Name: Yuxuan Zhang

Date: October 1, 2024

HW4

1. We know that $\mathbb{V}\text{ar}(c^T\hat{\beta}) = c^T\,\mathbb{V}\text{ar}(\hat{\beta})c$, and $\mathbb{V}\text{ar}(c^T\check{\beta}c) = c^T\,\mathbb{V}\text{ar}(\check{\beta})c$. Then, we have

$$\mathbb{V}\text{ar}(c^T\check{\beta}) - \mathbb{V}\text{ar}(c^T\hat{\beta}) = c^T\left(\mathbb{V}\text{ar}(\check{\beta}) - \mathbb{V}\text{ar}(\hat{\beta})\right)c.$$

Since we know that $\mathbb{V}\text{ar}(\check{\beta}) - \mathbb{V}\text{ar}(\hat{\beta})$ is a positively defined matrix, for any non-zero vector c, we have

$$c^T\left(\mathbb{V}\text{ar}(\check{\beta}) - \mathbb{V}\text{ar}(\hat{\beta})\right)c > 0.$$

This implies that

$$\begin{aligned}\mathbb{V}\text{ar}(c^T\check{\beta}) - \mathbb{V}\text{ar}(c^T\hat{\beta}) &> 0\\ \mathbb{V}\text{ar}(c^T\check{\beta}) &> \mathbb{V}\text{ar}(c^T\hat{\beta}).\end{aligned}$$

Thus, the variance of $c^T\hat{\beta}$ is smaller than the variance of $c^T\check{\beta}$. This completes the proof.

2. (a) We know that for a random variable $\mathbf{X} \sim \text{N}(\mu, \mathbf{\Sigma})$ who follows a multivariate normal distribution, where $\mu = \begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}$ and $\mathbf{\Sigma} = \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}$, then the conditional distribution of $X_1|X_2 = a \sim \text{N}(\check{\mu}, \check{\Sigma})$ where $\check{\mu} = \mu_1 + \mathbf{\Sigma_{12}}\mathbf{\Sigma_{22}}^{-1}(a - \mu_2)$, and $\check{\Sigma} = \mathbf{\Sigma_{11}} - \mathbf{\Sigma_{12}}\mathbf{\Sigma_{22}}^{-1}\mathbf{\Sigma_{21}}$, according to https://en.wikipedia.org/wiki/Multivariate_normal_distribution. Let's apply this to our problem. Let $\tilde{y} = y - X\beta = Za + \epsilon$. Notice that $\text{Cov}(a, \epsilon) = 0$ since a and ϵ are uncorrelated. Then,

$$\mathbb{E}[\tilde{y}] = \mathbb{E}[Za + \epsilon] = Z\,\mathbb{E}[a] + \mathbb{E}[\epsilon] = 0,$$

$$\mathbb{V}\text{ar}[\tilde{y}] = \text{Cov}(Za + \epsilon) = Z\text{Cov}(a)Z^T + \text{Cov}(\epsilon) = Z\Psi Z^T + \sigma^2 I_n.$$

Also, notice that

$$\text{Cov}(\tilde{y}, a) = \text{Cov}(Za + \epsilon, a) = Z\text{Cov}(a) = Z\Psi.$$

We also know from the problem that $a \sim \text{N}(0, \Psi)$. Then, we can construct

$$\begin{bmatrix}a\\ \tilde{y}\end{bmatrix} \sim \text{N}\left(\begin{bmatrix}0\\ 0\end{bmatrix}, \begin{bmatrix}\Psi & \Psi Z^T\\ Z\Psi & Z\Psi Z^T + \sigma^2 I_n\end{bmatrix}\right).$$

Using the theorem mentioned earlier, we have

$$\mathbb{E}[a|\tilde{y}] = 0 + \Psi Z^T(Z\Psi Z^T + \sigma^2 I_n)^{-1}(y - X\beta - 0),$$

$$\text{Cov}(a|\tilde{y}) = \Psi - \Psi Z^T(Z\Psi Z^T + \sigma^2 I_n)^{-1}Z\Psi.$$

Thus, we have

$$a|y, \sigma^2, \Psi, \epsilon \sim \text{N}\left(\Psi Z^T(Z\Psi Z^T + \sigma^2 I_n)^{-1}(y - X\beta), \Psi - \Psi Z^T(Z\Psi Z^T + \sigma^2 I_n)^{-1}Z\Psi\right).$$

(b) $\hat{a}$ is the conditional expectation, which is the posterior mean of a, which we derived as

$$\hat{a} = \Psi Z^T (Z\Psi Z^T + \sigma^2 I_n)^{-1}(y - X\beta).$$

Substitute $y = X\beta + Za + \epsilon$, we get

$$\begin{aligned}\hat{a} &= \Psi Z^T (Z\Psi Z^T + \sigma^2 I_n)^{-1}(X\beta + Za + \epsilon - X\beta) \\ &= \Psi Z^T (Z\Psi Z^T + \sigma^2 I_n)^{-1}(Za + \epsilon).\end{aligned}$$

Taking the expectation conditional on a, since $\mathbb{E}[\epsilon] = 0$, we get

$$\begin{aligned}\mathbb{E}[\hat{a}|a] &= \mathbb{E}\left[\Psi Z^T (Z\Psi Z^T + \sigma^2 I_n)^{-1}(Za + \epsilon)\right] \\ &= \Psi Z^T (Z\Psi Z^T + \sigma^2 I_n)^{-1} Za + \Psi Z^T (Z\Psi Z^T + \sigma^2 I_n)^{-1}\,\mathbb{E}[\epsilon] \\ &= \Psi Z^T (Z\Psi Z^T + \sigma^2 I_n)^{-1} Za.\end{aligned}$$

The variance of $\hat{a}$ conditional on a, we have

$$\begin{aligned}\mathbb{V}\text{ar}(\hat{a}|a) &= \mathbb{V}\text{ar}(\Psi Z^T (Z\Psi Z^T + \sigma^2 I_n)^{-1}(Za + \epsilon)) \\ &= 0 + \mathbb{V}\text{ar}(\Psi Z^T (Z\Psi Z^T + \sigma^2 I_n)^{-1}\epsilon \\ &= \Psi Z^T (Z\Psi Z^T + \sigma^2 I_n)^{-1}\sigma^2 (Z\Psi Z^T + \sigma^2 I_n)^{-1} Z\Psi \\ &= \sigma^2 \Psi Z^T (Z\Psi Z^T + \sigma^2 I_n)^{-2} Z\Psi.\end{aligned}$$

For $\check{a}$, we substitute $y = X\beta + Za + \epsilon$ and get

$$\begin{aligned}\check{a} &= (Z^T Z)^{-1} Z^T (y - X\beta) \\ &= (Z^T Z)^{-1} Z^T (X\beta + Za + \epsilon - X\beta) \\ &= (Z^T Z)^{-1} Z^T (Za + \epsilon) \\ &= (Z^T Z)^{-1}(Z^T Z)a + (Z^T Z)^{-1} Z^T \epsilon \\ &= a + (Z^T Z)^{-1} Z^T \epsilon.\end{aligned}$$

Since $\epsilon \sim \mathrm{N}(0, \sigma^2 I_n)$, $\mathbb{E}[\check{a}|a] = a + \mathbb{E}[\epsilon] = a$. The variance of $\check{a}|a$ is

$$\begin{aligned}\mathbb{V}\text{ar}(\check{a}|a) &= \mathbb{V}\text{ar}(a + (Z^T Z)^{-1} Z^T \epsilon) \\ &= 0 + \mathbb{V}\text{ar}((Z^T Z)^{-1} Z^T \epsilon) \\ &= (Z^T Z)^{-1} Z^T \sigma^2 Z^T (Z^T Z)^{-1} \\ &= \sigma^2 (Z^T Z)^{-1}.\end{aligned}$$

(c) Same as the last question, we know from part (a) that

$$\hat{a} = \Psi Z^T (Z\Psi Z^T + \sigma^2 I_n)^{-1}(y - X\beta).$$

Then,

$$\begin{aligned}\mathbb{E}[\hat{a}] &= \mathbb{E}[\Psi Z^T (Z\Psi Z^T + \sigma^2 I_n)^{-1}(y - X\beta)] \\ &= \Psi Z^T (Z\Psi Z^T + \sigma^2 I_n)^{-1}\,\mathbb{E}[y - X\beta] \\ &= \Psi Z^T (Z\Psi Z^T + \sigma^2 I_n)^{-1} \cdot 0 \\ &= 0.\end{aligned}$$

The variance of $\hat{a}$ is

$$\begin{aligned}\mathbb{V}\text{ar}(\hat{a}) &= \mathbb{E}[(\hat{a}-\mathbb{E}[\hat{a}])(\hat{a}-\mathbb{E}[\hat{a}])^T] \\ &= \mathbb{E}[aa^T] \\ &= \Psi Z^T(Z\Psi Z^T+\sigma^2 I_n)^{-1}\,\mathbb{E}[(Za+\epsilon)(Za+\epsilon)^T](Z\Psi Z^T+\sigma^2 I_n)^{-1}Z\Psi^T.\end{aligned}$$

Notice that $\mathbb{E}[(Za+\epsilon)(Za+\epsilon)^T] = Z\,\mathbb{E}[aa^T]Z^T + \mathbb{E}[\epsilon\epsilon^T] = Z\Psi Z^T + \sigma^2 I_n$. Then, we have

$$\begin{aligned}\mathbb{V}\text{ar}(\hat{a}) &= \Psi Z^T(Z\Psi Z^T+\sigma^2 I_n)^{-1}(Z\Psi Z^T+\sigma^2 I_n)(Z\Psi Z^T+\sigma^2 I_n)^{-1}Z\Psi^T \\ &= \Psi Z^T(Z\Psi Z^T+\sigma^2 I_n)^{-1}Z\Psi.\end{aligned}$$

For $\check{a} = (Z^TZ)^{-1}Z^T(y-X\beta)$, substitute $y - X\beta = Za+\epsilon$, we have

$$\check{a} = (Z^TZ)^{-1}Z^T(Za+\epsilon).$$

Then, the expectation is

$$\begin{aligned}\mathbb{E}[\check{a}] &= \mathbb{E}[(Z^TZ)^{-1}Z^T(Za+\epsilon)] \\ &= (Z^TZ)^{-1}Z^T(Z\,\mathbb{E}[a]+\mathbb{E}[\epsilon]) \\ &= 0.\end{aligned}$$

The variance of $\check{a}$ is given by

$$\mathbb{V}\text{ar}(\check{a}) = \mathbb{E}[(\check{a}-\mathbb{E}[\check{a}])(\check{a}-\mathbb{E}[\check{a}])^T].$$

Again, since $\mathbb{E}[\check{a}]=0$, this simplifies to

$$\mathbb{V}\text{ar}(\check{a}) = \mathbb{E}[\check{a}\check{a}^T].$$

Substitute $\check{a} = (Z^TZ)^{-1}Z^T(Za+\epsilon)$, we have

$$\begin{aligned}\mathbb{V}\text{ar}(\check{a}) &= \mathbb{E}[(Z^TZ)^{-1}Z^T(Za+\epsilon)(Za+\epsilon)^TZ(Z^TZ)^{-1}] \\ &= (Z^TZ)^{-1}Z^T\,\mathbb{E}[(Za+\epsilon)(Za+\epsilon)^T]Z(Z^TZ)^{-1} \\ &= (Z^TZ)^{-1}Z^T(Z\Psi Z^T+\sigma^2 I_n)Z(Z^TZ)^{-1} \\ &= (Z^TZ)^{-1}Z^TZ\Psi Z^TZ(Z^TZ)^{-1} + \sigma^2(Z^TZ)^{-1}.\end{aligned}$$