

# Midterm 1

Yanlin Li

## Read in the data

```
library(tidyverse)
library(tidymodels)
library(car)

beijing <- read_csv("beijing.csv")
```

## Exercise 1

```
m1 <- lm(SO2 ~ month + TEMP + PRES + DEWP + RAIN + wd + WSPM, data = beijing)
summary(m1)
```

Call:

```
lm(formula = SO2 ~ month + TEMP + PRES + DEWP + RAIN + wd + WSPM,
    data = beijing)
```

Residuals:

Min	1Q	Median	3Q	Max
-45.67	-11.62	-4.27	4.68	365.05

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	586.28527	21.88485	26.790	< 2e-16 ***
month	-0.95756	0.03719	-25.749	< 2e-16 ***
TEMP	-0.90373	0.02412	-37.473	< 2e-16 ***
PRES	-0.53682	0.02154	-24.923	< 2e-16 ***
DEWP	-0.27520	0.01965	-14.005	< 2e-16 ***

RAIN	-0.53035	0.14389	-3.686	0.000228	***
wdENE	-3.41085	0.55682	-6.126	9.14e-10	***
wdESE	-1.01756	0.64698	-1.573	0.115779	
wdN	-10.11872	0.74412	-13.598	< 2e-16	***
wdNE	-5.38173	0.53928	-9.980	< 2e-16	***
wdNNE	-6.63680	0.67431	-9.842	< 2e-16	***
wdNNW	-11.46124	0.75044	-15.273	< 2e-16	***
wdNW	-9.18869	0.60411	-15.210	< 2e-16	***
wdS	3.36289	0.68493	4.910	9.16e-07	***
wdSE	-0.59650	0.72146	-0.827	0.408361	
wdSSE	1.82725	0.78152	2.338	0.019389	*
wdSSW	3.96496	0.59601	6.652	2.93e-11	***
wdSW	2.09717	0.55018	3.812	0.000138	***
wdW	-0.46689	0.59409	-0.786	0.431939	
wdWNW	-5.53336	0.60107	-9.206	< 2e-16	***
wdWSW	2.05522	0.58754	3.498	0.000469	***
WSPM	-2.48948	0.11972	-20.794	< 2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21.15 on 34289 degrees of freedom  
(753 observations deleted due to missingness)

Multiple R-squared: 0.2209, Adjusted R-squared: 0.2204

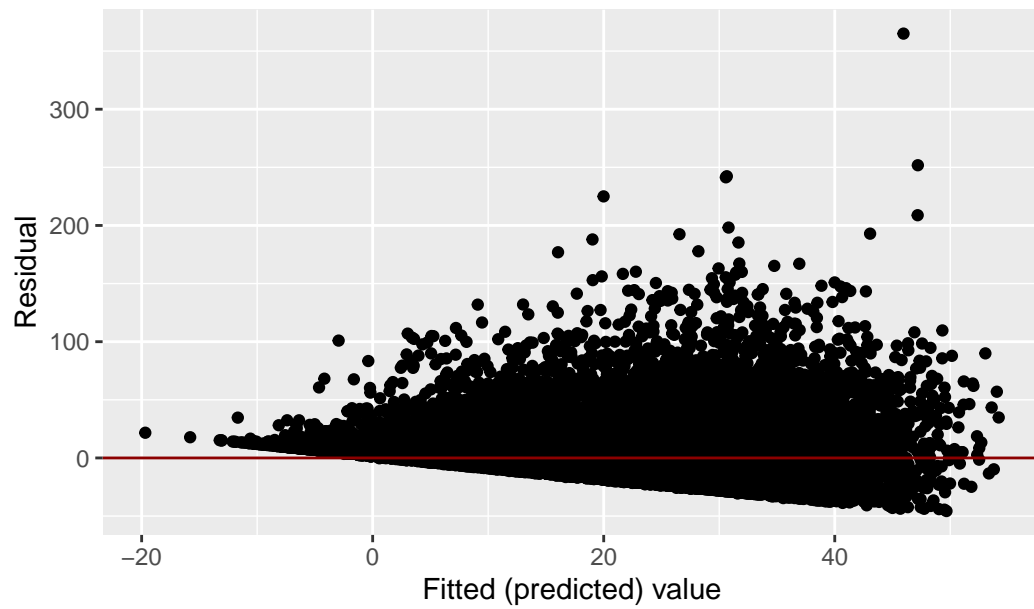
F-statistic: 462.9 on 21 and 34289 DF, p-value: < 2.2e-16

## Exercise 5

```
m1_augmented <- augment(m1)

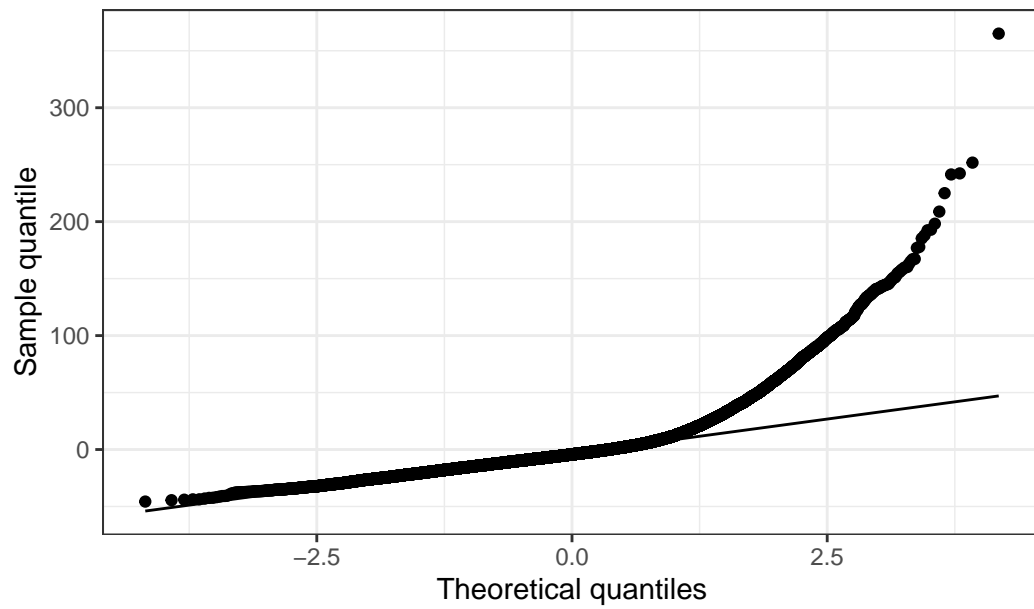
ggplot(m1_augmented, aes(x= .fitted, y = .resid)) +
  geom_point()+
  geom_hline(yintercept = 0, color = "darkred") +
  labs(x = "Fitted (predicted) value", y = "Residual") +
  ggtitle("Bad linearity and constant variance")
```

### Bad linearity and constant variance



```
ggplot(m1_augmented, aes (sample = .resid)) +  
  stat_qq() +  
  stat_qq_line() +  
  theme_bw() +  
  labs(x= "Theoretical quantiles",  
       y = "Sample quantile") +  
  ggtitle("Not Normal Distribution in the Q-Q Plot")
```

Not Normal Distribution in the Q–Q Plot



### Exercise 6

```
586.28527 - 0.95756 * 2 - 0.90373 * 1.6 - 0.53682 * 1013.7 - 0.27520 * (-3.7) - 0.53035 *
```

```
[1] 37.27851
```

```
15 - 37.27851
```

```
[1] -22.27851
```

### Exercise 7

### Exercise 8