

# ethanol

*Chris Parrish*

*January 21, 2016*

ethanol

reference:

- Cannon, et al., chapter 6, case study

Import the data.

```
data <- read.table("Ethanol.csv", header=TRUE, sep=",")
head(data)
```

```
##      Sugar O2Conc Ethanol
## 1 Glucose     0     59
## 2 Glucose     0     30
## 3 Glucose    46     44
## 4 Glucose    46     18
## 5 Glucose    92     22
## 6 Glucose    92     23
```

```
dim(data)
```

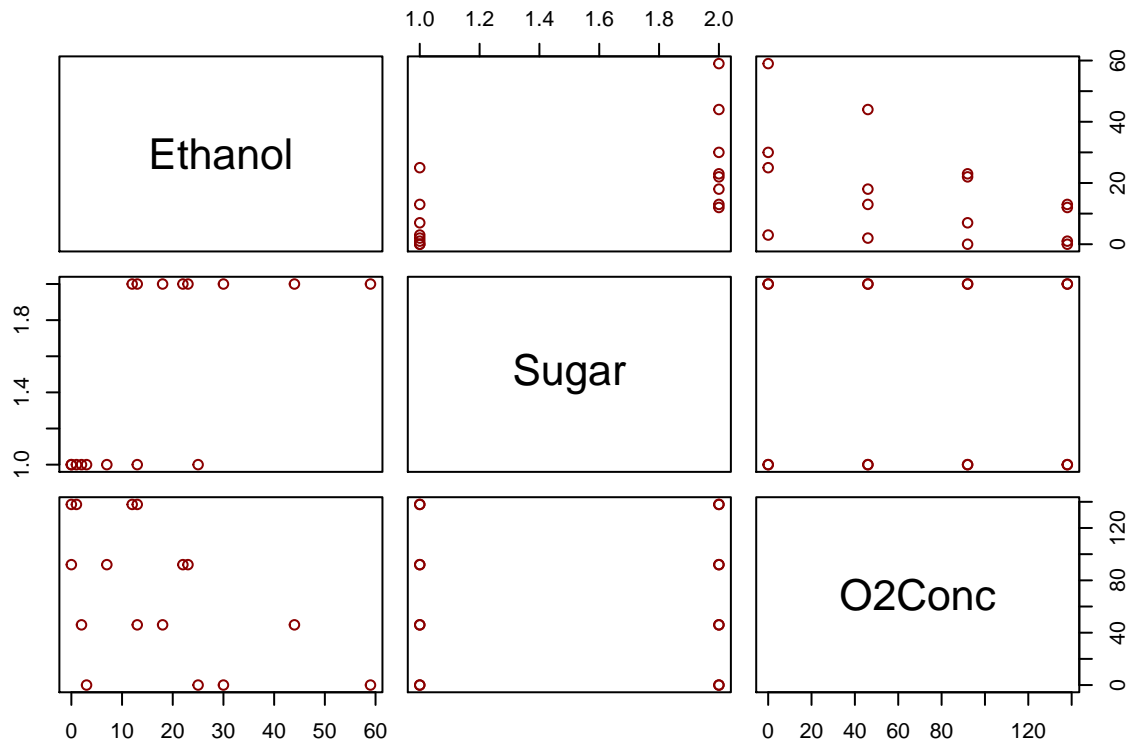
```
## [1] 16  3
```

```
str(data) # O2Conc is a numerical variable
```

```
## 'data.frame':  16 obs. of  3 variables:
## $ Sugar  : Factor w/ 2 levels "Galactose","Glucose": 2 2 2 2 2 2 2 1 1 ...
## $ O2Conc : int  0 0 46 46 92 92 138 138 0 0 ...
## $ Ethanol: int  59 30 44 18 22 23 12 13 25 3 ...
```

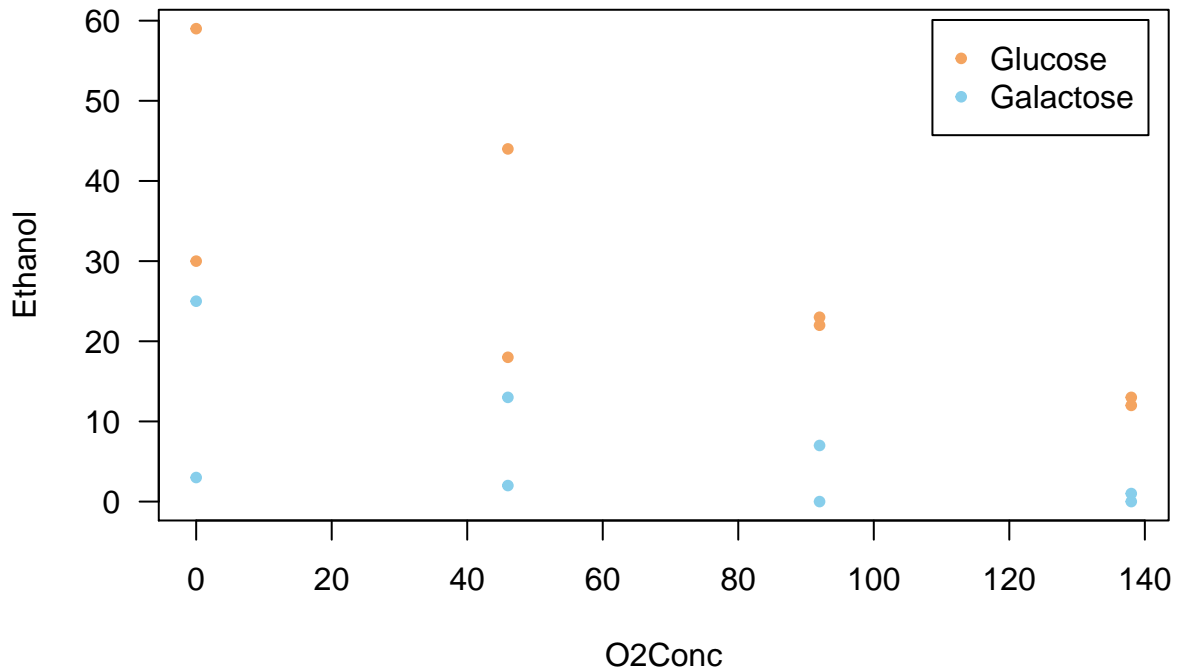
Scatterplot matrix.

```
pairs(~ Ethanol + Sugar + O2Conc, data=data, col="darkred")
```



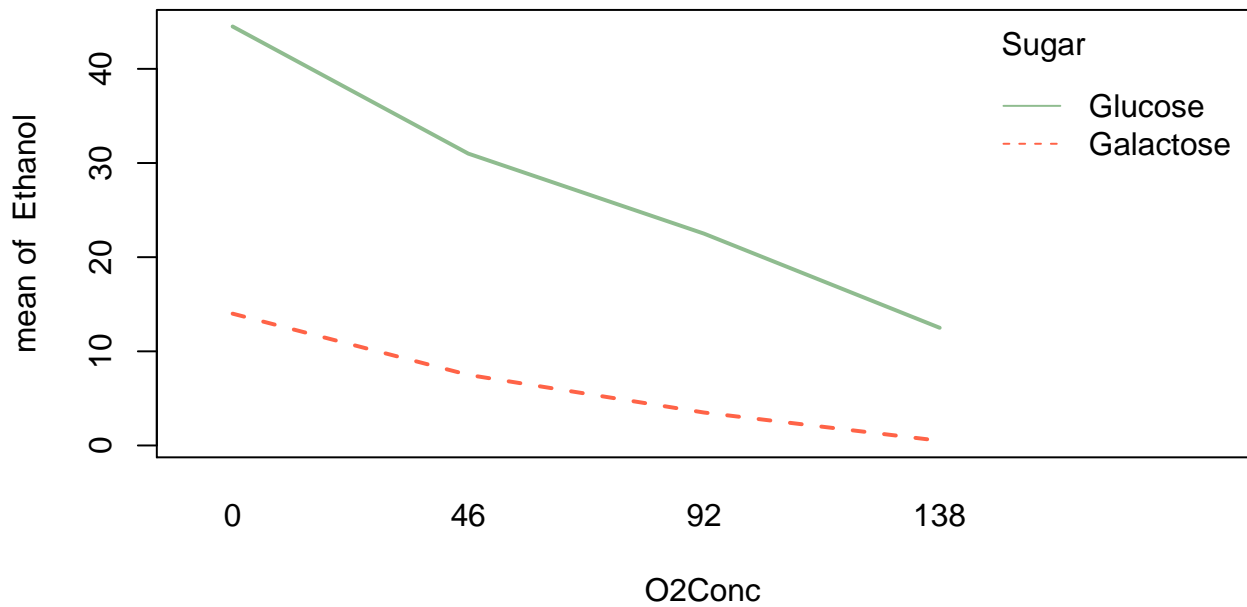
Plot Ethanol ~ O2Conc | Sugar

```
plot(Ethanol ~ O2Conc, data=data,
     las=1, type="n")
glucose <- data[data$Sugar=="Glucose", c(2, 3)]
galactose <- data[data$Sugar=="Galactose", c(2, 3)]
points(glucose, pch=20, col="sandybrown")
points(galactose, pch=20, col="skyblue")
legend("topright", pch=20, inset=0.02,
      legend=c("Glucose", "Galactose"),
      col=c("sandybrown", "skyblue"))
```

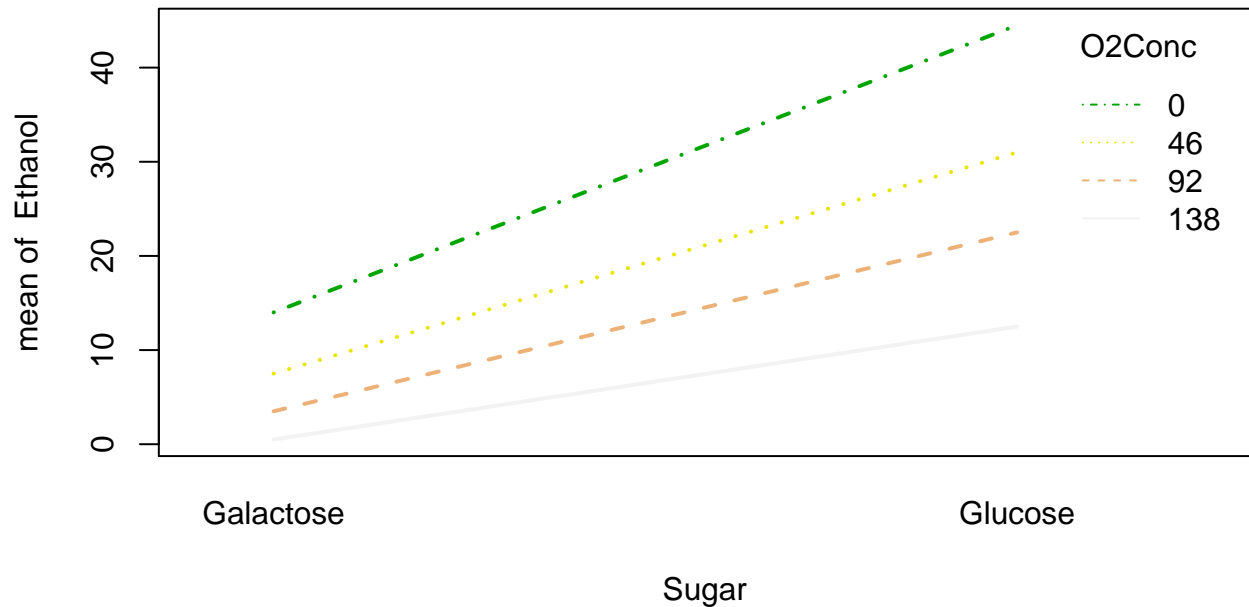


Interaction plots.

```
with(data,
      interaction.plot(O2Conc, Sugar, Ethanol,
                     col=c("tomato", "darkseagreen"), lwd=2))
```



```
with(data,
      interaction.plot(Sugar, O2Conc, Ethanol,
                     col=terrain.colors(4), lwd=2))
```



Two-way ANOVA with interaction.

Note that this display does not at all match the display on p.304. Here, O2Conc is a numerical variable, but on p.307 the authors mention that O2Conc is a categorical factor, hence their three degrees of freedom.

```
ethanol.aov<- aov(Ethanol ~ O2Conc + Sugar + O2Conc:Sugar, data=data)
options(show.signif.stars=FALSE)
anova(ethanol.aov)
```

```
## Analysis of Variance Table
##
## Response: Ethanol
##          Df Sum Sq Mean Sq F value    Pr(>F)
## O2Conc     1  1110.0   1110.05  12.0690 0.0045965
## Sugar      1  1806.2   1806.25  19.6385 0.0008189
## O2Conc:Sugar 1   180.0    180.00   1.9571 0.1871409
## Residuals  12  1103.7    91.97
```

Change O2Conc to a categorical factor. The ANOVA now matches the display on p.304.

```
data.mod <- data
data.mod$O2Conc <- factor(data.mod$O2Conc)
str(data.mod)
```

```
## 'data.frame':  16 obs. of  3 variables:
## $ Sugar : Factor w/ 2 levels "Galactose","Glucose": 2 2 2 2 2 2 2 2 1 1 ...
## $ O2Conc : Factor w/ 4 levels "0","46","92",...: 1 1 2 2 3 3 4 4 1 1 ...
## $ Ethanol: int  59 30 44 18 22 23 12 13 25 3 ...
```

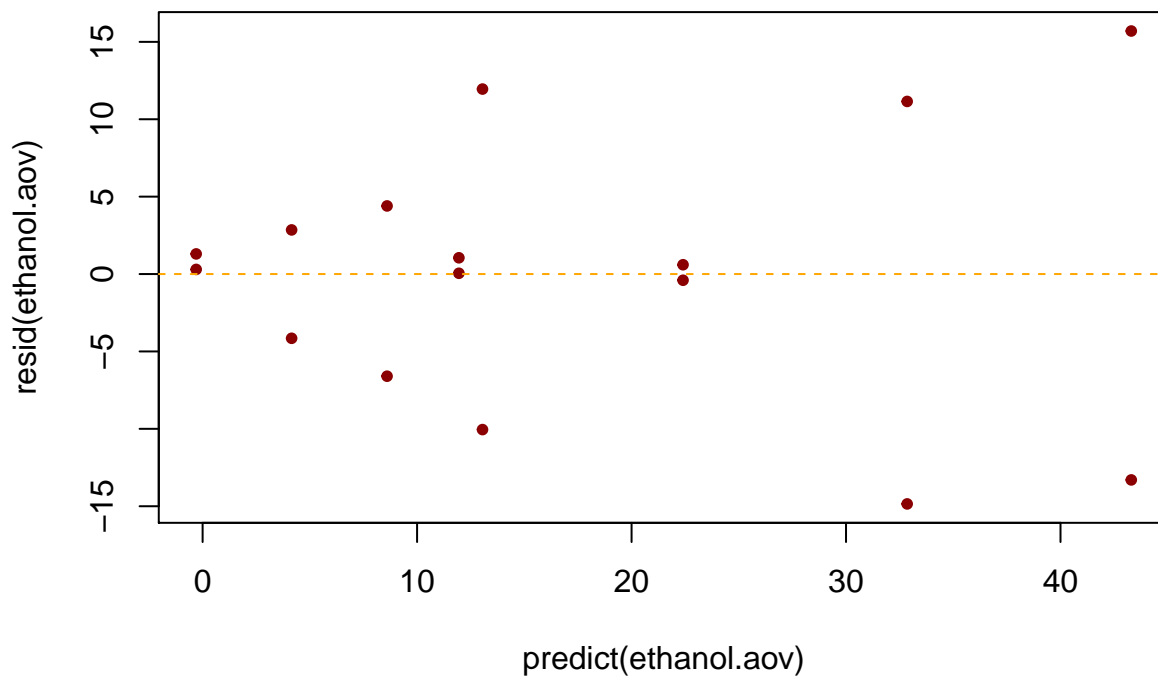
```
ethanol.mod.aov<- aov(Ethanol ~ Sugar + O2Conc + Sugar:O2Conc, data=data.mod)
anova(ethanol.mod.aov)
```

```
## Analysis of Variance Table
```

```
##  
## Response: Ethanol  
##           Df Sum Sq Mean Sq F value Pr(>F)  
## Sugar      1 1806.25 1806.25 13.2935 0.006532  
## O2Conc     3 1125.50  375.17  2.7611 0.111472  
## Sugar:O2Conc 3  181.25   60.42  0.4446 0.727690  
## Residuals  8 1087.00  135.87
```

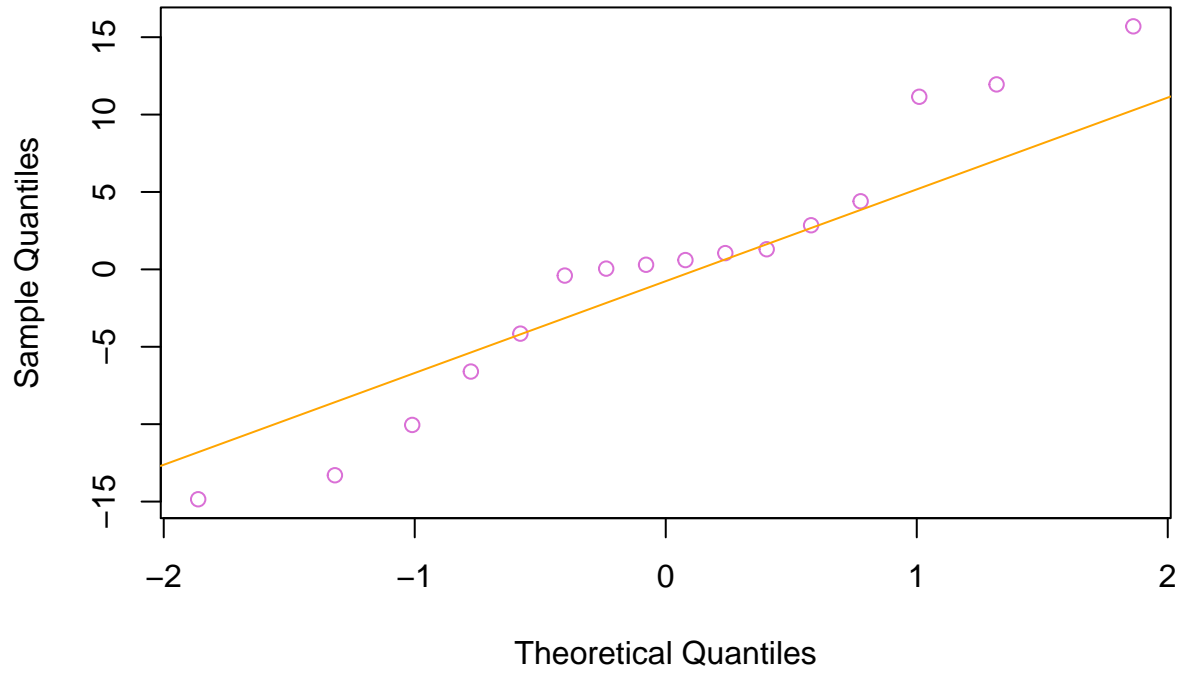
Residuals.

```
plot(predict(ethanol.aov), resid(ethanol.aov),  
      pch=20, col="darkred")  
abline(h=0, col="orange", lty="dashed")
```



```
qqnorm(resid(ethanol.aov), col="orchid")  
qqline(resid(ethanol.aov), col="orange")
```

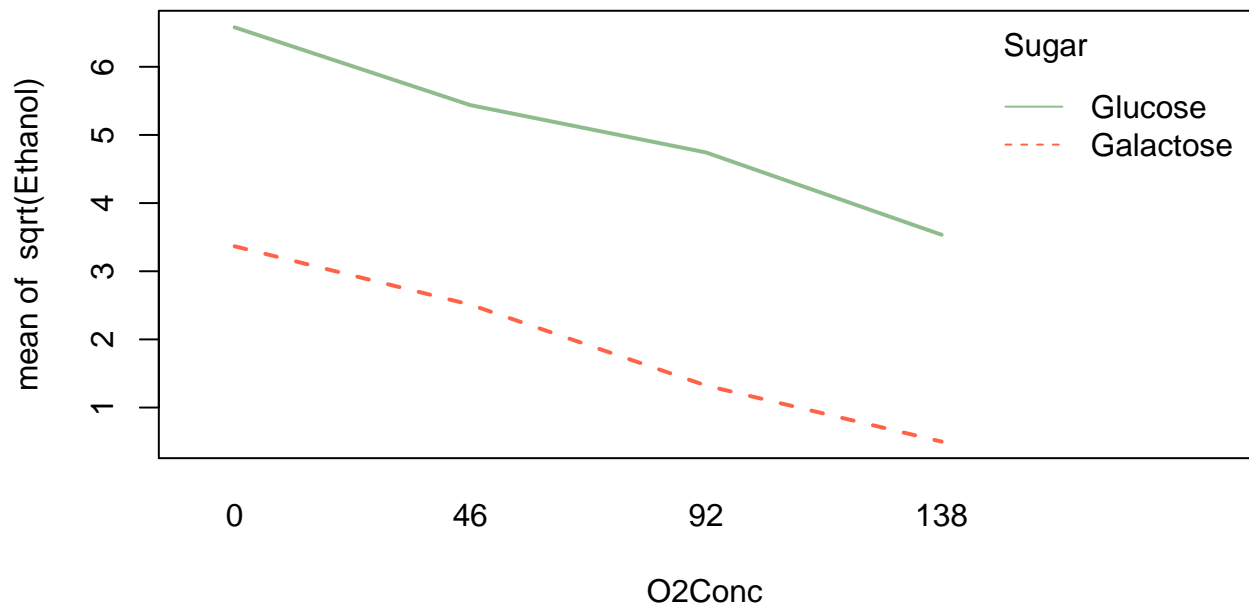
## Normal Q-Q Plot



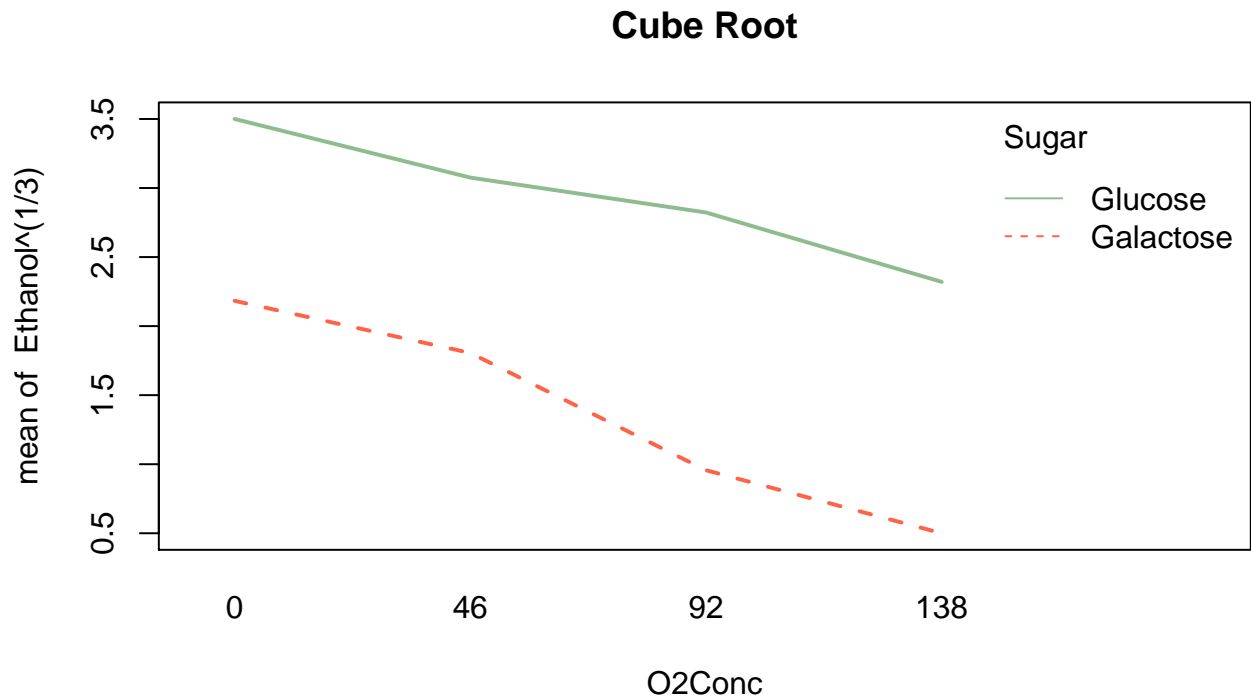
Interaction plots for square root, cube root, and log.

```
with(data,  
      interaction.plot(O2Conc, Sugar, sqrt(Ethanol),  
                      col=c("tomato", "darkseagreen"), lwd=2, main="Square Root"))
```

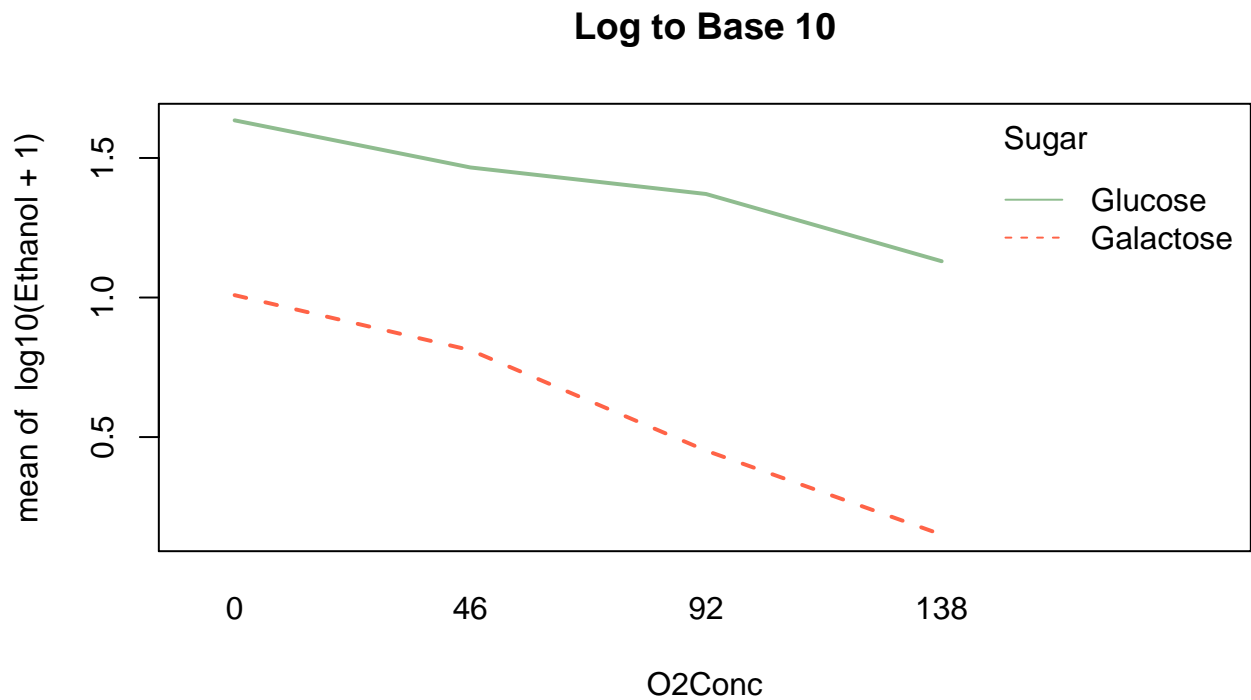
## Square Root



```
with(data,
      interaction.plot(O2Conc, Sugar, Ethanol^(1/3),
                     col=c("tomato", "darkseagreen"), lwd=2, main="Cube Root"))
```



```
with(data,
      interaction.plot(O2Conc, Sugar, log10(Ethanol + 1),
                     col=c("tomato", "darkseagreen"), lwd=2, main="Log to Base 10"))
```



Model 1: original scale (for ethanol concentration)

```
ethanol.lm1<- lm(Ethanol ~ Sugar + O2Conc, data=data)
summary(ethanol.lm1)
```

```
##
## Call:
## lm(formula = Ethanol ~ Sugar + O2Conc, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.550  -5.362  -1.400   5.050  20.200
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.5500     4.8427   3.624 0.003087
## SugarGlucose  21.2500     4.9685   4.277 0.000901
## O2Conc        -0.1620     0.0483  -3.353 0.005192
##
## Residual standard error: 9.937 on 13 degrees of freedom
## Multiple R-squared:  0.6944, Adjusted R-squared:  0.6473
## F-statistic: 14.77 on 2 and 13 DF,  p-value: 0.0004507
```

Model 2: square roots

```
ethanol.lm2<- lm(sqrt(Ethanol) ~ Sugar + O2Conc, data=data)
summary(ethanol.lm2)
```

```
##
## Call:
## lm(formula = sqrt(Ethanol) ~ Sugar + O2Conc, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.66362 -1.01748  0.05486  1.08597  1.60433
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.395670   0.560331   6.060 4.03e-05
## SugarGlucose  3.149074   0.574887   5.478 0.000106
## O2Conc        -0.021318   0.005589  -3.814 0.002148
##
## Residual standard error: 1.15 on 13 degrees of freedom
## Multiple R-squared:  0.7741, Adjusted R-squared:  0.7394
## F-statistic: 22.28 on 2 and 13 DF,  p-value: 6.311e-05
```

Model 3: cube roots

```
ethanol.lm3<- lm(Ethanol^(1/3) ~ Sugar + O2Conc, data=data)
summary(ethanol.lm3)
```

```
##
## Call:
```



```
## lm(formula = Ethanol^(1/3) ~ Sugar + O2Conc, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1191 -0.5498  0.1315  0.3601  0.8359
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.088083   0.304309   6.862 1.15e-05
## SugarGlucose  1.568442   0.312215   5.024 0.000233
## O2Conc       -0.010533   0.003035  -3.470 0.004145
##
## Residual standard error: 0.6244 on 13 degrees of freedom
## Multiple R-squared:  0.7414, Adjusted R-squared:  0.7017
## F-statistic: 18.64 on 2 and 13 DF,  p-value: 0.0001519
```

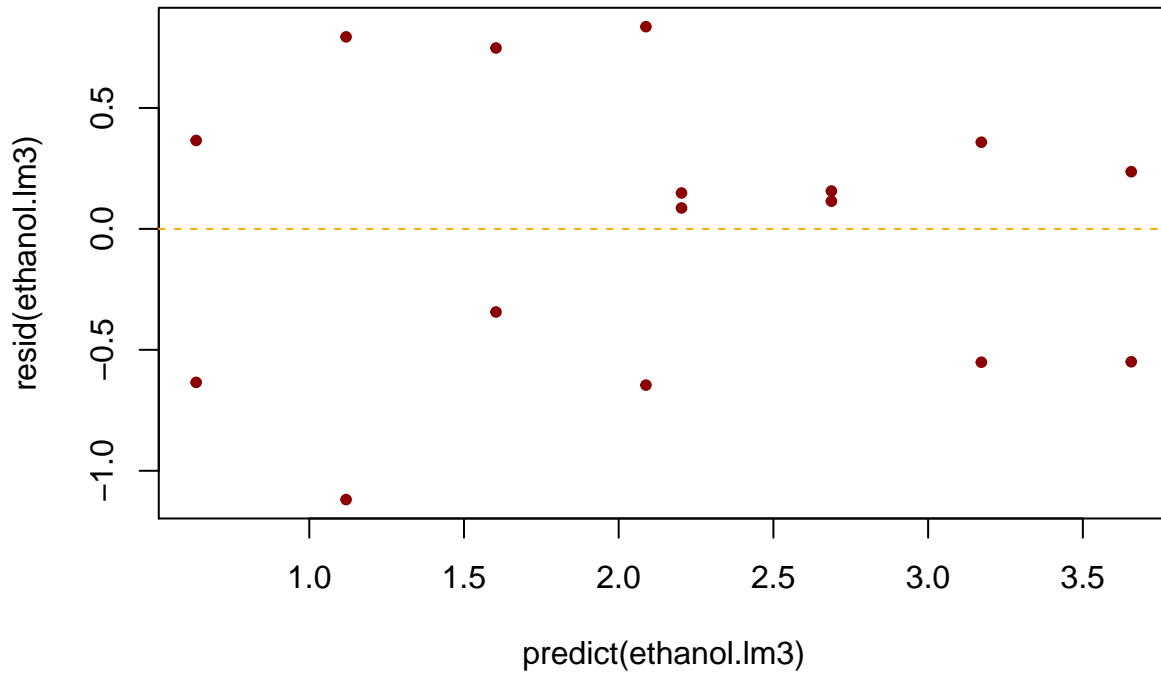
Model 4: log

```
ethanol.lm4<- lm(log(Ethanol + 1) ~ Sugar + O2Conc, data=data)
summary(ethanol.lm4)
```

```
##
## Call:
## lm(formula = log(Ethanol + 1) ~ Sugar + O2Conc, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1328 -0.5617  0.1050  0.2414  1.0792
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.178924   0.344433   6.326 2.63e-05
## SugarGlucose  1.830292   0.353380   5.179 0.000177
## O2Conc       -0.011371   0.003436  -3.310 0.005641
##
## Residual standard error: 0.7068 on 13 degrees of freedom
## Multiple R-squared:  0.744, Adjusted R-squared:  0.7046
## F-statistic: 18.89 on 2 and 13 DF,  p-value: 0.0001424
```

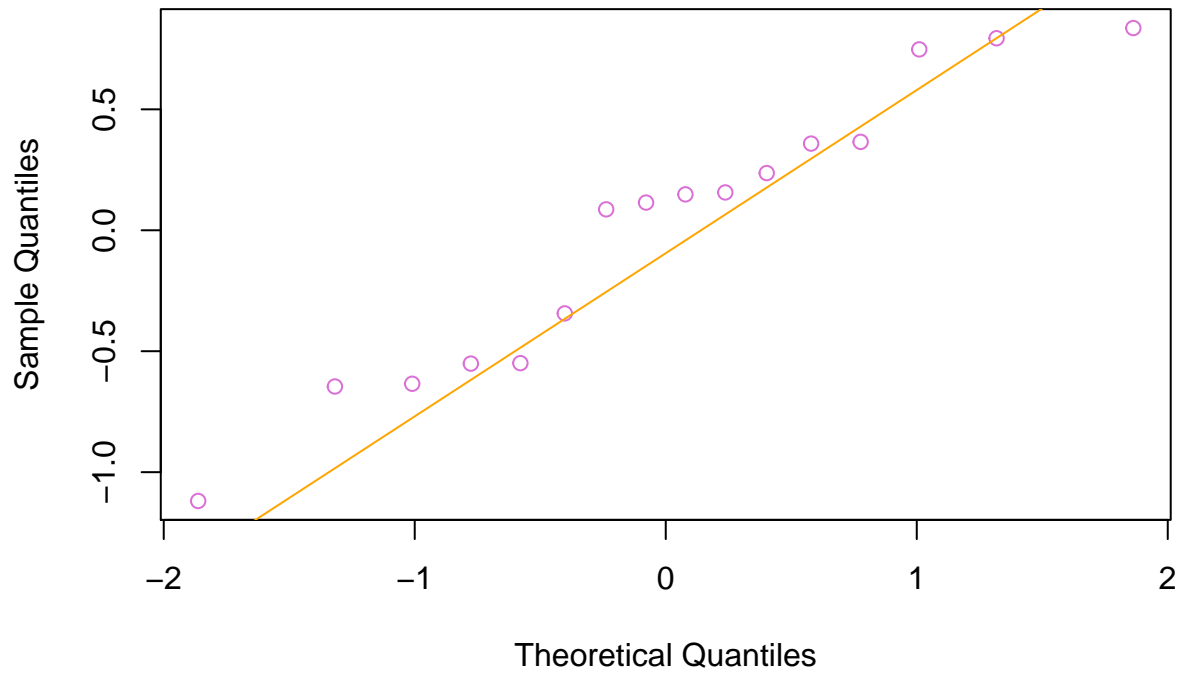
Residuals for model 3.

```
plot(predict(ethanol.lm3), resid(ethanol.lm3),
      pch=20, col="darkred")
abline(h=0, col="orange", lty="dashed")
```



```
qqnorm(resid(ethanol.lm3), col="orchid")
qqline(resid(ethanol.lm3), col="orange")
```

### Normal Q-Q Plot

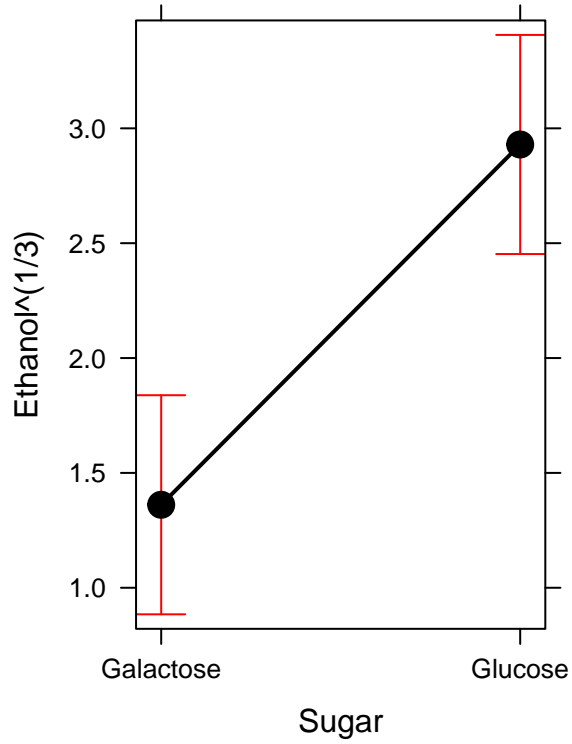


Effects plots for model 3.

```
ethanol.lm3<- lm(Ethanol^(1/3) ~ Sugar + O2Conc, data=data)
library(alr4)
```

```
plot(allEffects(ethanol.lm3))
```

**Sugar effect plot**



**O2Conc effect plot**

