

Stat 6021: Guided Question Set 7 Solutions

1. The model with x_2, x_7, x_8 as predictors has the lowest BIC. These predictors are the season passing yards of the team, the percent of plays that are rushes by the team, and opponents' season rushing yards.

2.

```
PRESS <- function(linear.model) {  
  ## calculate the predictive residuals  
  pr <- residuals(linear.model)/(1-lm.influence(linear.model)$hat)  
  ## calculate the PRESS  
  PRESS <- sum(pr^2)  
  
  return(PRESS)  
}
```

3. The PRESS statistic is 87.46123. The $R^2_{Prediction}$ is

$$\begin{aligned} R^2_{Prediction} &= 1 - \frac{PRESS}{SS_T} \\ &= 1 - \frac{87.46123}{326.9643} \\ &= 0.7325052 \end{aligned}$$

The model might be able to explain 73.25% of the variability in the new observations. The R^2 of this model is 0.8425. Both values are fairly high, so the model has good predictive ability.

4. Answer would vary but $R^2_{Prediction}$ with half the observations will generally be smaller.
5. The standard errors of the estimated coefficients are larger when half the dataset is used.
6. Answer would vary but SS_{res} for half the dataset should be larger.