

pset6

Peter

2023-10-20

```
library(Stat2Data)
library(mosaic)
```

```
## Registered S3 method overwritten by 'mosaic':
##   method                from
##   fortify.SpatialPolygonsDataFrame ggplot2

##
## The 'mosaic' package masks several functions from core packages in order to add
## additional features. The original behavior of these functions should not be affected by this.

##
## Attaching package: 'mosaic'

## The following objects are masked from 'package:dplyr':
##
##   count, do, tally

## The following object is masked from 'package:Matrix':
##
##   mean

## The following object is masked from 'package:ggplot2':
##
##   stat

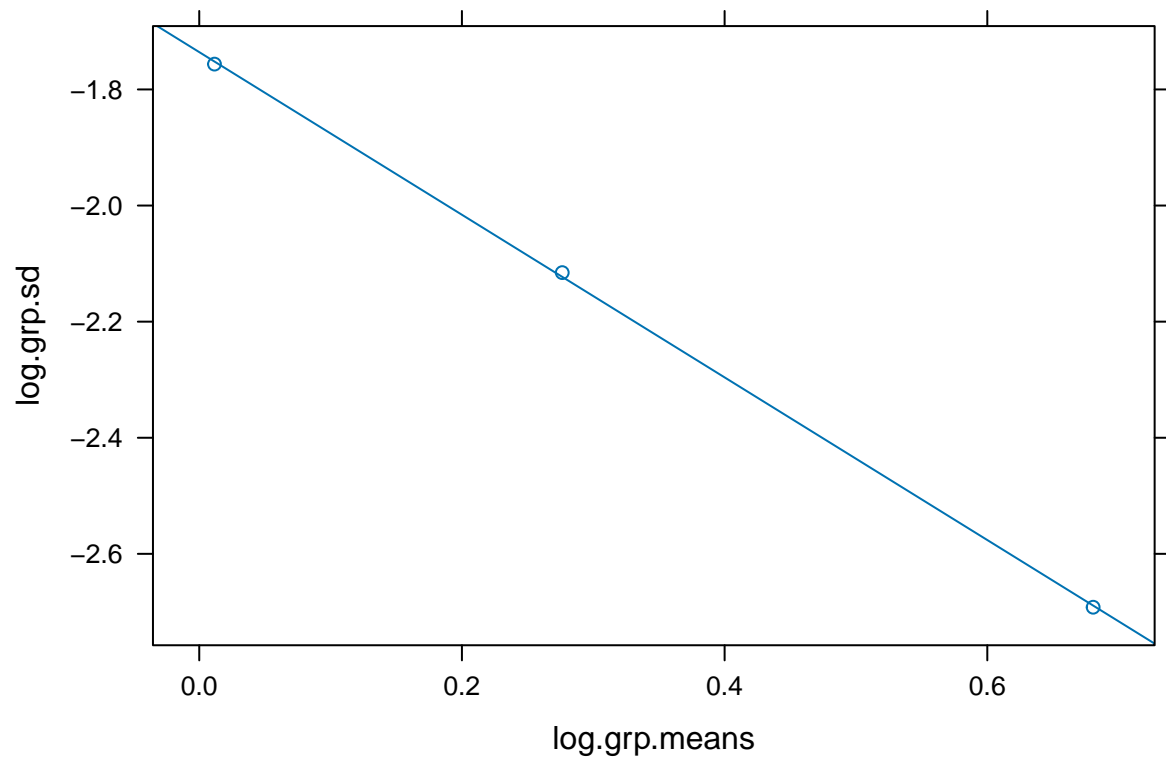
## The following objects are masked from 'package:stats':
##
##   binom.test, cor, cor.test, cov, fivenum, IQR, median, prop.test,
##   quantile, sd, t.test, var

## The following objects are masked from 'package:base':
##
##   max, mean, min, prod, range, sample, sum
```

```
#5.44c:
data("Olives")

log.grp.means = log(mean(Fenthion~Time, data = Olives))
```

```
log.grp.sd = log(sd(Fenthion~Time, data = Olives))
xyplot(log.grp.sd~log.grp.means, type=c("p","r"))
```



```
(trnsline = lm(log.grp.sd~log.grp.means))
```

```
##
## Call:
## lm(formula = log.grp.sd ~ log.grp.means)
##
## Coefficients:
## (Intercept) log.grp.means
##      -1.736      -1.401
```

```
#if 1 - slope =
#1 => no change to y
#.5 => sqrt y
#0 => log y
#-1 => 1/y
```

#The diagnostic plot suggests that we do not do any transformations on Y. $1 - -1.401 = 2.401$ which is g

```
#5.56:
```

```
data("Blood1")
```

#5.56a: Two sample t tests are used to determine if two population means are equal, in this case we have

```
#5.56b:
```

```
?Blood1 #Overwt 0=normal, 1=overweight, or 2=obese
```

```
BPcalc <- lm(SystolicBP~Overwt, data = Blood1) #factor method for more than one numeric  
anova(BPcalc)
```

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

```
##
```

```
## Response: SystolicBP
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
```

```
## Overwt      1  27788 27788.5   38.093 1.399e-09 ***
```

```
## Residuals 498 363286    729.5
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#With $f = 38.093$, $df = (1,498)$, and a $pval < 0.05$, we can conclude that at least one α_i is $\neq 0$.

#Yes, there is a significant difference in blood pressure based on weight.

```
#5.56c:
```

```
Overwt1 <- factor(Blood1$Overwt)
```

```
BPcalc1 <- aov(SystolicBP ~ Overwt1, data = Blood1)
```

```
library(asbio)
```

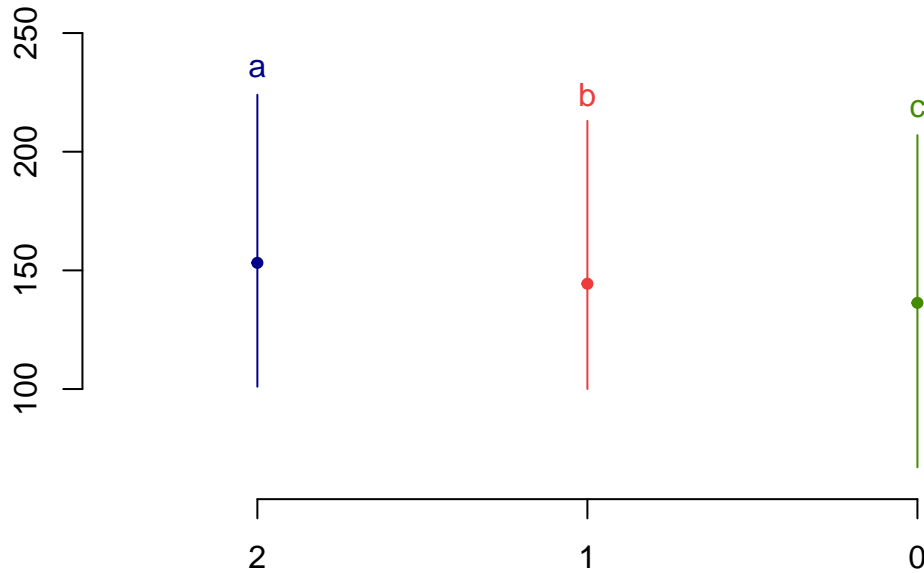
```
## Loading required package: tcltk
```

```
library(agricolae)
```

```
LSDout <- LSD.test(BPcalc1, "Overwt1", group = TRUE, p.adj = "none")
```

```
plot(LSDout)
```

Groups and Range



LSDout\$groups

```
## SystolicBP groups
## 2 153.1814 a
## 1 144.3670 b
## 0 136.3155 c
```

#6.2:

#6.2a: Response: Height of Tomato Plants

#6.2b: Factors: Type of Fertilizer (Factor 1), Amount of Water (Factor 2)

#6.2c: Factor 1: Experimental, 3 levels (Regular Fertilizer, Organic Fertilizer, No Fertilizer)

#Factor 2: Experimental, 2 levels (Normal Watering, Reduced Watering)

#6.2d: Complete Block Design: No

#6.10:

#a. Degrees of Freedom (DF) for Factor A:

#DF Factor A = $(3 - 1) = 2$

#b. Degrees of Freedom (DF) for Factor B:

#DF Factor B = $(3 - 1) = 2$

#c. Degrees of Freedom (DF) for Error:

#DF Error = $(1 - (2 + 2)) = (1 - 4) = -3$

#6.12:

*#If your dataset has one facotr of interest and one blocking factor, and you carry
#out a two-way ANOVA, your SSE will be too SMALL (small/large) and you will be MORE
#(more, less) likely to detect real differences than would a one-way ANOVA.*

*#6.29a: You could measure each rat twice, one time for each hormone injection (leptin and insulin), and
#like we did in AP Stats:
#Select male rats and female rats.
#Randomly assign each rat to one of two time slots.
#Time 1, inject hormones (e.g., leptin) in all rats.
#Time 2, inject hormones (e.g., insulin) in all rats.
#Measure amount eaten in Time 1 and Time 2 for each rat.*

#6.29b:

*#Reasons to not use blocks:
#takes much longer to do both treatments to all the rats, can't run simultaneously,
#and also uses more reasources.
#blocking minimizes variability, which may not be a good thing in this case.*

#6.36a:

```
data("Alfalfa")
```

```
favstats(Ht4~Row, data = Alfalfa)
```

```
##   Row min    Q1 median    Q3 max mean      sd n missing
## 1  a 1.00 1.015   1.03 1.240 1.45 1.16 0.2515949 3      0
## 2  b 0.70 0.960   1.22 2.005 2.79 1.57 1.0880717 3      0
## 3  c 0.45 0.910   1.37 1.650 1.93 1.25 0.7472617 3      0
## 4  d 1.65 1.990   2.33 2.565 2.80 2.26 0.5781868 3      0
## 5  e 1.07 1.265   1.46 3.155 4.85 2.46 2.0789661 3      0
```

#mean a: 1.16

#mean b: 1.57

#mean c: 1.25

#mean d: 2.26

#mean e: 2.46

```
favstats(Ht4 ~ 1, data = Alfalfa)
```

```
##   1 min    Q1 median    Q3 max mean      sd n missing
## 1 1 0.45 1.05   1.45 2.13 4.85 1.74 1.105396 15      0
```

#mean for all 15: 1.74

#SD for all 15: 1.105396

#6.36b:

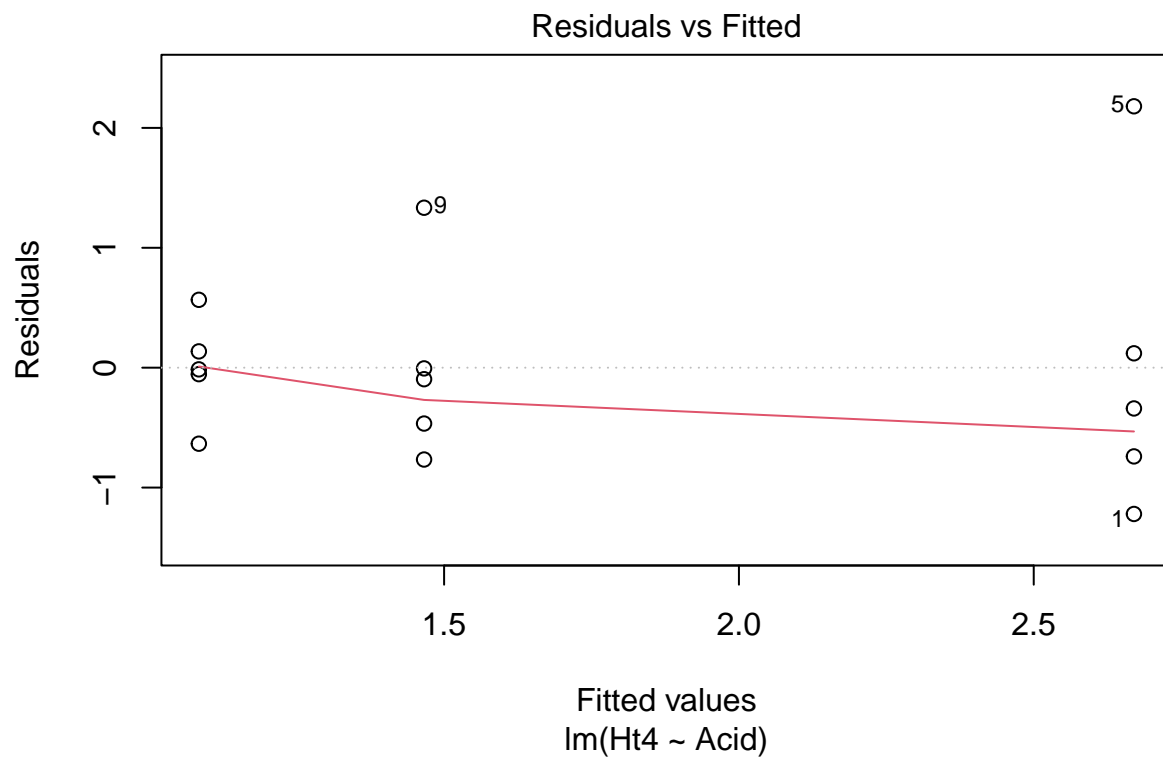
```
out = aov(Ht4~Acid + Row, data=Alfalfa)
anova(out)
```

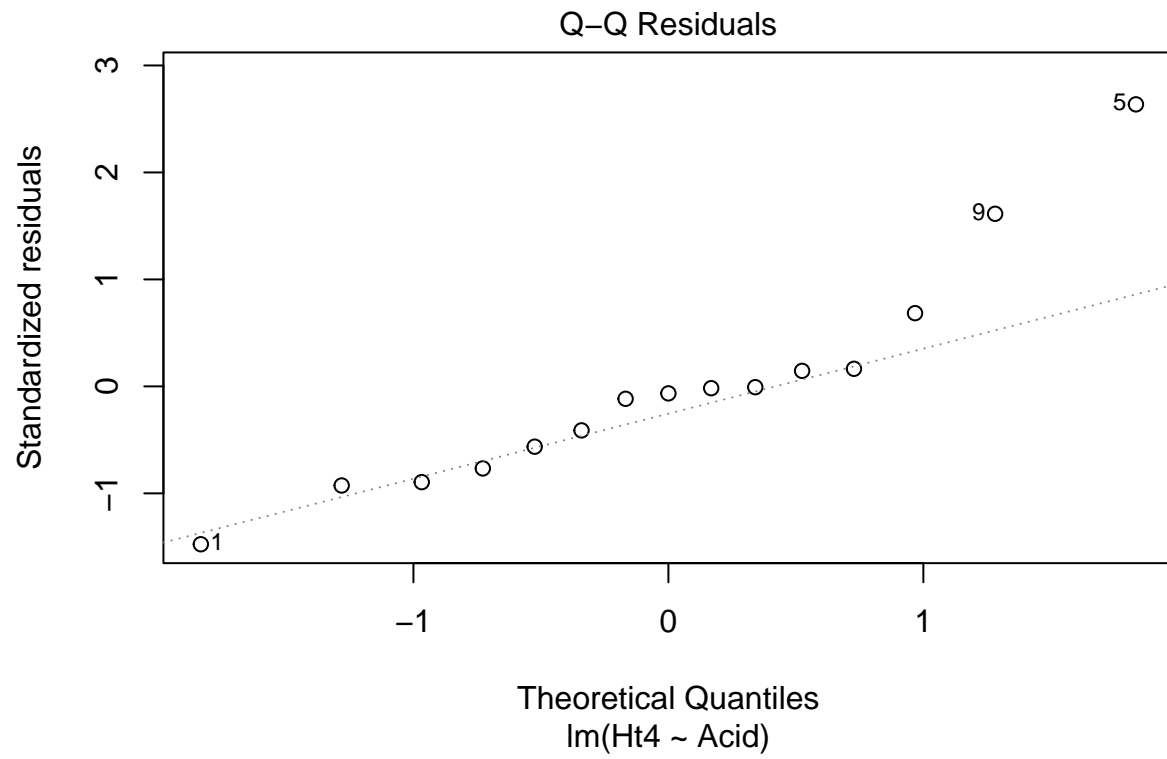
```
## Analysis of Variance Table
```

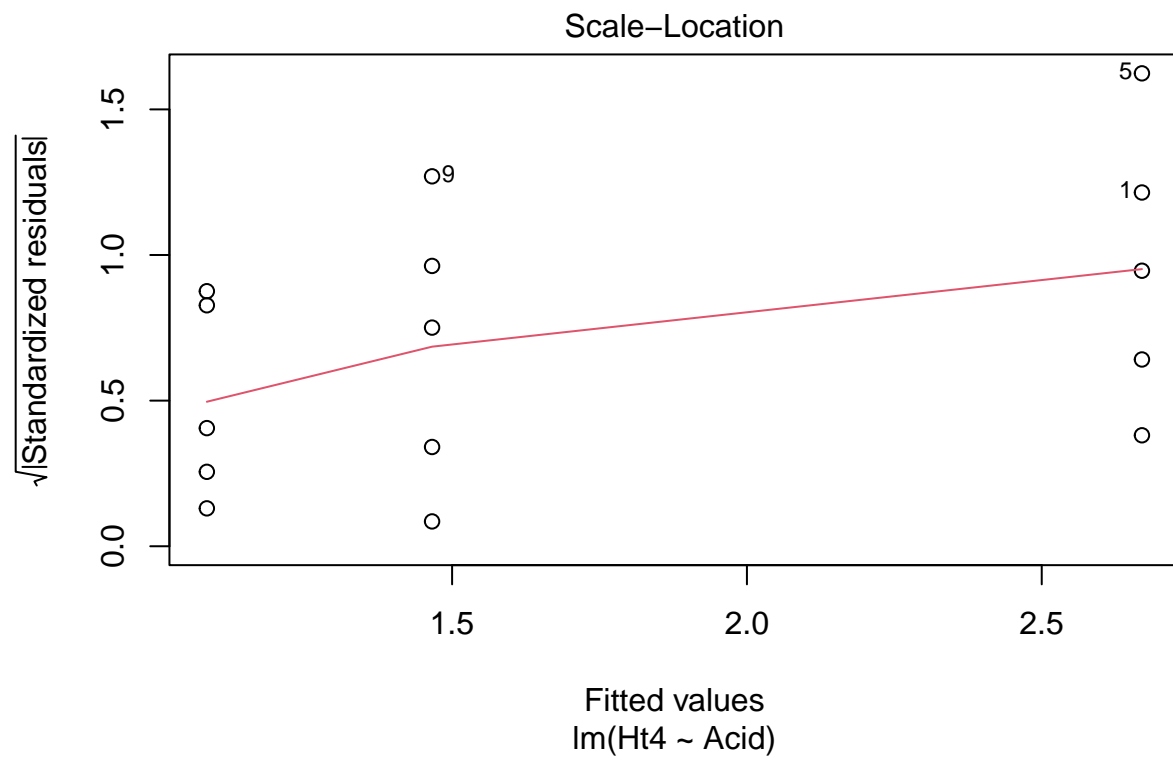
```
##
## Response: Ht4
##          Df Sum Sq Mean Sq F value    Pr(>F)
## Acid       2  6.8516   3.4258   4.5132 0.04874 *
## Row        4  4.1826   1.0456   1.3776 0.32352
## Residuals   8  6.0724   0.7591
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

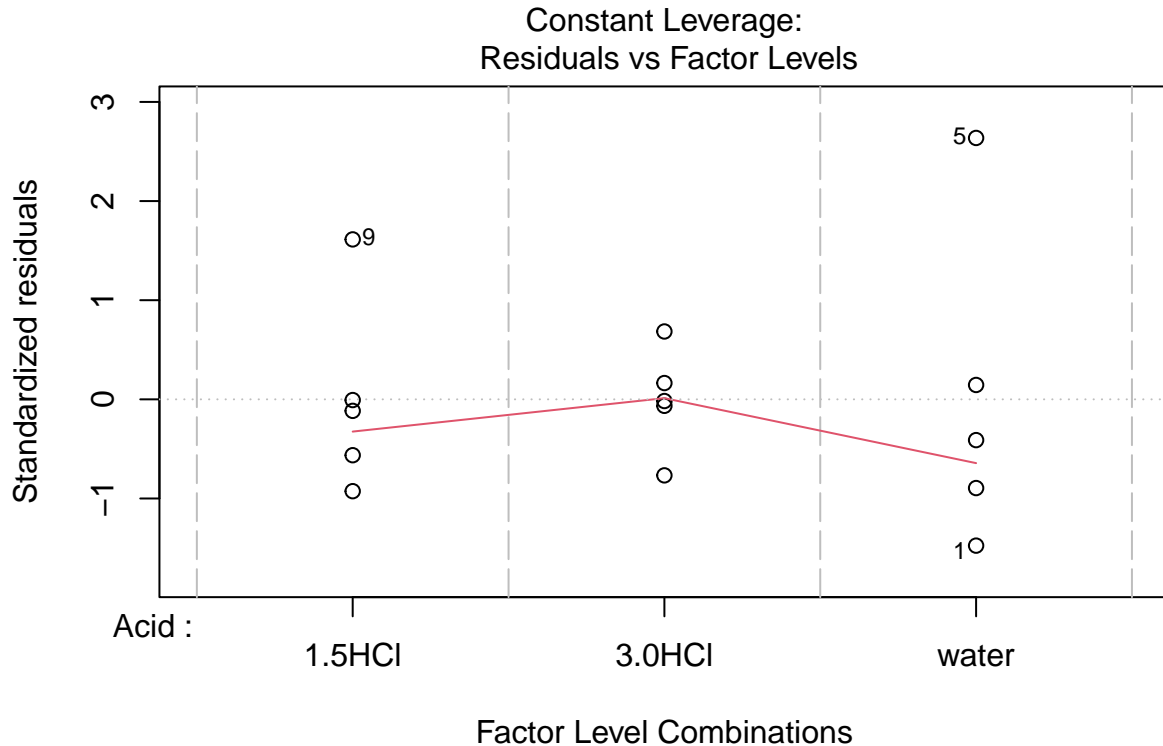
#6.36c:

```
time <- lm(Ht4~Acid, data = Alfalfa)
plot(time)
```









```
favstats(Ht4~Acid, data = Alfalfa)
```

##	Acid	min	Q1	median	Q3	max	mean	sd	n	missing
## 1	1.5HCl	0.70	1.00	1.37	1.46	2.80	1.466	0.8050963	5	0
## 2	3.0HCl	0.45	1.03	1.07	1.22	1.65	1.084	0.4312540	5	0
## 3	water	1.45	1.93	2.33	2.79	4.85	2.670	1.3151426	5	0

#Constant and additive: Yes the treatment are constant and additive based on the context.

#Errors zero-mean: According to the resids vs fits plot, yes the errors are 0 mean visually.

*#Errors have same variance: This condition is not met, our lowest SD (.4312) * 2 is #NOT greater than the largest SD (1.31). The errors do NOT have the same variance.*

#Errors are normal: With so few observations it's hard to tell, but there is a very #concerning curve upwards on the right hand side. points 9 and 5 are especially concerning.

#errors are independent: The errors are independent. The collection of the data was #done in a good environment and no contamination could be found.

#6.36d: With an $F=4.5132$, $df = (2, 8)$, $p \text{ val} < 0.05$, there is significant evidence to #conclude that at least one α_i is not 0, so there is a significant difference in #the average growth due to the treatments.

#6.36e: With an $F = 1.3776$, $df = (4,8)$, $p \text{ val } .32 > 0.05$, I can not conclude that at least one α_i is not equal to 0, so I can't conclude that there is not a significant difference in average growth due to distance from the window.

#6.37:

```
LSDalf <- aov(Ht4~Acid, data=Alfalfa)
```

```
library(asbio)
```

```
library(agricolae)
```

```
pairwise.t.test(Alfalfa$Ht4, Alfalfa$Acid, p.adj="none")
```

```
##
```

```
## Pairwise comparisons using t tests with pooled SD
```

```
##
```

```
## data: Alfalfa$Ht4 and Alfalfa$Acid
```

```
##
```

```
##      1.5HC1 3.0HC1
```

```
## 3.0HC1 0.526 -
```

```
## water 0.062 0.019
```

```
##
```

```
## P value adjustment method: none
```

```
(LSDout = pairw.anova(Alfalfa$Ht4,Alfalfa$Acid,method="lsd"))
```

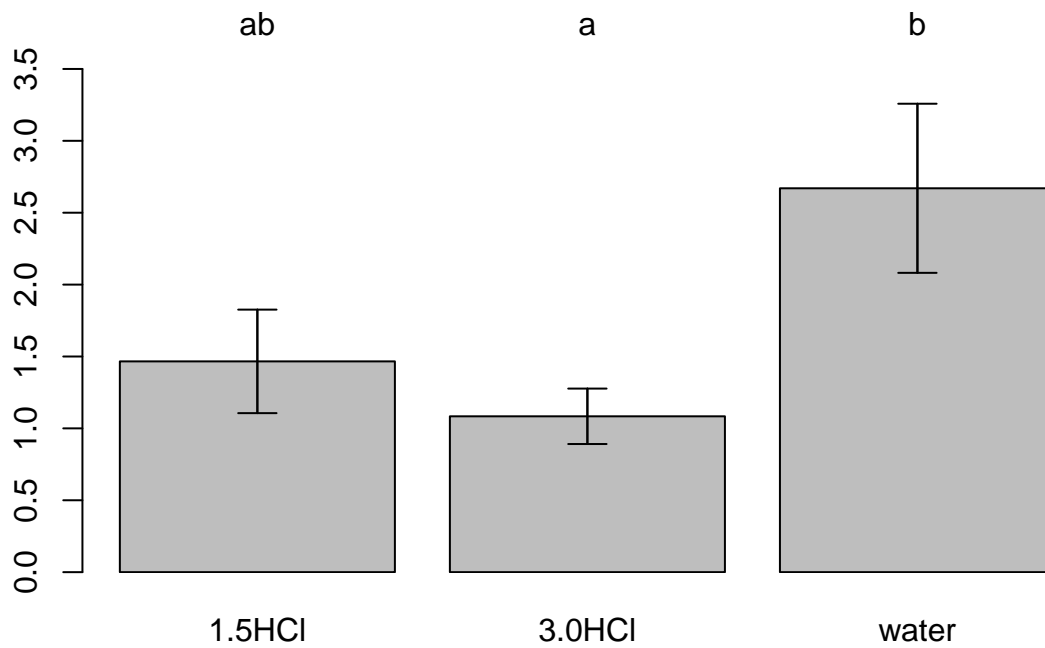
```
##
```

```
## 95% LSD confidence intervals
```

```
##
```

	LSD	Diff	Lower	Upper	Decision	Adj. p-value
## mu1.5HC1-mu3.0HC1	1.27388	0.382	-0.89188	1.65588	FTR H0	0.52584
## mu1.5HC1-muwater	1.27388	-1.204	-2.47788	0.06988	FTR H0	0.06185
## mu3.0HC1-muwater	1.27388	-1.586	-2.85988	-0.31212	Reject H0	0.01886

```
plot(LSDout)
```



```
##  
## Bars are means. Errors are SEs.  
##  
## The population means of factor levels with the same letter are not  
## significantly different at alpha = 0.05 using the Fisher LSD method.
```

```
plot(LSDout, type=2)
```

95% LSD confidence intervals

