Homework 6

Due Thursday, 10/24/19

1. (What is a "hat" matrix?) Throughout this problem, let $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ and $\boldsymbol{H} \in \mathbb{R}^{n \times n}$ be such that

   (i) $\boldsymbol{H}^T = \boldsymbol{H}$.

   (ii) $\boldsymbol{H}^2 = \boldsymbol{H}$.

   (iii) $\operatorname{im}(\boldsymbol{H}) = \operatorname{im}(\boldsymbol{X})$.

   If $\boldsymbol{H}$ satisfies properties (i), (ii) and (iii), we will call $\boldsymbol{H}$ the "hat" matrix for $\boldsymbol{X}$.

   (a) Show that $\boldsymbol{H}$ is an orthogonal projection matrix that projects vectors onto the image of $\boldsymbol{X}$. That is, show that for any $\boldsymbol{v} \in \mathbb{R}^n$

   $$(\boldsymbol{v} - \boldsymbol{H}\boldsymbol{v})^T (\boldsymbol{H}\boldsymbol{v}) = 0$$

   and

   $$\boldsymbol{H}\boldsymbol{v} = \underset{\boldsymbol{u} \in \operatorname{im}(\boldsymbol{X})}{\arg\min} \|\boldsymbol{v} - \boldsymbol{u}\|_2^2.$$

   (The proof should be identical to 2c on HW 4.)

   (b) Show that if $\boldsymbol{H}$ satisfies properties (i), (ii) and (iii), it is unique. That is, if $\boldsymbol{P} \in \mathbb{R}^{n \times n}$ is another matrix such that

   (i) $\boldsymbol{P}^T = \boldsymbol{P}$.

   (ii) $\boldsymbol{P}^2 = \boldsymbol{P}$.

   (iii) $\operatorname{im}(\boldsymbol{P}) = \operatorname{im}(\boldsymbol{X})$,

   then $\boldsymbol{H} = \boldsymbol{P}$.

   (c) Define $\boldsymbol{H} = \boldsymbol{X}\boldsymbol{X}^\dagger$. Use properties 1-4 of the Moore-Penrose pseudoinverse to show that $\boldsymbol{H}$ is the hat matrix for $\boldsymbol{X}$.

2. (The Gauss-Markov Theorem) Suppose $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ is a non-random, full rank design matrix and $\boldsymbol{\beta} \in \mathbb{R}^p$ is unknown. You will prove the **Gauss-Markov Theorem**:

   > Suppose $\mathbb{E}(\boldsymbol{\epsilon}) = \boldsymbol{0}_n$ and $\operatorname{Var}(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}_n$. If $\hat{\boldsymbol{\beta}}$ is the ordinary least squares estimator and $\tilde{\boldsymbol{\beta}}$ is any other linear unbiased estimator of $\boldsymbol{\beta}$, then $\operatorname{Var}\left(\tilde{\boldsymbol{\beta}}\right) = \operatorname{Var}\left(\hat{\boldsymbol{\beta}}\right) + \boldsymbol{M}$ for some symmetric and positive semi-definite matrix $\boldsymbol{M}$.

   $\hat{\boldsymbol{\beta}}$ is also called the B.L.U.E (Best Linear Unbiased Estimator). The proof when $\boldsymbol{X}$ is not full rank is almost identical to what you will show below.

   (a) What is the ordinary least squares estimator for $\boldsymbol{\beta}$? What is its variance? What is the hat matrix, $\boldsymbol{H}$?

(b) Now let $\tilde{\boldsymbol{\beta}} = \boldsymbol{A}^T\boldsymbol{Y}$ be another linear unbiased estimator for $\boldsymbol{\beta}$, where $\boldsymbol{A} \in \mathbb{R}^{n\times p}$. Since $\tilde{\boldsymbol{\beta}}$ must be an unbiased estimator regardless of the value for $\boldsymbol{\beta}$, show that

$$\boldsymbol{A}^T\boldsymbol{X} = \boldsymbol{I}_p.$$

What is $\mathrm{Var}\left(\tilde{\boldsymbol{\beta}}\right)$?

(c) Let $\hat{\boldsymbol{\beta}}$ be the ordinary least squares estimator from part (i). Show that

$$\mathrm{Var}\left(\tilde{\boldsymbol{\beta}}\right) = \mathrm{Var}\left(\hat{\boldsymbol{\beta}}\right) + \boldsymbol{M},$$

where $\boldsymbol{M}$ is a symmetric and positive semi-definite matrix. (Hint: $\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{A}^T\boldsymbol{H}\boldsymbol{A} + \boldsymbol{A}^T(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{A}$)

(d) For any $\boldsymbol{q} \in \mathbb{R}^p$, show that part (c) implies

$$\mathrm{Var}\left(\boldsymbol{q}^T\tilde{\boldsymbol{\beta}}\right) \geq \mathrm{Var}\left(\boldsymbol{q}^T\hat{\boldsymbol{\beta}}\right).$$

(e) Now suppose the true model for $\boldsymbol{Y}$ is

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$
$$\mathbb{E}(\boldsymbol{\epsilon}) = \boldsymbol{0}_n, \quad \mathrm{Var}(\boldsymbol{\epsilon}) = \sigma^2\boldsymbol{\Sigma},$$

where $\boldsymbol{\Sigma}$ is a known, invertible matrix (you saw an example when $\boldsymbol{\Sigma}$ was a diagonal matrix on the midterm). Let $\boldsymbol{R}$ be an invertible matrix such that $\boldsymbol{\Sigma} = \boldsymbol{R}\boldsymbol{R}^T$ (such an $\boldsymbol{R}$ is always guaranteed to exist).

(i) Let $\tilde{\boldsymbol{Y}} = \boldsymbol{R}^{-1}\boldsymbol{Y}$ and $\tilde{\boldsymbol{X}} = \boldsymbol{R}^{-1}\boldsymbol{X}$. What is $\mathbb{E}\left(\tilde{\boldsymbol{Y}}\right)$? What is $\mathrm{Var}\left(\tilde{\boldsymbol{Y}}\right)$?

(ii) What is the B.L.U.E for $\boldsymbol{\beta}$ under this new model for $\boldsymbol{Y}$ in terms of $\boldsymbol{X}, \boldsymbol{\Sigma}$ and $\boldsymbol{Y}$? This is called the **generalized least squares** estimate for $\boldsymbol{\beta}$.

3. KNNL: 6.10 a, c and 7.4 b (in their notation, b_j is $\hat{\beta}_j$).

4. KNNL 6.16 a, b, c