Sampling: HW7
Leslie Gains-Germain

1. The estimated regression equation I would propose is:

$$\hat{fish} = \hat{\beta}_0 + \hat{\beta}_1 x_1$$

where $\hat{fish}$ is the estimated mean number of fish caught by anglers on this lake in August after fishing x_1 hours. $\hat{\beta}_0$ is the estimated mean number of fish caught after fishing 0 hours (we would expect this estimate to be near 0). $\hat{\beta}_1$ is the estimated mean number of fish caught per hour by anglers on this lake in August. $\hat{\beta}_1$ is the quantity of interest.

2. The estimated regression equation I would propose is:

$$\hat{p}_{sports} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta_{11}} x_1^2 + \hat{\beta}_2 x_2$$

where $\hat{p}_{sports}$ is the estimated proportion of time in television news broadcasts in Bozeman that is devoted to sports in month x_1, with x_2 breaking news events. x_1 is the month of the year ($x_1 \in (1, 2, ..., 12)$), and x_2 is the number of breaking news events that occur each month. I would expect the proportion of time devoted to sports to peak in February because of the superbowl, and then decrease after that. I have included a squared term in the estimated regression equation to allow for this curvature in the relationship. For the second predictor, I would expect months with greater numbers of breaking news events to have lower proportions of time devoted to sports because more time will be devoted to these breaking news events.

3. (a) I fit the three regression models in the code below.

```
lm.1 <- lm(volume ~ diameter, data = cherry)
lm.2 <- lm(volume ~ height, data = cherry)
lm.3 <- lm(volume ~ diameter + height, data = cherry)
```

(b) Using the first regression model, the total volume for all black cherry trees in the forest is estimated to be 102318.9 cubic feet, with a 95% confidence interval from 97709.0 to 106928.7 cubic feet. My work is shown below.

```
n <- 31
N <- 2967
ybar <- mean(cherry$volume)
xbard <- mean(cherry$diameter)
txd <- 41835
thaty1 <- N * ybar + coef(lm.1)[2] * (txd - N * xbard)
c <- N * (N - n) / n
mse1 <- anova(lm.1)[2, 3]
var.that.y1 <- c * mse1
tstar <- qt(0.975, 29)
ci1 <- c(thaty1 - tstar * sqrt(var.that.y1), thaty1 + tstar * sqrt(var.that.y1))
```

Using the third regression model, the total volume for all black cherry trees in the forest is estimated to be 52884.3 cubic feet, with a 95% confidence interval from 38359.7 to 67408.9 cubic feet. My work is shown below.

```
xbarh <- mean(cherry$height)
txh <- 201756
thaty2 <- N * ybar + coef(lm.2)[2] * (txh - N * xbarh)
mse2 <- anova(lm.2)[2, 3]
var.that.y2 <- c * mse2
ci2 <- c(thaty2 - tstar * sqrt(var.that.y2), thaty2 + tstar * sqrt(var.that.y2))
```

Using the third regression model, the total volume for all black cherry trees in the forest is estimated to be 93362.5 cubic feet, with a 95% confidence interval from 89153.9 to 97571.1 cubic feet. My work is shown below.

```
thaty3 <- N * ybar + coef(lm.3)[2] * (txd - N * xbard) +
                    coef(lm.3)[3] * (txh - N * xbarh)
mse3 <- anova(lm.3)[3, 3]
var.that.y3 <- c * mse3
ci3 <- c(thaty3 - tstar * sqrt(var.that.y3), thaty3 + tstar * sqrt(var.that.y3))
```

(c) I prefer model 1 because the proportion of the variability in tree volume explained by the regression on diameter is estimated to be 0.9353, while the proportion of the variability in tree volume explained by the regression on height is only 0.3579. As a result, the confidence interval for the total tree volume obtained from model 1 is narrower than the confidence interval obtained from model 2.

(d) There is evidence that including both variables provides an improved model for estimating the total volume. There is strong evidence that model 3 is an improvement over model 1 (p-value = 0.014 from extra-sums-of-squares F-stat = 6.79 on 1 and 28 df). I also noticed that the adjusted R^2 for model 3 is larger than the adjusted R^2 for model 1. As a result, the confidence interval for the true total tree volume is narrower for model 3.

```
anova(lm.1, lm.3)

## Analysis of Variance Table
##
## Model 1: volume ~ diameter
## Model 2: volume ~ diameter + height
##   Res.Df RSS Df Sum of Sq    F Pr(>F)
## 1     29 524
## 2     28 422  1       102 6.79  0.014

ci3[2]-ci3[1]

## diameter
##     8417

ci1[2]-ci1[1]

## diameter
##     9220
```

(e) Assuming the total diameter and total height of black cherry trees in the forest is 41835 and 201756 inches, respectively, we are 95% confident that the true total black cherry tree volume in this forest is between 89153.9 and 97571.1 cubic feet.