

Hatch Rate Analysis 2016

June 19, 2017

Setup

```
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 3.3.2
```

```
Sys.setenv(RSTUDIO_PDFLATEX = "/Library/TeX/texbin/pdflatex")
```

Function that allows line wrapping of long code

```
hook_output = knitr_hooks$get('output')
knitr_hooks$set(output = function(x, options) {
  # this hook is used only when the linewidth option is not NULL
  if (!is.null(n <- options$linewidth)) {
    x = knitr::split_lines(x)
    # any lines wider than n should be wrapped
    if (any(nchar(x) > n)) x = strwrap(x, width = n)
    x = paste(x, collapse = '\n')
  }
  hook_output(x, options)
})
```

Set up Environment:

```
library('reshape2')
library('scales')
```

```
## Warning: package 'scales' was built under R version 3.3.2
```

```
library('stringr')
```

```
## Warning: package 'stringr' was built under R version 3.3.2
```

```
library('ggplot2')
```

```
## Warning: package 'ggplot2' was built under R version 3.3.2
```

```
setwd('/Users/Tessa/Dropbox/squid/squid_R_work/squid_hatching/hatching_rmd')
source('helpful_summary_functions.R') #load helper functions
hr <- read.csv('hatch_rate_transposed_updated.csv', header=T)
hr$Date <- as.Date(hr$Date, "%m/%d/%y") # format Date properly
```

Cumulative hatching over time:

Calculate cumulative hatching over time, per capsule ('2' = by column):

```
cumSumHR <- apply(hr[, -which(names(hr) == "Date")], 2, cumsum) #don't sum date column
```

Join to original dataframe so date column can correspond

```
#add 'cumul' to column names to diff from orig count info
colnames(cumSumHR) <- paste(colnames(cumSumHR), 'cumul', sep = "_")
#join DF's
fullHR_DF <- cbind(hr, cumSumHR)
```

```
#could add dates here if we want
#cumSumHR_wDates <- fullHR_DF[, grepl("cumul", colnames(fullHR_DF))]
#cumSumHR_wDates$Date <- fullHR_DF$Date
```

Melt to long-form; add classifying variables

```
meltedFullDF <- melt(fullHR_DF, id='Date')
meltedFullDF$Tank <- ""
meltedFullDF[grepl("1", meltedFullDF$variable),]$Tank <- 'Tank 1'
meltedFullDF[grepl("2", meltedFullDF$variable),]$Tank <- 'Tank 2'
meltedFullDF[grepl("3", meltedFullDF$variable),]$Tank <- 'Tank 3'
meltedFullDF[grepl("4", meltedFullDF$variable),]$Tank <- 'Tank 4'
meltedFullDF$Condition <- ""
meltedFullDF[grepl("1|2", meltedFullDF$variable),]$Condition <- 'pH 7.57, 97 uM/kg O2'
meltedFullDF[grepl("3|4", meltedFullDF$variable),]$Condition <- 'pH 7.89, 227 uM/kg O2'
```

Split cumulative counts; original counts -> two DF's

```
melted_CumOnly <- meltedFullDF[grepl("cumul", meltedFullDF$variable),]
melted_countsOnly <- meltedFullDF[!grepl("cumul", meltedFullDF$variable),]
```

Summarize cumulative count data; get Date in appropriate form for graphing

```
summCond_Cumul <- summarySE(melted_CumOnly, measurevar="value", groupvars=c("Condition", "Date"))
```

Loading required package: plyr

```
summCond_Cumul$Date <- as.Date(summCond_Cumul$Date)
```

Read in unhatched count data; test for diffs btwn treatments

```
unhatched_embs <- read.csv('unhatched_embryos.csv', header=T)
```

```
#unhatchMelt <- melt(unhatched_embs, id="Date")
#unhatchMelt$Tank <- ""
#unhatchMelt[grepl("1", unhatchMelt$variable),]$Tank <- 'Tank 1'
#unhatchMelt[grepl("2", unhatchMelt$variable),]$Tank <- 'Tank 2'
#unhatchMelt[grepl("3", unhatchMelt$variable),]$Tank <- 'Tank 3'
#unhatchMelt[grepl("4", unhatchMelt$variable),]$Tank <- 'Tank 4'
#unhatchMelt$Condition <- ""
#unhatchMelt[grepl("1|2", unhatchMelt$variable),]$Condition <- 'pH 7.57, 97 uM/kg O2'
#unhatchMelt[grepl("3|4", unhatchMelt$variable),]$Condition <- 'pH 7.89, 227 uM/kg O2'

#unhatchSum_cond <- summarySE(unhatchMelt, measurevar="value", groupvars=c("Condition"))
#unhatchMelt$Condition <- as.factor(unhatchMelt$Condition)

#Get total embs per capsule
totalHatched = colSums(hr[, -1])
totalUnhatched =unhatched_embs[, -1]
embs_per_cap = totalHatched + totalUnhatched
```

Is %hatch diff btwn treatments?

```
percentHatch = (totalUnhatched/embs_per_cap)*100
meltedPhatch = melt(percentHatch)
```

No id variables; using all as measure variables

```

meltedPhatch$Tank <- ""
meltedPhatch[grepl("1", meltedPhatch$variable),]$Tank <- 'Tank 1'
meltedPhatch[grepl("2", meltedPhatch$variable),]$Tank <- 'Tank 2'
meltedPhatch[grepl("3", meltedPhatch$variable),]$Tank <- 'Tank 3'
meltedPhatch[grepl("4", meltedPhatch$variable),]$Tank <- 'Tank 4'
meltedPhatch$Condition <- ""
meltedPhatch[grepl("1|2", meltedPhatch$variable),]$Condition <- 'pH 7.57, 97 uM/kg O2'
meltedPhatch[grepl("3|4", meltedPhatch$variable),]$Condition <- 'pH 7.89, 227 uM/kg O2'
meltedPhatch$percentHatched <- 100 - meltedPhatch$value
pHatched_treat =meltedPhatch[grepl("1|2", meltedPhatch$variable),]$percentHatched
pHatched_ctrl =meltedPhatch[grepl("3|4", meltedPhatch$variable),]$percentHatched
t.test(pHatched_ctrl, pHatched_treat, alternative="greater")

##
## Welch Two Sample t-test
##
## data: pHatched_ctrl and pHatched_treat
## t = 3.5011, df = 7.0533, p-value = 0.004929
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## 2.009755      Inf
## sample estimates:
## mean of x mean of y
## 99.88299 95.50893

```

Plot percent hatching over time

first, get total embs per cap into the cumulative dataframe

```

#cap name --> column
melted_CumOnly$Capsule <- str_match(melted_CumOnly$variable, '(X\\d\\S)_cumul')[,2]
#get # per capsule into right form
embs_per_capMelt <- melt(embs_per_cap)

```

No id variables; using all as measure variables

```

names(embs_per_capMelt) <- c('Capsule', 'embTotal')
#merge the two dataframes by intersecting the capsule name columns
melted_CumOnly_wTotals = merge(melted_CumOnly, embs_per_capMelt, by = intersect("Capsule", "Capsule"))
melted_CumOnly_wTotals$percentTotalHatched = melted_CumOnly_wTotals$value/melted_CumOnly_wTotals$embTot

```

Get summary of % Hatching

```

summCond_PercentCumul <- summarySE(melted_CumOnly_wTotals, measurevar="percentTotalHatched", groupvar="Date")
summCond_PercentCumul$Date <- as.Date(summCond_PercentCumul$Date)

```

Include incubation time; x-axis -> days, instead of 'Date'

```

# add in inc time:
inc_dates <- seq.Date(as.Date("2013/06/02"), as.Date("2013/07/02"), by = "days")
zero_hatching <- rep(0, length(inc_dates))
zero_hatch_df <- as.data.frame(matrix(zero_hatching, nrow = length(zero_hatching),
ncol = 16))

```

```

inc_only <- cbind(inc_dates, zero_hatch_df)
colnames(inc_only) <- colnames(hr)
hr_with_incubation <- merge(hr, inc_only, on = "Date", all = T)
hr_with_incubation$days = hr_with_incubation$Date - as.Date("2013/06/02")

cumSumHR_w_inc <- apply(hr_with_incubation[, -which(names(hr_with_incubation) ==
  c("Date", "days"))], 2, cumsum)
colnames(cumSumHR_w_inc) <- paste(colnames(cumSumHR_w_inc), "cumul", sep = "_")

fullHR_DF_w_inc <- cbind(hr_with_incubation, cumSumHR_w_inc)
meltedFullDF_w_inc <- melt(fullHR_DF_w_inc, id = c("Date", "days"))
meltedFullDF_w_inc$Tank <- ""
meltedFullDF_w_inc[grepl("1", meltedFullDF_w_inc$variable), ]$Tank <- "Tank 1"
meltedFullDF_w_inc[grepl("2", meltedFullDF_w_inc$variable), ]$Tank <- "Tank 2"
meltedFullDF_w_inc[grepl("3", meltedFullDF_w_inc$variable), ]$Tank <- "Tank 3"
meltedFullDF_w_inc[grepl("4", meltedFullDF_w_inc$variable), ]$Tank <- "Tank 4"
meltedFullDF_w_inc$Condition <- ""
meltedFullDF_w_inc[grepl("1|2", meltedFullDF_w_inc$variable), ]$Condition <- "pH 7.57, 97 uM/kg 02"
meltedFullDF_w_inc[grepl("3|4", meltedFullDF_w_inc$variable), ]$Condition <- "pH 7.89, 227 uM/kg 02"

melted_CumOnly_w_inc <- meltedFullDF_w_inc[grepl("cumul", meltedFullDF_w_inc$variable),
]
melted_countsOnly_w_inc <- meltedFullDF_w_inc[!grepl("cumul", meltedFullDF_w_inc$variable),
]

melted_CumOnly_w_inc$Capsule <- str_match(melted_CumOnly_w_inc$variable, "(X\\d\\S)_cumul")[,
2]
melted_CumOnly_w_incT = merge(melted_CumOnly_w_inc, embs_per_capMelt, by = intersect("Capsule",
"Capsule"))
melted_CumOnly_w_incT$percentTotalHatched = melted_CumOnly_w_incT$value/melted_CumOnly_w_incT$embTotal

summCond_PercentCumul_w_incT <- summarySE(melted_CumOnly_w_incT, measurevar = "percentTotalHatched",
groupvars = c("Condition", "days"))

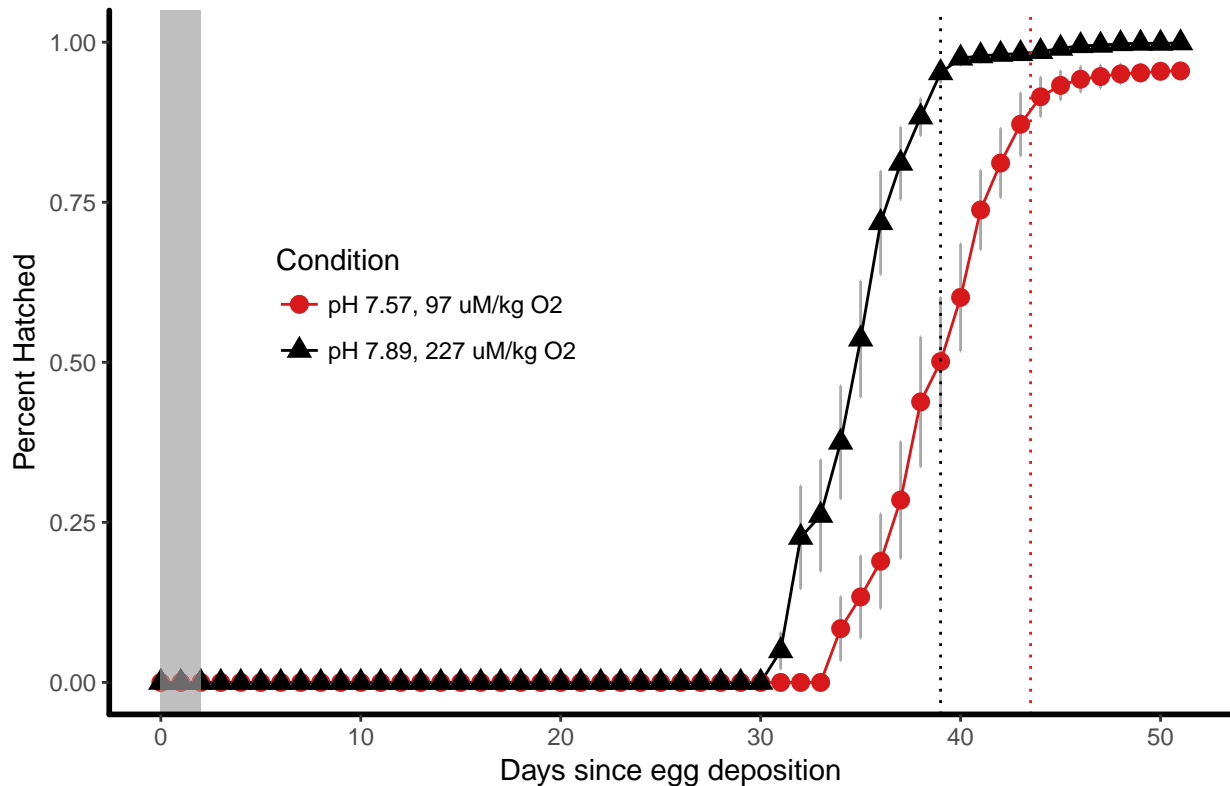
# pdf(file='percent_hatch_by_condition_w100line_w_incTime_brcolor.pdf',width=7,height=4.5)
# pdf(file='percent_hatch_by_condition_w90phatchlines_wincTime_brcolor.pdf',width=7,height=4.5)
# pdf(file='percent_hatch_by_condition_w90phatchlines_wincTime_brcolor.pdf',width=7,height=4.5)
# pdf(file='percent_hatch_by_condition_w90phatchlines_wincTime_color_notreatbox_realtime.pdf',width=7,height=4.5)
ggplot(summCond_PercentCumul_w_incT, aes(x = days, y = percentTotalHatched,
  shape = Condition, color = Condition)) + geom_errorbar(aes(ymin = percentTotalHatched -
se, ymax = percentTotalHatched + se), color = "darkgrey", width = 0.05) +
geom_line() + geom_point(size = 3) + scale_color_manual(values = c("#d7191c",
"black")) + xlab("Days since egg deposition") + ylab("Percent Hatched") +
ggtitle(expression(paste("Cumulative Percent Hatching of ", italic("Doryteuthis opalescens"),
" Embryos")) + theme_bw() + theme(plot.background = element_blank(),
panel.grid.major = element_blank(), panel.grid.minor = element_blank()) +
theme(panel.border = element_blank()) + theme(legend.key = element_blank(),
legend.justification = c(0.05, 0.85), legend.position = c(0.15, 0.65)) +
geom_rect(xmin = 0, xmax = 2, ymin = -Inf, ymax = +Inf, fill = "gray75",
  alpha = 0.01, inherit.aes = FALSE) + geom_vline(xintercept = 39, linetype = "dotted") +
geom_vline(xintercept = 43.5, linetype = "dotted", color = "#e41a1c") +

```

```
theme(axis.line.x = element_line(color = "black", size = 1), axis.line.y = element_line(color = "black", size = 1))
```

```
## Don't know how to automatically pick scale for object of type difftime. Defaulting to continuous.
```

Cumulative Percent Hatching of *Doryteuthis opalescens* Embryos



```
# dev.off()

# for 90% of each total Hatched () 90% of 99.8% and 90% of 94.5%: add
# horizontal line at 90 or 100 + geom_hline(yintercept=.9,
# linetype='dotted') doesn't look great. instead, let's try adding vertical
# lines at incubation time = 90% hatch (39d, 43.5d) +
# geom_vline(xintercept=39, linetype='dotted') + geom_vline(xintercept=43.5,
# linetype='dotted')

# try adding shaded rect for the time we didn't have embryos:
#+ geom_rect(xmin=0, xmax=3, ymin=-Inf, ymax=+Inf), fill='grey', alpha=0.2)

melted_CumOnly_w_incT[grepl("1|2", melted_CumOnly_w_incT$variable), ]$Condition <- "pH0x"
melted_CumOnly_w_incT[grepl("3|4", melted_CumOnly_w_incT$variable), ]$Condition <- "control"

summTank_PercentCumul_w_incT <- summarySE(melted_CumOnly_w_incT, measurevar = "percentTotalHatched",
groupvars = c("Tank", "days"))
```

Calculate # for 90% hatching (see Kaplan et al)

```
hatch90percent = data.frame((.9*totalHatched))
colnames(hatch90percent) <- c('hatch90percent')
hatch90percent$Capsule <- rownames(hatch90percent)
```

Is incubation time diff btwn treatments?

```
melted_CumOnly_w90p = merge(melted_CumOnly,hatch90percent, by = intersect("Capsule", "Capsule"))
over90phatched <- melted_CumOnly_w90p[melted_CumOnly_w90p$value >= melted_CumOnly_w90p$hatch90percent, ]
```

```
incubation90<- ddply(over90phatched, .(Capsule), summarize, end_incubation=min(Date))
incubation90$inc_start <- as.Date('2013-06-01')
incubation90$days <- incubation90$end_incubation - incubation90$inc_start
#add metadata
incubation90$Tank <- ""
incubation90[grepl("1", incubation90$Capsule),]$Tank <- 'Tank 1'
incubation90[grepl("2", incubation90$Capsule),]$Tank <- 'Tank 2'
incubation90[grepl("3", incubation90$Capsule),]$Tank <- 'Tank 3'
incubation90[grepl("4", incubation90$Capsule),]$Tank <- 'Tank 4'
incubation90$Condition <- ""
incubation90[grepl("1|2", incubation90$Capsule),]$Condition <- 'pH 7.57, 97 uM/kg 02'
incubation90[grepl("3|4", incubation90$Capsule),]$Condition <- 'pH 7.89, 227 uM/kg 02'

#t-test
inc90_treat =incubation90[grepl("1|2", incubation90$Capsule),]$days
inc90_ctrl =incubation90[grepl("3|4", incubation90$Capsule),]$days
t.test(inc90_ctrl, inc90_treat, alternative="less")
```

```
##
## Welch Two Sample t-test
##
## data: inc90_ctrl and inc90_treat
## t = -4.5826 days, df = 12.007, p-value = 0.0003143 days
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -2.749914
## sample estimates:
## Time differences in days
## mean of x mean of y
##      39.0      43.5
```

SE of mean of incubation90 values

```
se <- function(x) sqrt(var(x)/length(x))
se_inc90_ctrl <- se(inc90_ctrl)
se_inc90_treat <- se(inc90_treat)
mean_inc90_ctrl <- mean(inc90_ctrl)
mean_inc90_treat <- mean(inc90_treat)
```

Duration of hatching (reach 90% hatching as incubation time)

```
incubation90$hatch_start <- as.Date('2013-07-03')
incubation90[grepl("1|2", incubation90$Capsule),]$hatch_start <- as.Date('2013-07-06')
incubation90$hatching_duration <- incubation90$end_incubation - incubation90$hatch_start
```

```

hdur90_treat =incubation90[grepl("1|2", incubation90$Capsule),]$hatching_duration
hdur90_ctrl =incubation90[grepl("3|4", incubation90$Capsule),]$hatching_duration
t.test(hdur90_ctrl, hdur90_treat, alternative="less")

```

```

##
## Welch Two Sample t-test
##
## data: hdur90_ctrl and hdur90_treat
## t = -1.5275 days, df = 12.007, p-value = 0.07627 days
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf 0.2500859
## sample estimates:
## Time differences in days
## mean of x mean of y
##      7.0      8.5

```

Chi-Sq to compare two distributions:

<http://thebiobucket.blogspot.com/2011/08/comparing-two-distributions.html#more>

As in Kaplan et al, test if the distributions of hatching over time is the same for all treatments. Use COUNT data:

```

hr_for_ch <- hr
hr_for_ch$lowpH0x <- rowSums(hr_for_ch[, grepl('1|2', colnames(hr_for_ch))])
hr_for_ch$highpH0x <- rowSums(hr_for_ch[, grepl('3|4', colnames(hr_for_ch))])

chtable <- hr_for_ch[, grepl('pH0x', colnames(hr_for_ch))]
#chisq.test(chtable)

chisq.test(t(chtable)) # treatments need to be rows

```

```

## Warning in chisq.test(t(chtable)): Chi-squared approximation may be
## incorrect

```

```

##
## Pearson's Chi-squared test
##
## data: t(chtable)
## X-squared = 1192.1, df = 20, p-value < 2.2e-16

```

Check temperature ~ percent hatching:

```

pHatched_13.5 =meltedPhatch[grepl("1|3", meltedPhatch$variable),]$percentHatched
pHatched_12 =meltedPhatch[grepl("2|4", meltedPhatch$variable),]$percentHatched
t.test(pHatched_12, pHatched_13.5, alternative = 'less')

```

```

##
## Welch Two Sample t-test
##
## data: pHatched_12 and pHatched_13.5
## t = -0.41138, df = 12.394, p-value = 0.3439
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:

```

```
##      -Inf 2.323544
## sample estimates:
## mean of x mean of y
##  97.34613  98.04579
""
```