STAT5030 QE (2019-2020)

1. i) The loss function can be re-written as

$$L(\beta) = (Y - X\beta)^T(Y - X\beta) + \lambda\beta^T\beta.$$

Take derivative w.r.t. β and set it to 0, derive,

$$\frac{\partial L(\beta)}{\partial\beta} = -2X^T(Y - X\beta) + 2\lambda\beta = 0$$

$$\hat{\beta}_\lambda^{ridge} = (X^TX + \lambda I)^{-1}X^TY$$ (for simplicity, denoted as $\hat{\beta}$ in the following part)

For any $\beta \in R^p$, we have

$$\cancel{L(\beta) - L(\hat{\beta}_\lambda^{ridge}) = (Y - X\beta)^T(Y - X\beta) + \lambda\beta^T\beta - (Y - X\hat{\beta}_\lambda^{ridge})^T(Y - \hat{\beta}_\lambda^{ridge})\ \lambda(\hat{\beta}_\lambda^{ridge})^T\hat{\beta}^{ridge}}$$

$$\cancel{= (Y - X\hat{\beta}}$$

$$L(\beta) - L(\hat{\beta}) = (Y - X\beta)^T(Y - X\beta) + \lambda\beta^T\beta - (Y - X\hat{\beta})^T(Y - X\hat{\beta}) - \lambda\hat{\beta}^T\hat{\beta}$$

$$= (Y - X\hat{\beta} + X\hat{\beta} - X\beta)^T(Y - X\hat{\beta} + X\hat{\beta} - X\beta) + \lambda(\beta - \hat{\beta} + \hat{\beta})^T(\beta - \hat{\beta} + \hat{\beta})$$
$$- (Y - X\hat{\beta})^T(Y - X\hat{\beta}) - \lambda\hat{\beta}^T\hat{\beta}$$

$$= (\hat{\beta} - \beta)^TX^TX(\hat{\beta} - \beta) + \lambda(\beta - \hat{\beta})^T(\beta - \hat{\beta})$$

$$\geq 0$$

Therefore, $\hat{\beta} = (X^TX + \lambda I)^{-1}X^TY$ minimizes the loss function.

ii) By singular decomposition, we have

$$X = UDV^T,$$

where $U = [u_1 \cdots u_p]$, $V = [v_1 \cdots v_p]$ ~~are~~ contains orthonormal vectors, D is a $p \times p$ diagonal matrix whose ~~non-zero~~ diagonal elements are singular value of X.

Then plug-in the equation of $\hat{\beta}$, derive

$$\hat{\beta} = [V(D^2 + \lambda I_p)V^T]^{-1}VDU^TY$$

$$= V\,\mathrm{diag}\left(\frac{d_i}{d_i^2 + \lambda}\right)U^TY,$$

where $\mathrm{diag}(v)$ is a diagonal matrix whose ith diagonal element ~~element in the diagonal~~ is ith element of v, v is a vector.

(iii) What we want to prove is $\exists \lambda \in \mathbb{R}$, s.t.

$$MSE(\hat{\beta}_\lambda^{ridge}) < MSE(\hat{\beta}_{OLS}),$$

where $MSE(\hat{\beta}) = E((\hat{\beta}-\beta_0)^T(\hat{\beta}-\beta_0))$, β_0 is true parameters.

~~For any orthonormal matrix P, we have~~

Since X^TX is a square symmetric matrix, there exists a orthonormal matrix P s.t.

$$X^TX = P^TSP.$$

where $S = diag(s_1, \cdots, s_p)$

Let $X^* = XP^T$ and $\alpha = P\beta$, then the model can be written as

$$Y = X^*\alpha + \varepsilon$$

We know for any orthonormal matrix P, we have

$$\begin{aligned} MSE(\hat{\alpha}) = E((\hat{\alpha}-\alpha_0)^T(\hat{\alpha}-\alpha_0)) &= E[(P\hat{\beta}-P\beta_0)^T(P\hat{\beta}-P\beta_0)] \\ &= E[(\hat{\beta}-\beta_0)^TP^TP(\hat{\beta}-\beta_0)] \\ &= MSE(\hat{\beta}) \end{aligned}$$

Therefore, we can transform the original question to prove the following one

$\exists \lambda \in \mathbb{R}$, s.t. $\quad MSE(\hat{\alpha}_\lambda^{ridge}) < MSE(\hat{\alpha}_{OLS})$

It's easy to show that,

$$(X^*)^TX^* = S,$$

$$\hat{\alpha}_\lambda^{ridge} = [S+\lambda I_p]^{-1}(X^*)^TY$$

Set $Z = (I+\lambda S^{-1})^{-1}$, then $\hat{\alpha}_\lambda^{ridge} = Z\hat{\alpha}_{ols}$, where $\hat{\alpha}_{ols} = S^{-1}(X^*)^TY$

$$\begin{aligned} MSE(\hat{\alpha}_\lambda^{ridge}) &= E((\hat{\alpha}_\lambda^{ridge}-\alpha_0)^T(\hat{\alpha}_\lambda^{ridge}-\alpha_0)) \\ &= E((Z\hat{\alpha}_{ols}-\alpha_0)^T(Z\hat{\alpha}_{ols}-\alpha_0)) \\ &= \underbrace{E[(Z\hat{\alpha}_{ols}-Z\alpha_0)^T(Z\hat{\alpha}_{ols}-Z\alpha_0)]}_{①} + \underbrace{E[(Z\alpha_0-\alpha_0)^T(Z\alpha_0-\alpha_0)]}_{②} \end{aligned}$$

$$\begin{aligned} ① &= tr(E[(Z\hat{\alpha}_{ols}-Z\alpha_0)(Z\hat{\alpha}_{ols}-Z\alpha_0)]^T) \\ &= tr(ZE[(\hat{\alpha}_{ols}-\alpha_0)(\hat{\alpha}_{ols}-\alpha_0)^T]Z^T) \\ &= tr(Z\sigma^2S^{-1}Z^T) \\ &= \sigma^2 tr((I+\lambda S^{-1})^{-1}S^{-1}(I+\lambda S^{-1})^{-1}) \\ &= \sigma^2\sum[(\frac{1}{s_i})(1+\frac{\lambda}{s_i})^{-2}] \\ &= \sigma^2\sum\frac{s_i}{(s_i+\lambda)^2} \end{aligned}$$

$$② = \alpha_0^T\left[\left((I+\lambda S^{-1})^{-1}-I\right)^T\left((I+\lambda S^{-1})^{-1}-I\right)\right]\alpha_0$$
$$= \sum\left[\left(1+\frac{\lambda}{s_i}\right)^{-1}-1\right]^2\alpha_{0i}^2 \quad \text{, where } \alpha_{0i} \text{ is ith element of } \alpha_0.$$
$$= \sum\frac{\lambda^2\alpha_{0i}^2}{(s_i+\lambda)^2}$$

Therefore, $MSE(\hat{\alpha}_\lambda^{ridge}) = \sigma^2\sum\frac{s_i}{(s_i+\lambda)^2} + \sum\frac{\lambda^2\alpha_{0i}^2}{(\lambda+s_i)^2}$, and denoted as $f(\lambda)$ in the following part.

$$f'(\lambda) = \frac{df(\lambda)}{d\lambda} = \sum\frac{2s_i(\lambda\alpha_{0i}^2-\sigma^2)}{(\lambda+s_i)^3}$$

When $0\le\lambda<\min\left(\frac{\sigma^2}{\alpha_{01}^2}, \dots, \frac{\sigma^2}{\alpha_{0p}^2}\right)$, $\lambda\alpha_{0i}^2-\sigma^2<0$, then we have $f'(\lambda)<0$. It means $f(\lambda)$ is a monotonically decreasing function if $0\le\lambda<\min(\cdots)$.

So there must exists $\lambda\in\mathbb{R}$ s.t. $MSE(\hat{\beta}_\lambda^{ridge}) < MSE(\hat{\beta}_{ols})$.

2.

i) $Y = Zr + X\tau + \varepsilon$,

where $Z = \begin{bmatrix}1&0\\1&0\\1&0\\1&0\\0&1\\0&1\\0&1\\0&1\end{bmatrix}$, $r=\begin{bmatrix}r_1\\r_2\end{bmatrix}$, $X=\begin{bmatrix}1&0\\1&0\\0&1\\0&1\\1&0\\1&0\\0&1\\0&1\end{bmatrix}$, $\tau=\begin{bmatrix}\tau_1\\\tau_2\end{bmatrix}$.

$$cov(Zr+\varepsilon) = Z\,cov(r)Z^T + cov(\varepsilon)$$
$$= Z\sigma_r^2 I_2 Z^T + \sigma_E^2 I_2$$
$$= \sigma_r^2 ZZ^T + \sigma_E^2 I_8$$

ii) $E(Y) = X\tau$, therefore, we have
$$Y\sim N(X\tau, \sigma_r^2 ZZ^T+\sigma_E^2 I_2)$$

Set c_i^T is ith row vector of C^T, we know when there exists q_i^T s.t.
$$c_i^T\tau = q_i^T E(Y), \tag{1}$$
c_i^T is estimable. C^T is of full row rank, and $C^T\tau$ is estimable when $c_i^T\tau$ is estimable for all $i=1,2,\cdots,q$. Therefore, we need
$$C^T\tau = Q^T E(Y), \tag{2}$$
where Q^T is a q matrix whose ith row vector is q_i^T.

Equation (2) means
$$C^T\tau = Q^T X\tau \tag{3}$$
Equation (3) holds for all $\tau\in\mathbb{R}^2$, then we have
$$C^T = Q^T X$$

From the lecture notes, the best linear unbiased estimator of τ is $(X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}Y$, where $\Sigma = \sigma_r^2 zz^T + \sigma_E^2 I_8$.

iii) The ~~best~~ statistic is proposed as follows:

$$F = \frac{Q/q}{SSE/(8-2)} = \frac{Q/q}{SSE/6}.$$

where $Q = (C'\hat{\tau} - d)^T (C'(X^T\Sigma^{-1}X)^{-1}C)^{-1}(C'\hat{\tau} - d)$,

$SSE = (Y - X\hat{\tau})^T \Sigma^{-1} (Y - X\hat{\tau})$

And $F \sim F_{q,6}$ under H_0.

$\cancel{E(Q/q) = \hat{\tau} = HY}$, ~~where~~ $\cancel{H = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}}$

$\cancel{\sim N(\tau, H)}$

$$\hat{\tau} = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}Y$$

$$\sim N(\tau, (X^T\Sigma^{-1}X)^{-1})$$

$$C'\hat{\tau} - d \sim N(C'\tau - d, C'(X^T\Sigma^{-1}X)^{-1}C)$$

$$[C'(X^T\Sigma^{-1}X)^{-1}C]^{-\frac{1}{2}}(C'\hat{\tau} - d) \sim N([C'(X^T\Sigma^{-1}X)^{-1}C]^{-\frac{1}{2}}(C'\tau - d), I)$$

Therefore, $\cancel{Q \sim \chi^2}$

$$E(Q) = q + (C'\tau - d)^T(C'(X^T\Sigma^{-1}X)^{-1}C)^{-1}(C'\tau - d).$$

$$E\left(\frac{Q}{q}\right) = 1 + \frac{1}{q}(C'\tau - d)^T(C'(X^T\Sigma^{-1}X)^{-1}C)^{-1}(C'\tau - d)$$

iv) ~~Sep~~ $\Sigma = \sigma_r^2 zz^T + \sigma_E^2 I_8$

$$= \sigma_E^2\left(\phi \begin{bmatrix} 1 & 0\\ 1 & 0\\ 1 & 0\\ 1 & 0\\ 0 & 1\\ 0 & 1\\ 0 & 1\\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1&1&1&1&0&0&0&0\\ 0&0&0&0&1&1&1&1 \end{bmatrix} + I_8\right)$$

$$= \sigma_E^2\left(\phi \begin{bmatrix} J_4 & 0\\ 0 & J_4 \end{bmatrix} + \begin{bmatrix} I_4 & 0\\ 0 & I_4 \end{bmatrix}\right)$$

$$= \sigma_E^2 \left(\phi \begin{bmatrix} \cancel{\phi} I_4 + \phi J_4 & 0\\ 0 & I_4 + \phi J_4 \end{bmatrix}\right.$$

It's easy to prove if A^{-1}, B^{-1} exists, then inverse matrix of $\begin{bmatrix} A & 0\\ 0 & B \end{bmatrix}$ is $\begin{bmatrix} A^{-1} & 0\\ 0 & B^{-1} \end{bmatrix}$,

Therefore, $\Sigma^{-1} = (\sigma_E^2)^{-1} \begin{bmatrix} I_4 - \frac{\phi}{1+4\phi}J_4 & 0\\ 0 & I_4 - \frac{\phi}{1+4\phi}J_4 \end{bmatrix} = (\sigma_E^2)^{-1}\begin{bmatrix} 1-\frac{\phi}{1+4\phi} & -\frac{\phi}{1+4\phi} & -\frac{\phi}{1+4\phi} & -\frac{\phi}{1+4\phi} & \cdots \\ -\frac{\phi}{1+4\phi} & 1-\frac{\phi}{1+4\phi} & \cdots & & \\ \cdots & & & & \end{bmatrix}$

$$X^T\Sigma^{-1}X = (\sigma_E^2)^{-1}\begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1-\frac{\phi}{1+4\phi} & -\frac{\phi}{1+4\phi} & \cdots & & \\ -\frac{\phi}{1+4\phi} & 1-\frac{\phi}{1+4\phi} & \cdots & & 0 \\ & & & & \\ & & & 1-\frac{\phi}{1+4\phi} & -\frac{\phi}{1+4\phi} \\ 0 & & & \cdots & \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

$$= (\sigma_E^2)^{-1}\begin{bmatrix} 1-2\frac{\phi}{1+4\phi} & 1-2\frac{\phi}{1+4\phi} & -2\frac{\phi}{1+4\phi} & \cdots \\ -2\frac{\phi}{1+4\phi} & -2\frac{\phi}{1+4\phi} & 1-2\frac{\phi}{1+4\phi} & \cdots \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

$$= (\sigma_E^2)^{-1}\begin{bmatrix} 4\left(1-2\frac{\phi}{1+4\phi}\right) & 4\left(-2\frac{\phi}{1+4\phi}\right) \\ 4\left(-2\frac{\phi}{1+4\phi}\right) & 4\left(1-2\frac{\phi}{1+4\phi}\right) \end{bmatrix}$$

$\cancel{(X^T\Sigma^{-1}X)^{-1}}$ $\quad \cancel{=4(\sigma_E^2)^{-1}\,[\,I_1}$

$$= 4(\sigma_E^2)^{-1}\left(I_2 + \left(-2\frac{\phi}{1+4\phi}\right)J_2\right) = 4(\sigma_E^2)^{-1}(I_2 + wJ_2), \quad w = -\frac{2\phi}{1+4\phi}$$

Therefore, using the fact, we obtain

$$(X^T\Sigma^{-1}X)^{-1} = \frac{\sigma_E^2}{4}\left(I_2 - \frac{w}{1+2w}J_2\right)$$

$$C'(X^T\Sigma^{-1}X)^{-1}C = \frac{\sigma_E^2}{4}\begin{bmatrix} 1 & -1 \end{bmatrix}\begin{bmatrix} 1-\frac{w}{1+2w} & -\frac{w}{1+2w} \\ -\frac{w}{1+2w} & 1-\frac{w}{1+2w} \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$= \frac{\sigma_E^2}{4} \times 2$$

$$= \frac{\sigma_E^2}{2}$$, doesn't depend on w, and thus ϕ.

Therefore, test statistic in part (c) doesn't depend on ϕ.