QUESTION 2 :

We want to predict X_{T-h} based on info set $Z = \{X_T, X_{T-1}, \dots, X_{T-p}\}$
$\Rightarrow$ Estimator $\hat{X}_{T-h} = F(Z)$

Def. The best estimator is that one that minimizes

$$\mathbb{E}\left[\mathcal{L}(X_{T+h}, \hat{X}_{T+h})\right] = \mathbb{E}\left[(X_{T+h} - \hat{X}_{T+h})^2\right]$$

RELEVANT THEOREM : The best estimator / forecast of X_{T+h} is its conditional mean given the information available at time $T \Rightarrow \mathbb{E}(X_{T+h} | Z)$

In the case that each X_i follows a normal distribution, we can write down the conditional distribution from the definition of GP.

Joint distribution :

$$\begin{bmatrix} X_{T+h} \\ Z \end{bmatrix} = \begin{bmatrix} X_{T+h} \\ X_T \\ X_{T-1} \\ \vdots \\ X_{T-p} \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_x \\ \mu_z \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xz} \\ \Sigma_{zx} & \Sigma_{zz} \end{bmatrix} \right)$$

Conditional distribution :

$$P(X_{T+h} | Z) \sim N(\Sigma_{zx} \Sigma_{xx}^{-1} Z, \; \Sigma_{zz} \Sigma_{xx}^{-1} \Sigma_{xz})$$

and since the expected value of a gaussian it's its mean :

$$\mathbb{E}(X_{T+h} | Z) = \alpha Z$$

where $\alpha = \Sigma_{zx} \Sigma_{xx}^{-1}$ are the kernel weights in whatever kernel is used to model the covariance matrix.

$$\alpha_0 = \Sigma_{x_T x} \cdot \Sigma_{xx}^{-1}$$
$$\alpha_1 = \Sigma_{x_{T-1} x} \cdot \Sigma_{xx}^{-1}$$
$$\vdots$$