

Airbnb Project

Keqin Cao

5/4/2020

```
#Check model assumptions: constant variance
```

```
airbnb <- read.csv("airbnb.csv")
airbnb$last_scraped<- NULL
airbnb$amenities<- NULL
airbnb$id<- NULL
```

```
head(airbnb)
```

```
##   host_total_listings_count host_identity_verified neighbourhood_cleansed
## 1                           36                               t Downtown/Civic Center
## 2                           62                               t Downtown/Civic Center
## 3                           65                               f Downtown/Civic Center
## 4                           4                               t Downtown/Civic Center
## 5                           1                               f Downtown/Civic Center
## 6                           1                               f Downtown/Civic Center
##   latitude longitude  property_type      room_type accommodates bathrooms
## 1 37.78806 -122.4137 Condominium Entire home/apt          4         1
## 2 37.78821 -122.4110 Condominium Entire home/apt          4         1
## 3 37.78740 -122.4078 Apartment    Private room          2         1
## 4 37.78909 -122.4099 Boutique hotel Private room          2         1
## 5 37.78129 -122.4160 Apartment    Entire home/apt        4         1
## 6 37.78770 -122.4131 Apartment    Entire home/apt        3         1
##   bedrooms beds bed_type minimum_nights cancellation_policy cleaning_fee price
## 1       1     2 Real Bed           3             strict        100     599
## 2       0     2 Real Bed           1             strict         50     209
## 3       1     1 Real Bed           1            flexible        25      89
## 4       1     2 Real Bed           5            flexible        50     125
## 5       1     3 Real Bed           1             strict        50     200
## 6       0     2 Real Bed           2            flexible        25     110
```

```
linearfit<- lm(price~host_total_listings_count+latitude +longitude+factor(accommodates)+factor(bathrooms))
anova(linearfit)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: price
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
## host_total_listings_count	1	42669329	42669329	2751.179	< 2.2e-16 ***
## latitude	1	192133613	192133613	12388.147	< 2.2e-16 ***
## longitude	1	8469054	8469054	546.057	< 2.2e-16 ***
## factor(accommodates)	17	2230831309	131225371	8460.983	< 2.2e-16 ***
## factor(bathrooms)	7	200010203	28572886	1842.286	< 2.2e-16 ***
## factor(bedrooms)	7	233663157	33380451	2152.262	< 2.2e-16 ***

```

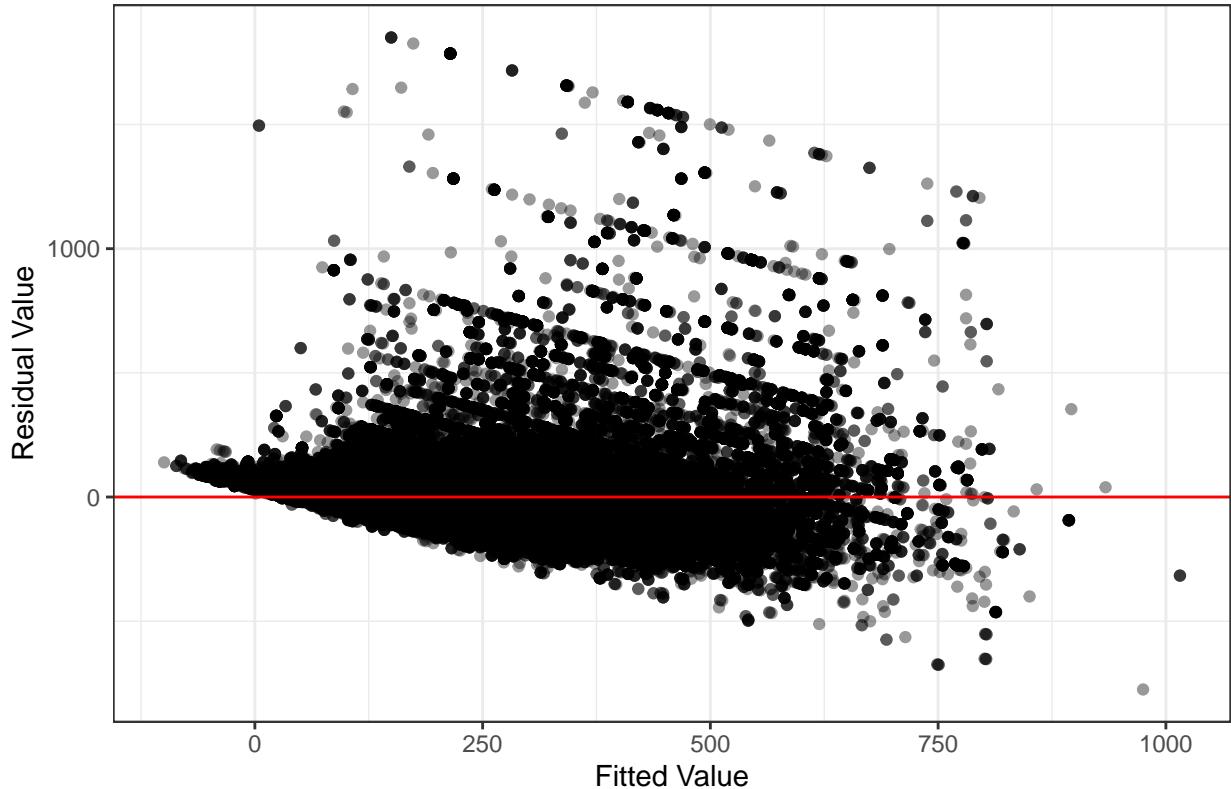
## factor(beds)          12   48308833   4025736   259.566 < 2.2e-16 ***
## factor(minimum_nights) 62   53863946   868773    56.016 < 2.2e-16 ***
## cleaning_fee           1   239288095 239288095 15428.514 < 2.2e-16 ***
## Residuals              215032 3335032637    15509
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

library(ggplot2)

fitted_residual<-ggplot(data=airbnb,aes(x=fitted(linearfit), y = resid(linearfit))) +geom_jitter( alpha=0.5)+theme_bw() +xlab("Fitted Value")+
ylab("Residual Value")+
ggtitle("Residual vs. Fitted Plot")+
geom_hline(aes(yintercept=0), color="red")
fitted_residual

```

Residual vs. Fitted Plot



Check the price itself if it is normally distributed

```
qqnorm(airbnb$price)
```

Normal Q-Q Plot

