

Script05: ANOVA stats

Sara Beier

2024-11-06

Contents

1	Clean and setup working space	2
2	Load and transform input data	2
3	ANOVA statistics Incubation Day 5	3
3.1	Two-way ANOVAS with Tukey Posthoc testing	3
3.2	Create Boxplot Graphics	5
4	Growth curves	6
4.1	Mixed model ANOVA with Posthoc tests for pairwise comparisons	6
4.2	Create Plot Growth Curves	8

1 Clean and setup working space

```
rm(list=ls())
loaded_packages <- setdiff(loadedNamespaces(),
                           c("base", "compiler", "datasets", "graphics", "grDevices",
                             "grid", "methods", "parallel", "splines", "stats",
                             "stats4", "tcltk", "tools", "utils"))
for (pkg in loaded_packages) {
  try(detach(paste0("package:", pkg), unload = TRUE, character.only = TRUE),
      silent = TRUE)
}

# Load packages
library(tidyverse)
library(car) #Levene's test
library(readxl) # read excel files (to read in Table S)
library(lme4) # mixed models via lmer
library(lmerTest) # mixed models via lmer + p-values
library(multcomp) # posthoc multiple comparisons after lmer
library(kableExtra)

#define functions
## Function for obtaining standard error, modified from
## (https://stackoverflow.com/questions/30485154/how-to-write-summaryse-function-in-r)
summarySE <- function(data=NULL, measurevar, groupvars=NULL, na.rm=FALSE,
                      conf.interval=.95, .drop=TRUE) {
  library(plyr)
  length2 <- function (x, na.rm=FALSE) {
    if (na.rm) sum(!is.na(x))
    else      length(x)
  }
  datac <- ddply(data, groupvars, .drop=.drop,
                .fun = function(xx, col) {
                  c(N      = length2(xx[[col]], na.rm=na.rm),
                    mean   = mean  (xx[[col]], na.rm=na.rm),
                    sd     = sd    (xx[[col]], na.rm=na.rm)
                  )
                },
                measurevar)

  datac <- rename(datac, c("mean" = measurevar))
  datac$se <- datac$sd / sqrt(datac$N)
  return(datac)
}
```

2 Load and transform input data

```
# Load data from Table S3
dat <- read_xlsx("../Data/Stables.xlsx",
                 sheet = "Table_S3", .name_repair = "universal") %>%
  mutate_at(c(5:10), as.numeric)
```

```
# Transform data to improve match of residual to normal distribution
dat_trans <- dat %>%
  mutate(Cellcounts=log(Cellcounts)) %>%
  mutate(Bacterial.production=Bacterial.production) %>%
  mutate(Respiration=Respiration) %>%
  mutate(ASV.richness=1/ASV.richness) %>%
  mutate(Shannon.entropy=log(Shannon.entropy)) %>%
  mutate(Pilou.s.evenness=Pilou.s.evenness)
```

3 ANOVA statistics Incubation Day 5

3.1 Two-way ANOVAS with Tukey Posthoc testing

Two way ANOVAS with posthoc testing for pairwise comparisons are performed after inspecting normal distribution of residuals and testing for homogeneity of variances (levene's test)

```
par(mfrow = c(3, 2))
for(i in seq(5,dim(dat_trans)[2])) {
  dat.i <- dat_trans[dat_trans$Time=='T5',c(1:4,i)] %>%
    na.omit()
  x <- dat.i[,5, drop=T]
  res.aov <- aov(x ~ Source.community*Medium, data = dat.i)
  qqnorm(resid(res.aov))
  legend("bottomright", c(names(dat)[i]),,bty="n")
}

stats.out <- data.frame()

for(i in seq(5,dim(dat_trans)[2])) {
  dat.i <- dat_trans[dat_trans$Time=='T5',c(1:4,i)] %>%
    na.omit()
  x <- dat.i[,5, drop=T]
  levene.i <- leveneTest(x ~ Source.community*Medium, data = dat.i)[[3]][[1]]

  res.aov <- aov(x ~ Source.community*Medium, data = dat.i)

  stats.i <- data.frame(names = c("C-S", "CS-C", "CS-S"),
    p.adj = c(TukeyHSD(res.aov)[[1]][[11]],
              TukeyHSD(res.aov)[[1]][[10]],
              TukeyHSD(res.aov)[[1]][[12]]),
    diff = c(TukeyHSD(res.aov)[[1]][[2]],
             TukeyHSD(res.aov)[[1]][[1]],
             TukeyHSD(res.aov)[[1]][[3]]),
    test= rep(c(names(dat)[i]),3),
    p.source = rep(summary (res.aov)[[1]][[5]][1],3),
    F.source = rep(summary (res.aov)[[1]][[4]][1],3),
    p.DOM = rep(summary (res.aov)[[1]][[5]][2],3),
    F.DOM = rep(summary (res.aov)[[1]][[4]][2],3),
    p.source.DOM = rep(summary (res.aov)[[1]][[5]][3],3),
    F.source.DOM = rep(summary (res.aov)[[1]][[4]][3],3),
    levene = rep(levene.i,3))

  stats.out <- rbind(stats.out, stats.i)
}
```

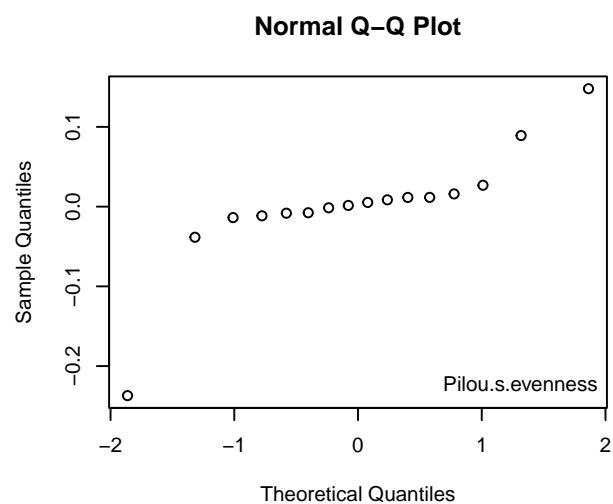
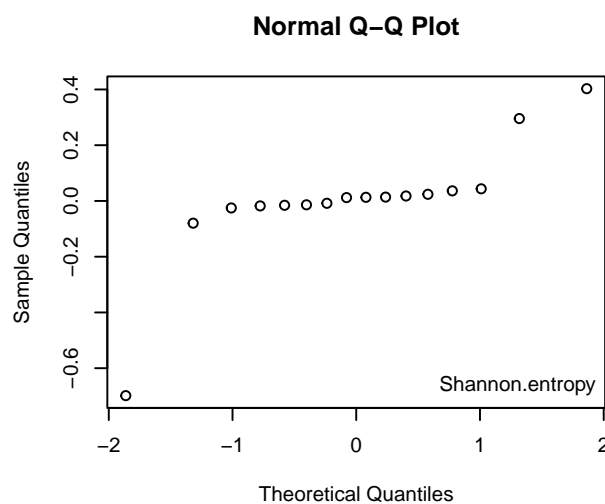
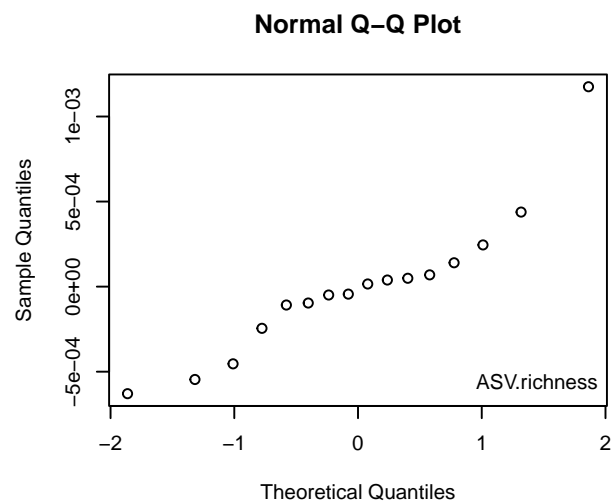
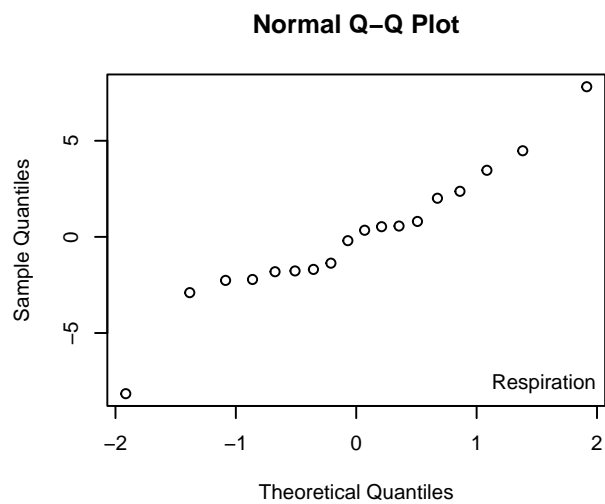
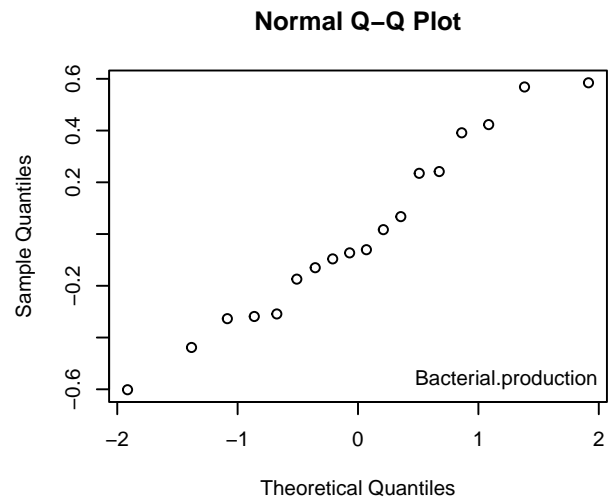
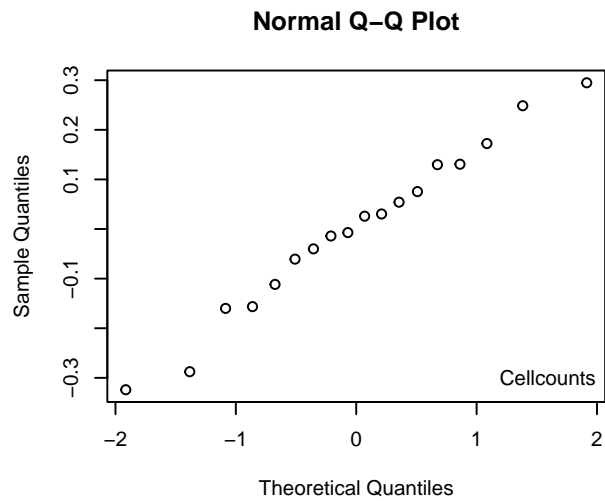


Figure 1: Inspect normality of residuals.

names	Tukey Posthoc		Two-way ANOVA							HOV
	p.adj	diff	test	p.source	F.source	p.DOM	F.DOM	p.source.DOM	F.source.DOM	levene
C-S	0.000	-0.861	Cellcounts	0.000	37.816	0.054	4.544	0.114	2.616	0.628
CS-C	0.992	0.014	Cellcounts	0.000	37.816	0.054	4.544	0.114	2.616	0.628
CS-S	0.000	-0.875	Cellcounts	0.000	37.816	0.054	4.544	0.114	2.616	0.628
C-S	0.015	-0.792	Bacterial.production	0.006	8.249	0.845	0.040	0.241	1.603	0.845
CS-C	0.933	0.085	Bacterial.production	0.006	8.249	0.845	0.040	0.241	1.603	0.845
CS-S	0.008	-0.876	Bacterial.production	0.006	8.249	0.845	0.040	0.241	1.603	0.845
C-S	0.957	-0.667	Respiration	0.010	6.854	0.417	0.708	0.756	0.287	0.489
CS-C	0.025	7.206	Respiration	0.010	6.854	0.417	0.708	0.756	0.287	0.489
CS-S	0.015	-7.873	Respiration	0.010	6.854	0.417	0.708	0.756	0.287	0.489
C-S	0.065	-0.001	ASV.richness	0.008	7.982	0.104	3.203	0.331	1.237	0.679
CS-C	0.008	-0.001	ASV.richness	0.008	7.982	0.104	3.203	0.331	1.237	0.679
CS-S	0.302	0.001	ASV.richness	0.008	7.982	0.104	3.203	0.331	1.237	0.679
C-S	0.179	-0.306	Shannon.entropy	0.149	2.318	0.655	0.212	0.779	0.256	0.478
CS-C	1.000	-0.005	Shannon.entropy	0.149	2.318	0.655	0.212	0.779	0.256	0.478
CS-S	0.250	-0.302	Shannon.entropy	0.149	2.318	0.655	0.212	0.779	0.256	0.478
C-S	0.043	-0.155	Pilou.s.evenness	0.044	4.340	0.960	0.003	0.909	0.096	0.417
CS-C	0.850	-0.033	Pilou.s.evenness	0.044	4.340	0.960	0.003	0.909	0.096	0.417
CS-S	0.166	-0.121	Pilou.s.evenness	0.044	4.340	0.960	0.003	0.909	0.096	0.417

```
kbl(stats.out,digits = 3) %>%
kable_classic() %>%
add_header_above(c(" " = 1, " Tukey Posthoc " = 2, " Two-way ANOVA " = 7, "HOV" =1))%>%
  kable_styling(latex_options="scale_down")
```

```
#print(format(stats.out,scientific = FALSE, digits = 6))
```

3.2 Create Boxplot Graphics

```
color=c('#7F95AA','#7F95AA','#7F95AA','darkolivegreen1','darkolivegreen1',
        'darkolivegreen1')
color2 = rep(c("black", "#CF5053", "black"),2)

par(mfrow=c(2,3))
par(mar=c(2, 2.2, 2, 0.8))
boxplot(Shannon.entropy ~ interaction(Source.community,Medium),
        data = dat[dat$Time=='T5',], col=color, xaxt="n")
title('A) Shannon entropy', adj=0, line=0.8)
legend('bottomleft', c('S-DOM', 'SW-DOM'), col = c('#7F95AA','darkolivegreen1'),
       pch=c(15,15), bty='n', pt.cex=2)
boxplot(Pilou.s.evenness ~ interaction(Source.community,Medium),
        data = dat[dat$Time=='T5',], col=color, xaxt="n")
title('B) Pielou\'s evenness', adj=0, line=0.8)
boxplot(ASV.richness ~ interaction(Source.community,Medium),
        data = dat[dat$Time=='T5',], col=color, xaxt="n")
title('C) ASV richness', adj=0, line=0.8)

boxplot(Cellcounts/1000000 ~ interaction(Source.community,Medium),
        data = dat[dat$Time=='T5',], col=color, xaxt="n")
axis(1, at = c(1:6), labels = FALSE)
mtext(c("C", "H", "S", "C", "H", "S"), side = 1, line =1, at = pretty(1:6),
      col=color2, cex=0.8)
title('D) Cell densities', adj=0, line=0.8)
```

```

boxplot(Bacterial.production ~ interaction(Source.community,Medium),
        data = dat[dat$Time=='T5',], col=color, xaxt="n")
axis(1, at = c(1:6), labels = FALSE)
mtext(c("C", "H", "S", "C", "H", "S"), side = 1, line = 1, at = pretty(1:6),
      col=color2, cex=0.8)
title('E) Productivity', adj=0, line=0.8)
boxplot(Respiration ~ interaction(Source.community,Medium),
        data = dat[dat$Time=='T5',], col=color, xaxt="n")
axis(1, at = c(1:6), labels = FALSE)
mtext(c("C", "H", "S", "C", "H", "S"), side = 1, line = 1, at = pretty(1:6),
      col=color2, cex=0.8)
title('F) Respiration', adj=0, line=0.8)

```

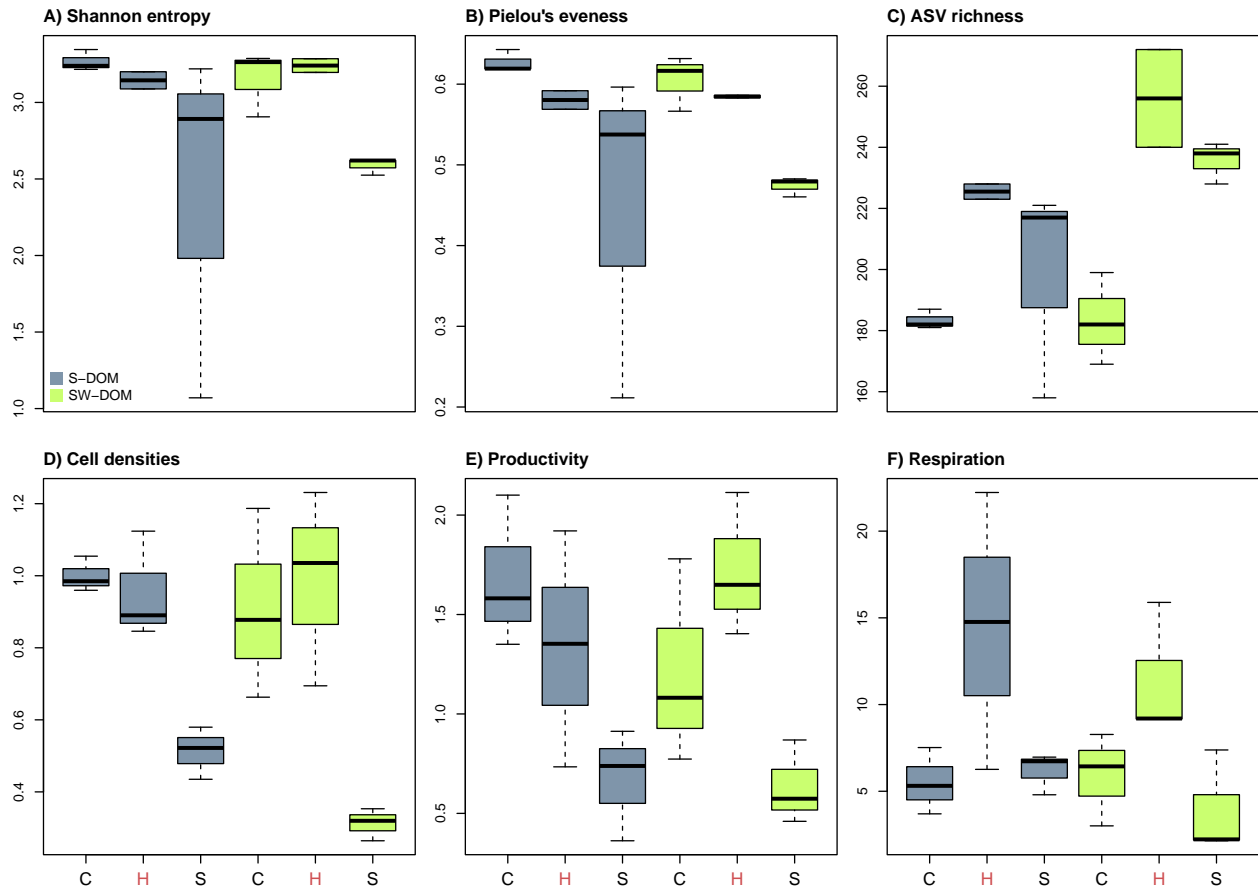


Figure 2: Boxplot figures illustration diversity parameters (A-C) and functional characteristics (D-F) of experimental communities at day 5. C, H, and S indicate Canet control, hybrid and SOLA control communities.

4 Growth curves

4.1 Mixed model ANOVA with Posthoc tests for pairwise comparisons

```

m <- lme4::lmer(log(Cellcounts) ~ Source.community*Medium + (1|Time), data = dat)
qqnorm(resid(m))

```

Normal Q-Q Plot

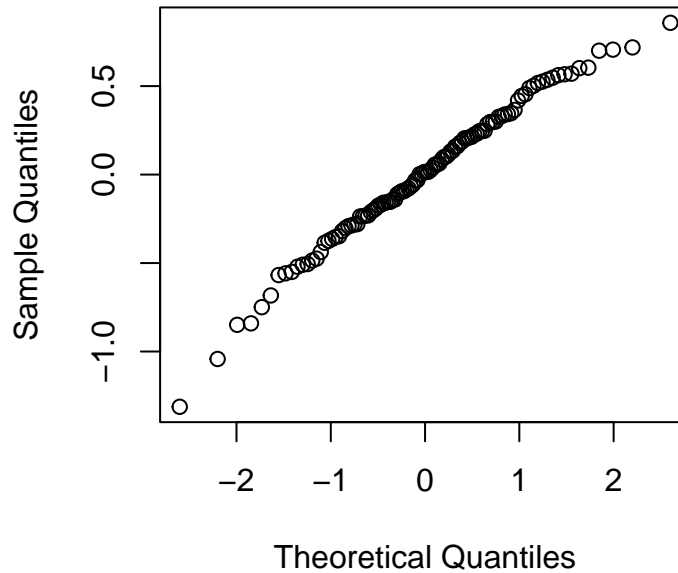


Figure 3: Inspect normality of residuals

```
mt <- lmerTest::lmer(log(Cellcounts) ~ Source.community*Medium + (1|Time), data = dat)
anova(mt)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##               Sum Sq Mean Sq NumDF DenDF  F value Pr(>F)
## Source.community    51.096  25.5481     2    97  143.5841 <2e-16 ***
## Medium              0.233   0.2332     1    97   1.3105 0.2551
## Source.community:Medium 0.225   0.1127     2    97   0.6333 0.5330
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
summary(multcomp::glht(m, linfct=mcp(Source.community="Tukey")))
```

```
##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lme4::lmer(formula = log(Cellcounts) ~ Source.community * Medium +
## (1 | Time), data = dat)
##
## Linear Hypotheses:
##               Estimate Std. Error z value Pr(>|z|)
## CS - C == 0  -0.3466    0.1406  -2.465   0.0365 *
## S - C == 0   -1.5848    0.1406 -11.271  <0.001 ***
## S - CS == 0  -1.2382    0.1406  -8.806  <0.001 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

4.2 Create Plot Growth Curves

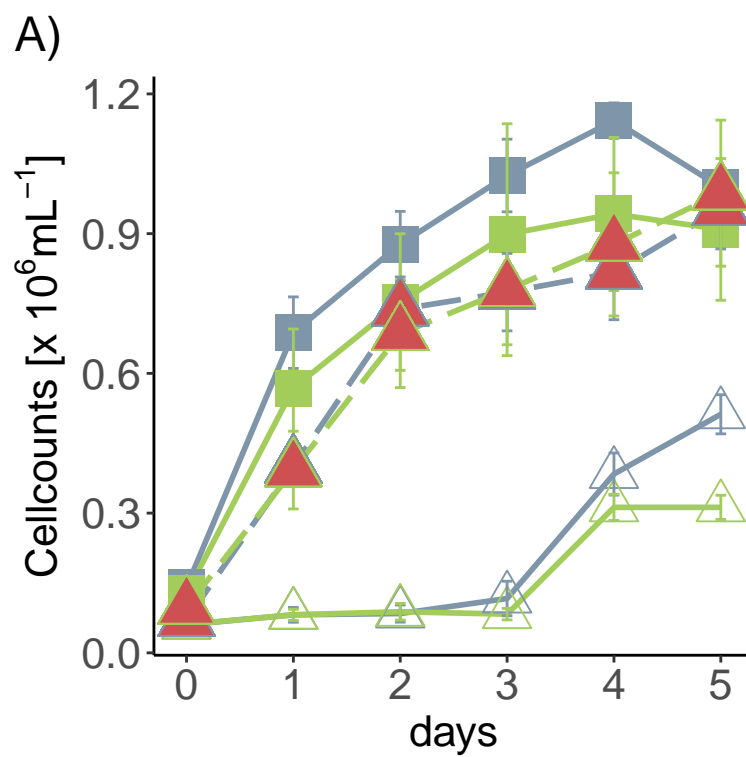


Figure 4: Growth curves of resuscitated communities (error bars indicate standard errors of triplicate incubations).