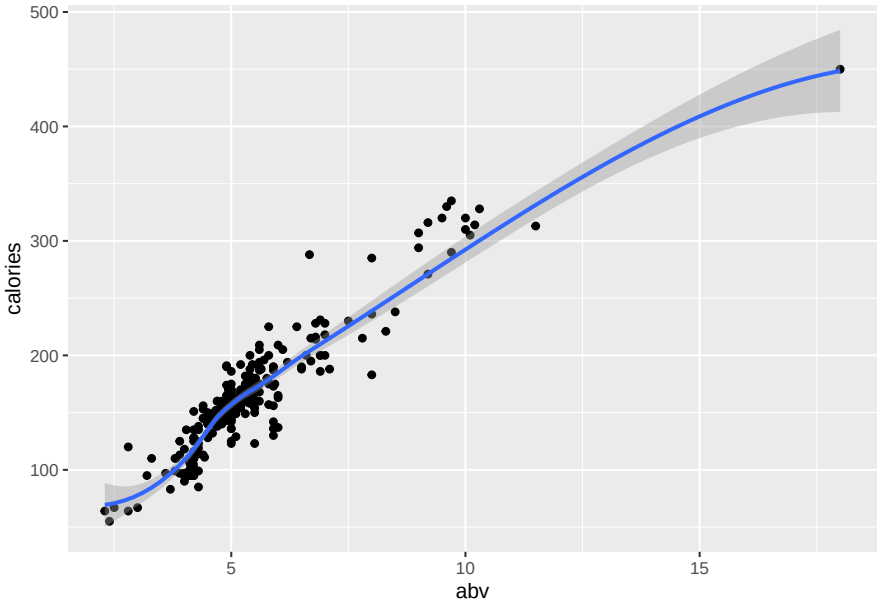
# STAT120 Wk 98 Group Activity

Victor Huang, Neda Tehrani, Aisha Dem

5/26/2021

## Problem 1




(a) An increase in abv leads to an increase in calories

(b) The data appears to be distributed fairly evenly around the linear model. As such, it is appropriate to use a linear model to make predictions.

```
##
## Call:
## lm(formula = calories ~ abv, data = beer)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -98.616  -9.626   1.168  11.260  85.186
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.7597     4.2129   -0.18    0.857
## abv          30.5209     0.7537   40.49   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.91 on 267 degrees of freedom
## Multiple R-squared:   0.86,  Adjusted R-squared:  0.8595
```

```
## F-statistic:  1640 on 1 and 267 DF,  p-value: < 2.2e-16
```

c.i) What is the equation of the least squares regression line for the sample?

calories = -0.7597 + 30.5209(abv)

c.ii) If we are testing to see if the slope in the population is non-zero, what are the null and alternative hypotheses? What is the p-value? What conclusion should we make?

$H_0$: The slope is zero vs. $H_a$: The slope is non-zero. We get a p-value of 2.2e-16, which is much smaller than the sig value of 0.05. As such, we can reject the null hypothesis and have evidence that the slope is non-zero

c.iii) Give a 95% confidence interval for the population slope, and explain what it means in the context of this problem.

We get a 95% confidence interval of (29.037, 32.0049). We are 95% confident that the slope of the linear regression for the whole population falls between the interval of (29.037, 32.0049).

d.i) a 95% confidence interval for the mean number of calories for all beers which have 5% abv.

```
##        fit      lwr     upr
## 1 151.8445 151.6961 151.993
```

We get a 95% confidence interval of (151.6961, 151.993). We are 95% confident that the mean number of calories for all beers with 5% abv falls between the interval of (151.6961, 151.993).

d.ii) a 95% prediction interval for the number of calories for all beers which have 5% abv.

```
##        fit      lwr      upr
## 1 151.8445 149.4618 154.2273
```

We get a 95% prediction interval of (149.4618 , 154.2273). 95% of all beers with 5% abv have calorie counts between the interval of (149.4618 , 154.2273).

d.iii) a 95% confidence interval for the mean number of calories for all beers which have 9% abv.

```
##        fit     lwr      upr
## 1 273.9279 273.569 274.2869
```

We get a 95% confidence interval of (273.569, 274.2869). We are 95% confident that the mean number of calories for all beers with 9% abv falls between the interval of (273.569, 274.2869).

d.iv) a 95% prediction interval for the number of calories for all beers which have 9% abv.

```
##        fit      lwr     upr
## 1 273.9279 271.5229 276.333
```

We get a 95% prediction interval of (271.5229, 276.333). 95% of all beers with 9% abv have mean calorie counts between the interval of (271.5229, 276.333).

```
##
## Call:
## lm(formula = calories ~ abv, data = beer, subset = -which(beer$abv >
##     15))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -67.056 -11.117   1.599  10.788  81.637
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -12.7587     4.4335  -2.878  0.00433 **
## abv          32.8518     0.8074  40.689  < 2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.71 on 266 degrees of freedom
## Multiple R-squared:  0.8616, Adjusted R-squared:  0.8611
## F-statistic:  1656 on 1 and 266 DF,  p-value: < 2.2e-16
```

## Problem 2.

The linear equation changed from calories = -0.7597 + 30.5209(abv) to calories = -12.7587 + 32.8518(abv). However the p-value remains approximately the same. Since we got rid of the outlier, we have less variation, meaning that the confidence and prediciton intervals likely narrowed.