Chapter 8: Analysis of Variance

Alex Chi

Exercises

8.1 (Comparing average gas mileage).

A table of gas mileages on four new models of Japanese luxury cars is given in Larsen & Marx [30, Question 12.1.1], which is shown in Table 1.4, page 27. See Example 1.11 for a method of entering this data. Test if the four models (A, B, C, D) give the same gas mileage, on average, at $\alpha = 0.05$ significance.

```
y1 = c(22, 26)
y2 = c(28, 24, 29)
y3 = c(29, 32, 28)
y4 = c(23, 24)
y = c(y1, y2, y3, y4)
Model = c(rep("A", 2), rep("B", 3), rep("C", 3), rep("D", 2))
mileages = data.frame(y, Model)
```

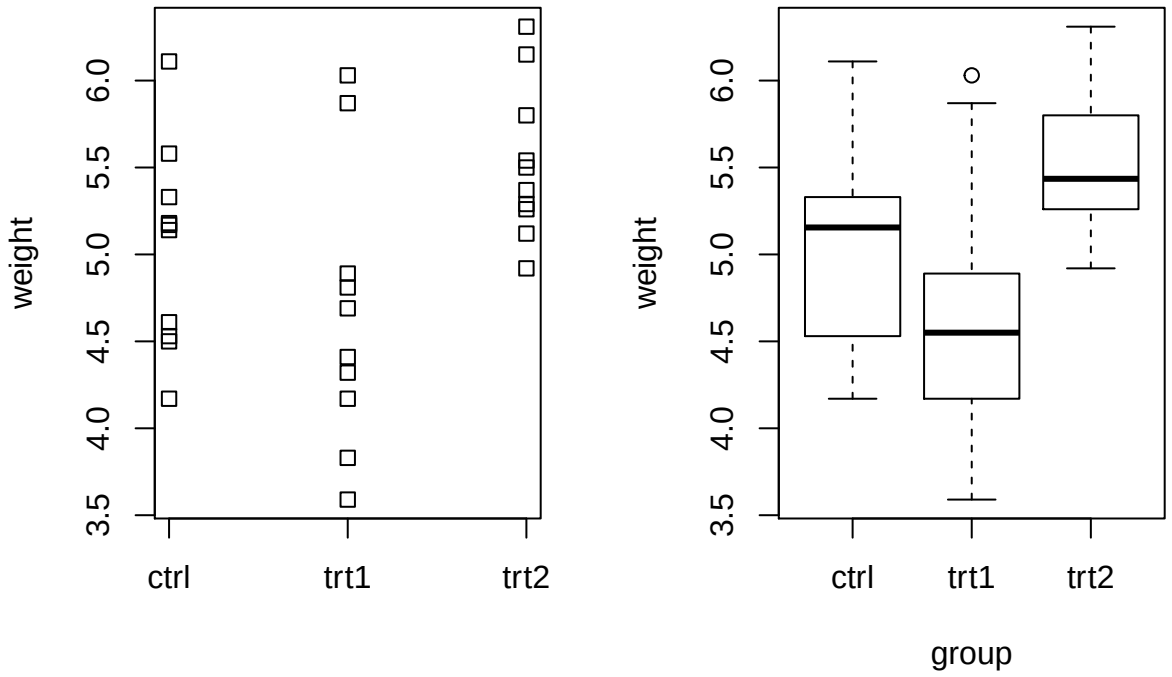
```
oneway.test(y ~ Model, mileages, var.equal=TRUE)
```

```
##
##  One-way analysis of means
##
## data:  y and Model
## F = 3.9358, num df = 3, denom df = 6, p-value = 0.07227
```

8.2 (Yields of plants).

The PlantGrowth data is an R dataset that contains results from an experiment on plant growth. The yield of a plant is measured by the dried weight of the plant. The experiment recorded yields of plants for a control group and two different treatments. After a preliminary exploratory data analysis, use one-way ANOVA to analyze the differences in mean yield for the three groups. Start with the exploratory data analysis, and also check model assumptions. What conclusions, if any, can be inferred from this sample data?

```
par(mfrow=c(1, 2))
stripchart(weight ~ group, PlantGrowth, vertical=TRUE)
boxplot(weight ~ group, PlantGrowth, vertical=TRUE)
```




```
M = aov(weight ~ group, PlantGrowth)
oneway.test(weight ~ group, PlantGrowth, var.equal=TRUE)
```

```
##
##  One-way analysis of means
##
## data:  weight and group
## F = 4.8461, num df = 2, denom df = 27, p-value = 0.01591
```

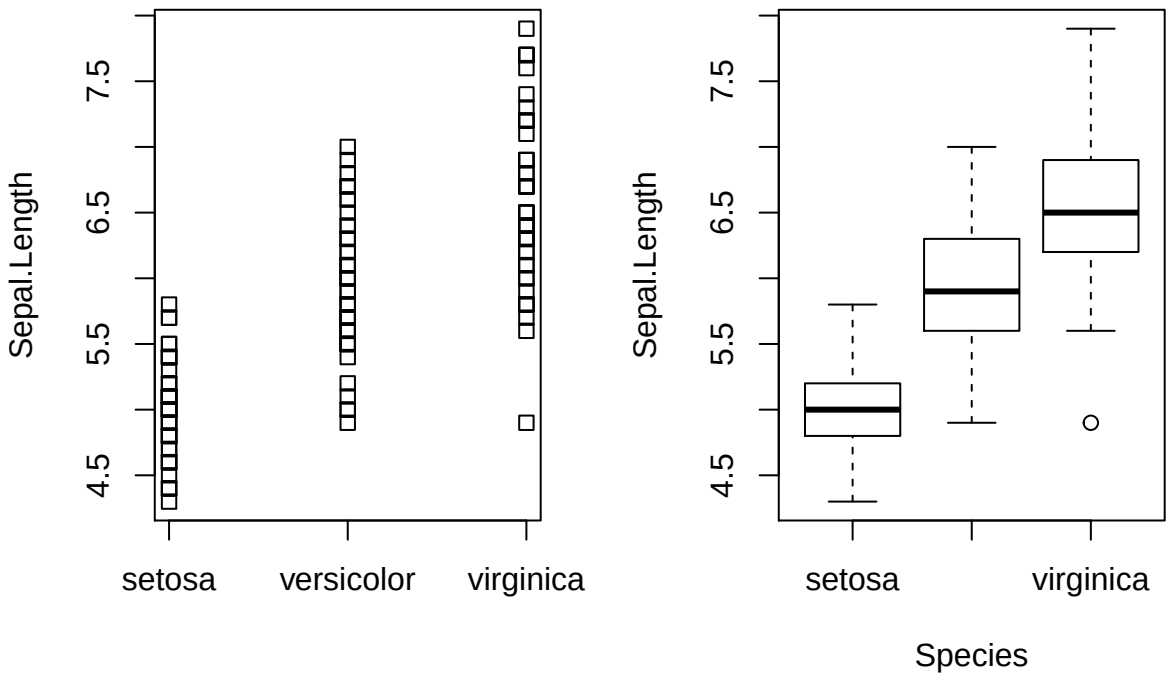
```
model.tables(M, type="means")
```

```
## Tables of means
## Grand mean
##
## 5.073
##
##  group
## group
##  ctrl  trt1  trt2
## 5.032 4.661 5.526
```

8.3 (Differences in iris by species).

The iris data has 50 observations of four measurements for each of three species of iris: setosa, versicolor, and virginica. We are interested in possible differences in the sepal length of iris among the three species. Perform a preliminary analysis as in Example 8.3. Write the effects model for a one-way ANOVA. What are the unknown parameters? Next fit a one-way ANOVA model for Sepal.Length by Species using lm. Display the ANOVA table. What are the parameter estimates?

```
par(mfrow=c(1, 2))
attach(iris)
stripchart(Sepal.Length ~ Species, vertical=TRUE)
boxplot(Sepal.Length ~ Species, vertical=TRUE)
```




```
meansd = function(x) c(mean=mean(x), sd=sd(x))
by(Sepal.Length, Species, FUN=meansd)
```

```
## Species: setosa
##      mean        sd
## 5.0060000 0.3524897
## ------------------------
## Species: versicolor
##      mean        sd
## 5.9360000 0.5161711
## ------------------------
## Species: virginica
##      mean        sd
## 6.5880000 0.6358796
```

```
L = lm(Sepal.Length ~ Species)
options(show.signif.stars=FALSE)
anova(L)
```

```
## Analysis of Variance Table
##
## Response: Sepal.Length
##              Df Sum Sq Mean Sq F value    Pr(>F)
```

```
## Species     2 63.212  31.606  119.26 < 2.2e-16
## Residuals 147 38.956   0.265
```

8.4 (Checking model assumptions).

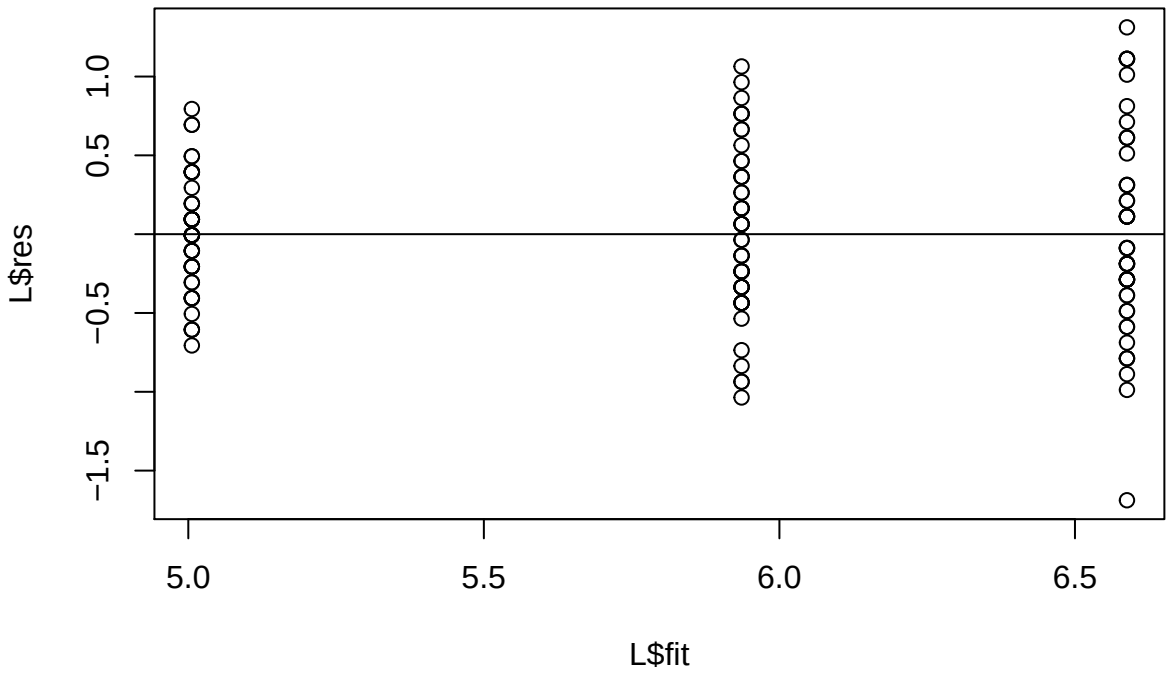
Refer to your results from Exercise 8.3. What are the assumptions required for inference? Analyze the residuals of the model to assess whether there is a serious departure from any of these assumptions. How can you check for normality of the error variable?

Answer: Equal variance assumption, samples are independent.

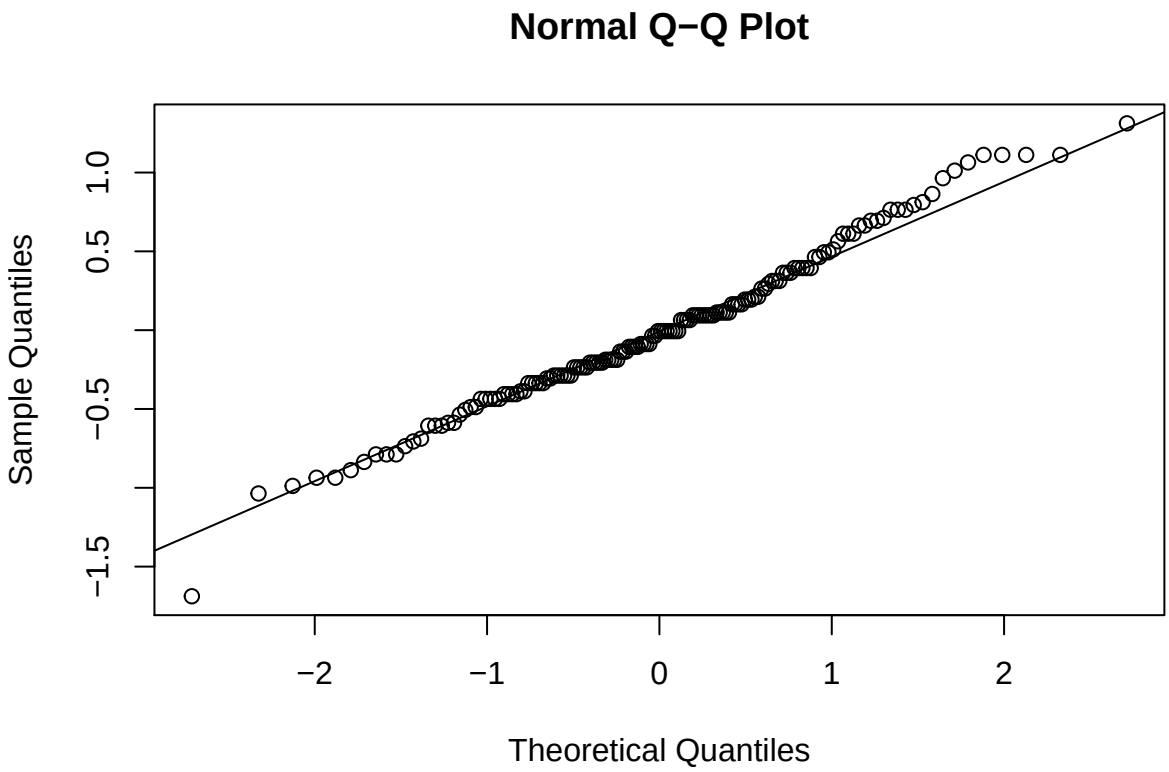
```
oneway.test(Sepal.Length ~ Species)
```

```
##
##  One-way analysis of means (not assuming equal variances)
##
## data:  Sepal.Length and Species
## F = 138.91, num df = 2.000, denom df = 92.211, p-value < 2.2e-16
```

```
plot(L$fit, L$res)
abline(h = 0)
```




```
qqnorm(L$res)
qqline(L$res)
```



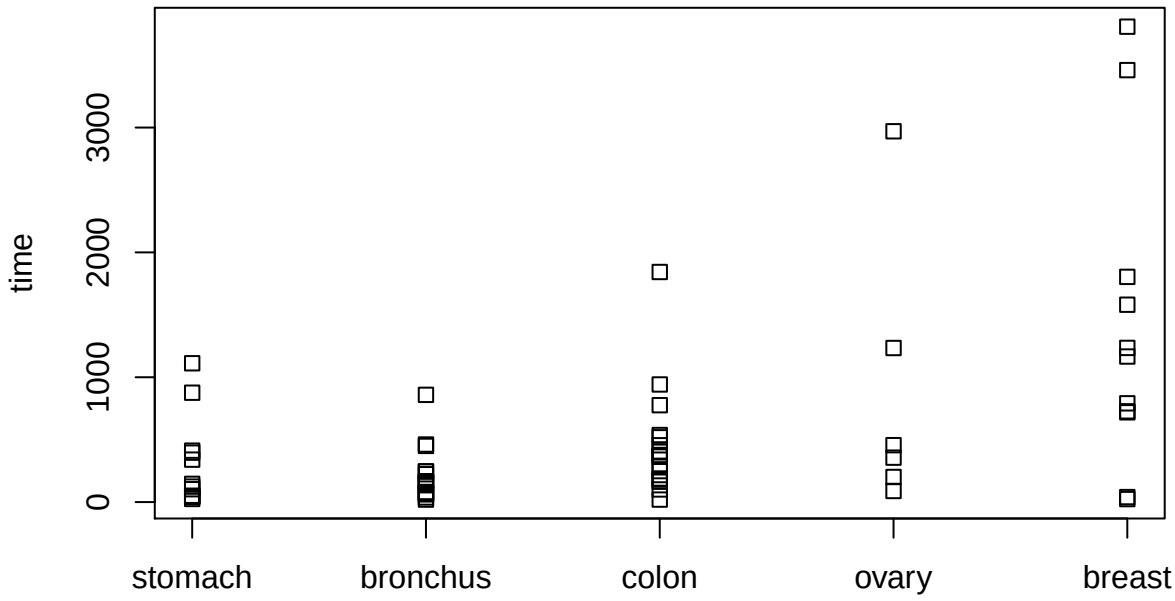
所以，前文的假设可以被验证，可以对这个模型做 ANOVA 分析。

8.5 (Cancer survival data).

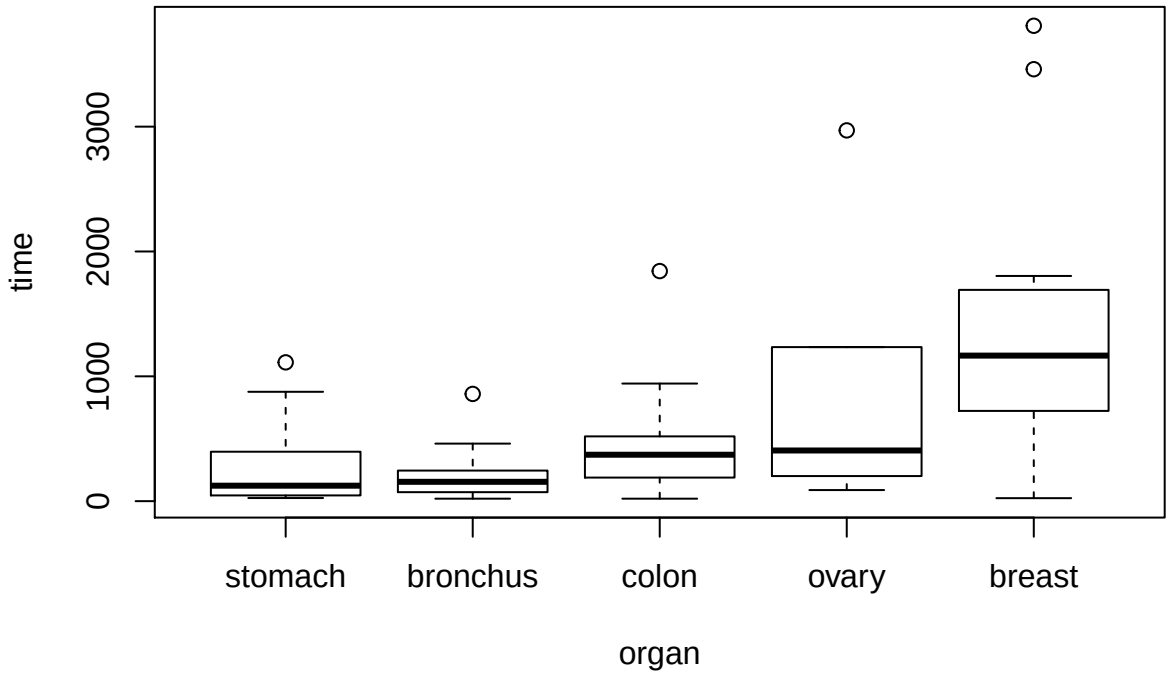
The cancer survival data “PATIENT.DAT”was introduced in Example 8.9. Start with the exploratory data analysis, and check for NID error model assumptions. Consider a transformation of the data if the assumptions for error are not satisfied. Complete a one-way ANOVA to determine whether mean survival times differ by organ. If there are significant differences, follow up with appropriate multiple comparisons to determine which means differ and describe how they differ.

Answer: 把 $N(\mu, \sigma^2)$ 全部减掉 μ。或者用 LSD。

```
times = read.table("Rx-Data/PATIENT.DAT", sep="\t")
names(times) = c("stomach","bronchus","colon","ovary","breast")
times1 = stack(times)
names(times1) = c("time", "organ")
times1 = na.omit(times1)
attach(times1)
stripchart(time ~ organ, vertical=TRUE)
```




```r
boxplot(time ~ organ, vertical=TRUE)
```




```r
oneway.test(time ~ organ)
```

```
##
##  One-way analysis of means (not assuming equal variances)
##
## data:  time and organ
## F = 3.5152, num df = 4.000, denom df = 19.862, p-value = 0.02514
```

```r
oneway.test(time ~ organ, var.eq=TRUE)
```

```
##
##  One-way analysis of means
##
## data:  time and organ
## F = 6.4334, num df = 4, denom df = 59, p-value = 0.0002295
```