STT863_HW4_Code

Yuhan

2023-10-09

```
#libraries
library(ggplot2)
library(lmtest)
```

```
## Loading required package: zoo
```

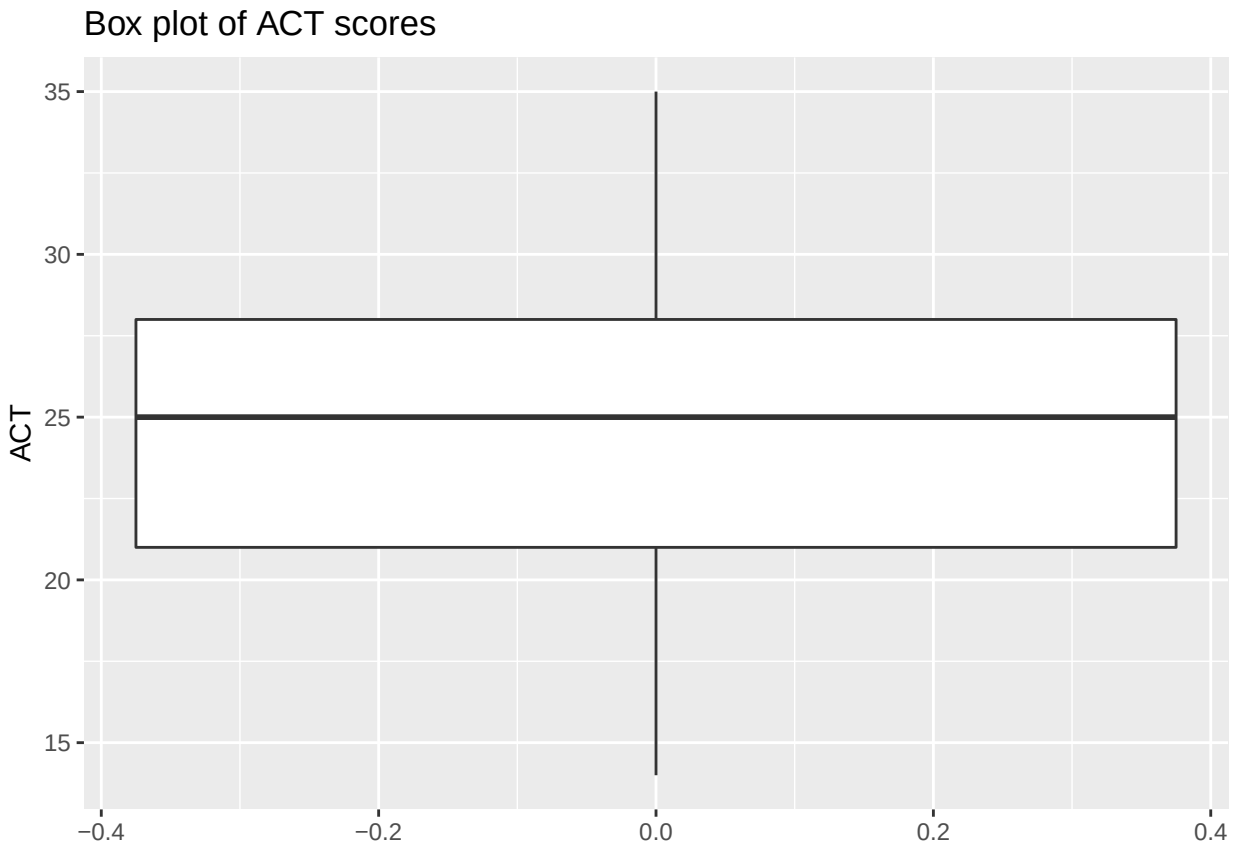
```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

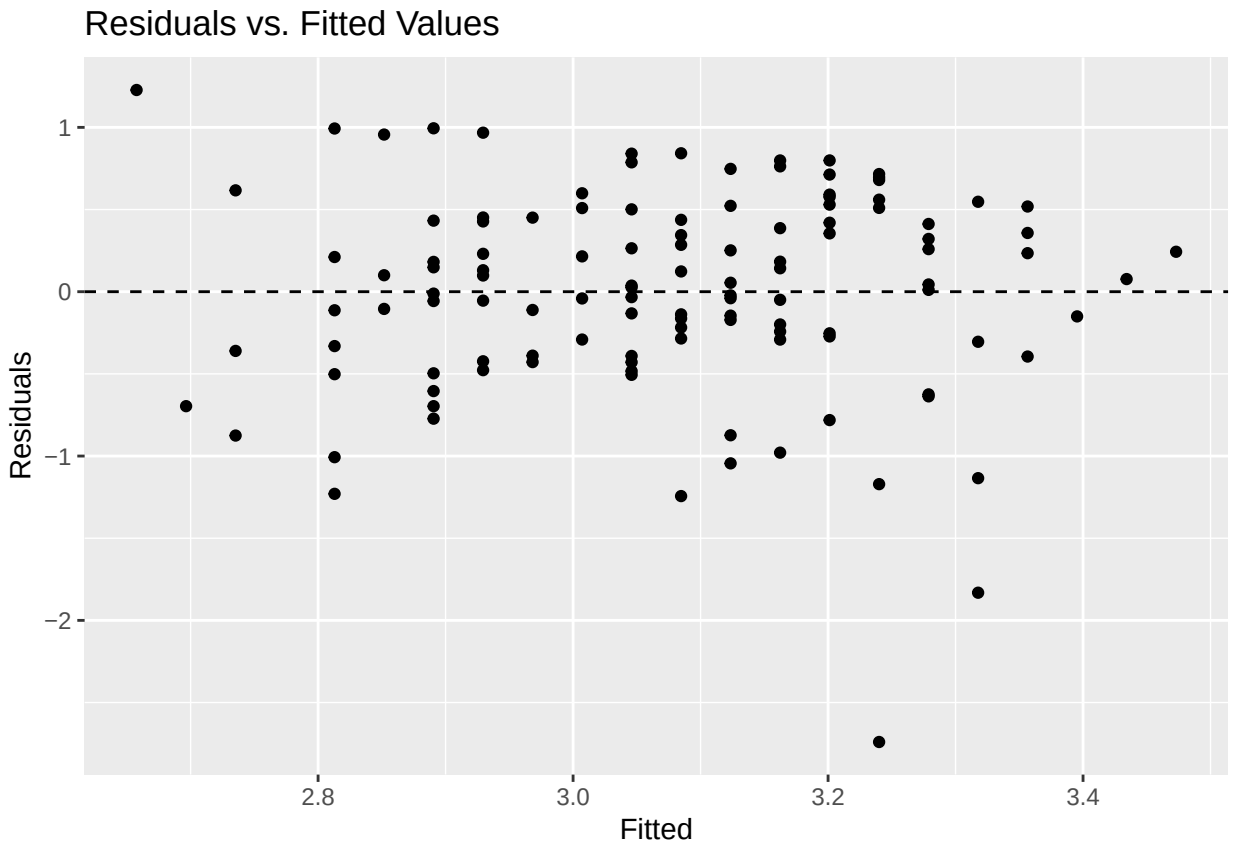
```
library(car)
```

```
## Loading required package: carData
```

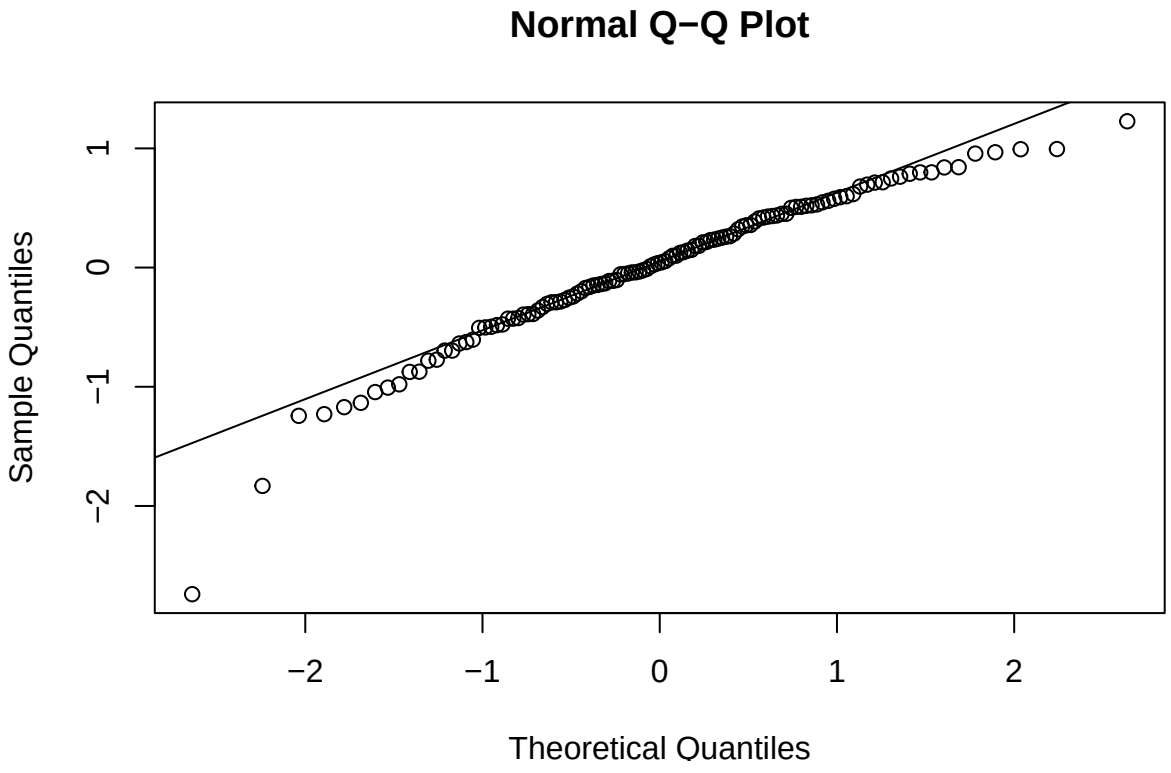
```
gpa_data <- read.table('CH01PR19.txt', header = FALSE, col.names = c('GPA', 'ACT'))
model <- lm(GPA ~ ACT, data = gpa_data)

# a Box plot
#gpa_data$GPA_Range <- cut(gpa_data$GPA, breaks = seq(0, 4, 1), include.lowest = TRUE, right = FALSE)
ggplot(gpa_data, aes(y = ACT)) +
  geom_boxplot() +
  ggtitle("Box plot of ACT scores")
```




```r
# c
ggplot(data.frame(Fitted = fitted(model), Residuals = resid(model)), aes(x = Fitted, y = Residuals)) +
  geom_point() +
  geom_hline(yintercept = 0, linetype = "dashed") +
  ggtitle("Residuals vs. Fitted Values")
```




```r
# d Normal probability plot
qqnorm(resid(model))
qqline(resid(model))
```



```r
# Calculating r
osm <- qqnorm(resid(model), plot.it = FALSE)$x
osr <- qqnorm(resid(model), plot.it = FALSE)$y
r <- cor(osm, osr)
print(r)
```

```
## [1] 0.9744497
```

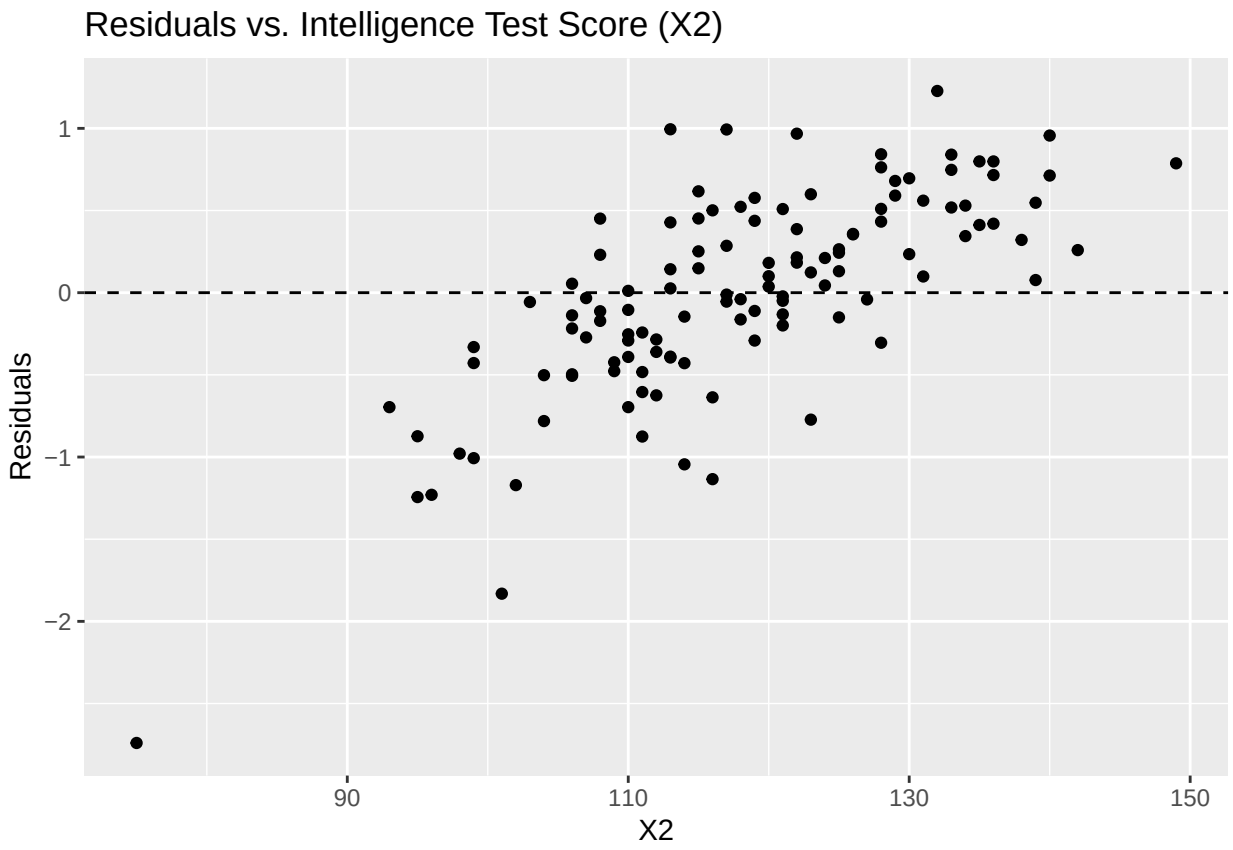
```r
# e brown-forsythe test
group <- ifelse(gpa_data$ACT < 26, "Group1", "Group2")
bf_test <- leveneTest(resid(model), group, center = median)
```

```
## Warning in leveneTest.default(resid(model), group, center = median): group
## coerced to factor.
```

```r
print(bf_test)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value Pr(>F)
## group   1  0.8042 0.3717
##       118
```

```r
# f esiduals plot
additional_data <- read.table('CH03PR03.txt', header = FALSE, col.names = c('GPA', 'ACT', 'X2', 'X3'))
ggplot(data.frame(X2 = additional_data$X2, Residuals = resid(model)), aes(x = X2, y = Residuals)) +
  geom_point() +
  geom_hline(yintercept = 0, linetype = "dashed") +
  ggtitle("Residuals vs. Intelligence Test Score (X2)")
```



```r
ggplot(data.frame(X3 = additional_data$X3, Residuals = resid(model)), aes(x = X3, y = Residuals)) +
  geom_point() +
  geom_hline(yintercept = 0, linetype = "dashed") +
  ggtitle("Residuals vs. High School Class Rank Percentile (X3)")
```