HW2 Solutions - Stats 500

Problem 1:

```
library(faraway)
data(teengamb)
```

```
gambmod = lm(gamble ~ sex + status + income + verbal, data = teengamb)
lmsum = summary(gambmod)
lmsum$r.squared
```

```
## [1] 0.5267234
```

(a).

The multiple R-squared is 0.5267, so about 53% of variation in response is explained by the predictors.

(b).

Observation with the largest residual:

```
res = residuals(gambmod)
unname(which.max(res))
```

```
## [1] 24
```

(c).

```
cat("The mean of residuals: ", mean(res), "\n")
```

```
## The mean of residuals:  -3.065293e-17
```

```
cat("And the median: ", median(res))
```

```
## And the median:  -1.451392
```

(d).

```
cor(res, fitted.values(gambmod))
```

```
## [1] -1.070659e-16
```

(e).

```
cor(res, teengamb$income)
```

```
## [1] -7.242382e-17
```

(f).

Since sex is coded as 0 = male, 1 = female, assuming all other variables are held constant, the predicted difference (male - female) is

```
(0 - 1) * coef(gambmod)[2]
```

```
##      sex
## 22.11833
```

Problem 7:

```
library(faraway)
data(wafer)
wafmod = lm(resist ~ x1 + x2 + x3 + x4, data = wafer)
```

```
X = model.matrix(wafmod)
head(X)
```

```
##   (Intercept) x1+ x2+ x3+ x4+
## 1           1   0   0   0   0
## 2           1   1   0   0   0
## 3           1   0   1   0   0
## 4           1   1   1   0   0
## 5           1   0   0   1   0
## 6           1   1   0   1   0
```

Comaring this with the original data we find that the predictors have been coded as - : 0 , + : 1.

(b).

```
cor(X)
```

```
## Warning in cor(X): the standard deviation is zero
```

```
##             (Intercept) x1+ x2+ x3+ x4+
## (Intercept)           1  NA  NA  NA  NA
## x1+                  NA   1   0   0   0
## x2+                  NA   0   1   0   0
## x3+                  NA   0   0   1   0
## x4+                  NA   0   0   0   1
```

NA's appear in the first column/row because the intercept column is constant, so it has zero variance and undefined correlation with other columns.

(c).

The expected difference is

```
(1 - 0) * unname(coefficients(wafmod)[2])
```

```
## [1] 25.7625
```

(d).

```
wafmod2 = lm(resist ~ x1 + x2 + x3, data = wafer)
summary(wafmod)
```

```
##
## Call:
## lm(formula = resist ~ x1 + x2 + x3 + x4, data = wafer)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -43.381 -17.119   4.825  16.644  33.769
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   236.78      14.77  16.032 5.65e-09 ***
## x1+            25.76      13.21   1.950 0.077085 .
## x2+           -69.89      13.21  -5.291 0.000256 ***
## x3+            43.59      13.21   3.300 0.007083 **
## x4+           -14.49      13.21  -1.097 0.296193
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 26.42 on 11 degrees of freedom
## Multiple R-squared:  0.7996, Adjusted R-squared:  0.7267
## F-statistic: 10.97 on 4 and 11 DF,  p-value: 0.0007815
```

```
summary(wafmod2)
```

```
##
## Call:
## lm(formula = resist ~ x1 + x2 + x3, data = wafer)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -36.137 -20.550   3.575  18.462  41.013
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   229.54      13.32  17.231 7.88e-10 ***
## x1+            25.76      13.32   1.934 0.077047 .
## x2+           -69.89      13.32  -5.246 0.000206 ***
## x3+            43.59      13.32   3.272 0.006677 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 26.64 on 12 degrees of freedom
## Multiple R-squared:  0.7777, Adjusted R-squared:  0.7221
## F-statistic: 13.99 on 3 and 12 DF,  p-value: 0.0003187
```

Comparing the two fitted models, we observe that: 1. The coefficients for x1, x2 and x3 remain unchanged, while the intercept has increased. 2. The standard errors for x1, x2 and x3 have increased slightly, but that of the intercept has decreased. 3. As we would expect, we have a lower R-squared value for the new model

(e).

Since x1, x2, x3 are are all uncorrelated with x4, adding or removing x4 to the model affects only the intercept and the other coefficients remain unchanged.