Homework 6

蒋翌坤 20307100013

1. (a) We can take a look at the second derivative. $\frac{d^2}{db^2}\sum_{i=1}^n (y_i - x_i^T b)^2 = 2\sum_{i=1}^n x_i^2 > 0$ ($x_i \neq 0$, $x_i^2 = (x_{i1}^2, \dots, x_{ip}^2)^T$). Since the second derivative is a positive constant, which means the function is convex. Therefore, the minimum is unique at the point where the first derivative is 0, thus guaranteed $\hat{\beta}$ to be unique.

(b) The geometric interpretation of $\hat{\beta}$ is that under the transformation X, vector $\hat{\beta}$ is closest to vector y.

When solving the least squares problem, we get a system of linear equations by letting derivative equal zero: $X^T y - X^T X\hat{\beta} = 0$. For this system of linear equations to have an existing and unique solution, we need $X^T X$ to have full rank, that is $\text{rank}(X^T X) = p$. So we need to assume $\text{rank}(X^T X) = \text{rank}(X) = p$ to guarantee the existence and uniqueness of $\hat{\beta}$.

2. Let $X^T X = PDP^T$, P orthogonal and $D = \text{diag}(\lambda_1, \dots, \lambda_p)$, λ_i is eigenvalue of $X^T X$, so $\text{Cov}(\hat{\beta}) = \sigma^2 (X^T X)^{-1} = \sigma^2 PD^{-1}P^T$. Suppose λ_a is very small, then λ_a^{-1} is very large. So, $\text{Var}(\hat{\beta}_i) = \sum_{j=1}^p p_{ij}^2 \lambda_j^{-1}$ is very large. Therefore, the effect of a small eigenvalue is that every $\hat{\beta}_i$ has a large variance.

3. Suppose there are another linear unbiased estimator $\tilde{\beta}$, we can write it as $\tilde{\beta} = Ay = Cy + \hat{\beta}$ for some $A = C + (X^T X)^{-1}X^T$. For $\tilde{\beta}$ to be unbiased, we need $\mathbb{E}(\tilde{\beta}) = C\mathbb{E}(y) + \beta = \beta \Rightarrow CX\beta = 0 \Rightarrow CX = 0$.

$\text{Cov}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$

$\text{Cov}(\tilde{\beta}) = \text{Cov}(Cy + \hat{\beta}) = \text{Cov}((C + (X^T X)^{-1}X^T)y) = \sigma^2 (C + (X^T X)^{-1}X^T)(C + (X^T X)^{-1}X^T)^T = \sigma^2 (CC^T + CX(X^T X)^{-1} + (X^T X)^{-1}X^T C^T + (X^T X)^{-1}) = \sigma^2 CC^T + \text{Cov}(\hat{\beta})$

For any constant v, $\text{Var}(v^T\tilde{\beta}) = v^T\text{Cov}(\tilde{\beta})v = \sigma^2 v^T CC^T v + \text{Var}(v^T\hat{\beta})$. $v^T CC^T v = (C^T v)^T (C^T v) \geq 0$, it equals 0 iff $C^T v = 0 \Rightarrow C = 0$. so for any $\tilde{\beta} \neq \hat{\beta}$, $C \neq 0$, $\text{Var}(v^T\tilde{\beta}) > \text{Var}(v^T\hat{\beta})$. Therefore, $\hat{\beta}$ is the best linear unbiased estimator.

□

4. $\mathbb{E}(e^T e) = \mathbb{E}(y^T (I-H)^T (I-H) y) = \beta^T X^T (I-H)^T (I-H) X\beta + \text{tr}((I-H)^T (I-H)\sigma^2 I)$

$(I-H)X\beta = (I - X(X^T X)^{-1}X^T)X\beta = X\beta - X(X^T X)^{-1}X^T X\beta = 0,$

$H^T = (X(X^T X)^{-1}X^T)^T = X(X^T X)^{-1}X^T = H,$

$H^2 = X(X^T X)^{-1}X^T X(X^T X)^{-1}X^T = X(X^T X)^{-1}X^T = H,$

$(I-H)^T(I-H) = (I-H)^2 = I - H,$

$\text{tr}(H) = \text{tr}(X(X^T X)^{-1}X^T) = \text{tr}(X^T X(X^T X)^{-1}) = \text{tr}(I_p) = p,$

so $\text{tr}((I-H)^T(I-H)\sigma^2 I) = \sigma^2\text{tr}(I-H) = \sigma^2(n - \text{tr}(H)) = \sigma^2(n-p)$

Therefore, $\frac{\mathbb{E}(e^T e)}{n-p} = \sigma^2$

□

5. For every b, we can order $e_i = y_i - x_i^T b$ and get $e_{(1)}, \dots, e_{(n)}$, suppose $e_{(m)} < 0 \leq e_{(m+1)}$ for some m (if $m = 0$, then $e_{(1)} \geq 0$; if $m = n$, then $e_{(n)} < 0$), then $\sum_{i=1}^n |e_i| = \sum_{j=m+1}^n e_{(j)} - \sum_{i=1}^m e_{(i)}$. Under this condition, $\arg\min_b$ can easily be found since it is a linear function of b. Note that we can divide $\mathbb{R}^p$ into $\mathcal{S}_0, \mathcal{S}_1, \dots, \mathcal{S}_n$ such that when $b \in \mathcal{S}_m$, $e_{(m)} < 0 \leq e_{(m+1)}$. Therefore, we can find the overall $\arg\min_b$ by checking all $b \in \mathcal{S}_0, \mathcal{S}_1, \dots, \mathcal{S}_n$.