STAT3008 Exercise 12 Solutions

(2011-2012 2^{nd} Semester)

Q1. (10.2)

R Codes:

```
###Forward Selection###
library(alr3)
m=lm(Y~1,data=mantel)
X1=mantel$X1
X2=mantel$X2
X3=mantel$X3
step(m,scope=~X1+X2+X3,direction="forward")
```

By using backward selection in R, we can find out that $X_A = \{X_1\}$ gives the smallest AIC which is equivalent to $C(p)$ in this case. The corresponding AIC is -0.309.

R Codes:

```
###Backward Elimination###
library(alr3)
m=lm(Y~X1+X2+X3,data=mantel)
X1=mantel$X1
X2=mantel$X2
X3=mantel$X3
step(m,scope=~1,direction="backward")
```

By using backward selection in R, we can find out that $X_A = \{X_1, X_2\}$ gives the smallest AIC which is equivalent to $C(p)$ in this case. The corresponding AIC is -316.2. It can be shown that $X_A = \{X_1, X_2\}$ is a perfect fit for response Y.

Therefore, $X_A = \{X_1, X_2\}$ should be adopted.

Q2. (10.8)

$$E(Y \mid X_C) = E\left[E(Y \mid X) \mid X_C\right]$$
$$= E\left[\left(\beta_A^T X_A + \beta_I^T X_I\right) \mid X_C\right]$$
$$= \beta_A^T E(X_A \mid X_C) + \beta_I^T E(X_I \mid X_C)$$
$$= \beta_A^T E(X_A \mid X_C) + 0$$
$$= \beta_A^T E(X_A \mid X_C)$$

Therefore, we will have a linear mean function if the regressions of each of variables in X_A on X_C have linear mean function.

Q3. (10.9)

The data are not experimental but rather observational, we cannot infer causation from these data, and so the negative coefficient estimate does not necessarily imply that higher speed limits causes lower accident rates.

Q4. (10.10)

$$C_p = (k' - p)(F_p - 1) + p$$

Where k' is the number of terms in the full mean function, p is the number of terms in the candidate for the active subset, F_p is the F-statistic for comparing these two mean functions. For a fixed value of p, orders mean functions in the same wat as F_p. will be smaller than p only if $F_p < 1$; under H_0 that the smaller mean function is appropriate, the expected value of F_p is close to one, so subsets with values of C_p substantially smaller than p are not clearly superior to subsets with $C_p \approx p$.