BACS HW12

Bijon Setyawan Raya

5/12/2021

```
cars <- Auto
head(cars)
```

```
##   mpg cylinders displacement horsepower weight acceleration year origin
## 1  18         8          307        130   3504         12.0   70      1
## 2  15         8          350        165   3693         11.5   70      1
## 3  18         8          318        150   3436         11.0   70      1
## 4  16         8          304        150   3433         12.0   70      1
## 5  17         8          302        140   3449         10.5   70      1
## 6  15         8          429        198   4341         10.0   70      1
##                        name
## 1 chevrolet chevelle malibu
## 2         buick skylark 320
## 3        plymouth satellite
## 4             amc rebel sst
## 5               ford torino
## 6          ford galaxie 500
```

Question 1

```
cars_log <- with(cars, data.frame(log(mpg), log(cylinders), log(displacement),
log(horsepower), log(weight), log(acceleration), year, origin))

names(cars_log) <- names(cars)[1:8] # rename the columns
head(cars_log)
```

```
##        mpg cylinders displacement horsepower   weight acceleration year origin
## 1 2.890372  2.079442     5.726848   4.867534 8.161660     2.484907   70      1
## 2 2.708050  2.079442     5.857933   5.105945 8.214194     2.442347   70      1
## 3 2.890372  2.079442     5.762051   5.010635 8.142063     2.397895   70      1
## 4 2.772589  2.079442     5.717028   5.010635 8.141190     2.484907   70      1
## 5 2.833213  2.079442     5.710427   4.941642 8.145840     2.351375   70      1
## 6 2.708050  2.079442     6.061457   5.288267 8.375860     2.302585   70      1
```

a. Run a new regression on the cars_log dataset, with mpg.log. dependent on all other variables

```
cars_regr <-
  lm(
    mpg ~
      cylinders +
      displacement +
      horsepower +
      weight +
      acceleration +
      year +
      factor(origin),
    data = cars_log
  )
summary(cars_regr)
```

```
##
## Call:
## lm(formula = mpg ~ cylinders + displacement + horsepower + weight +
##     acceleration + year + factor(origin), data = cars_log)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.39727 -0.06880  0.00450  0.06356  0.38542
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      7.301938   0.361777  20.184  < 2e-16 ***
## cylinders       -0.081915   0.061116  -1.340  0.18094
## displacement     0.020387   0.058369   0.349  0.72707
## horsepower      -0.284751   0.057945  -4.914 1.32e-06 ***
## weight          -0.592955   0.085165  -6.962 1.46e-11 ***
## acceleration    -0.169673   0.059649  -2.845  0.00469 **
## year             0.030239   0.001771  17.078  < 2e-16 ***
## factor(origin)2  0.050717   0.020920   2.424  0.01580 *
## factor(origin)3  0.047215   0.020622   2.290  0.02259 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.113 on 383 degrees of freedom
## Multiple R-squared:  0.8919, Adjusted R-squared:  0.8897
## F-statistic:   395 on 8 and 383 DF,  p-value: < 2.2e-16
```

i. Which log-transformed factors have a significant effect on log.mpg. at 10% significance?

`horsepower`, `weight`, `acceleration`, `year`, `factor(origin)2`, and `factor(origin)3`.

ii. Do some new factors now have effects on mpg, and why might this be?

`acceleration` and `horsepower` suddenly became significant in this case, which they weren't in the previous homework.

iii. Which factors still have insignificant or opposite (from correlation) effects on mpg? Why might this be?

Only `cylinders`. The more cylinders cars have, the higher the gas consumption.

b. Let's take a closer look at weight, because it seems to be a major explanation of mpg

i. Create a regression (call it regr_wt) of mpg on weight from the original cars dataset

```
regr_wt <- lm(mpg ~ weight, data = Auto)
```

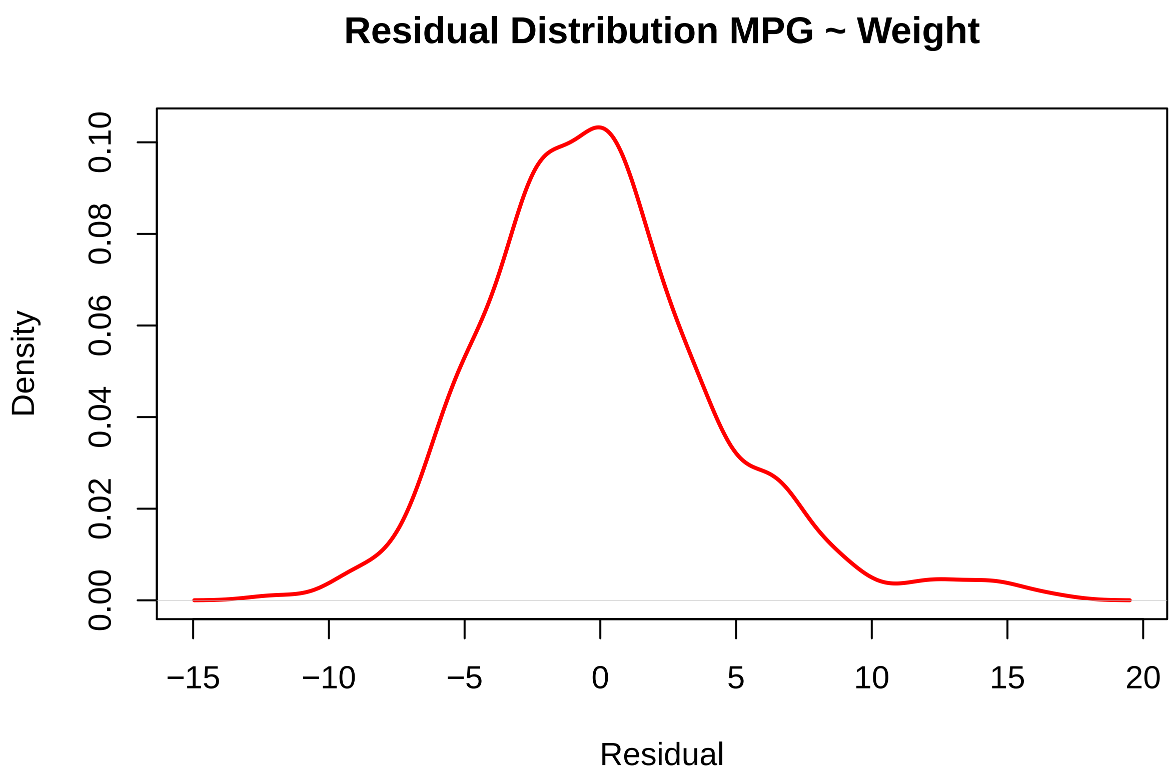
ii. Create a regression (call it regr_wt_log) of log.mpg. on log.weight. from cars_log

```
regr_wt_log <- lm(mpg ~ weight, data = cars_log)
```

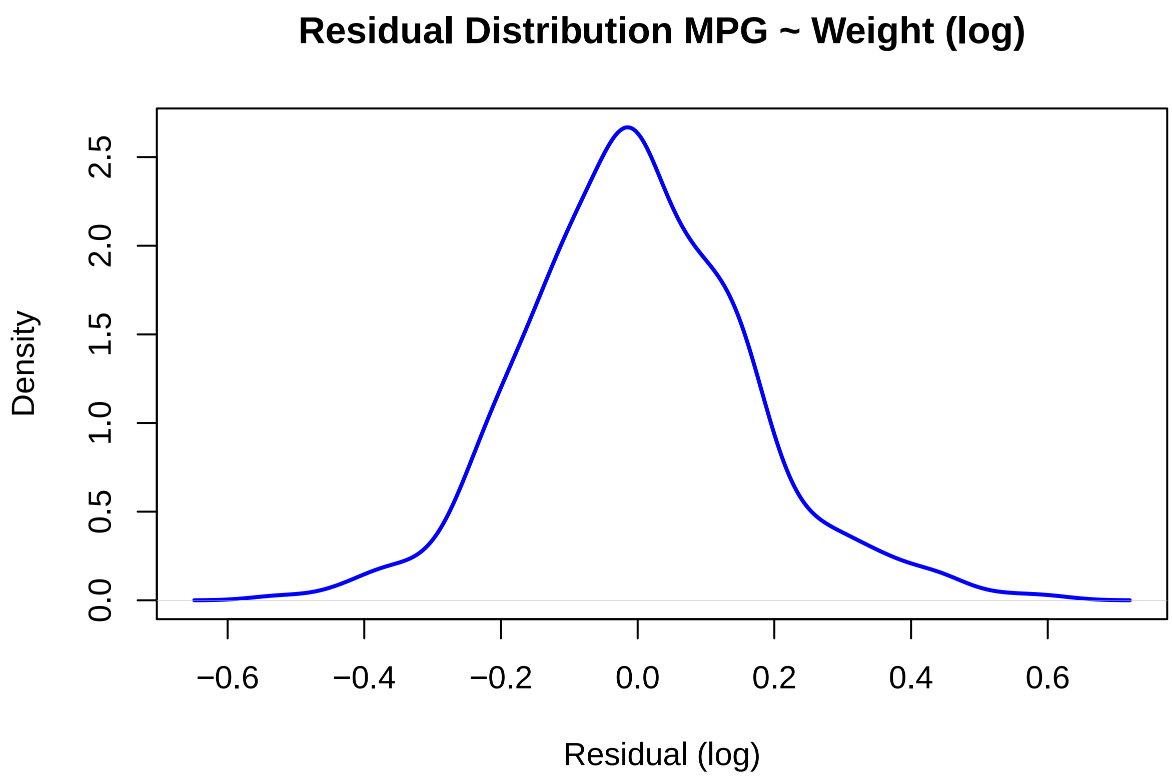
iii. visualize the residuals of both regression models

1. density plots of residuals

```
plot(
  density(resid(regr_wt)),
  main = "Residual Distribution MPG ~ Weight",
  lwd = 2,
  col = "red",
  xlab = "Residual"
)
```

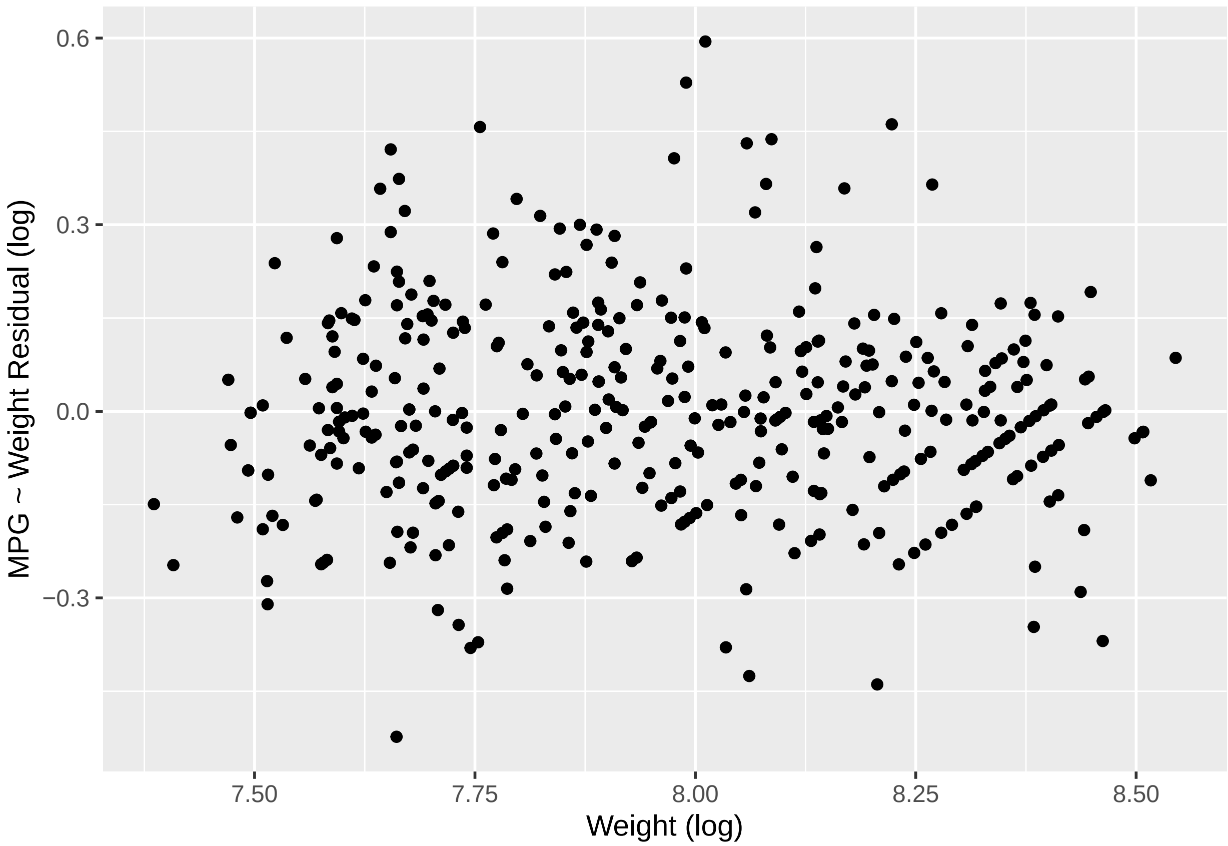



```
plot(
  density(resid(regr_wt_log)),
  main = "Residual Distribution MPG ~ Weight (log)",
  lwd = 2,
  col = "blue",
  xlab = "Residual (log)"
)
```



2. scatterplot of log.weight. vs. residuals

```
library(ggplot2)
ggplot(
  cars_log,
  aes(
    x = weight,
    y = lm(mpg ~ weight)$residual)
  ) +
  geom_point() +
  xlab("Weight (log)") +
  ylab("MPG ~ Weight Residual (log)")
```



c. Let's examine the 95% confidence interval of the slope of log.weight. vs. log.mpg.