HW7 Solutions

Problem 10.2

(a). Setting significance level at $\alpha = 0.05$, we keep eliminating the predictor with the largest p-value, until it becomes smaller than α. We get the following sequence of dropped variables:

```
lmod <- lm(lpsa~ ., data = prostate)
pval <- head(sort(summary(lmod)$coef[-1,4], decreasing = TRUE),1)
lmod2 <- update(lmod, .~.-gleason)
pval <- c(pval,head(sort(summary(lmod2)$coef[-1,4], decreasing = TRUE),1))
lmod3 <- update(lmod2, .~.-lcp)
pval <- c(pval, head(sort(summary(lmod3)$coef[-1,4], decreasing = TRUE),1))
lmod4 <- update(lmod3, .~.-pgg45)
pval <- c(pval, head(sort(summary(lmod4)$coef[-1,4],decreasing = TRUE),1))
lmod5 <- update(lmod4, .~.-age)
pval <- c(pval, head(sort(summary(lmod5)$coef[-1,4],decreasing = TRUE),1))
lmod6 <- update(lmod5, .~.-lbph)
pval <- c(pval, head(sort(summary(lmod6)$coef[-1,4],decreasing = TRUE),1))
pval
```

```
##     gleason         lcp       pgg45         age        lbph         svi
## 0.775032844 0.251268799 0.253309151 0.169528154 0.112129523 0.002029012
```

Thus the model selected by backward elimination has lcavol, lweight and svi as predictors. R^2 has dropped by ~0.028.

(b).

```
require(leaps)
```

```
## Loading required package: leaps
```

```
## Warning: package 'leaps' was built under R version 3.4.2
```

```
rs <- summary(regsubsets(lpsa ~ ., data = prostate))
AIC <- 97*log(rs$rss/97) + 2*seq(2,9)
which(rs$which[which.min(AIC),])
```

```
## (Intercept)      lcavol     lweight         age        lbph         svi
##           1           2           3           4           5           6
```

So the optimal model according to the AIC criterion has lcavol, lweight, age, lbph and svi as predictors.

(c).

```
which(rs$which[which.max(rs$adjr2),])
```

```
## (Intercept)      lcavol     lweight         age        lbph         svi
##           1           2           3           4           5           6
##         lcp       pgg45
##           7           9
```

We can see that R_a^2 only leaves gleason out of the model.

(c).

```
which(rs$which[which.min(rs$cp),])
```

```
## (Intercept)      lcavol     lweight        lbph         svi
```

```
##             1             2             3             5             6
```

C_p statistics leaves age, lcp and gleason out of the model.