Homework 1.

1. (a) Explain why the sample mean $\hat{\mu}_i = \frac{1}{n}\sum_{j=1}^{n} X_{ji}$ is an "appropriate" estimator of the population mean μ_i.

$$\hat{\mu}_i = \frac{1}{n}\sum_{j=1}^{n} X_{ji}$$

$$E(\hat{\mu}_i) = E(\frac{1}{n}\sum_{j=1}^{n} X_{ji})$$

$$E(\hat{\mu}_i) = \frac{1}{n}\sum_{j=1}^{n} E(X_{ji})$$

$$E(\hat{\mu}_i) = \frac{1}{n} n\mu_i$$

$$E(\hat{\mu}_i) = \mu_i$$

The sample mean, $E(\hat{\mu}_i)$ is an unbiased estimator of poulation mean, μ_i, so it is an appropriate estimator.

The sample covariance matrix $\hat{\Sigma}_{ij} = \frac{1}{n}\sum_{t=1}^{n}(X_{ti} - \hat{\mu}_i)(X_{tj} - \hat{\mu}_j)$ is appropriate because:

$$E[\hat{\Sigma}_{ij}] = E\left[\frac{1}{n}\sum_{t=1}^{n}(X_{ti} - \hat{\mu}_i)(X_{tj} - \hat{\mu}_j)\right] \tag{1}$$

$$= \frac{1}{n}\sum_{t=1}^{n} E[(X_{ti} - \hat{\mu}_i)(X_{tj} - \hat{\mu}_j)] \tag{2}$$

$$\approx \frac{1}{n}\sum_{t=1}^{n} E[(X_{ti} - \mu_i)(X_{tj} - \mu_j)] \tag{3}$$

$$= \frac{1}{n}\sum_{t=1}^{n} \Sigma_{ij} \tag{4}$$

$$= \Sigma_{ij} \tag{5}$$

Therefore $E[\hat{\Sigma}] = \Sigma$, making it an **unbiased estimator** of the true covariance matrix.

Additionally, by the **Law of Large Numbers**, as $n \to \infty$:

$$\hat{\Sigma}_{ij} \to \Sigma_{ij} \text{ (convergence in probability)}$$

This makes it a **consistent estimator**.

(b) Assume that $\hat{\mu} = 0$, show that $\hat{\Sigma} = \hat{X}^T\hat{X}$

$$\hat{\Sigma}_{ij} = \frac{1}{n}\sum_{l=1}^{n}(\hat{X}_{li} - \hat{\mu}_i)(\hat{X}_{lj} - \hat{\mu}_j)$$
$$\hat{\Sigma}_{ij} = \frac{1}{n}\sum_{l=1}^{n}(\hat{X}_{li})(\hat{X}_{lj})$$
$$\hat{\Sigma}_{ij} = \frac{1}{n}\sum_{l=1}^{n}(\hat{X}^T)_{il}(\hat{X}_{lj})$$
$$\hat{\Sigma} = \frac{1}{n}\hat{X}^T\hat{X}$$

(c) Don't assume that $\hat{\mu} = 0$, show that $\hat{\Sigma} = \frac{1}{n}(\hat{X} - 1\hat{\mu}^T)^T(\hat{X} - 1\hat{\mu}^T)$

$$\hat{\Sigma}_{ij} = \frac{1}{n}\sum_{l=1}^{n}(\hat{X}_{li} - \hat{\mu}_i)(\hat{X}_{lj} - \hat{\mu}_j)$$
$$\hat{\Sigma}_{ij} = \frac{1}{n}\sum_{l=1}^{n}(\hat{X}_{li}\hat{X}_{lj}) - (\hat{\mu}_i\hat{X}_{lj}) - (\hat{\mu}_j\hat{X}_{li}) - (\hat{\mu}_i\hat{\mu}_j)$$
$$\hat{\Sigma}_{ij} = \frac{1}{n}\sum_{l=1}^{n}(\hat{X}_{li}\hat{X}_{lj}) - \frac{1}{n}\sum_{l=1}^{n}(\hat{\mu}_i\hat{X}_{lj}) - \frac{1}{n}\sum_{l=1}^{n}(\hat{\mu}_j\hat{X}_{li}) - \frac{1}{n}\sum_{l=1}^{n}(\hat{\mu}_i\hat{\mu}_j)$$
$$\hat{\Sigma}_{ij} = \frac{1}{n}\sum_{l=1}^{n}(\hat{X}_{li}\hat{X}_{lj}) - (\hat{\mu}_i\hat{\mu}_j) - (\hat{\mu}_j\hat{\mu}_i) - (\hat{\mu}_i\hat{\mu}_j)$$
$$\hat{\Sigma}_{ij} = \frac{1}{n}\sum_{l=1}^{n}(\hat{X}_{li}\hat{X}_{lj}) - (\hat{\mu}_j\hat{\mu}_i)$$
$$\hat{\Sigma} = \frac{1}{n}\hat{X}^T\hat{X} - (\hat{\mu}^T\hat{\mu})$$
$$\hat{\Sigma} = \frac{1}{n}(\hat{X}^T - 1\hat{\mu}^T)(\hat{X} - 1\hat{\mu}^T)$$

(d) Why the projection length is $|\tilde{X}^{(i)} \cdot w|$ and why $(\tilde{X}^{(i)} \cdot w)w$ is the projection vector:

Why the projection length is $|\tilde{X}^{(i)} \cdot w|$

$$
\begin{aligned}
\text{Let } l &= \text{length of the projection} \\
\cos(\theta) &= \frac{l}{\|\tilde{X}^{(i)}\|} \\
l &= \|\tilde{X}^{(i)}\| \cos(\theta) \\
\tilde{X}^{(i)} \cdot w &= \|\tilde{X}^{(i)}\| \|w\| \cos(\theta) \\
\tilde{X}^{(i)} \cdot w &= \|\tilde{X}^{(i)}\| \|w\| \left(\frac{l}{\|\tilde{X}^{(i)}\|} \right) \\
\tilde{X}^{(i)} \cdot w &= \|w\| \cdot l \\
\tilde{X}^{(i)} \cdot w &= l \quad (\text{since } \|w\| = 1)
\end{aligned}
$$

Therefore, $|\tilde{X}^{(i)} \cdot w| = |l| = l$, which is the length of the projection.

Why $(\tilde{X}^{(i)} \cdot w)w$ is the projection vector:

$$
\begin{aligned}
\text{We know: } \tilde{X}^{(i)} \cdot w &= l \text{ (length of projection)} \qquad (6) \\
\text{The projection vector} &= l \cdot w \qquad (7) \\
&= (\tilde{X}^{(i)} \cdot w)w \qquad (8)
\end{aligned}
$$

The scalar $\tilde{X}^{(i)} \cdot w$ tells us how far to go, and the unit vector w tells us which direction. Multiplying them gives the projection vector.