

# Class 4 - Statistics part II

## Activity 4

### a) Hypothesis tests for 3 or more populations:

load feeders.RData that you saved in Class2/ folder and repeat the selection indicated in the following box

```
load("../PdABiostatistics/Class2/feeders.RData")
```

```
feeders_select <- feeders[feeders$SITE %in% c("Glide","Rita") & feeders$DurGROWTH == "7", ]
```

- What is the size of the pH and OD\_avg samples for each treatment?

```
# to do this we can write the code to select only the values that we want,  
# and apply length().  
# however, since we have NAs we need to ignore them
```

```
# To count the number of cells from each condition containing non-NA values  
# you can use function length() and inside use function na.omit()
```

```
length( na.omit( feeders_select[feeders_select$TREATMENT == "net","OD_avg"]) )
```

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

```
length( na.omit( feeders_select[feeders_select$TREATMENT == "open","OD_avg"]) )
```

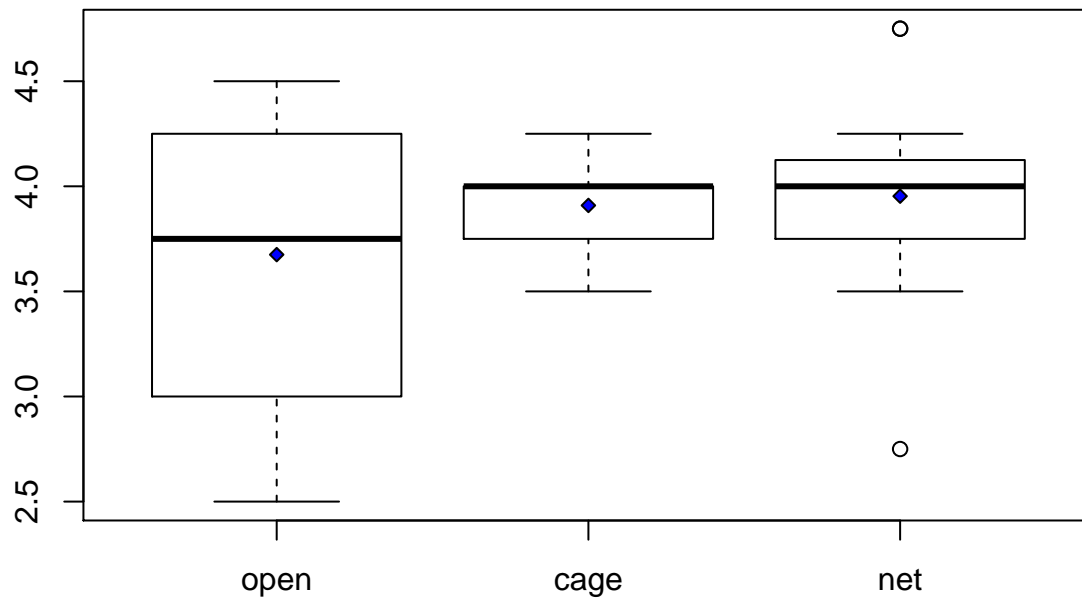
```
## [1] 10
```

```
length( na.omit( feeders_select[feeders_select$TREATMENT == "cage","OD_avg"])))
```

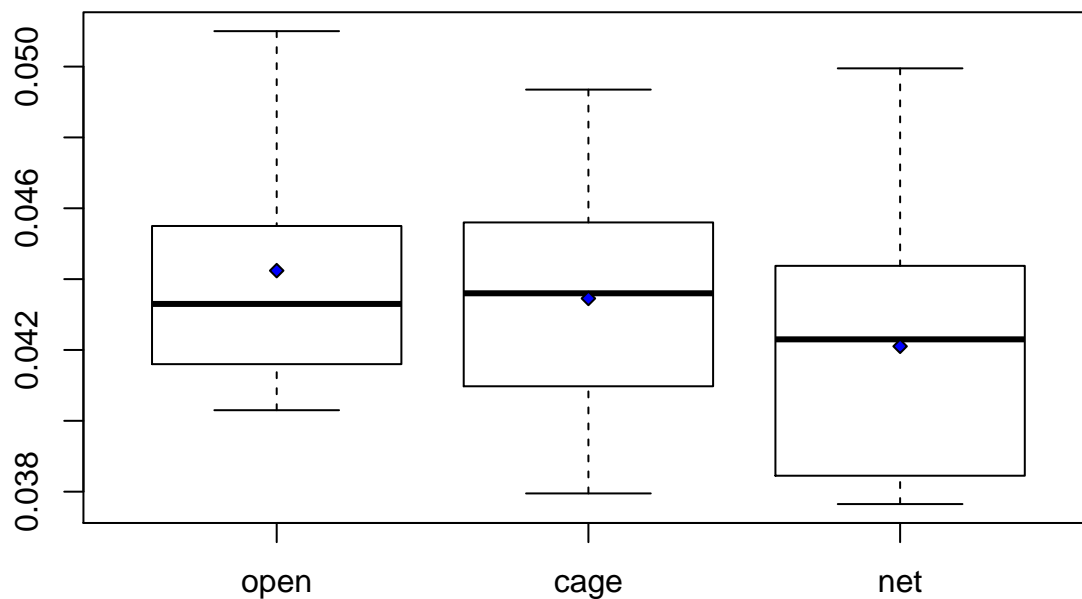
```
## [1] 11
```

- a.2) Create a boxplot for pH and another for OD\_avg in function of TREATMENT, and plot the corresponding mean values as blue dots

```
mean_pH <- as.vector(by(feeders_select$pH, feeders_select$TREATMENT, mean, na.rm = TRUE))  
boxplot(formula = pH ~ TREATMENT, data = feeders_select )  
points(1:3, mean_pH, pch = 23, cex = 0.75,  
       bg = "blue")
```



```
mean_OD_avg <- as.vector(by(feeders_select$OD_avg, feeders_select$TREATMENT, mean, na.rm = TRUE))
boxplot(formula = OD_avg ~ TREATMENT, data = feeders_select )
points(1:3, mean_OD_avg, pch = 23, cex = 0.75,
      bg = "blue")
```



- a.2) Run the the Shapiro-wilk test and the Bartlett test for OD\_avg for all the three TREATMENT variables.

```
by(feeders_select$OD_avg, feeders_select$TREATMENT, shapiro.test)
```

```
## feeders_select$TREATMENT: open
##
##  Shapiro-Wilk normality test
##
## data:  dd[, ]
## W = 0.90586, p-value = 0.2537
```

```
##
## -----
## feeders_select$TREATMENT: cage
##
## Shapiro-Wilk normality test
##
## data:  dd[x, ]
## W = 0.97949, p-value = 0.9633
##
## -----
## feeders_select$TREATMENT: net
##
## Shapiro-Wilk normality test
##
## data:  dd[x, ]
## W = 0.92909, p-value = 0.2358
# All TREATMENT samples follow a normal distribution

bartlett.test(OD_avg ~ TREATMENT, data = feeders_select)

##
## Bartlett test of homogeneity of variances
##
## data:  OD_avg by TREATMENT
## Bartlett's K-squared = 0.040905, df = 2, p-value = 0.9798
# samples variances are not homogeneous
```

- a.3) Run the adequate ANOVA test to investigate if there are any differences between TREATMENTS for OD\_avg

```
# ANOVA test with no assumption of equal variances
oneway.test(OD_avg ~ TREATMENT, data = feeders_select)
```

```
##
## One-way analysis of means (not assuming equal variances)
##
## data:  OD_avg and TREATMENT
## F = 1.2099, num df = 2.000, denom df = 20.562, p-value = 0.3186
```

For a p-value of 0.3186 we conclude that sample means for OD\_avg among the three treatments are similar (can't reject H0)

- a.4) Run the following command. What does function table() do to the data provided ?

```
tbl_1 <- table(feeders$OBSERVER, feeders$SITE)

tbl_2 <- table(feeders$OBSERVER, feeders$TREATMENT)

# Generates a frequency table. In the example, two vectors are provided
# resulting in a 2x2 or 2x3 contingency table. You can also use this function
# for one single vector and also for numeric data.
```

- a.4) Assume that the OBSERVER were suppose to be homogeneously distributed in their task to take the samples in all the three types of cages (TREATMENTS). Was this true? Which would be the expected frequencies for the most even distribution?

```
obs_chsq <- chisq.test(tbl_2)
obs_chsq
```

```
##
## Pearson's Chi-squared test
##
## data:  tbl_2
## X-squared = 0.45402, df = 2, p-value = 0.7969
```

```
obs_chsq$expected
```

```
##
##      open    cage    net
## CGL 63.92818 47.58564 62.48619
## TH  69.07182 51.41436 67.51381
```

```
obs_chsq$observed
```

```
##
##      open cage net
## CGL    67   46  61
## TH     66   53  69
```

```
# The distribution of observations was homogeneous between the two OBSERVERS
```