

Day15

Olivia Wu

2024-04-23

### **Problem 5.6**

B, error terms do not need to be all positive

### **Problem 5.7**

D

### **Problem 5.11**

a) It is impossible to assign a race of the mother to each baby. Thus, we are not certain about the independent condition, and cannot move forward with inference.

b) We would need the 10% condition to be satisfied, or we want each group's sample size to be at most 10% of the population size for each race.

### **Problem 5.12**

a) This is an observational study, so we cannot infer cause-and-effect.

b) We can say that there is an association between those who preferred heavy metal and those who drive fast.

### **Problem 5.14**

a) The number of groups does not matter in an ANOVA test

b) Car age is quantitative.

c) It is not a condition that the sample sizes must be the same across all groups.

d) We can generalize because there was random sampling.

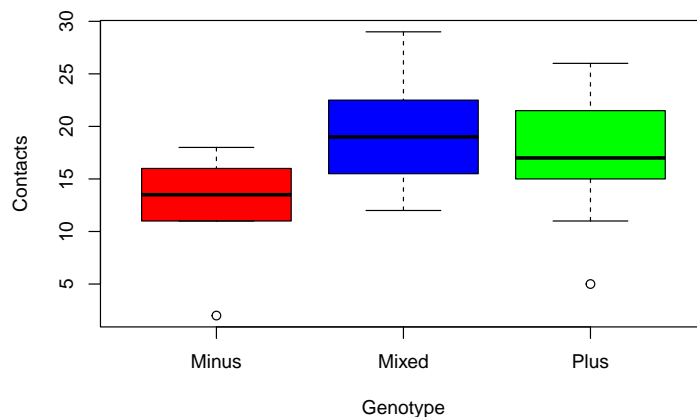
## Problem 5.22

- a) This is not possible, because the model should be adjusted so that all the residuals can be lowered by a constant, thereby achieving a smaller SSE.
- b) This could happen if the negative residual is pretty extreme.
- c) This could happen if my predicted result is closer to my actual score than your predicted result was to your actual score.
- d) This could happen because of the same reason in c).

## Problem 5.27

- a) The standard deviations are all similar. Minus has fewer observations and a smaller mean.

```
boxplot(Contacts~Genotype,data=Mouse, col=c("red","blue","green"))
```



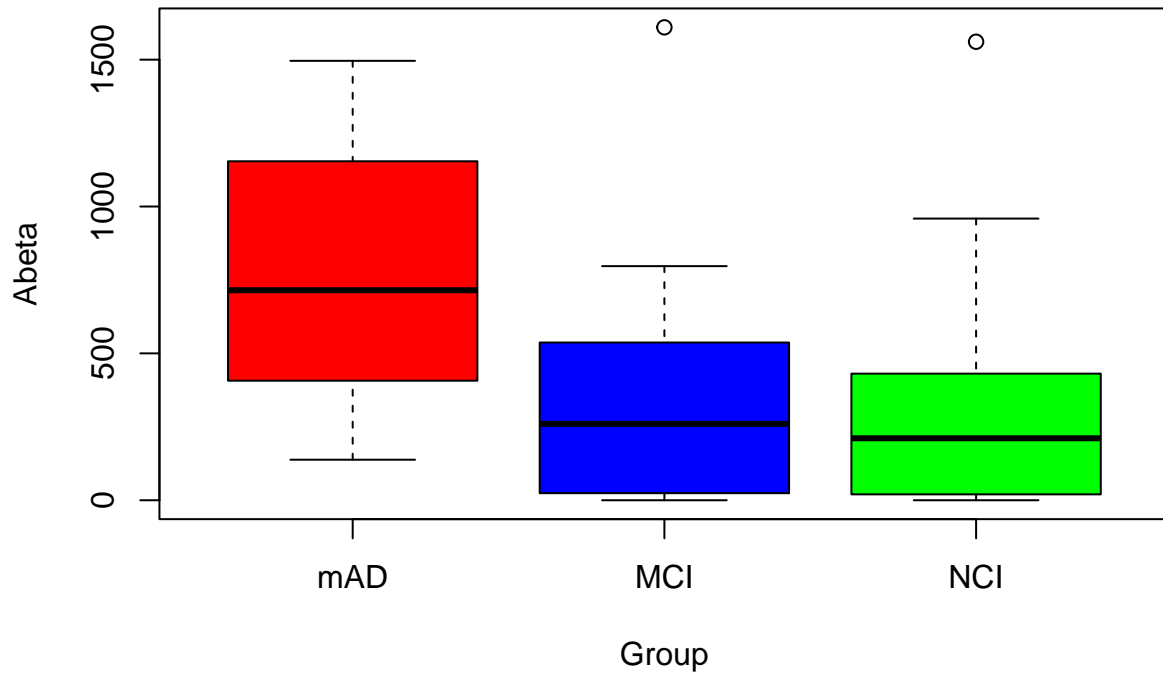
```
favstats(Contacts~Genotype,data=Mouse)
```

##	Genotype	min	Q1	median	Q3	max	mean	sd	n	missing
## 1	Minus	2	11.0	13.5	15.75	18	12.70000	4.522782	10	0
## 2	Mixed	12	15.5	19.0	22.50	29	19.26316	4.794002	19	0
## 3	Plus	5	15.0	17.0	21.50	26	17.52632	5.460651	19	0

- b) The standard deviation of residuals are similar across all groups. The residuals are roughly symmetric for each group. Independence is met because no mouse should affect another mouse.

## Problem 5.31

- a) The lengths of NCI, MCI, and mAD are 19, 21, and 17, respectively. Their means are 336.263, 341.05, and 761.29. Their standard deviations are 435.61, 406.41, and 426.69.
- b) The boxplot for NCI is skewed left, so the normal condition is not met.



### Problem 5.33

- a) We calculate the  $df(4)$  and  $SS(10998)$  of Occupation by subtracting the  $df$  and  $SS$  of Error from the Total. The  $F$ -statistic is equal to  $\frac{2749}{202} = 13.609$
- b) Since  $df = 4$ , we know there are 5 groups.
- c) Since  $p = 0 < 0.05$ , we reject the null hypothesis. There is enough evidence to suggest that the mean lifetimes are different across the occupation categories.

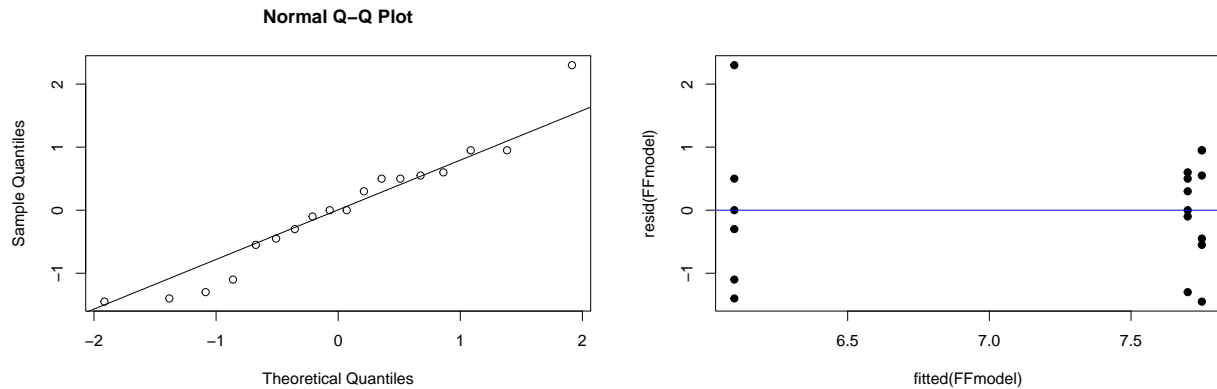
### Problem 5.38

- a)  $H_0 : \mu_1 = \mu_2 = \mu_3$ , where  $\mu_1, \mu_2, \mu_3$  are the mean stiffness responses for the vertical suture, meniscus arrow, and FasT-Fix methods.

\$H\_a\$: \$at least one of them is not equal

- b) Random assignment was used, so we have met the independence condition. The residuals are roughly normal, and the variance is roughly constant. The standard deviations are close enough (0.693, 1.327).

```
##          1          2          3
## 0.9710819 1.3266499 0.6928203
```



c)  $p = 0.02 < 0.05$ , so we reject the null hypothesis. There is enough evidence that there is a difference in mean stiffness across methods.

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Method      2  10.57   5.285    4.981 0.0219 *
## Residuals   15  15.91   1.061
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

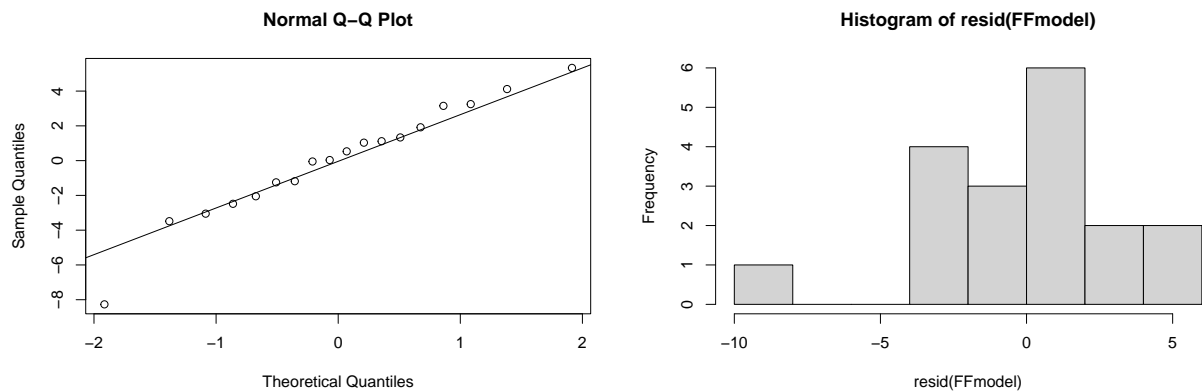
## Problem 5.40

a)  $H_0 : \mu_1 = \mu_2 = \mu_3$ , where  $\mu_1, \mu_2, \mu_3$  are the mean displacement responses for the vertical suture, meniscus arrow, and FasT-Fix methods.

\$H\_a\$: \$at least one of them is not equal

b) The residuals are roughly normal, and the variance is roughly constant. The standard deviations are close enough (2.66 vs 4.47).

```
##           1           2           3
## 2.666646 2.883343 4.470198
```



c)  $p = 0.014 < 0.05$ , so we reject the null hypothesis. There is enough evidence that there is a difference in mean displacement response across methods.

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Method      2  136.5   68.26   5.784 0.0137 *
## Residuals   15  177.0   11.80
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```