1. Suppose we have two covariates $X_1$ and $X_2$, and are interested in models of the form

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_1 X_2.$$ 

We then transform the covariates by adding a constant to each of them, yielding $\tilde{X}_1 = X_1 + c_1$ and $\tilde{X}_2 = X_2 + c_2$. This gives us a new working model

$$\hat{Y} = \tilde{\beta}_0 + \tilde{\beta}_1 \tilde{X}_1 + \tilde{\beta}_2 \tilde{X}_2 + \tilde{\beta}_3 \tilde{X}_1 \tilde{X}_2.$$ 

(a) Derive a simplified expression for $E[\hat{\beta}_1] - E[\tilde{\beta}_1]$.

**Solution:** Since the columnspaces of the two design matrices are the same, the fitted values must be identical:

$$\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_1 X_2 = \tilde{\beta}_0 + \tilde{\beta}_1 \tilde{X}_1 + \tilde{\beta}_2 \tilde{X}_2 + \tilde{\beta}_3 \tilde{X}_1 \tilde{X}_2$$

$$= \hat{\beta}_0 + \hat{\beta}_1 (X_1 + c_1) +$$

$$\tilde{\beta}_2 (X_2 + c_2) + \tilde{\beta}_3 (X_1 + c_1)(X_2 + c_2)$$

$$= \hat{\beta}_0 + \hat{\beta}_1 c_1 + \hat{\beta}_2 c_2 + \hat{\beta}_3 c_1 c_2 +$$

$$\tilde{\beta}_1 X_1 + \tilde{\beta}_2 X_2 + \tilde{\beta}_3 c_2 X_1 + \tilde{\beta}_3 c_1 X_2 +$$

$$\tilde{\beta}_3 X_1 X_2.$$ 

Equating the coefficients of $X_1$ yields

$$\hat{\beta}_1 = \tilde{\beta}_1 + c_2 \tilde{\beta}_3,$$

so $E[\hat{\beta}_1] - E[\tilde{\beta}_1] = c_2 E[\tilde{\beta}_3] = c_2 \beta_3$ (since $\tilde{\beta}_3 \equiv \beta_3$, see part b).

(b) Derive a simplified expression for $E[\hat{\beta}_3] - E[\tilde{\beta}_3]$.

**Solution:** $\hat{\beta}_3 \equiv \tilde{\beta}_3$, so their expected values are identical.

(c) Are $Z[\hat{\beta}_3]$ and $Z[\tilde{\beta}_3]$ equal (where $Z[\hat{\theta}] = \hat{\theta} / SD(\hat{\theta})$ is the Z-score)? You do not need to derive an explicit expression for their difference, but you must explain your reasoning as to why they are equal or different.

**Solution:** $\hat{\beta}_3 \equiv \tilde{\beta}_3$, so their expected values, standard deviations, and Z-scores are identical.
(d) Suppose we fit the two models without the interaction (i.e. \( \hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 \) and \( \tilde{Y} = \tilde{\beta}_0 + \tilde{\beta}_1 \tilde{X}_1 + \tilde{\beta}_2 \tilde{X}_2 \)). Derive an expression for \( E[\hat{\beta}_1] - E[\tilde{\beta}_1] \) in this setting.

**Solution:** If we don’t include the interaction term, \( \hat{\beta}_1 \equiv \tilde{\beta}_1 \), so \( E[\hat{\beta}_1] - E[\tilde{\beta}_1] = 0 \) in this situation.

2. Suppose we have a sample of size \( n \), where \( n \) is even, and we use ordinary least squares to fit a working model \( \hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 \), relating a standardized covariate \( X_1 \) to an outcome \( Y \). The observations are correlated, so that \( \text{cor}(Y_{2j}, Y_{2j+1}|X) = r \) for \( j = 1, 2, \ldots, n/2 \).

(a) Derive an expression for the variance of \( \hat{\beta}_1 \).

**Solution:**

\[
\text{cov}(\hat{\beta}) = (X'X)^{-1}X'SX(X'X)^{-1},
\]

where \( S \) is a matrix with 2 \( \times \) 2 blocks of the form

\[
\sigma^2 \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}
\]

along the main diagonal. Under the conditions given, \( (X'X)^{-1} = n^{-1}I_2 \). Therefore

\[
[X'SX]_{ij} = \sum_{kk'} X_{ki}S_{kk'}X_{kj}
\]

\[
= \sigma^2 \sum_k X_{ki}X_{kj} + r \left( \sum_{k \text{ odd}} X_{ki}X_{k+1,j} + X_{k+1,i}X_{kj} \right)
\]

If \( i = j = 1 \), then

\[
\text{cov}(\hat{\beta}_1) = n^{-2}(n\sigma^2 + r \left( \sum_{k \text{ odd}} X_{k1}X_{k+1,1} + X_{k+1,1}X_{k1} \right))
\]

(b) Under what conditions on \( X \) will the variance of \( \hat{\beta}_1 \) not depend on \( r \)? Interpret the meaning of this condition.

**Solution:** If \( \sum_{k \text{ odd}} X_{k1}X_{k+1,1} + X_{k+1,1}X_{k1} = 0 \), then the variance of \( \hat{\beta}_1 \) does not depend on \( r \). The condition resembles a correlation coefficient between the \( X_1 \) values within each correlated pair of observations.
3. Suppose we have a \( n \times 3 \) design matrix that satisfies

\[
(n(X'X))^{-1} = \begin{pmatrix} 1 & 0.5 & 0.2 \\ 0.5 & 1 & 0.5 \\ 0.2 & 0.5 & 1 \end{pmatrix}.
\]

Our goal is to estimate \( \beta_2 - \beta_1 \), where the working model is \( \hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 \).

(a) Suppose that the model is correctly specified, so that \( E[Y|X] = X\beta \) and \( \text{var}(Y|X) = \sigma^2 I \). Thus the coefficient estimates are unbiased, and the standard errors can be estimated using the usual approach. What sample size is required so that the standard error of \( \hat{\beta}_2 - \hat{\beta}_1 \) is less than 0.1?

**Solution:** We are working with a contrast having coefficients \((0, -1, 1)\). Thus the standard error is \( \sigma^2 / n \), and the sample size is \( n = 100\sigma^2 \).

(b) Now suppose that the working model is incorrect in such a way that the coefficient estimates are biased, with \( E[\hat{\beta}_2 - \hat{\beta}_1] = \beta_2 - \beta_1 + \theta \). However the standard error still can be calculated in the usual way. Derive an expression for the approximate coverage probability of the usual 95% confidence interval for \( \beta_2 - \beta_1 \). Then, using the standard normal tail probabilities given below, determine a specific condition such that the coverage probability is around 80%.

<table>
<thead>
<tr>
<th>( x )</th>
<th>( P(Z \leq x) )</th>
<th>( x )</th>
<th>( P(Z \leq x) )</th>
<th>( x )</th>
<th>( P(Z \leq x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.500</td>
<td>1.0</td>
<td>0.841</td>
<td>2.0</td>
<td>0.977</td>
</tr>
<tr>
<td>0.1</td>
<td>0.540</td>
<td>1.1</td>
<td>0.864</td>
<td>2.1</td>
<td>0.982</td>
</tr>
<tr>
<td>0.2</td>
<td>0.579</td>
<td>1.2</td>
<td>0.885</td>
<td>2.2</td>
<td>0.986</td>
</tr>
<tr>
<td>0.3</td>
<td>0.618</td>
<td>1.3</td>
<td>0.903</td>
<td>2.3</td>
<td>0.989</td>
</tr>
<tr>
<td>0.4</td>
<td>0.655</td>
<td>1.4</td>
<td>0.919</td>
<td>2.4</td>
<td>0.992</td>
</tr>
<tr>
<td>0.5</td>
<td>0.691</td>
<td>1.5</td>
<td>0.933</td>
<td>2.5</td>
<td>0.994</td>
</tr>
<tr>
<td>0.6</td>
<td>0.726</td>
<td>1.6</td>
<td>0.945</td>
<td>2.6</td>
<td>0.995</td>
</tr>
<tr>
<td>0.7</td>
<td>0.758</td>
<td>1.7</td>
<td>0.955</td>
<td>2.7</td>
<td>0.997</td>
</tr>
<tr>
<td>0.8</td>
<td>0.788</td>
<td>1.8</td>
<td>0.964</td>
<td>2.8</td>
<td>0.997</td>
</tr>
<tr>
<td>0.9</td>
<td>0.816</td>
<td>1.9</td>
<td>0.971</td>
<td>2.9</td>
<td>0.998</td>
</tr>
</tbody>
</table>

**Solution:**

\[
P(\hat{\beta}_2 - \hat{\beta}_1 - 2\text{SE} \leq \beta_2 - \beta_1 \leq \hat{\beta}_2 - \hat{\beta}_1 + 2\text{SE}) =
P(-2 \leq (\hat{\beta}_2 - \hat{\beta}_1 - (\beta_2 - \beta_1)) / \text{SE} \leq 2) =
P(-2 - \theta / \text{SE} \leq Z \leq 2 - \theta / \text{SE}).
\]

If \( 2 - \theta / \text{SE} \approx 0.9 \) then the coverage probability is around 80%.
4. Suppose we take the OLS estimate \( \hat{\beta} \) of a vector of regression coefficients, and form a new estimate \( \tilde{\beta} = f\hat{\beta} \), where \( 0 \leq f \leq 1 \) is a scalar. Derive the MSPE for \( \tilde{\beta} \), and derive an expression for the value of \( f \) that minimizes the MSPE. You can calculate the MSPE using the same design matrix as in the training set.

Solution:

\[
\begin{align*}
n^{-1}\|Y^*-X\tilde{\beta}\|^2 &= n^{-1}\|Y^*-X\beta\|^2 + n^{-1}\|X\beta - X\tilde{\beta}\|^2 \\
&= \sigma^2 + \|X\beta - X\tilde{\beta}\|^2/n \\
&= \sigma^2 + n^{-1}\|X(\beta - f\beta)\|^2 + n^{-1}\|X(f\beta - f\hat{\beta})\|^2 \\
&= \sigma^2 + (1-f)^2n^{-1}\|X\beta\|^2 + f^2n^{-1}\|X(\hat{\beta} - \beta)\|^2
\end{align*}
\]

The MSPE is the expected value:

\[E[n^{-1}\|Y^*-X\tilde{\beta}\|^2] = \sigma^2 + (1-f)^2n^{-1}\|X\beta\|^2 + f^2p\sigma^2/n.\]

The optimal value of \( f \) is

\[\|X\beta\|^2/(\|X\beta\|^2 + p\sigma^2).\]

5. Suppose we have a single standardized covariate \( X \in \mathbb{R}^n \) and a response variable \( Y \in \mathbb{R}^n \). We aim to fit a regression without an intercept by minimizing the loss function

\[\|Y - \beta X\|^2 + \lambda|\beta|\]

over \( \beta \in \mathcal{R} \), where \( \lambda \geq 0 \) is a scalar.

(a) Is the loss function convex? Explain your reasoning.

Solution: The two summands are convex, and the sum of two convex functions is convex. Thus the loss function is convex.

(b) Derive an expression for a value \( K \) such that if \( \lambda \geq K \) then the value of \( \beta \) that minimizes the loss function is zero.

Solution: Minimizing the loss function is equivalent to minimizing

\[L(\beta) = -2nr\beta + n\beta^2 + \lambda|\beta|,\]
where $Y'X = nr$. Since $L(\beta) = 0$ and $L$ is convex, we can conclude that $\beta$ is the minimizer as long as $L$ is decreasing for $\beta < 0$ and increasing for $\beta > 0$. The derivative of $L$ for $\beta > 0$ is $-2nr + 2n\beta + \lambda$, and the derivative of $L$ for $\beta > 0$ is $-2nr + 2n\beta - \lambda$. Thus if $\lambda > 2n|r|$ then the minimizer is $\beta = 0$. 