Prediction:

True model

we fit the model $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon$ and estimate $\mathbf{b}$ of $\boldsymbol{\beta}$. Using OLS.

Let $\mathbf{x}_0' = [1, x_{01}, ..., x_{0k}]$ represent a set of explanatory-variable scores for which a prediction is desired

a) If we use $\hat{Y}_0 = \mathbf{x}_0'\mathbf{b}$ to estimate $E(Y_0)$, then the error in estimation is $\delta \equiv \hat{Y}_0 - E(Y_0)$. Show that if the model is correct, then $E(\delta) = 0$ (i.e. $\hat{Y}_0$ is an unbiased estimator of $E(Y_0)$) and that $V(\delta) = \sigma_\varepsilon^2 \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0$.

$E(\delta) = E[\hat{Y}_0 - \underbrace{E(Y_0)}_{\text{constant}}]$

$= E(\hat{Y}_0) - E(Y_0)$ Linearity of E

$= x_0' E(b) - E(\underbrace{x_0\beta}_{\text{True}} + \underbrace{\varepsilon_0}_{\text{RV}})$

$= x_0' E(\beta) - x_0\beta$

$= (x_0' - x_0)\beta. = 0$

since $x_0' = x_0$ True

$V(\delta) = V(\hat{Y}_0 - E(Y_0))$

$= V(\hat{Y}_0)$

$= x_0^T V(\beta) x_0$

$V(\beta) = \sigma^2 (X^T X)^{-1}$

$= \sigma^2 x_0^T (X^T X)^{-1} x_0$

b) We may be interested not in estimating the *expected* value of Y_0 but in predicting or forecasting the *actual* value of $Y_0 = \mathbf{x}_0'\boldsymbol{\beta} + \varepsilon_0$ that will be observed. The error in the forecast is then

$E(Y_0)$

$$D \equiv \hat{Y}_0 - Y_0 = \mathbf{x}_0'\mathbf{b} - (\mathbf{x}_0'\boldsymbol{\beta} + \varepsilon_0) = \mathbf{x}_0'(\mathbf{b} - \boldsymbol{\beta}) - \varepsilon_0$$

Show that $E(D) = 0$ and that $V(D) = \sigma_\varepsilon^2[1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0]$. Why is the forecast error D greater than the variance of δ found in part (a)?

$E(D) = E(\hat{Y}_0 - Y_0)$

$= E(x_0'(b-\beta) - \varepsilon_0)$

$= 0$

$V(D) = V(\hat{Y}_0 - Y_0)$

$= V(x_0'(b - \underbrace{\beta}_{\text{constant}}) - \varepsilon_0)$

$= x_0^T V(b) x_0 + V(\varepsilon_0)$

$= \sigma^2 x_0 (X^T X)^{-1} x_0 + \sigma^2$

[2 Pts.] *TRUE / FALSE Under the assumptions $E(Y) = X\beta$ and $Cov(Y) = \sigma^2 I$, all the fitted values have the same variance.*

FALSE. The variance-covariance matrix of the fitted values is $\sigma^2 H$ where H is the projection matrix $X(X^T X)^{-1} X^T$. The variances of the fitted values are the diagonal terms of this matrix, which need not all be the same.

5. [2 Pts.] *TRUE/FALSE If m is a submodel of a full model M, then $ErrSS(m) - ErrSS(M)$ and $ErrSS(M)$ are statistically independent under the assumption of normality.*

   TRUE. $ErrSS(m) - ErrSS(M) = y^T[(I - H_m) - (I - H_M)]y = y^T(H_M - H_m)y$, while $ErrSS(M) = y^T(I - H_M)y$. Since $(I - H_M)y$ and $(H_M - H_m)y$ are both multivariate normal functions of y under the normality assumption, and since they are also orthogonal, their quadratic forms must be independent.