Assumptions of error terms

In ANOVA model, we assume the error terms
are independent and normally distributed with constant variance σ^2.
For example, with one-way ANOVA model:
$y_{ij} = \mu_i + \epsilon_{ij}$
We assume $\epsilon_{ij} \sim Normal(0, \sigma^2)$ and are independent.

Estimate of errors (residual): $e_{ij} = y_{ij} - \bar{y}_{i\cdot}$ where $\bar{y}_{i\cdot}$ is the fitted value of y_{ij}.
In general, residual = observed value -fitted value
Residulas and fitted values can be extracted from ANOVA output with

```
output$residuals
output$fitted.values
```

estimate errors in two-way ANOVA

Two factor design with interaction:

$y_{ijk} = \mu_{ij} + \epsilon_{ijk}$

$e_{ijk} = y_{ijk} - \bar{y}_{ij\cdot}$

Blocking desgin:

$e_{ij} = y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot}$

check independence of error terms

likely cause of dependence: experimental units are close in time or space.
We can plot residuals vs time or space if the time order (or space order) of the experiment is available.

R : plot y variable vs x variable:

```
plot(x,y)
```

checking homogeneous error variances

1. Compare s for different treatment groups. $\max s \leq 3 \min s$.
2. Plot y values vs the treatments.
3. plot residuals vs or treatments or fitted values.

```
> battery=read.table("http://educ.jmu.edu/~chen3lx/
                      math321/battery.txt",header=T)
> output=aov(Lifetime~factor(Type),data=battery)
```

Make boxplot of the response: Lifetime

```
> boxplot(battery$Lifetime~ battery$Type)
```

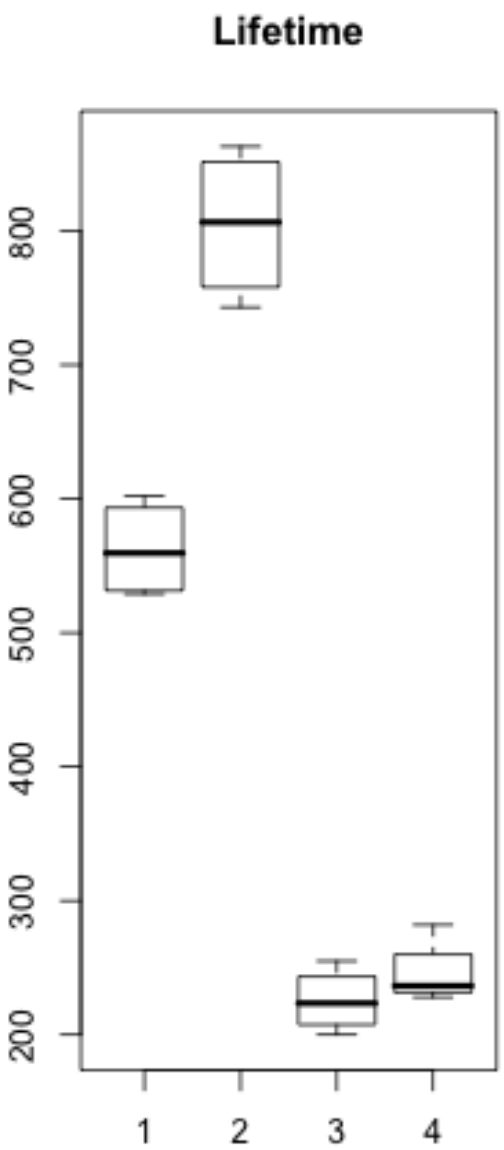
Make residual plot:
plot of residuals vs timeorder (checking for independence)
boxplot of residuals or plot of residuals vs fitted values (checking for constant variance)

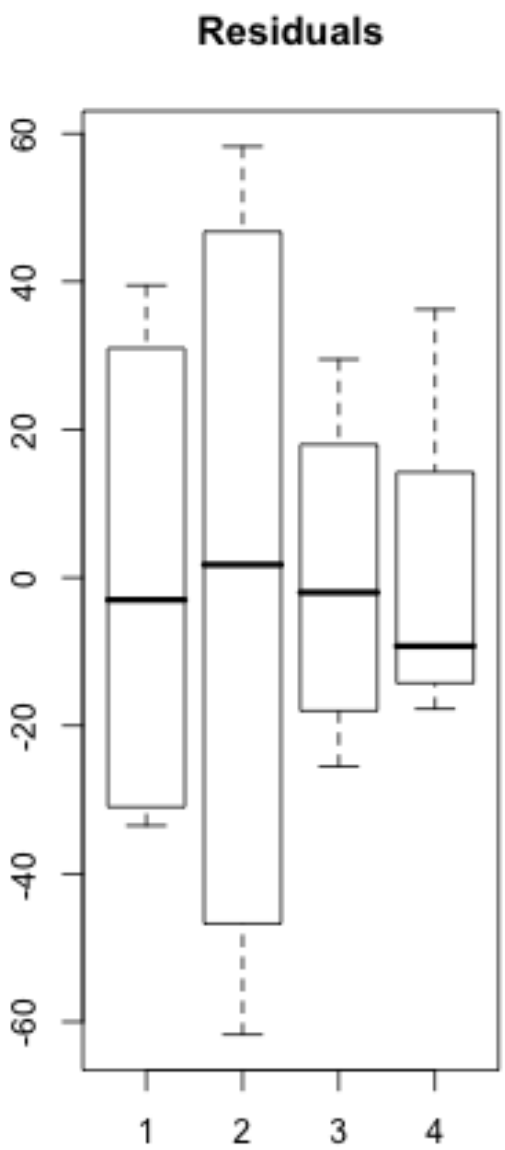
```
> plot(battery$TimeOrder,output$residuals)
```

```
> plot(output$fitted.values, output$residuals)
```

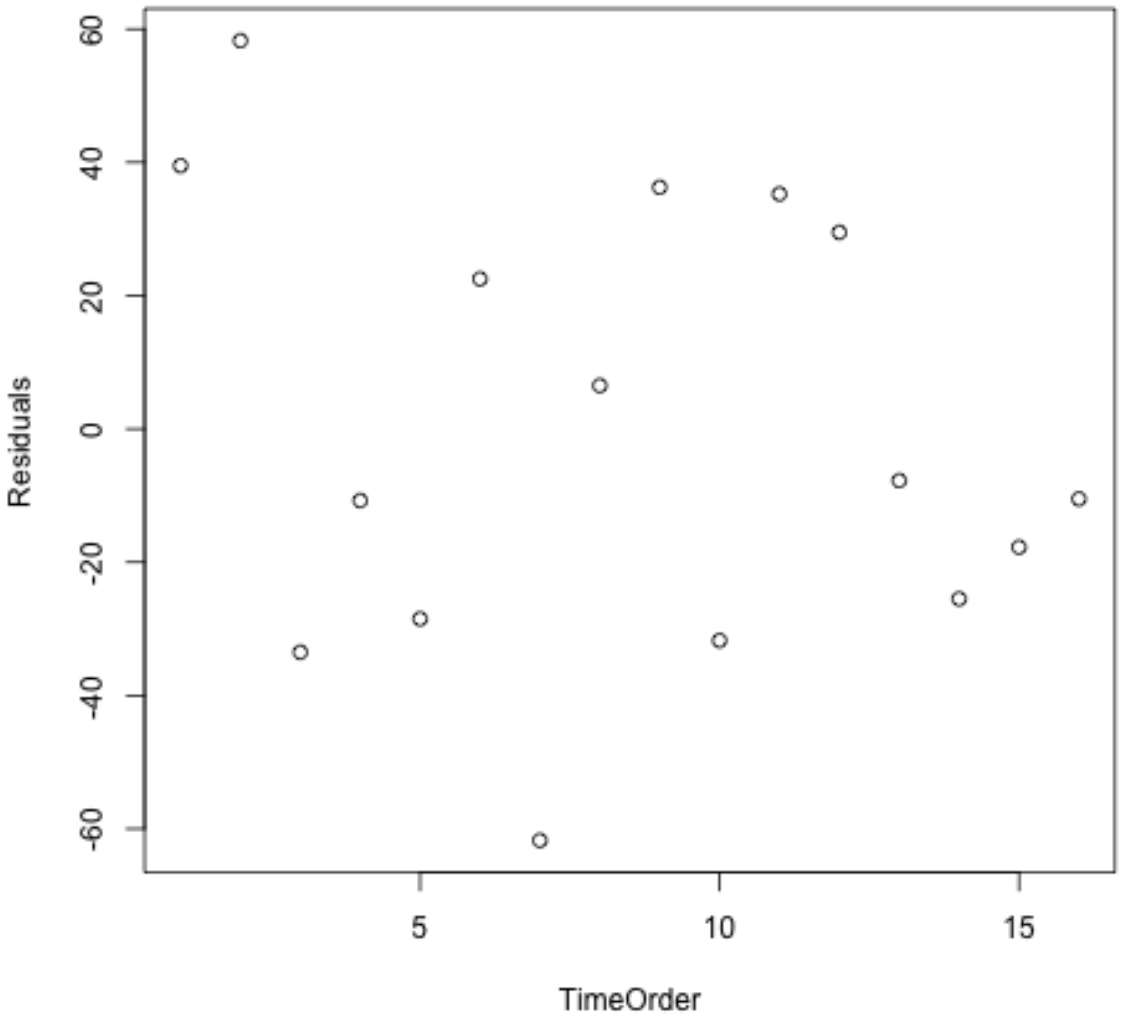
```
> boxplot(output$residuals~ battery$Type)
```

Boxplot

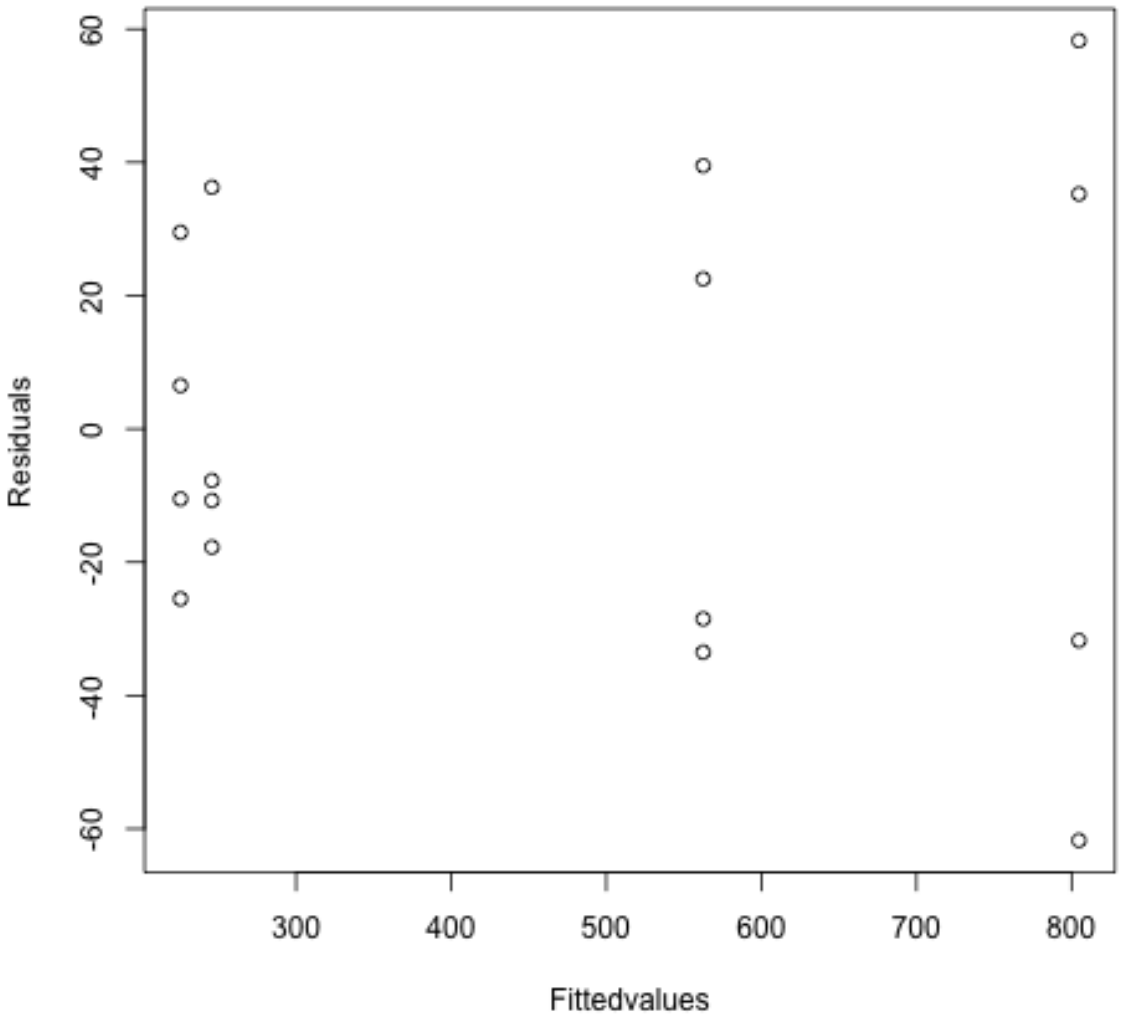

Lifetime
200
300
400
500
600
700
800
1
2
3
4


Residuals
-60
-40
-20
0
20
40
60
1
2
3
4

Residual vs time order



Residual vs fitted value



Check nomality of errors

Histogram or stem-and-leaf plot of residuals.

```
> hist(output$residuals)
> stem(output$residuals)
```

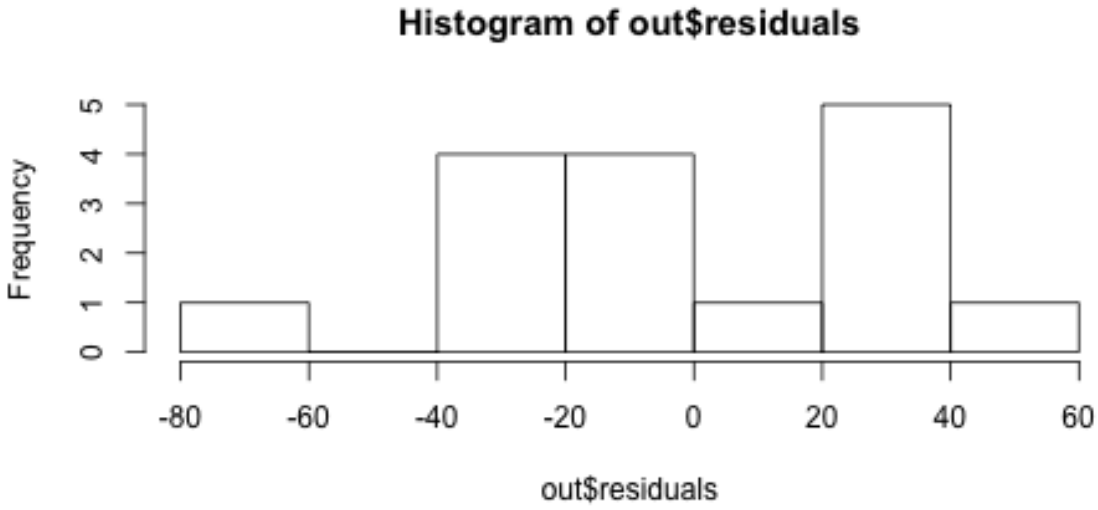
QQ plot: a plot of r_i vs z_i where $r_1, r_2, \cdots, r_n$ are the ordered residuals and z_i is the expected quantiles from the standard normal distribution.

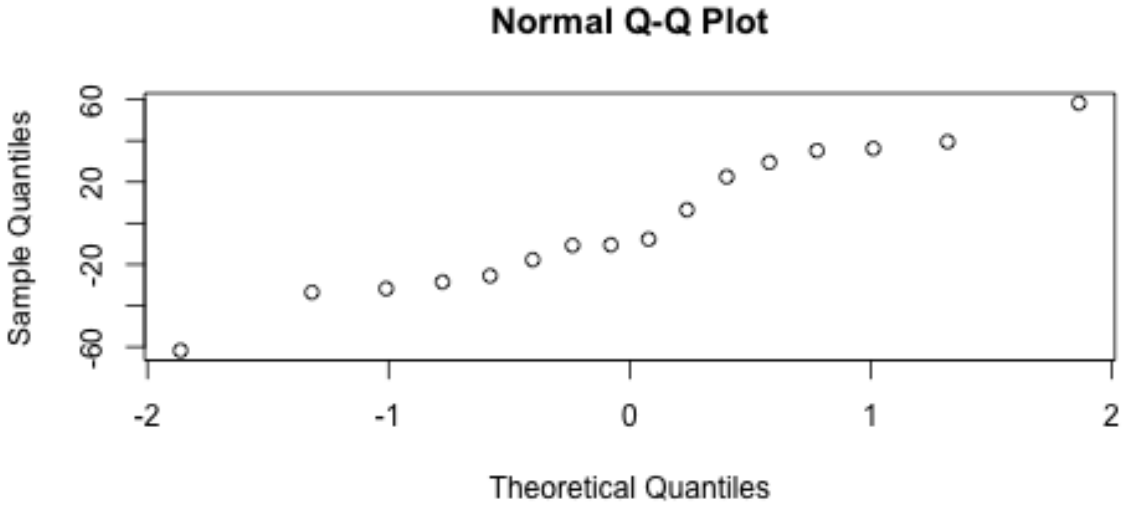
```
> sort(residuals) # random sample from z distribution
 -1.8053348 -1.2823375 -1.1972396 -0.6608950 -0.459794
 -0.2891898 0.3517191  0.4342224  1.1808055  1.2619139

> qnorm(0.05+0.1*(0:9)) #theoretical quantiles
 -1.6448536 -1.0364334 -0.6744898 -0.3853205 -0.125661
 0.1256613 0.3853205  0.6744898  1.0364334  1.6448536
```

R code for making QQ plot:

```
> qqnorm(output$residuals)
```

Histogram and QQ plot





Exercise

Use the candle.txt data on our course website. Fit an one-way ANOVA model to the data.
Make the proper plots to check the following assumptions:
Are the error terms independent (treat the order variable in the data as the time order by which each candle was burned)?
Do the error terms have homogeneous variance?
Are the errors normally distributed?