1. (a) Minimize $||\beta||^2 = \beta' \beta$ with $X\beta - Y = 0$

Using the method of Lagrange multipliers, we obtain the function $L(\beta, \lambda) = \beta' \beta + \lambda (X\beta - Y)$. Differentiating we obtain:

(1) $\frac{\partial L}{\partial \beta} = 2\beta + X'\lambda$ and

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

From (1) we solve for $\beta$ to obtain $\beta = -\frac{X'\lambda}{2}$ and solve the equation $X\beta - Y = 0$ for $\lambda$.

$$X\beta - Y = \frac{-1}{2}XX'\lambda - Y = 0$$

$$\lambda = -2(XX')^{-1}Y$$

Which implies that

$\hat{\beta} = X'(XX')^{-1}Y$

(b) First we find $E[\hat{\beta}]$:

$$E[\hat{\beta}] = E[X'(XX')^{-1}(XB + \epsilon)]$$

$$= E[X'(XX')^{-1}X\beta + X'(XX')^{-1}\epsilon]$$

$$= E[X'(XX')^{-1}X\beta]$$

$$= X'(XX')^{-1}X\beta$$

So $\hat{\beta}$ is unbiased when $X'(XX')^{-1}X\beta = \beta$, and since $X'(XX')^{-1}X$ is a projection matrix onto the row space of $X$, $\hat{\beta}$ is unbiased when $\beta$ is in the row space of $X$.

(c)

$$cov(\hat{\beta}|X) = cov(X'(XX')^{-1}Y)$$

$$= cov(X'(XX')^{-1}(X\beta + \epsilon))$$

$$= cov(X'(XX')^{-1}(X\beta)) + cov(X'(XX')^{-1}\epsilon))$$

$$= 0 + X'(XX')^{-1}cov(\epsilon)(XX')^{-1}X$$

$$= X'(XX')^{-1}\sigma^2(XX')^{-1}X$$

(d)

$$E||\hat{Y} - Y||^2 = E||XX'(XX')^{-1}Y - Y||^2 = E||Y - Y||^2 = 0$$
(e) \[ E||\hat{Y} - E[Y]||^2 = E||\hat{Y} - X\beta||^2 \]
\[ = E[(\hat{Y} - X\beta)'(\hat{Y} - X\beta)] \]
\[ = E[\hat{Y}' - \beta'X')(\hat{Y} - X\beta)] \]
\[ = E[(Y'Y - Y'X\beta - (X\beta)'Y + (X\beta)'X\beta)] \text{ since } \hat{Y} = Y \]
\[ = E[Y'Y] - 2E[Y'X\beta] + (X\beta)'(X\beta) \]
\[ = 2(X\beta)'(X\beta) + n\sigma^2 - 2E[(X\beta)'(X\beta + \epsilon)] \]
\[ = n\sigma^2 \]

So \[ E||\hat{Y} - E[Y]||^2/n = \sigma^2. \]

(f) \[ Y^* = X\beta + \epsilon^*; E[\epsilon^*] = 0 \text{ and } var(\epsilon^*) = \sigma^2 \text{ and } Y^* \perp Y. \]
\[ E||Y^* - \hat{Y}||^2 = E||Y^* - Y||^2 \]
\[ = E[(Y^* - Y)'(Y^* - Y)] \]
\[ = E[(Y^*)'Y^* - (Y^*)'Y - Y'(Y^*) + Y'Y] \]
\[ = E[(Y^*)'Y^*] + E[Y'Y] - 2E[(Y^*)'Y] \]

Note \[ E[(Y^*)'Y] = E[Y^*]'E[Y] \text{ since } Y^* \perp Y \text{ and that } E[(Y^*)'Y^*] = nE[y_i^2] \]
\[ = nE[y_i^2] + nE[y_i^2] - 2nE[y_i^2] \]
\[ = 2n(E[y_i^2] - E[y_i]^2) \]
\[ = 2n(var(y)) \]
\[ = 2n\sigma^2. \]

So \[ E||Y^* - \hat{Y}||^2/n = 2\sigma^2 \]

2. (a)

Because the two design matrices \( X_1 \) and \( X_2 \) are independent, \( cov(\hat{\beta}_1, \hat{\beta}_2) = 0 \). This implies that:

\[ \text{var}(\frac{\hat{\beta}_1 + \hat{\beta}_2}{2}) = (1/4)(\text{var}(\hat{\beta}_1) + \text{var}(\hat{\beta}_2)) \]
\[ = (1/4)(\sum_{x_i \in X_1} \frac{\sigma^2}{(x_i - \bar{x})^2} + \sum_{x_i \in X_2} \frac{\sigma^2}{(x_i - \bar{x})^2}) \]
We want to see when this expression is equal to \( \text{var}(\hat{\beta}) \):

\[
(1/4)\left( \sum_{x_i \in X_1} \sigma^2 (x_i - \bar{x})^2 + \sum_{x_i \in X_2} \sigma^2 (x_i - \bar{x})^2 \right) = \text{var}(\hat{\beta})
\]

\[
(1/4)\left( \frac{\sum_{x_i \in X_1} \sigma^2 (x_i - \bar{x})^2 + \sum_{x_i \in X_2} \sigma^2 (x_i - \bar{x})^2}{\sum_{x_i = 1}^{n} (x_i - \bar{x})^2} \right) = \frac{\sigma^2}{\sum_{x_i = 1}^{n} (x_i - \bar{x})^2}
\]

\[
\frac{1}{\sum_{x_i \in X_1} (x_i - \bar{x})^2} + \frac{1}{\sum_{x_i \in X_2} (x_i - \bar{x})^2} = \frac{4}{\sum_{x_i = 1}^{n} (x_i - \bar{x})^2}
\]

Note that \( \sum_{x_i \in X_1} (x_i - \bar{x})^2 + \sum_{x_i \in X_2} (x_i - \bar{x})^2 = \sum_{x_i = 1}^{n} (x_i - \bar{x})^2 \Rightarrow \)

\[
\sum_{x_i \in X_1} (x_i - \bar{x})^2 + \sum_{x_i \in X_2} (x_i - \bar{x})^2 = \frac{4 \left( \sum_{x_i \in X_1} (x_i - \bar{x})^2 \right) \left( \sum_{x_i \in X_2} (x_i - \bar{x})^2 \right)}{\sum_{x_i \in X_1} (x_i - \bar{x})^2 + \sum_{x_i \in X_2} (x_i - \bar{x})^2}
\]

\[
\left( \sum_{x_i \in X_1} (x_i - \bar{x})^2 + \sum_{x_i \in X_2} (x_i - \bar{x})^2 \right)^2 = 4 \left( \sum_{x_i \in X_1} (x_i - \bar{x})^2 \right) \left( \sum_{x_i \in X_2} (x_i - \bar{x})^2 \right)
\]

\[
\left( \sum_{x_i \in X_1} (x_i - \bar{x})^2 - \sum_{x_i \in X_2} (x_i - \bar{x})^2 \right)^2 = 0
\]

That is, \( \bar{\beta} \) has the same variance as the OLS estimate when the two resulting sum of squares for \( X_1 \) and \( X_2 \) are equal. In this case, \( \bar{\beta} = \hat{\beta} \) from OLS, since

\[
\bar{\beta} = \frac{\hat{\beta}_1 + \hat{\beta}_2}{2} = (1/2) \sum_{x_i \in X_1} (x_i - \bar{x})
\]

\[
= \sum_{x_i \in X_1} (x_i - \bar{x})(y_i - \bar{y}) + \sum_{x_i \in X_2} (x_i - \bar{x})(y_i - \bar{y})
\]

\[
= \frac{\sum_{x_i \in X_1} (x_i - \bar{x})(y_i - \bar{y}) + \sum_{x_i \in X_2} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{x_i = 1}^{n} n(x_i - \bar{x})^2}
\]

\[
= \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}
\]

3
Which is the $\hat{\beta}$ from OLS.

(b) Let $ss_1$ denote $\sum_{x_i \in X_1} (x_i - \bar{x}_1)^2$ and $ss_2$ denote $\sum_{x_i \in X_2} (x_i - \bar{x}_2)^2$. It is an algebraic fact that, when $\bar{x}_1 \neq \bar{x}$,

$$ss_1 < \sum_{x_i \in X_1} (x_i - \bar{x})^2.$$ 

And so we have

$$\text{var}(\hat{\beta}) = \frac{1}{4}(\frac{\sigma^2}{ss_1} + \frac{\sigma^2}{ss_2}) > \frac{1}{4}(\frac{\sigma^2}{\sum_{x_i \in X_1} (x_i - \bar{x})^2} + \frac{\sigma^2}{\sum_{x_i \in X_2} (x_i - \bar{x})^2}) \geq \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x}_2)^2} = \text{var}(\hat{\beta}).$$

The difference is exactly

$$\frac{1}{ss_1} + \frac{1}{ss_2} - \frac{1}{\sum_{i=1}^{n} (x_i - \bar{x})^2}.$$ 

3. First note that the point on the line $(\bar{x}_i, y_i)$ for the horizontal residual of $(x_i, y_i)$ is given by $(\frac{y_i - \beta_0}{\beta_1}, y_i)$. We want to show that

$$\sum_{i=1}^{n} (x_i - \frac{y_i - \beta_0}{\beta_1}) = 0.$$ 

$$\sum_{i=1}^{n} (x_i - \frac{y_i - \beta_0}{\beta_1}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \frac{y_i}{\beta_1} + \sum_{i=1}^{n} \frac{\beta_0}{\beta_1}$$

$$= n\bar{x} - n\frac{\bar{y}}{\beta_1} + n\frac{\bar{y} - \beta_1\bar{x}}{\beta_1}$$

$$= n\bar{x} - n\frac{\bar{y}}{\beta_1} + n\frac{\bar{y}}{\beta_1} - n\bar{x} = 0$$

4. (a) We want $(I + FF')(I + \lambda FF') = I$

$$(I + FF')(I + \lambda FF') = I + \lambda F F' + F F' + \lambda F F' F F'$$

$$= I + (\lambda + 1 + \lambda F' F) F F'$$

We need $(\lambda + 1 + \lambda F' F) = 0 \rightarrow$

$$(\lambda + 1 + \lambda F' F) = 0 \rightarrow \lambda = \frac{-1}{1 + F' F}$$

4
\[(I - FF')(I + \lambda FF') = I + \lambda FF' - FF' = \lambda FF' FF'\]
\[= I + (\lambda - 1 - \lambda F'F)FF'\]

We need \((\lambda - 1 - \lambda F'F) = 0\) \(\Rightarrow\)
\[\lambda = \frac{1}{1 - F'F}\]

So \((I + FF')^{-1} = (I + \lambda FF')\) with \(\lambda = \frac{-1}{1 + F'F}\)
and
\((I - FF')^{-1} = (I + \lambda FF')\) with \(\lambda = \frac{1}{1 - F'F}\).

(b)

Write \(X'X = \sum_{i=1}^{n} x'_i x_i\) where \(x_i\) is the \(i^{th}\) row of \(X\). We know that \(X'X = I\) and that
\((\sigma^2(X'X)^{-1})_{ii} = \sigma^2 I_{ii} = \text{var}(\hat{\beta}_i)\).
Denote \(X_0\) as \(X\) with the new row \(x_0\) added. So we have \((\sigma^2(X'_0X_0)^{-1})_{ii} = \text{var}(\hat{\beta}_i)\).

Solve for \((X'_0X_0)^{-1}\):

\[(X'_0X_0)^{-1} = \left(\sum_{i=1}^{n} x'_i x_i + x'_0 x_0\right)^{-1}\]
\[= (I + \sum_{i=1}^{n} x'_i x_i)^{-1}\]
\[= (I - \frac{1}{1 + ||x_0||} x'_0 x_0)\text{ from part (a)}\]
\[= (I - \frac{1}{2} x'_0 x_0)\]

So \(\text{var}(\hat{\beta}_i) = \sigma^2(1 - \frac{1}{2} x^2_{0,i})\), where \(x^2_{0,i}\) is the \(i^{th}\) element of \(x_0\). To minimize this for all \(\hat{\beta}_i\), each \(x_{0,i}\) must be constant across \(i\), so choose each \(x_{0,i} = \sqrt{p + 1}\) for all \(i\).

5. (a)

\[
\text{cov}(Y, \hat{Y}) = \text{cov}(Y, X(X'X)^{-1}X'Y) \\
= X(X'X)^{-1}X'\text{cov}(Y, Y) \\
= X(X'X)^{-1}X'\sigma^2.
\]
(b)

\[
cov(Y, \hat{Y}) = \frac{1}{n-1} \sum_{i=1}^{n} Y_i \hat{Y}_i - n \bar{Y} \bar{\hat{Y}}
\]

\[
E[cov(Y, \hat{Y})] = E[\frac{1}{n-1} \sum_{i=1}^{n} Y_i \hat{Y}_i - n \bar{Y} \bar{\hat{Y}}]
\]

\[
= E[\frac{1}{n-1} Y'PY] - nE[(\frac{1}{n}1'Y')(\frac{1}{n}1'PY))]
\]

\[
= \frac{1}{n-1}E[tr(YY'P)] - \frac{1}{n}E[tr((1'Y')(1'PY))]
\]

First we simplify the left term:

\[
E[tr(YY'P)] = E[tr((X\beta + \epsilon)(X\beta + \epsilon)'P)]
\]

\[
= E[tr((X\beta'X' + \sigma^2I)P)]
\]

\[
= E[tr((X\beta'X')P)] + E[tr((\sigma^2I)P)]
\]

\[
= (X\beta)'X\beta + \sigma^2p \text{ where } p \text{ is the rank of the projection matrix}
\]

And the second term is:

\[
\frac{1}{n}E[tr((1'Y')(1'PY)))] = \frac{1}{n}E[tr((1'Y')(PY)'1)]
\]

\[
= \frac{1}{n}E[tr(1'(X\beta + \epsilon)(\beta'X' + \epsilon')1)]
\]

\[
= \frac{1}{n}E[tr(1'(X\beta'X' + \sigma^2I)1)]
\]

\[
= \frac{1}{n}E[tr(1'(X\beta'X') + 1'\sigma^2)]
\]

\[
= \frac{1}{n}((X\beta)'1)^2 + n\sigma^2)
\]

Combining terms, we get:
\[ E \hat{\text{cov}}(Y, \hat{Y}) = \frac{1}{n-1}[(X\beta)'X\beta + \sigma^2 p - \frac{1}{n}((X\beta)'1)^2 + n\sigma^2)] \]

Now let \( X\beta = \tilde{Y} \)
\[
= \frac{1}{n-1}[(\tilde{Y}'\tilde{Y} + (p-1)\sigma^2 - \frac{1}{n} (\tilde{Y}'1)^2] \]
\[
= \frac{(p-1)\sigma^2}{n-1} + \frac{1}{n-1} \sum_{i=1}^{n} \tilde{y}_i^2 - \frac{1}{n(n-1)} (\sum_{i=1}^{n} \tilde{y}_i)^2 \]
\[
= \frac{(p-1)\sigma^2}{n-1} + \frac{1}{n-1} \sum_{i=1}^{n} \tilde{y}_i^2 - \frac{1}{n-1} \frac{n(\tilde{Y})^2}{n} \]
\[
= \frac{(p-1)\sigma^2}{n-1} + \text{var}(\tilde{Y}). \]

So \( E \hat{\text{cov}}(Y, \hat{Y}) \) is non-negative.

6. (a)
\[
d(Q, l) = d((x_0, y_0), L(x_0, y_0)) = \frac{[\beta x_0 - y_0 + \alpha]}{\sqrt{\beta^2 + 1}} \]

and the loss function is given by:
\[
L(\beta, \alpha) = \sum_i \frac{(\beta x_i - y_i + \alpha)^2}{\beta^2 + 1}. \]

(b)
\[
d(Q, l) = |B'(Q - W)| \]

and the loss function is given by:
\[
L(B, W) = \sum_Q (B'(Q - W))(B'(Q - W))'. \]

(c)
Differentiating with respect to \( W \), we obtain the expression
\[
\sum_{Q} -2[B'(Q - w)B] = 0
\]
\[
= \sum_{Q} B'QB - B'WB = 0
\]
\[
\sum_{Q} B'QB = \sum_{Q} B'WB
\]

Which implies that
\[W = (\bar{x}, \bar{y})'\]

since
\[
\sum_{i} B'(x_i, y_i)B = nB'(\bar{x}, \bar{y})'B
\]

(d) Let \(S = (\bar{x}, \bar{y})'\). Then we can write the quadratic form as
\[
\sum_{Q} B'(Q - S)(B'(Q - S))'.
\]

7. (a)
Here we need \(V_n \to 0\) and \(\frac{\sigma^2}{(n-1)V_n} \to 0\).
So \(V_n\) must approach 0 slower than \(\frac{1}{n-1}\). Then, the fastest rate at which \(V_n \to 0\) is \(\frac{1}{n^{p-1}}\) for \(p < 1\).
(b) Here we need \(r_n \to 1\) and \(\text{var}(\hat{\beta}_1), \text{var}(\hat{\beta}_2) = \frac{1}{n(1-r_n)} \to 0\).
So we need \(1 - r_n^2\) to approach 0 slower than \(\frac{1}{n}\) or equivalently, we need \((1 - r_n)(1 + r_n) = 2(1 - r_n)\) to approach 0 slower than \((1/n)\). This implies that \(r_n\) can approach 1 no faster than \((1/n^p)\) for \(p < 1\).

8. In this setting,
\[
\text{var}(\hat{\beta}) = \frac{\sigma^2(1 - p_n)}{(n-1)\sigma_x^2} + \frac{k_n\sigma^2(p_n)}{(n-1)\sigma_x^2}
\]
So for (i) to hold, we need \(p_n + k_n p_n \to 0\) or \(p_n(k_n - 1) \to t\). Where \(t < \infty\).
For (ii) to hold, we need
\[
\text{var}(\hat{\beta}) = \frac{\sigma^2(1 - p_n)}{(n-1)\sigma_x^2} + \frac{k_n\sigma^2(p_n)}{(n-1)\sigma_x^2} = \frac{\sigma^2}{(n-1)\sigma_x^2}
\]
So we need the term

\[
\frac{\sigma^2(1 - p_n + k_n p_n)}{(n - 1)\sigma_x^2}
\]

to go to

\[
\frac{\sigma^2}{(n - 1)\sigma_x^2} \Rightarrow
\]

\[p_n + k_n p_n \to 0 \text{ or } p_n (k_n - 1) \to 0.\]