Stat 344 – HW 9

Trey Tipton

March 29, 2022

Problem 6.37

d.)

```
act.model <- lm(GPA ~ ACT, data = ACTgpa)
new_data <- data.frame(ACT = 25)
conf_int <- predict(act.model, newdata = new_data,
        interval = 'confidence',
        level = 0.95)
conf_int
```

```
##        fit      lwr      upr
## 1 3.288243 3.166694 3.409792
```

A 95% confidence interval for the average GPA for any students who scored 25 on the ACT is (3.17, 3.41).

e.)

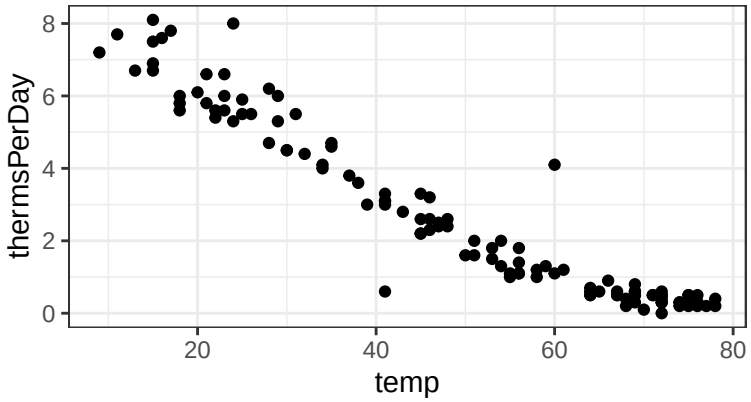
```
new_data <- data.frame(ACT = 30)
conf_int <- predict(act.model, newdata = new_data,
        interval = 'prediction',
        level = 0.95)
conf_int
```

```
##        fit      lwr      upr
## 1 3.723365 3.100775 4.345955
```

A 95% prediction interval interval for the GPA for a student who scored 30 on the ACT is (3.10, 4.35).

Problem 6.43

```
utilities <- Utilities
gf_point(thermsPerDay ~ temp, data = utilities)
```



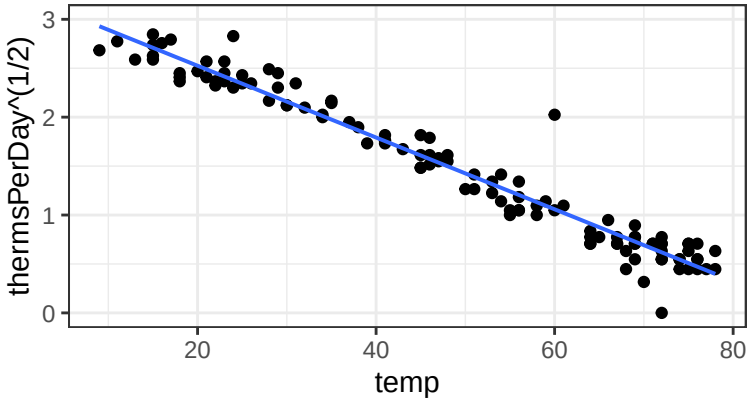
a.) We should remove the fourth observation from the data before doing the analysis for two reasons: It appears to be an outlier of our otherwise linear dataset in the scatterplot, and the notes for that day say "bad meter reading." There is only one other observation that appears to be an outlier, however, I do not see any reason to remove it because there is no notes that say it was a bad reading.

```
utilities_filt <- utilities %>%
  filter(! (temp == 41 & thermsPerDay < 1))
```

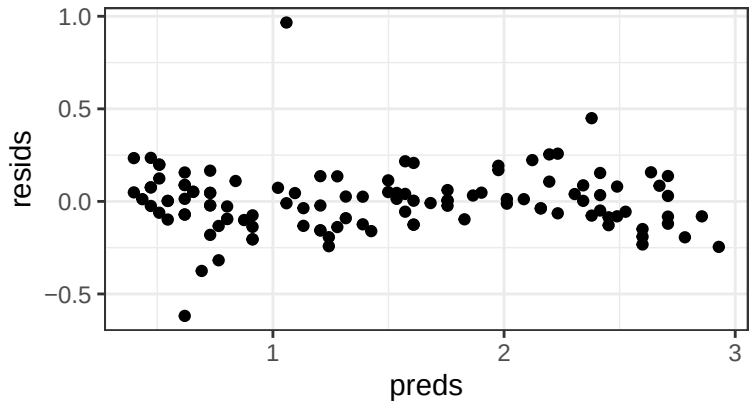
b.) The model is mostly linear, but has a slight curve that would require a x down or y down transformation. After looking at the scatterplot, moving y down (sqrt(thermsPerDay)) seemed to be the best transformation to make it look linear. It also fixed any problems with conditions, especially the ACF plot and the predictions vs. residuals plot.

c.)

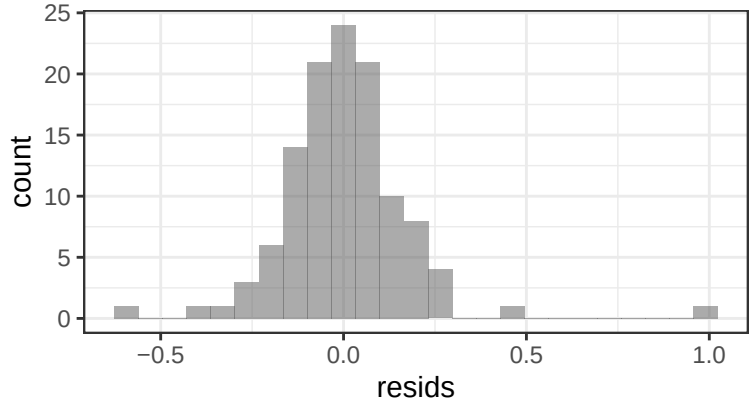
```
util.model <- lm(thermsPerDay^(1/2) ~ temp, data = utilities_filt)

util_model <- utilities_filt %>%
  mutate(preds = predict(util.model), resids = resid(util.model))

gf_point(thermsPerDay^(1/2) ~ temp, data = utilities_filt)%>%
  gf_lm()
```



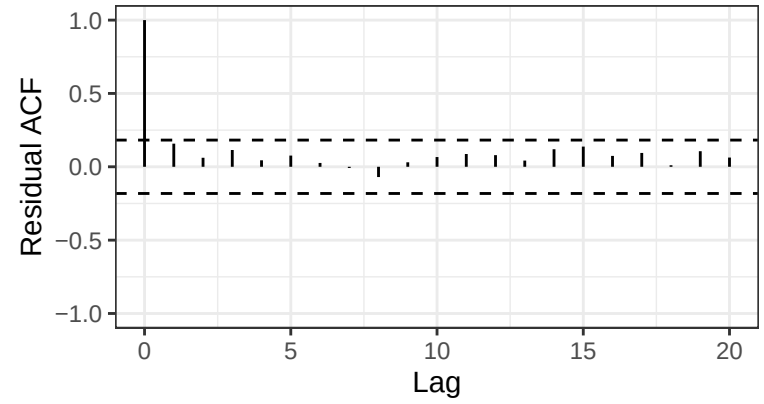

```
gf_point(resids ~ preds, data = util_model)
```




```
gf_histogram(~resids, data = util_model)
```




```
s245::gf_acf(~util.model) %>%
  gf_lims(y = c(-1, 1))
```



This model seems to pass all of the conditions. For independence conditions, the ACF is all within the limits of 1 and -1 (which it was not before the transformation). For normality of ressiduals, the histogram shows that the residuals are normally distributed. For residual variance, the predictions vs. residuals plot does not have any trends, so the model passes all of the conditions. There does not appear to be any reason for concern, and I am fairly happy with this model and the scatterplot showed that the data is a better fit for a linear model after the transformation we made.