Regression_Cars

Damian

June 11, 2016

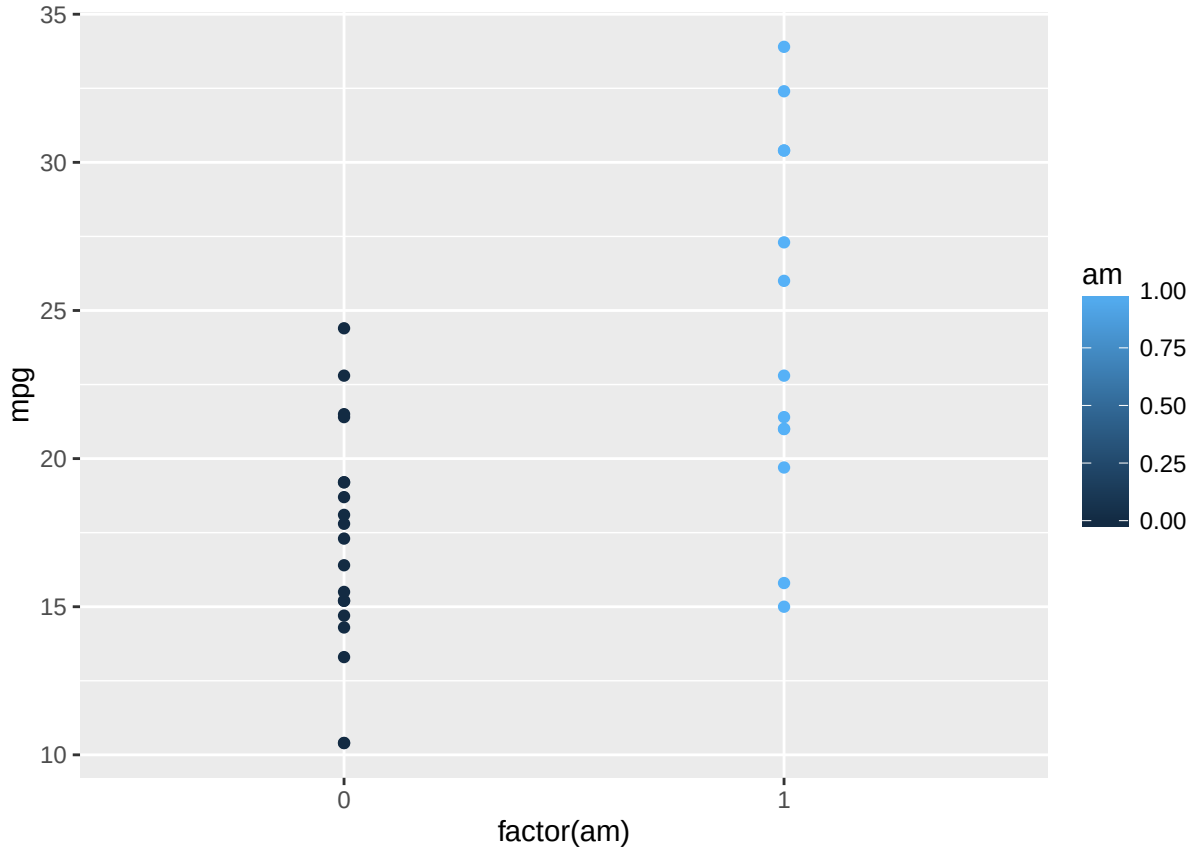
```
library(ggplot2)
library(datasets)
data(mtcars)
```

does transmission type have an effect on mpg?

We will answer this question!

exploratory analysis

```
ggplot(data = mtcars, aes(x = factor(am), y = mpg, color = am)) + geom_point()
```



Seems like cars with manual transmission have higher mpg

```
auto_mean <- mean(mtcars$mpg[mtcars$am == 0])
man_mean <- mean(mtcars$mpg[mtcars$am == 1])
```

Mean mpg by transmission type Automatic : 17.1473684 Manual : 24.3923077

Linear Models

I fit three models * One predicting mpg only using transmission type * One predicting mpg only using weight * One predicting mpg using weight and transmission

transmission type alone

```
model1 <- lm(data = mtcars, formula = mpg~am)
summary(model1)
```

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## am             7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

transmission and weight

```
model2 <- lm(data = mtcars, formula = mpg ~ am + wt)
summary(model2)
```

```
##
## Call:
## lm(formula = mpg ~ am + wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5295 -2.3619 -0.1317  1.4025  6.8782
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 37.32155    3.05464  12.218 5.84e-13 ***
## am          -0.02362    1.54565  -0.015    0.988
## wt          -5.35281    0.78824  -6.791 1.87e-07 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.098 on 29 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7358
## F-statistic: 44.17 on 2 and 29 DF,  p-value: 1.579e-09
```

weight alone

```
model3 <- lm(data = mtcars, formula = mpg ~ wt)
summary(model3)
```

```
##
## Call:
## lm(formula = mpg ~ wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5432 -2.3647 -0.1252  1.4096  6.8727
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.2851     1.8776  19.858  < 2e-16 ***
## wt           -5.3445     0.5591  -9.559 1.29e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.046 on 30 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7446
## F-statistic: 91.38 on 1 and 30 DF,  p-value: 1.294e-10
```

What does it mean?

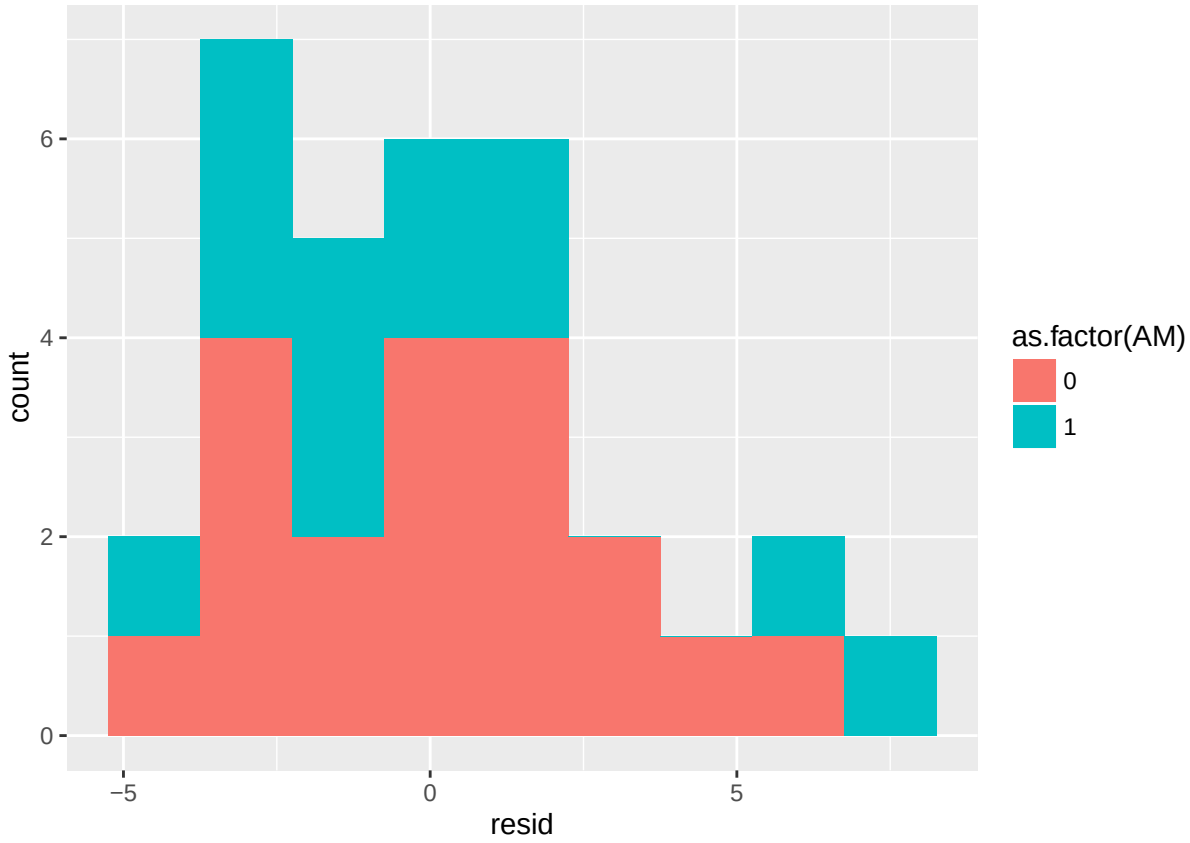
```
p_am <- summary(model1)$coefficients[2,4]
p_am_wt <- summary(model2)$coefficients[2,4]
```

Transmission type is a statistically significant predictor when used alone. The p-value : 2.8502074×10^{-4} can be interpreted as saying that there's a 2.8% chance of observing a correlation of this magnitude if there was no relationship between mpg and transmission type.

However, when weight is included in the model, transmission type is far from predictive, with p-value equal to 0.9879146.

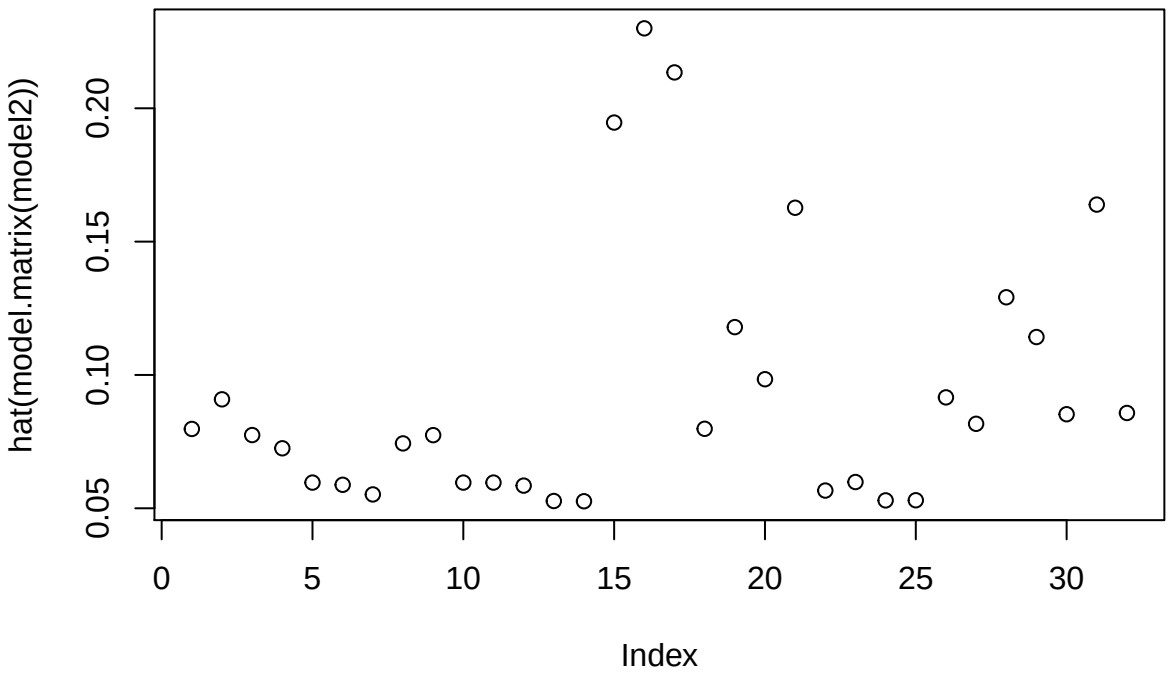
residual plot

```
residual_df <- data.frame(row.names=names(model3$residuals), resid = model3$residuals)
residual_df$AM <- mtcars$am
ggplot(data = residual_df, aes(x=resid, fill = as.factor(AM))) + geom_histogram(binwidth=1.5)
```



leverage

```
plot(hat(model.matrix(model2)))
```


hat(model.matrix(model2))
0.05
0.10
0.15
0.20
0
5
10
15
20
25
30
Index