row space of $\mathbf{X}$, where $\mathbf{e}_{\mathbf{i}}$ is the $p$-vector with 0 in every position except for a 1 in the $i^{\text {th }}$ position. Therefore every $\beta_{i}$ is estimable.

