Regression Analysis Assignment1 Solution

November 1, 2021

Problem1. Let $\mathrm{L}(\boldsymbol{\beta}^t, \lambda^t) = ||\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}||^2 + \sum_{j=1}^{q} \lambda_j (\mathbf{A}^t \boldsymbol{\beta} - \boldsymbol{c})_j$,where $\lambda = (\lambda_1, \ldots, \lambda_q)^t \in \mathbb{R}^q$. Suppose $\boldsymbol{\gamma} \in \mathbb{R}^{p+1}$ and $u = (u_1, u_2, \ldots u_q)^t \in \mathbb{R}^q$ satisfy following:

$$\left.\frac{\partial \mathrm{L}}{\partial \boldsymbol{\beta}_i}\right|_{\boldsymbol{\beta}=\boldsymbol{\gamma}, \lambda=u} = 0, \quad \text{for } i = 1, 2, \ldots, p+1 \tag{1}$$

$$\left.\frac{\partial \mathrm{L}}{\partial \lambda_i}\right|_{\boldsymbol{\beta}=\boldsymbol{\gamma}, \lambda=u} = 0 \quad \text{for } j = 1, 2, \ldots, q \tag{2}$$

Write $\mathbf{A} = (\mathbf{A_1}|\mathbf{A_2}|\cdots|\mathbf{A_q})$. Solving the equation (1) gives

$$\begin{aligned}\left.\frac{\partial \mathrm{L}}{\partial \boldsymbol{\beta}_i}\right|_{\boldsymbol{\beta}=\boldsymbol{\gamma}, \lambda=u} &= 2\sum_{k=1}^{p+1} (\mathbf{X^t X})_{ik} \gamma_k - 2(\mathbf{X^t Y})_i + \sum_{l=1}^{q} u_l (\mathbf{A}_l^t)_i \\ &= 2\sum_{k=1}^{p+1} (\mathbf{X^t X})_{ik} \gamma_k - 2(\mathbf{X^t Y})_i + \sum_{l=1}^{q} u_l \mathbf{A}_{il} = 0, \text{ for } i = 1, 2, \ldots p+1\end{aligned}$$

Reformulating this, we have $2(\mathbf{X^t X})\boldsymbol{\gamma} - 2\mathbf{X^t Y} + \mathbf{A}u = 0 \cdots (*)$. Solving equation (2) gives

$$\left.\frac{\partial \mathrm{L}}{\partial \lambda_i}\right|_{\boldsymbol{\beta}=\boldsymbol{\gamma}, \lambda=u} = \mathbf{A}_j^t \gamma_j - \mathbf{c}_j = 0, \mathbf{A}_j^t \gamma_j = \mathbf{c}_j \text{ for } j = 1, 2, \ldots, q.$$

Again, reformulation of this equation is given by $\mathbf{A^t}\boldsymbol{\gamma} = \mathbf{c} \cdots (**)$, which is our given constraint. To solve (*),

$$2(\mathbf{X^t X})\boldsymbol{\gamma} = 2\mathbf{X^t Y} - \mathbf{A}u, \quad , \boldsymbol{\gamma} = (\mathbf{X^t X})^{-1}\mathbf{X^t Y} - \frac{1}{2}(\mathbf{X^t X})^{-1}\mathbf{A}u$$

Plug in this into (**). Then we obtain

$$\begin{aligned}\mathbf{A}^t((\mathbf{X^t X})^{-1}\mathbf{X^t Y} - \frac{1}{2}(\mathbf{X^t X})^{-1}\mathbf{A}u) &= \mathbf{c} \\ \mathbf{A}^t(\mathbf{X^t X})^{-1}\mathbf{X^t Y} - \frac{1}{2}\mathbf{A^t}(\mathbf{X^t X})^{-1}\mathbf{A}u &= \mathbf{c} \\ -\frac{1}{2}\mathbf{A^t}(\mathbf{X^t X})^{-1}\mathbf{A}u &= -\mathbf{A}^t(\mathbf{X^t X})^{-1}\mathbf{X^t Y} + \mathbf{c}.\end{aligned}$$

So, $u = 2[\mathbf{A^t}(\mathbf{X^t X})^{-1}\mathbf{A}]^{-1}(\mathbf{A}^t(\mathbf{X^t X})^{-1}\mathbf{X^t Y} - \mathbf{c})$. Note that as $\mathbf{A}$ is full rank, $\mathbf{A^t}(\mathbf{X^t X})^{\mathbf{-1}}\mathbf{A}$ is indeed invertible. Thus we have

$$\boldsymbol{\gamma} = (\mathbf{X^t X})^{-1}\mathbf{X^t Y} - (\mathbf{X^t X})^{-1}\mathbf{A}[\mathbf{A^t}(\mathbf{X^t X})^{-1}\mathbf{A}]^{-1}(\mathbf{A}^t(\mathbf{X^t X})^{-1}\mathbf{X^t Y} - \mathbf{c})$$

By KKT condition, we already have $\hat{\boldsymbol{\beta}}_r = \boldsymbol{\gamma}$. Alternatively, one can easily verify $\boldsymbol{\gamma} = \underset{\boldsymbol{A^t \beta = c}}{\arg\min} ||\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}||$ by direct computation.

Problem2. Write $\mathbf{X}_{(k)} = (\mathbf{X}_1^t|\mathbf{X}_2^t|\cdots|\mathbf{X}_k^t)$ and $\mathbf{X}_{(-)} = (\mathbf{X}_{k+1}^t|\cdots|\mathbf{X}_n^t)$, so that $\mathbf{X} = (\mathbf{X}_{(k)}^t|\mathbf{X}_{(-)}^t)^t$. Similarly, write $\mathbf{Y}_{(k)} = (Y_1, \ldots, Y_k)^t$ and $\mathbf{Y}_{(-)} = (Y_{k+1}, \ldots, Y_n)^t$, so that $\mathbf{Y} = (\mathbf{Y}_{(k)}^t, \mathbf{Y}_{(-)}^t)^t$. Note that $\hat{\boldsymbol{\beta}}_{(-)} = (\mathbf{X}_{(-)}^t\mathbf{X}_{(-)})^{-1}\mathbf{X}_{(-)}^t\mathbf{Y}_{(-)}$. Since $||\mathbf{Y}_{(k)} - \mathbf{X}_{(k)}\boldsymbol{\beta}|| \geq 0$ for all $\boldsymbol{\beta} \in \mathbb{R}^{p+1}$ and the equality holds when $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$, by the uniqueness of minimizer of $g(\boldsymbol{\beta}) = ||\mathbf{Y}_{(k)} - \mathbf{X}_{(k)}\boldsymbol{\beta}||$, one can see that $\hat{\boldsymbol{\beta}} = (\mathbf{X}_{(k)}^t\mathbf{X}_{(k)})^{-1}\mathbf{X}_{(k)}^t\mathbf{Y}_{(k)}$. Because $\mathbf{X}$ is full rank, so are $\mathbf{X}_{(k)}$ and $\mathbf{X}_{(-)}$. Hence, $\mathbf{X}_{(k)}^t\mathbf{X}_{(k)}$ and $\mathbf{X}_{(-)}^t\mathbf{X}_{(-)}$ are invertible. Since $\mathbf{X}^t\mathbf{X} = \mathbf{X}_{(k)}^t\mathbf{X}_{(k)} + \mathbf{X}_{(-)}^t\mathbf{X}_{(-)}$,

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y} = (\mathbf{X}_{(k)}^t\mathbf{X}_{(k)} + \mathbf{X}_{(-)}^t\mathbf{X}_{(-)})^{-1}(\mathbf{X}_{(k)}^t|\mathbf{X}_{(-)}^t)\mathbf{Y} \\ &= (\mathbf{I}_{p+1} + (\mathbf{X}_{(k)}^t\mathbf{X}_{(k)})^{-1}\mathbf{X}_{(-)}^t\mathbf{X}_{(-)})^{-1}(\mathbf{X}_{(k)}^t\mathbf{X}_{(k)})^{-1}(\mathbf{X}_{(k)}^t|\mathbf{X}_{(-)}^t)\mathbf{Y}\end{aligned}$$

Lemma. *For $m \times m$ matrices $\mathbf{A}, \mathbf{B}$ and $m \times m$ identity matrix $\mathbf{I}$, $\mathbf{I} - \mathbf{AB}$ is invertible if and only if $\mathbf{I} - \mathbf{BA}$ is invertible.*

Proof. Suppose $\mathbf{I} - \mathbf{AB}$ is invertible. Suppose for $\mathbf{x} \in \mathbb{R}^m$, $(\mathbf{I} - \mathbf{BA})\mathbf{x} = \mathbf{0}$, $\mathbf{x} = \mathbf{BAx}$. Then, $\mathbf{Ax} = \mathbf{ABAX}$ and so $(\mathbf{I} - \mathbf{AB})\mathbf{Ax} = \mathbf{0}$. Because $\mathbf{I} - \mathbf{AB}$ is invertible, $\mathbf{Ax} = 0$ and thus $\mathbf{x} = \mathbf{0}$.($\because$ $\mathbf{A}$ is invertible.) This implies $\mathbf{I} - \mathbf{BA}$ is invertible. By changing the role of $\mathbf{A}$ and $\mathbf{B}$, one can similarly show that "if" part also holds. $\square$

Using Woodbury's formula and the result of Lemma, we have

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{I}_{p+1} + (\mathbf{X}_{(k)}^t\mathbf{X}_{(k)})^{-1}\mathbf{X}_{(-)}^t\mathbf{X}_{(-)})^{-1}(\mathbf{X}_{(k)}^t\mathbf{X}_{(k)})^{-1}(\mathbf{X}_{(k)}^t|\mathbf{X}_{(-)}^t)\mathbf{Y} \\ &= (\mathbf{I}_{p+1} - (\mathbf{X}_{(k)}^t\mathbf{X}_{(k)})^{-1}(\mathbf{I}_{p+1} + \mathbf{X}_{(-)}^t\mathbf{X}_{(-)}(\mathbf{X}_{(k)}^t\mathbf{X}_{(k)})^{-1})^{-1}\mathbf{X}_{(-)}^t\mathbf{X}_{(-)})(\mathbf{X}_{(k)}^t\mathbf{X}_{(k)})^{-1}(\mathbf{X}_{(k)}^t|\mathbf{X}_{(-)}^t)\mathbf{Y} \\ &= (\mathbf{I}_{p+1} - (\mathbf{X}_{(k)}^t\mathbf{X}_{(k)})^{-1}(\mathbf{I}_{p+1} + \mathbf{X}_{(-)}^t\mathbf{X}_{(-)}(\mathbf{X}_{(k)}^t\mathbf{X}_{(k)})^{-1})^{-1}\mathbf{X}_{(-)}^t\mathbf{X}_{(-)})((\mathbf{X}_{(k)}^t\mathbf{X}_{(k)})^{-1}\mathbf{X}_{(k)}^t\mathbf{Y}_{(k)} \\ &\quad + (\mathbf{X}_{(k)}^t\mathbf{X}_{(k)})^{-1}\mathbf{X}_{(-)}^t\mathbf{Y}_{(-)}) \\ &= (\mathbf{I}_{p+1} - (\mathbf{X}_{(k)}^t\mathbf{X}_{(k)})^{-1}(\mathbf{I}_{p+1} + \mathbf{X}_{(-)}^t\mathbf{X}_{(-)}(\mathbf{X}_{(k)}^t\mathbf{X}_{(k)})^{-1})^{-1}\mathbf{X}_{(-)}^t\mathbf{X}_{(-)})(\hat{\boldsymbol{\beta}} + (\mathbf{X}_{(k)}^t\mathbf{X}_{(k)})^{-1}\mathbf{X}_{(-)}^t\mathbf{Y}_{(-)}) \\ &= \hat{\boldsymbol{\beta}} + (\mathbf{X}_{(k)}^t\mathbf{X}_{(k)})^{-1}\mathbf{X}_{(-)}^t\mathbf{Y}_{(-)} - (\mathbf{X}_{(k)}^t\mathbf{X}_{(k)})^{-1}(\mathbf{I}_{p+1} + \mathbf{X}_{(-)}^t\mathbf{X}_{(-)}(\mathbf{X}_{(k)}^t\mathbf{X}_{(k)})^{-1})^{-1}\mathbf{X}_{(-)}^t\mathbf{X}_{(-)}) \\ &\quad (\hat{\boldsymbol{\beta}} + (\mathbf{X}_{(k)}^t\mathbf{X}_{(k)})^{-1}\mathbf{X}_{(-)}^t\mathbf{Y}_{(-)})\end{aligned}$$

Deleting $\hat{\boldsymbol{\beta}}$ in both hand sides, one have

$$\begin{aligned}(\mathbf{X}_{(k)}^t\mathbf{X}_{(k)})^{-1}\mathbf{X}_{(-)}^t\mathbf{Y}_{(-)} &= (\mathbf{X}_{(k)}^t\mathbf{X}_{(k)})^{-1}(\mathbf{I}_{p+1} + \mathbf{X}_{(-)}^t\mathbf{X}_{(-)}(\mathbf{X}_{(k)}^t\mathbf{X}_{(k)})^{-1})^{-1}\mathbf{X}_{(-)}^t\mathbf{X}_{(-)}(\hat{\boldsymbol{\beta}} + (\mathbf{X}_{(k)}^t\mathbf{X}_{(k)})^{-1}\mathbf{X}_{(-)}^t\mathbf{Y}_{(-)}) \\ \mathbf{X}_{(-)}^t\mathbf{Y}_{(-)} &= (\mathbf{I}_{p+1} + \mathbf{X}_{(-)}^t\mathbf{X}_{(-)}(\mathbf{X}_{(k)}^t\mathbf{X}_{(k)})^{-1})^{-1}\mathbf{X}_{(-)}^t\mathbf{X}_{(-)}(\hat{\boldsymbol{\beta}} + (\mathbf{X}_{(k)}^t\mathbf{X}_{(k)})^{-1}\mathbf{X}_{(-)}^t\mathbf{Y}_{(-)})\end{aligned}$$

By mutliplying both hand sides by $\mathbf{I}_{p+1} + \mathbf{X}_{(-)}^t\mathbf{X}_{(-)}(\mathbf{X}_{(k)}^t\mathbf{X}_{(k)})^{-1}$ on the left,

$$\begin{aligned}\mathbf{X}_{(-)}^t\mathbf{Y}_{(-)} + \mathbf{X}_{(-)}^t\mathbf{X}_{(-)}(\mathbf{X}_{(k)}^t\mathbf{X}_{(k)})^{-1}\mathbf{X}_{(-)}^t\mathbf{Y}_{(-)} &= \mathbf{X}_{(-)}^t\mathbf{X}_{(-)}\hat{\boldsymbol{\beta}} + \mathbf{X}_{(-)}^t\mathbf{X}_{(-)}(\mathbf{X}_{(k)}^t\mathbf{X}_{(k)})^{-1}\mathbf{X}_{(-)}^t\mathbf{Y}_{(-)} \\ \mathbf{X}_{(-)}^t\mathbf{Y}_{(-)} &= \mathbf{X}_{(-)}^t\mathbf{X}_{(-)}\hat{\boldsymbol{\beta}}\end{aligned}$$

Therefore, $\hat{\boldsymbol{\beta}} = (\mathbf{X}_{(-)}^t\mathbf{X}_{(-)})^{-1}\mathbf{X}_{(-)}^t\mathbf{Y}_{(-)} = \hat{\boldsymbol{\beta}}_{(-)}$.

Problem3. We first derive $100(1-\alpha)\%$ confidence interval for $\mu_{\mathbf{z}} = \boldsymbol{\beta}_0 + \mathbf{z}^t\boldsymbol{\beta}_1$. Let $\hat{\mu}_{\mathbf{z}} = \hat{\boldsymbol{\beta}}_0 + \mathbf{z}^t\hat{\boldsymbol{\beta}}_1$. Recall that $\hat{\boldsymbol{\beta}}_0 = \bar{Y} - \frac{1}{n}\mathbf{1}^t\mathbf{X}_1\hat{\boldsymbol{\beta}}_1$. Note that $\mathbf{1}^t\mathbf{X}_{1,\perp} = \mathbf{1}^t(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^t)\mathbf{X}_1 = \mathbf{1}^t - \mathbf{1}^t)\mathbf{X}_1 = \mathbf{0}$. Let $\mathcal{P}_{\mathbf{z}} = \frac{1}{n}\mathbf{1}^t - \frac{1}{n}\mathbf{1}^t\mathbf{X}_1(\mathbf{X}_{1,\perp}^t\mathbf{X}_{1,\perp})^{-1}\mathbf{X}_{1,\perp}^t + \mathbf{z}^t(\mathbf{X}_{1,\perp}^t\mathbf{X}_{1,\perp})^{-1}\mathbf{X}_{1,\perp}^t$. Then,

$$
\begin{aligned}
\mathcal{P}_{\mathbf{z}}\mathcal{P}_{\mathbf{z}}^t &= \frac{1}{n^2}\mathbf{1}^t\mathbf{1} - \frac{1}{n^2}\mathbf{1}^t\mathbf{X}_{1,\perp}(\mathbf{X}_{1,\perp}^t\mathbf{X}_{1,\perp})^{-1}\mathbf{X}_1^t\mathbf{1} + \frac{1}{n^2}\mathbf{1}^t\mathbf{X}_{1,\perp}(\mathbf{X}_{1,\perp}^t\mathbf{X}_{1,\perp})^{-1}\mathbf{z} - \frac{1}{n}\mathbf{1}^t\mathbf{X}_1(\mathbf{X}_{1,\perp}^t\mathbf{X}_{1,\perp})^{-1}\mathbf{X}_{1,\perp}^t\mathbf{1} \\
&\quad + \frac{1}{n^2}\mathbf{1}^t\mathbf{X}_1(\mathbf{X}_{1,\perp}^t\mathbf{X}_{1,\perp})^{-1}\mathbf{X}_{1,\perp}^t\mathbf{X}_{1,\perp}(\mathbf{X}_{1,\perp}^t\mathbf{X}_{1,\perp})^{-1}\mathbf{X}_1^t\mathbf{1} - \frac{1}{n}\mathbf{1}^t\mathbf{X}_1(\mathbf{X}_{1,\perp}^t\mathbf{X}_{1,\perp})^{-1}\mathbf{X}_{1,\perp}^t\mathbf{X}_{1,\perp}(\mathbf{X}_{1,\perp}^t\mathbf{X}_{1,\perp})^{-1}\mathbf{z} \\
&\quad + \frac{1}{n}\mathbf{z}^t(\mathbf{X}_{1,\perp}^t\mathbf{X}_{1,\perp})^{-1}\mathbf{X}_{1,\perp}^t\mathbf{1} - \frac{1}{n}\mathbf{z}^t(\mathbf{X}_{1,\perp}^t\mathbf{X}_{1,\perp})^{-1}\mathbf{X}_{1,\perp}^t\mathbf{X}_{1,\perp}(\mathbf{X}_{1,\perp}^t\mathbf{X}_{1,\perp})^{-1}\mathbf{X}_1^t\mathbf{1} \\
&\quad + \mathbf{z}^t(\mathbf{X}_{1,\perp}^t\mathbf{X}_{1,\perp})^{-1}\mathbf{X}_{1,\perp}^t\mathbf{X}_{1,\perp}(\mathbf{X}_{1,\perp}^t\mathbf{X}_{1,\perp})^{-1}\mathbf{z} \\
&= \frac{1}{n} + \frac{1}{n^2}\mathbf{1}^t\mathbf{X}_1(\mathbf{X}_{1,\perp}^t\mathbf{X}_{1,\perp})^{-1}\mathbf{X}_1^t\mathbf{1} - \frac{1}{n}\mathbf{1}^t\mathbf{X}_1(\mathbf{X}_{1,\perp}^t\mathbf{X}_{1,\perp})^{-1}\mathbf{z} - \frac{1}{n}\mathbf{z}^t(\mathbf{X}_{1,\perp}^t\mathbf{X}_{1,\perp})^{-1}\mathbf{X}_1^t\mathbf{1} + \mathbf{z}^t(\mathbf{X}_{1,\perp}^t\mathbf{X}_{1,\perp})^{-1}\mathbf{z} \\
&= \frac{1}{n} + (\mathbf{z}^t - \frac{\mathbf{1}^t\mathbf{X}_1}{n})(\mathbf{X}_{1,\perp}^t\mathbf{X}_{1,\perp})^{-1}(\mathbf{z} - \frac{\mathbf{X}_1^t\mathbf{1}}{n}) = \mathbf{C}_{\mathbf{z}}.
\end{aligned}
$$

Hence the variance of $\hat{\mu}_{\mathbf{z}}$ is given by

$$
\begin{aligned}
\mathrm{Var}(\hat{\mu}_{\mathbf{z}}) &= \mathrm{Var}(\bar{Y} - \frac{1}{n}\mathbf{1}^t\mathbf{X}_1\hat{\boldsymbol{\beta}}_1 + \mathbf{z}^t(\mathbf{X}_{1,\perp}^t\mathbf{X}_{1,\perp})^{-1}\mathbf{X}_{1,\perp}^t\mathbf{Y}) \\
&= \mathrm{Var}(\frac{1}{n}\mathbf{1}^t\mathbf{Y} - \frac{1}{n}\mathbf{1}^t\mathbf{X}_1(\mathbf{X}_{1,\perp}^t\mathbf{X}_{1,\perp})^{-1}\mathbf{X}_{1,\perp}^t\mathbf{Y} + \mathbf{z}^t(\mathbf{X}_{1,\perp}^t\mathbf{X}_{1,\perp})^{-1}\mathbf{X}_{1,\perp}^t\mathbf{Y}) \\
&= \mathrm{Var}(\mathcal{P}_{\mathbf{z}}\mathbf{Y}) = \mathcal{P}_{\mathbf{z}}\mathrm{Var}(\mathbf{Y})\mathcal{P}_{\mathbf{z}}^t = \sigma^2\mathcal{P}_{\mathbf{z}}\mathcal{P}_{\mathbf{z}}^t = \sigma^2\mathbf{C}_{\mathbf{z}}
\end{aligned}
$$

Claim. $\hat{\boldsymbol{\beta}} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y}$ is independent of SSE$=\mathbf{Y}^t(\mathbf{I} - \mathbf{X}(\mathbf{X^tX})^{-1}\mathbf{X^t})\mathbf{Y}$ is independent.

Proof. Observe that $\mathbf{I} - \mathbf{X}(\mathbf{X^tX})^{-1}\mathbf{X^t}$ is symmetric, idempotent. Hence SSE$= ||(\mathbf{I} - \mathbf{X}(\mathbf{X^tX})^{-1}\mathbf{X^t})\mathbf{Y}||^2$. As SSE is a function of $(\mathbf{I} - \mathbf{X}(\mathbf{X^tX})^{-1}\mathbf{X^t})\mathbf{Y}$, it suffices to show that $\hat{\boldsymbol{\beta}}$ and $(\mathbf{I} - \mathbf{X}(\mathbf{X^tX})^{-1}\mathbf{X^t})\mathbf{Y}$ are independent. Under normality, this can be established if we show the covariance between them are $\mathbf{0}$.

$$
\begin{aligned}
\mathrm{Cov}(\hat{\boldsymbol{\beta}}, (\mathbf{I} - \mathbf{X}(\mathbf{X^tX})^{-1}\mathbf{X^t})\mathbf{Y}) &= (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathrm{Var}(\mathbf{Y})(\mathbf{I} - \mathbf{X}(\mathbf{X^tX})^{-1}\mathbf{X^t})^t \\
&= \sigma^2(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t(\mathbf{I} - \mathbf{X}(\mathbf{X^tX})^{-1}\mathbf{X^t}) \\
&= \sigma^2((\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t - (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t) = \mathbf{0}
\end{aligned}
$$

This proves the claim. □

Note that $\hat{\mu}_{\mathbf{z}}$ is a function of $\hat{\boldsymbol{\beta}}$ and $\mathrm{E}(\hat{\mu}_{\mathbf{z}}) = \mu_{\mathbf{z}}$. So $\hat{\mu}_{\mathbf{z}}$ is independent of SSE. And we know that $\frac{\mathrm{SSE}}{\sigma^2} \sim \chi^2(n-p-1)$. Let $\hat{\sigma}^2$ denote $\frac{\mathrm{SSE}}{n-p-1}$. Since $\frac{\hat{\mu}_{\mathbf{z}} - \mu_{\mathbf{z}}}{\sqrt{\sigma^2\mathbf{C}_{\mathbf{z}}}} \sim \mathrm{N}(0,1)$, we see that

$$
\frac{(\hat{\mu}_{\mathbf{z}} - \mu_{\mathbf{z}})/\sqrt{\sigma^2\mathbf{C}_{\mathbf{z}}}}{\sqrt{\hat{\sigma}^2/\sigma^2}} = \frac{\hat{\mu}_{\mathbf{z}} - \mu_{\mathbf{z}}}{\sqrt{\hat{\sigma}^2\mathbf{C}_{\mathbf{z}}}} \sim t(n-p-1).
$$

From this, one can deduce that $100(1-\alpha)\%$ confidence interval for $\mu_{\mathbf{z}}$ is

$$
\mu_{\mathbf{z}}: \quad \hat{\mu}_{\mathbf{z}} \pm t_{\alpha/2}(n-p-1)\sqrt{\hat{\sigma}^2\mathbf{C}_{\mathbf{z}}}.
$$

$100(1-\alpha)\%$ confidence interval for $\mathbf{Y}_{\mathbf{z}} = \boldsymbol{\beta}_0 + \mathbf{z}^t\boldsymbol{\beta}_1 + \epsilon$(out-of-sample response), the difference that for $\mu_{\mathbf{z}}$ is ϵ. In the regression, there is an implicit assumption that ϵ is independent of sample.(This is quite intuitive.) Also, since out-of-sample is mutually independent with in-sample, $\mathrm{Var}(\hat{\mathbf{Y}}_{\mathbf{z}} - \mathbf{Y}_{\mathbf{z}}) = \mathrm{Var}(\hat{\mathbf{Y}}_{\mathbf{z}}) + \mathrm{Var}(\mathbf{Y}_{\mathbf{z}}) = \sigma^2\mathbf{C}_{\mathbf{z}} + \sigma^2 = \sigma^2(1 + \mathbf{C}_{\mathbf{z}})$. Here $\mathbf{Y}_{\mathbf{z}} = \hat{\boldsymbol{\beta}}_0 + \mathbf{z}^t\hat{\boldsymbol{\beta}}_1$. Using the similar argument in the above, we see that100 $(1-\alpha)\%$ confidence interval for $\mathbf{Y}_{\mathbf{z}}$ is

$$
\mathbf{Y}_{\mathbf{z}}: \quad \hat{\mathbf{Y}}_{\mathbf{z}} \pm t_{\alpha/2}(n-p-1)\sqrt{\hat{\sigma}^2(1 + \mathbf{C}_{\mathbf{z}})}.
$$