Oliver
Jack

HW1

Ex. 1

1) First let us rewrite Y^TAY:

$$Y^TAY = (Y-\mu)^T AY + \mu^T AY$$
$$= (Y-\mu)^T A(Y-\mu) + \mu^T AY + (Y-\mu)^T A\mu$$

Then by taking the expectation we get:

$$\mathbb{E}[Y^TAY] = \mathbb{E}[(Y-\mu)^T A(Y-\mu)] + \mathbb{E}[\mu^T AY] + \mathbb{E}[(Y-\mu)^T A\mu]$$

We have:

$$\mathbb{E}[(Y-\mu)^T A\mu] = \sum_{i=1}^{n}\sum_{j=1}^{n} \underbrace{\mathbb{E}[(Y-\mu)_i]}_{=0} A_{ij}\mu_j = 0$$

$$\mathbb{E}[\mu^T AY] = \sum_{i=1}^{n}\sum_{j=1}^{n} \mu_i A_{ij} \underbrace{\mathbb{E}[Y_j]}_{=\mu_j} = \mu^T A\mu$$

Thus:

$$\mathbb{E}[Y^TAY] = \mathbb{E}[(Y-\mu)^T A(Y-\mu)] + \mu^T A\mu$$

Now let us show that $\mathbb{E}[(Y-\mu)^T A(Y-\mu)]$ is equal to $tr(A\Sigma)$:

$$\mathbb{E}[(Y-\mu)^T A(Y-\mu)]$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}[(Y-\mu)_i A_{ij} (Y-\mu)_j] \quad \Big\| \quad Cov[(Y-\mu)_i,(Y-\mu)_j] = \mathbb{E}[(Y-\mu)_i(Y-\mu)_j] - \underbrace{\mathbb{E}[(Y-\mu)_i]}_{=0}\underbrace{\mathbb{E}[(Y-\mu)_j]}_{=0}$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\, Cov[Y_i-\mu_i, Y_j-\mu_j]$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\, Cov(Y_i, Y_j)$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\,\Sigma_{ji} \quad \Big\| \quad \Sigma_{ij} = \Sigma_{ji}$$
$$= \sum_{i=1}^{n} (A\Sigma)_{ii}$$
$$= tr(A\Sigma) \quad \blacksquare$$

2) By the definition of the covariance we have:

$$cov[Y, Y^TAY] = \mathbb{E}[(Y-\mu)(Y^TAY - \mathbb{E}[Y^TAY])^T]$$
$$= \mathbb{E}[(Y-\mu)(Y^TAY - \mu^TA\mu - tr(A\Sigma))^T] \quad (*)$$

We have:

$Y^TAY - \mu^TA\mu$

similar to 1)

$= (Y-\mu)^T A (Y-\mu) + \mu^TAY + (Y-\mu)^TA\mu - \mu^TA\mu$

$= (Y-\mu)^TA(Y-\mu) + \mu^TA(Y-\mu) + (Y-\mu)^TA\mu$

We have:

$$\mu^TA(Y-\mu) = \sum_{i=1}^{n}\sum_{j=1}^{n} \mu_i A_{ij} (Y-\mu)_j$$

$$\overset{A \text{ symm.}}{=} \sum_{j=1}^{n}\sum_{i=1}^{n} (Y-\mu)_j A_{ji} \mu_i$$

$$= (Y-\mu)^TA\mu$$

So:

$= (Y-\mu)^TA(Y-\mu) + 2(Y-\mu)^TA\mu$

Thus:

$$(*) = \mathbb{E}[(Y-\mu)((Y-\mu)^TA(Y-\mu) + 2(Y-\mu)^TA\mu - tr(A\Sigma))^T]$$

$$= \mathbb{E}[(Y-\mu)((Y-\mu)^TA(Y-\mu))^T] + 2\mathbb{E}[(Y-\mu)((Y-\mu)^TA\mu)^T] - tr(A\Sigma)\underbrace{\mathbb{E}[Y-\mu]}_{=0}$$

$$= \mathbb{E}[(Y-\mu)(Y-\mu)^TA(Y-\mu)] + 2\mathbb{E}[(Y-\mu)\underbrace{\mu^TA(Y-\mu)}_{A \text{ is symmetric } = (Y-\mu)^TA\mu}]$$

$$= \mathbb{E}[(Y-\mu)(Y-\mu)^TA(Y-\mu)] + 2\underbrace{\mathbb{E}[(Y-\mu)(Y-\mu)^T]}_{=\Sigma}A\mu$$

Finally, we need to show that

$\mathbb{E}[(Y-\mu)(Y-\mu)^TA(Y-\mu)] = 0$

Since $(Y-\mu)$ is a centered Gaussian random vector, we have that

$Y-\mu \overset{\mathcal{L}}{=} \mu - Y$ (equal in distribution)

$\Rightarrow (Y-\mu)(Y-\mu)^TA(Y-\mu) \overset{\mathcal{L}}{=} (\mu-Y)(\mu-Y)^TA(\mu-Y)$

$$\Rightarrow \mathbb{E}[(Y-\mu)(Y-\mu)^TA(Y-\mu)] = \underbrace{\mathbb{E}[(\mu-Y)(\mu-Y)^TA(\mu-Y)]}_{-\mathbb{E}[(Y-\mu)(Y-\mu)^TA(Y-\mu)]}$$

$\Rightarrow \mathbb{E}[(Y-\mu)(Y-\mu)^TA(Y-\mu)] = 0$

So we have that:

$$\mathrm{cov}[Y, Y^T A Y] = 2\Sigma A \mu$$ $\boxtimes$

Ex. 3

The vector of residuals can be estimated by:

$$\begin{aligned}
\hat{\varepsilon} &= y - X\hat{b} \\
&= y - X(X^TX)^{-1}X^T y \\
&= (I_n - X(X^TX)^{-1}X^T)y \\
&= (I_n - X(X^TX)^{-1}X^T)X\beta + (I_n - X(X^TX)^{-1}X^T)\varepsilon \\
&= X\beta - X\underbrace{(X^TX)^{-1}X^TX}_{=I_k}\beta + (I_n - X(X^TX)^{-1}X^T)\varepsilon \\
&= \underbrace{(I_n - X(X^TX)^{-1}X^T)}_{=P}\varepsilon \sim \mathcal{N}(0, \sigma^2 P)
\end{aligned}$$

Let us show that P is an orthogonal projection:

$$\begin{aligned}
P^T &= (I_n - X(X^TX)^{-1}X^T)^T \\
&= I_n^T - (X^T)^T((X^TX)^T)^{-1}X^T \\
&= I_n - X(X^TX)^{-1}X^T \\
&= P
\end{aligned}$$

$P^2 = P$ (shown in class already)

Then by theorem 1.11. we have that all eigenvalues of P are in $\{0,1\}$, with rank(P) 1s. For any orthogonal projection it holds that:

$$\begin{aligned}
\mathrm{rank}(P) &= \mathrm{tr}(P) \\
&= \mathrm{tr}(I_n) - \mathrm{tr}(X(X^TX)^{-1}X^T) \\
\text{trace inv. under cyclic permutation} \quad &= n - \mathrm{tr}(\underbrace{(X^TX)^{-1}X^TX}_{=I_k}) \\
&= n - k
\end{aligned}$$

Furthermore, since $P^T = P$, $\exists$ a matrix $S \in \mathbb{R}^{n\times n}$ s.t. $S^TS = SS^T = I_n$ and $S^TPS = \mathrm{diag}(\underbrace{1, 1, \ldots 1}_{n-k \text{ times}}\underbrace{0, \ldots, 0}_{k \text{ times}})$

Next, set $Z = S\hat{\varepsilon}$.

Since $\hat{\varepsilon} \sim \mathcal{N}(0, \sigma^2 P)$, we have that:

$Z \sim N(0, \sigma^2 S^T P S)$

Since $\forall i \in \{n-k+1, \dots, n\}$: $(S^T P S)_{ii} = 0$, we have that $Z_i = 0$.

Hence:

$$\|Z\|_2^2 = \sum_{i=1}^{n-k} Z_i^2 \quad \text{where } Z_i \sim N(0, \sigma^2)$$

$$\Rightarrow \frac{\|Z\|_2^2}{\sigma^2} = \sum_{i=1}^{n-k} \underbrace{\left(\frac{Z_i}{\sigma}\right)}_{\sim N(0,1)}^2 \sim \chi^2_{n-k}$$

Since S is a unitary matrix, we get that:

$$\|\hat{\varepsilon}\|_2^2 = \|S^T Z\|_2^2 = \|Z\|_2^2$$

So:

$$\frac{\|\hat{\varepsilon}\|_2^2}{\sigma^2} = \frac{\|Y - X\hat{\ell}\|_2^2}{\sigma^2} \sim \chi^2_{n-k} \qquad \Big\| \begin{array}{l} X \sim \chi^2_n \\ \Rightarrow \mathbb{E}[X] = n \end{array}$$

$$\Rightarrow \mathbb{E}\left[\frac{\|Y - X\hat{\ell}\|_2^2}{\sigma^2}\right] = n - k$$

$$\Rightarrow \mathbb{E}\left[\frac{\|Y - X\hat{\ell}\|_2^2}{n-k}\right] = \sigma^2$$

$\Rightarrow \hat{\sigma}^2$ is unbiased $\blacksquare$