Problem set 1 Q4 Solution

TA Team

```
library(ISLR)
```

4(a) Fit a multiple regression model to predict Sales using Price, Urban, and US.

```
mod1 <- lm(data=Carseats, Sales ~ Price + Urban + US)
```

4(b) Provide an interpretation of each coefficient in the model. Be careful–some of the variables in the model are qualitative!

```
?Carseats ## gives us units of the variables
summary(mod1)
```

```
##
## Call:
## lm(formula = Sales ~ Price + Urban + US, data = Carseats)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.9206 -1.6220 -0.0564  1.5786  7.0581
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.043469   0.651012  20.036  < 2e-16 ***
## Price       -0.054459   0.005242 -10.389  < 2e-16 ***
## UrbanYes    -0.021916   0.271650  -0.081    0.936
## USYes        1.200573   0.259042   4.635 4.86e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.472 on 396 degrees of freedom
## Multiple R-squared:  0.2393, Adjusted R-squared:  0.2335
## F-statistic: 41.52 on 3 and 396 DF,  p-value: < 2.2e-16
```

Interpretation

1. Raising the price of carseats by 1 dollar is associated with 54.46 fewer car seats sold, with other variables held fixed.
2. A store being in an urban locations is associated with selling 21.92 fewer car seats than non-urban stores, with other variables held fixed.

3. A store being in the US is associated with selling 1200.57 more car seats than non-US stores, with other variables held fixed.

4(c) Write out the model in equation form, being careful to handle the qualitative variables properly.

$$\text{Sales}_i = 13.04 - 0.05 \times \text{Price}_i - 0.02 \times \mathbb{I}[\text{Urban}_i = \text{Yes}] + 1.20 \times \mathbb{I}[\text{US}_i = \text{Yes}]$$

4(d) For which of the predictors can you reject the null hypothesis H0 : Beta_j = 0? Use the significance level 0.05 for the hypothesis test.

We can reject the hypothesis of $\beta_j = 0$ at the 5% level for all β_j except for β_{Urban}.

4(e) On the basis of your response to question (d), fit a smaller model that only uses the predictors for which there is evidence of association with the outcome.

```
mod2 <- lm(data=Carseats, Sales ~ Price + US)
```

4(f) What are the value of R2 for models in (a) and (e)? Does larger R2 mean the model fit the data better?

```
summary(mod1)$r.squared
```

```
## [1] 0.2392754
```

```
summary(mod2)$r.squared
```

```
## [1] 0.2392629
```

Model 1 has a larger R^2 value.

No. R-squared measures the percent of variation in Y explained by variation in X, and it will always increase as we add more covariates into the model. A more appropriate criterion for model selection would be the adjusted R-squared which takes into account the model complexity.

4(g) Using the model from (e), construct the 95 % confidence interval(s) for the coefficient(s).

```
confint(mod2)
```

```
##                   2.5 %      97.5 %
## (Intercept) 11.79032020 14.27126531
## Price       -0.06475984 -0.04419543
## USYes        0.69151957  1.70776632
```

4(h) Fit a linear regression model in (e) with interaction effect(s). Provide an interpretation of each coefficient in the model.

```
mod3 <- lm(data=Carseats, Sales ~ Price*US)
summary(mod3)
```

```
##
## Call:
## lm(formula = Sales ~ Price * US, data = Carseats)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.9299 -1.6375 -0.0492  1.5765  7.0430
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.974798   0.953079  13.614  < 2e-16 ***
## Price       -0.053986   0.008163  -6.613 1.22e-10 ***
## USYes        1.295775   1.252146   1.035    0.301
## Price:USYes -0.000835   0.010641  -0.078    0.937
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.472 on 396 degrees of freedom
## Multiple R-squared:  0.2393, Adjusted R-squared:  0.2335
## F-statistic: 41.52 on 3 and 396 DF,  p-value: < 2.2e-16
```

Interpretation

On average:

1. Raising the price of carseats by 1 dollar is associated with 53.99 fewer car seats sold for a non-US store.
2. A store being in the US is associated with selling 1295.78 more car seats than non-US stores when the Price is zero.
3. For stores in US, raising the price of carseats by 1 dollar is associated with 0.84 fewer car seats sold comparing to the stores not in the US.