

# cancer

*Chris Parrish*

*January 19, 2016*

cancer

reference:

- Cannon, et al., Stat2, chapter 05, example 5.8

Import the data.

```
data <- read.csv("CancerSurvival.CSV", header=TRUE)
head(data, 3)
```

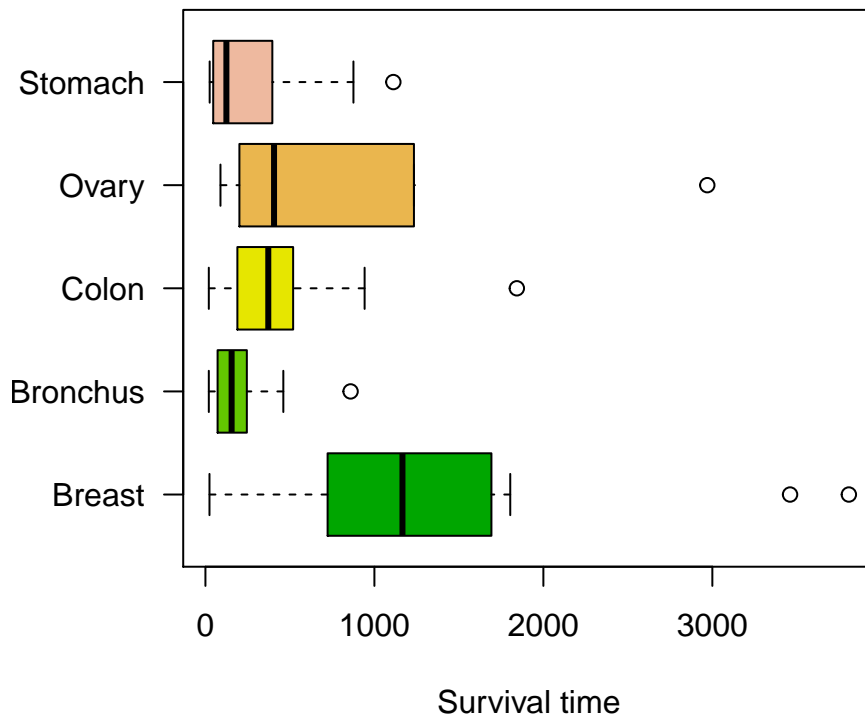
```
##   Survival   Organ
## 1     124 Stomach
## 2      42 Stomach
## 3      25 Stomach
```

```
dim(data)
```

```
## [1] 64  2
```

View the data.

```
old.par <- par(mar=c(4, 8, 4, 4))
plot(Survival ~ Organ, data=data, horizontal=TRUE,
     col=terrain.colors(6), las=1, ylab="Survival time", xlab="")
```



```
par(old.par)
```

Group statistics.

```
n <- with(data, tapply(Survival, Organ, length))
mean <- with(data, round(tapply(Survival, Organ, mean), 3))
sd <- with(data, round(tapply(Survival, Organ, sd), 3))
cancer.statistics <- cbind(n, mean, sd)
cancer.statistics
```

```
##          n      mean      sd
## Breast  11 1395.909 1238.967
## Bronchus 17  211.588  209.859
## Colon   17  457.412  427.169
## Ovary    6  884.333 1098.579
## Stomach 13  286.000  346.310
```

```
grand.mean <- cbind(n = length(data$Survival),
                    mean = mean(data$Survival),
                    sd = sd(data$Survival))
rownames(grand.mean) <- c("Total")
grand.mean <- round(grand.mean, 3)
grand.mean
```

```
##          n      mean      sd
## Total  64 558.625 776.479
```

First model: ANOVA with aov

```
cancer.aov1 <- aov(Survival ~ Organ, data=data)
cancer.aov1
```

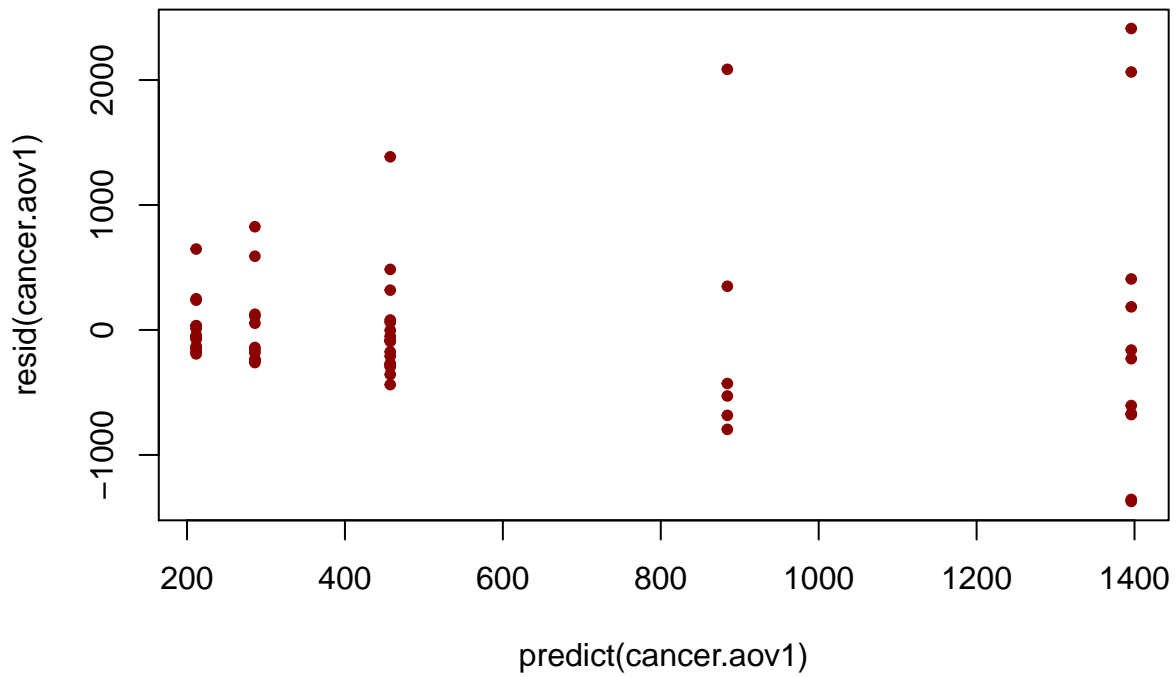
```
## Call:
## aov(formula = Survival ~ Organ, data = data)
##
## Terms:
##              Organ Residuals
## Sum of Squares 11535761 26448144
## Deg. of Freedom      4      59
##
## Residual standard error: 669.5324
## Estimated effects may be unbalanced
```

```
options(show.signif.stars=FALSE)
summary(cancer.aov1)
```

```
##          Df  Sum Sq Mean Sq F value  Pr(>F)
## Organ      4 11535761 2883940  6.433 0.000229
## Residuals 59 26448144  448274
```

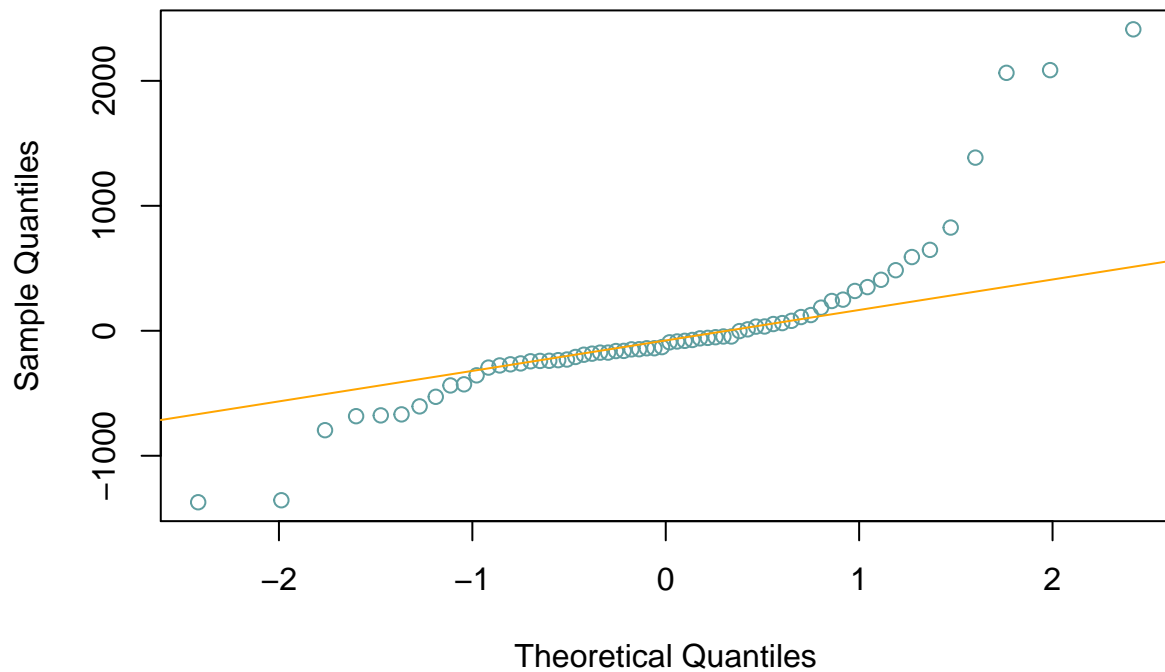
Residuals.

```
plot(predict(cancer.aov1), resid(cancer.aov1),  
     pch=20, col="darkred")
```

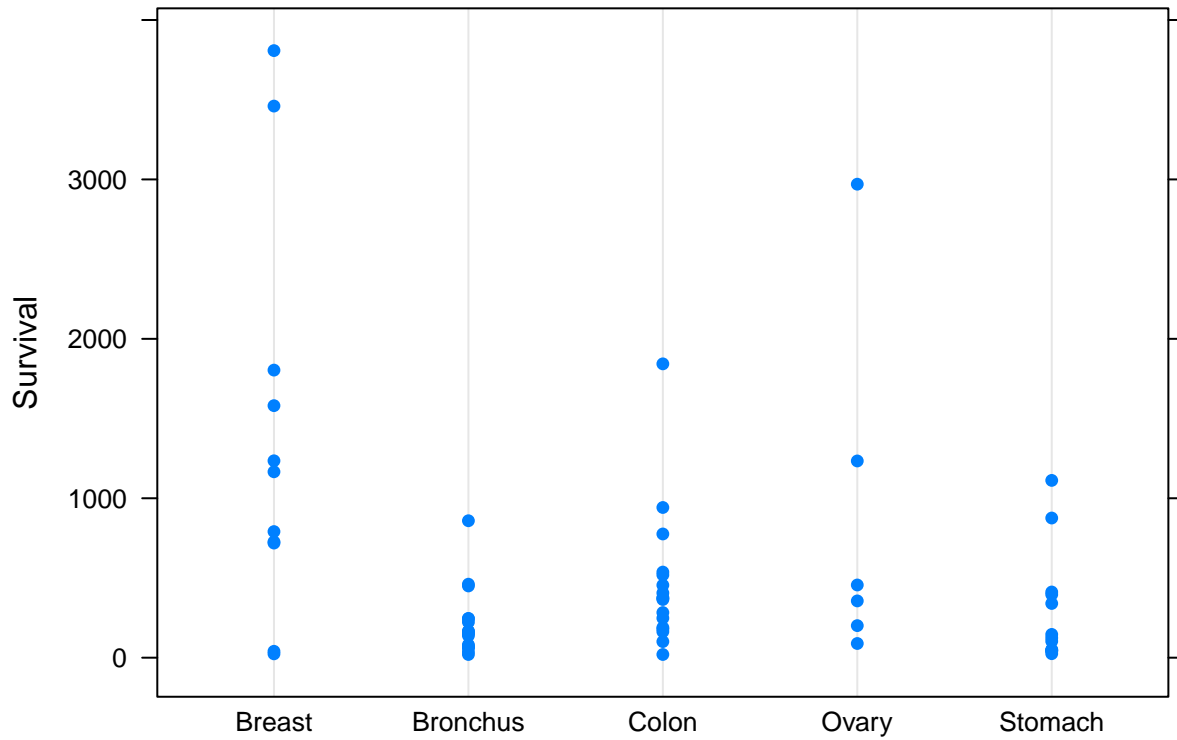


```
qqnorm(resid(cancer.aov1), col="cadetblue")  
qqline(resid(cancer.aov1), col="orange")  
library(lattice)
```

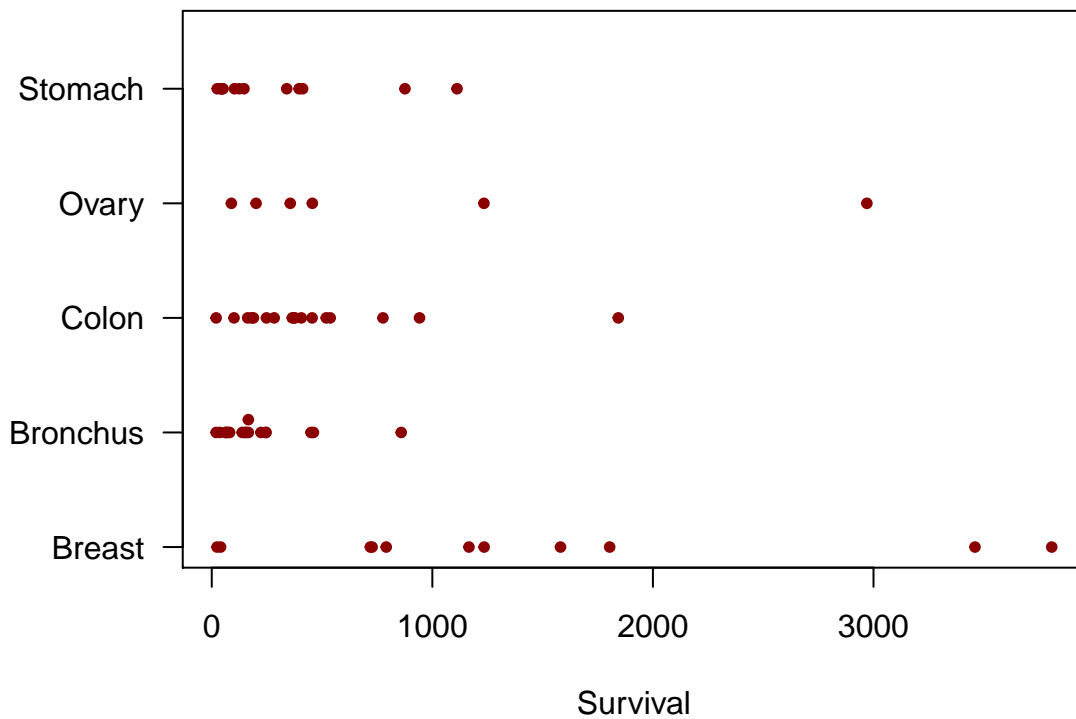
### Normal Q-Q Plot



```
dotplot(Survival ~ Organ, data=data, jitter=TRUE)
```



```
old.par <- par(mar=c(4, 6, 4, 3))  
stripchart(Survival ~ Organ, data=data,  
           pch=20, las=1, col="darkred", method="stack")
```



```
par(old.par)
std.dev <- cancer.statistics[ , 3]
std.dev
```

```
## Breast Bronchus Colon Ovary Stomach
## 1238.967 209.859 427.169 1098.579 346.310
```

```
ratio <- max(std.dev) / min(std.dev)
ratio
```

```
## [1] 5.903807
```

Second model: ANOVA with aov and log transformation

```
cancer.aov2 <- aov(log(Survival) ~ Organ, data=data)
cancer.aov2
```

```
## Call:
## aov(formula = log(Survival) ~ Organ, data = data)
##
## Terms:
##                Organ Residuals
## Sum of Squares 24.48656 84.26959
## Deg. of Freedom      4      59
##
## Residual standard error: 1.195114
## Estimated effects may be unbalanced
```

```
summary(cancer.aov2)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Organ      4  24.49   6.122   4.286 0.00412
## Residuals 59  84.27   1.428
```

```
cancer.aov2$coefficients
```

```
## (Intercept) OrganBronchus OrganColon OrganOvary OrganStomach
## 6.5586034 -1.6054332 -0.8094811 -0.4079870 -1.5906836
```

Compare with lm

aov is a wrapper for lm, but its results are displayed differently.

```
cancer.lm2 <- lm(log(Survival) ~ Organ, data=data)
summary(cancer.lm2)
```

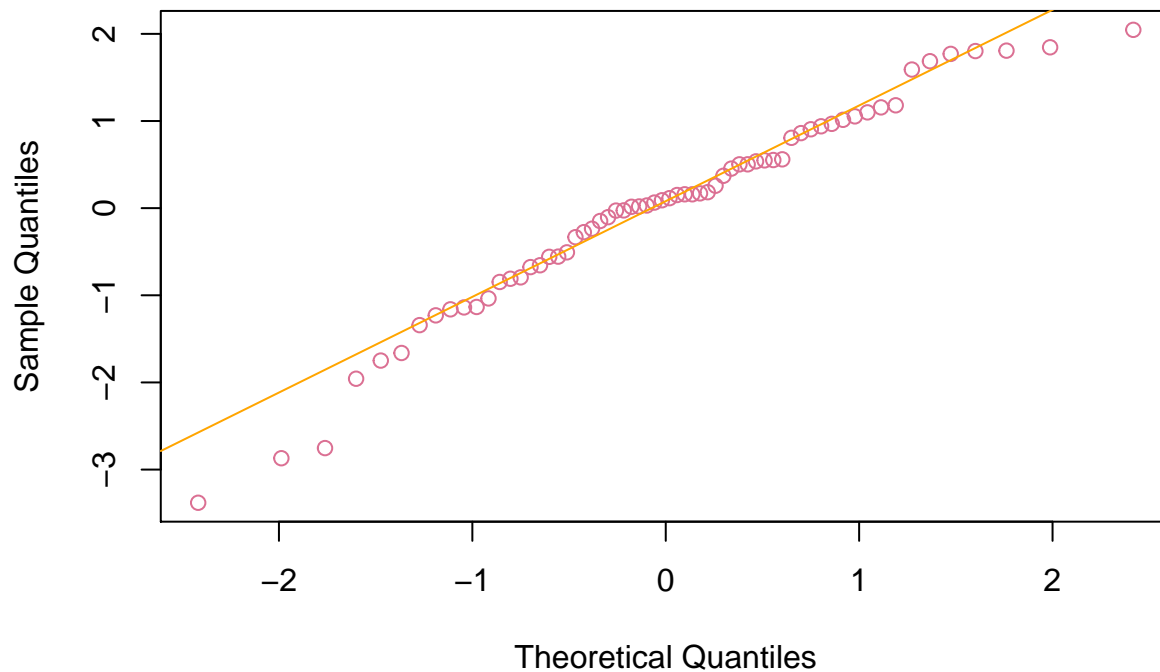
```
##
## Call:
## lm(formula = log(Survival) ~ Organ, data = data)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3805 -0.6607  0.1025  0.8207  2.0460
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    6.5586     0.3603  18.201 < 2e-16
## OrganBronchus -1.6054     0.4625  -3.472 0.000975
## OrganColon     -0.8095     0.4625  -1.750 0.085247
## OrganOvary     -0.4080     0.6065  -0.673 0.503801
## OrganStomach  -1.5907     0.4896  -3.249 0.001915
##
## Residual standard error: 1.195 on 59 degrees of freedom
## Multiple R-squared:  0.2252, Adjusted R-squared:  0.1726
## F-statistic: 4.286 on 4 and 59 DF,  p-value: 0.004122
```

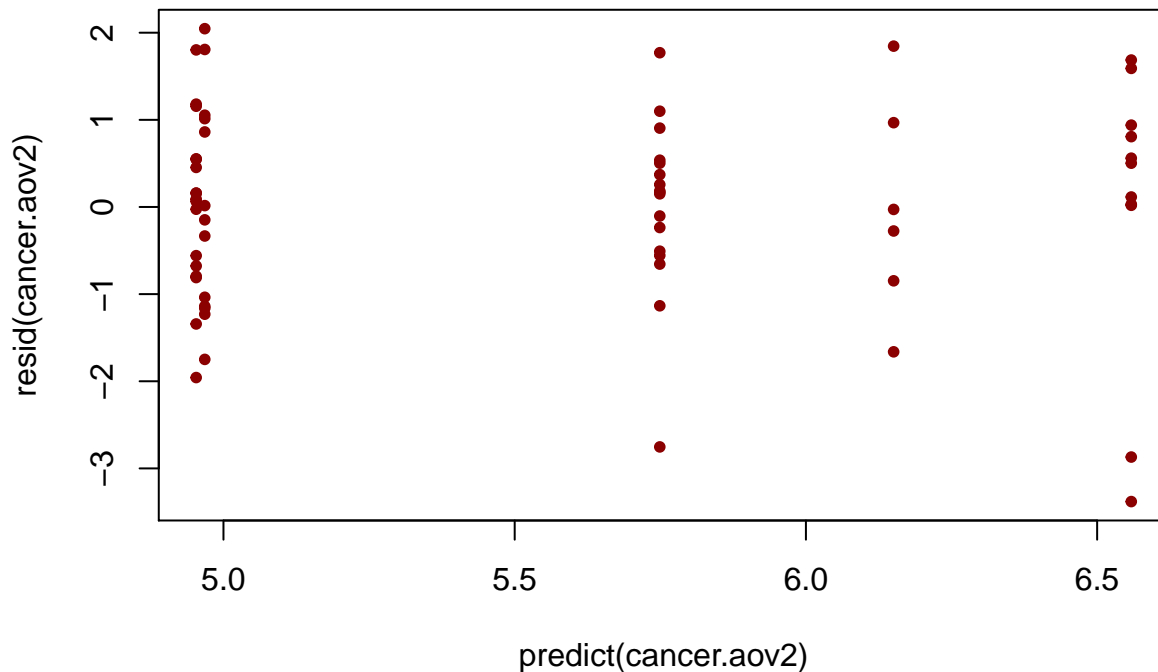
Residuals.

```
qqnorm(resid(cancer.aov2), col="palevioletred")
qqline(resid(cancer.aov2), col="orange")
```

**Normal Q-Q Plot**



```
plot(predict(cancer.aov2), resid(cancer.aov2),
      pch=20, col="darkred")
```



Group statistics.

```
n <- with(data, tapply(log(Survival), Organ, length))
mean <- with(data, round(tapply(log(Survival), Organ, mean), 3))
sd <- with(data, round(tapply(log(Survival), Organ, sd), 3))
cancer.statistics <- cbind(n, mean, sd)
cancer.statistics
```

```
##          n mean  sd
## Breast  11 6.559 1.648
## Bronchus 17 4.953 0.953
## Colon    17 5.749 0.997
## Ovary    6 6.151 1.257
## Stomach  13 4.968 1.250
```

```
grand.mean <- cbind(n = length(data$Survival),
                    mean = mean(log(data$Survival)),
                    sd = sd(log(data$Survival)))
rownames(grand.mean) <- c("Total")
grand.mean <- round(grand.mean, 3)
grand.mean
```

```
##          n mean  sd
## Total  64 5.556 1.314
```

```
std.dev <- cancer.statistics[ , 3]
std.dev
```

```
## Breast Bronchus Colon Ovary Stomach
## 1.648 0.953 0.997 1.257 1.250
```

```
ratio <- max(std.dev) / min(std.dev)
ratio
```

```
## [1] 1.729276
```