STAT 408/608 Homework 7 Solutions: Written Section

March 29, 2015

1. (a) In this case, X_2 is perfectly explained by X_1. This means that the residuals of X_2 regressed on X_1 are all exactly 0. So, the plotted points in the added variable plot will all have horizontal axis value 0. This means the plot will be a straight vertical line of points scattered in the vertical direction.

   (b) In this case, Y is perfectly explained by X_1. This means that the residuals of Y regressed on X_1 are all exactly 0. So, the plotted points in the added variable plot will all have vertical axis value 0. This means the plot will be a straight horizontal line of points scattered in the horizontal direction.

   (c) The plot will look like residual plot, scatter of data points. Since X_2 is not explained by X_1 at all, the residual of Y regressing on X_1 is orthogonal to the residual of X_2 regressing on X_1.

2. (a) $Var(\hat{\beta}|x) = \sigma^2(x'x)^{-1} = \begin{bmatrix} \sigma^2/n & 0 & 0 \\ 0 & \sigma^2/(1-\rho^2) & -\rho\sigma^2/(1-\rho^2) \\ 0 & -\rho\sigma^2/(1-\rho^2) & \sigma^2/(1-\rho^2) \end{bmatrix}$

   So, $Var(\hat{\beta}_1|x) = Var(\hat{\beta}_2|x) = \sigma^2/(1-\rho^2$
   if $\rho \to 0$, then $Var(\hat{\beta}_1|x) = Var(\hat{\beta}_2|x) \to \sigma^2$
   if $\mid \rho \mid \to 1$, then $Var(\hat{\beta}_1|x) = Var(\hat{\beta}_2|x) \to \infty$

   The $\mid \rho \mid \to 1$, the variables x_1 and x_2 become more correlated, which indicates both variable contain mostly the same information. This is multicollinearity problem which will lead to estimation of β being less accurate.

   (b) Because the mean of X is zero, so:
   $SXX = \sum(x_i - \bar{x})^2 = \sum x_i^2 = 1$
   $Var(\hat{\beta}_j = \dfrac{1}{1-\rho^2} \times \dfrac{\sigma^2}{(n-1)S^2_{x_j}} = \dfrac{1}{1-\rho^2} \times \dfrac{\sigma^2}{(n-1)\dfrac{SXX}{n-1}} = \dfrac{\sigma^2}{1-\rho^2}$
   which matches the VIF shown in page 203.

3. (a) The design matrix:

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 2 & 1 \\ 1 & 3 & 0 \\ 1 & 3 & 1 \end{bmatrix}$$

The vector $(x_1 - \bar{x}_1) = (-1, -1, 0, 0, 1, 1)'$ is orthogonal to vector $(x_2 - \bar{x}_2) = (-0.50.5, -0.50.5, -0.50.5,)'$, because their dot product is 0.

$VIF = \frac{1}{1 - r_{12}^2}$, where $r_{12} = corr(x_1, x_2) = \frac{cov(x_1, x_2)}{\sigma_{x_1}\sigma_{x_2}}$

$cov(x_1, x_2) = E[(x_1 - \bar{x}_1)(x_2 - \bar{x}_2)] = E[(x_1 - 2)(x_2 - 0.5)] = 0$, therefore, $VIF = 1$

(b) The design matrix:

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 3 & 0 \\ 1 & 2 & 1 \\ 1 & 3 & 1 \\ 1 & 4 & 1 \end{bmatrix}$$

$cov(x_1, x_2) = E[(x_1 - \bar{x}_1)(x_2 - \bar{x}_2)] = 0.3$, $\sigma_{x_1} = 1.4088$, and $\sigma_{x_2} = 0.5477$, therefore, $VIF = \frac{1}{1 - 0.273} = 1.375$

The variance inflation factor will be larger than part (a) because the two variables are now correlated.

(c) The variance inflation factors would all equal 1 if the predictors are all orthogonal to each other. In this case the $R^2 = 0$ which will result in $VIF = 1$.