020- SOD tests for significance and correlation

Directory and doc rules

```
knitr::opts_chunk$set(
  echo = TRUE,          # Display code chunks
  eval = TRUE,          # Evaluate code chunks
  warning = FALSE,      # Hide warnings
  message = FALSE,      # Hide messages
  fig.width = 20,        # Set plot width in inches
  fig.height = 9,       # Set plot height in inches
  fig.align = "center" # Align plots to the center
)
```

Load packages

```
library(tinytex)
library(tidyr)
library(tidyverse)
library(vegan)
```

Load data

```
#getwd()
#sdata<- read.csv("/Users/cmantegna/Documents/WDFWmussels/data/soddata.csv")
asdata<- read.csv("/Users/cmantegna/Documents/WDFWmussels/data/sod_analytes.csv")
```

```
# Data contains numbers below 0 that must be adjusted.These numbers represent samples whose values were


#replace any SOD values at or below 0 with half of the lower detection limit of .005 (.005*.5). Lower d
asdata$sod[asdata$sod <= 0] <- 0.0025
```

Check data

```
summary(asdata)
```

```
##   site_name            latitude       longitude      reporting_area
##  Length:311         Min.   :47.05   Min.   :-123.5   Min.   : 6.00
##  Class :character   1st Qu.:47.33   1st Qu.:-122.7   1st Qu.: 8.20
##  Mode  :character   Median :47.61   Median :-122.6   Median :10.00
##                     Mean   :47.71   Mean   :-122.6   Mean   :10.01
##                     3rd Qu.:48.02   3rd Qu.:-122.4   3rd Qu.:11.00
##                     Max.   :48.82   Max.   :-122.2   Max.   :13.00
```

```
##   site_number      sample_id          sod              mercury
##  Min.   : 1.00   Min.   :  1.0   Min.   : 0.0025   Min.   :-0.02570
##  1st Qu.:21.00   1st Qu.: 78.5   1st Qu.: 2.0003   1st Qu.: 0.03110
##  Median :40.00   Median :156.0   Median : 5.8743   Median : 0.03339
##  Mean   :39.74   Mean   :156.1   Mean   : 7.5653   Mean   : 0.03363
##  3rd Qu.:59.00   3rd Qu.:233.5   3rd Qu.:10.9163   3rd Qu.: 0.03925
##  Max.   :77.00   Max.   :312.0   Max.   :73.3449   Max.   : 0.05195
##     arsenic          cadmium          copper            lead
##  Min.   : 7.246   Min.   :1.537   Min.   : 4.771   Min.   :0.06993
##  1st Qu.: 8.121   1st Qu.:1.747   1st Qu.: 5.829   1st Qu.:0.19825
##  Median : 8.466   Median :1.862   Median : 6.550   Median :0.24936
##  Mean   : 8.510   Mean   :1.870   Mean   : 7.104   Mean   :0.28314
##  3rd Qu.: 8.835   3rd Qu.:1.962   3rd Qu.: 7.551   3rd Qu.:0.32670
##  Max.   :10.284   Max.   :2.312   Max.   :30.471   Max.   :0.88554
##       zinc            lmwPAH            PAH16             sumPCB
##  Min.   : 67.97   Min.   :  79.14   Min.   :  24.59   Min.   : 16.95
##  1st Qu.: 79.33   1st Qu.: 112.89   1st Qu.:  71.00   1st Qu.: 35.82
##  Median : 85.54   Median : 188.18   Median : 150.63   Median : 49.07
##  Mean   : 85.70   Mean   : 412.33   Mean   : 547.38   Mean   : 58.35
##  3rd Qu.: 91.81   3rd Qu.: 364.36   3rd Qu.: 292.07   3rd Qu.: 71.36
##  Max.   :110.78   Max.   :6125.29   Max.   :9800.46   Max.   :175.58
##      hmwPAH            sumPAH          PAHgroup3       PAHgroup4
##  Min.   :  11.61   Min.   :  97.4   Min.   :1.000   Min.   :1.000
##  1st Qu.:  74.83   1st Qu.: 179.8   1st Qu.:1.000   1st Qu.:1.500
##  Median : 156.65   Median :  356.9   Median :2.000   Median :2.000
##  Mean   : 580.13   Mean   :  989.1   Mean   :1.987   Mean   :2.492
##  3rd Qu.: 284.32   3rd Qu.:  694.8   3rd Qu.:3.000   3rd Qu.:3.000
##  Max.   :9394.33   Max.   :14700.7   Max.   :3.000   Max.   :4.000
##    PAHgroup5       PAHgroup6
##  Min.   :0.000   Min.   :0.000
##  1st Qu.:1.000   1st Qu.:1.000
##  Median :2.000   Median :3.000
##  Mean   :2.019   Mean   :2.537
##  3rd Qu.:3.000   3rd Qu.:4.000
##  Max.   :4.000   Max.   :5.000
```

Shapiro-Wilkes

```
#test for normality. Data is not normally distributed.
# *All analytes were determined not normally distributed in 010-p450.Rmd*

shapiro.test(asdata$sod)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  asdata$sod
## W = 0.76932, p-value < 2.2e-16
```

Kruskal-Wallis

site and reporting area

```
#test for significant interaction

# Change to character for the kw test AFTER keeping as numeric for the correlation tests - if necessary
#apdata$PAHgroup <- as.character(apdata$PAHgroup)
#apdata$PAHgroup2 <- as.character(apdata$PAHgroup2)
#apdata$PAHgroup3 <- as.character(apdata$PAHgroup3)

kruskal.test(sod ~ site_name, data = asdata)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  sod by site_name
## Kruskal-Wallis chi-squared = 114.05, df = 73, p-value = 0.001515
```

```
kruskal.test(sod ~ reporting_area, data = asdata)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  sod by reporting_area
## Kruskal-Wallis chi-squared = 9.8218, df = 8, p-value = 0.2778
```

PAH groups

```
kruskal.test(sod ~ PAHgroup3, data = asdata)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  sod by PAHgroup3
## Kruskal-Wallis chi-squared = 2.4174, df = 2, p-value = 0.2986
```

```
kruskal.test(sod ~ PAHgroup4, data = asdata)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  sod by PAHgroup4
## Kruskal-Wallis chi-squared = 12.01, df = 3, p-value = 0.007351
```

```
kruskal.test(sod ~ PAHgroup5, data = asdata)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  sod by PAHgroup5
## Kruskal-Wallis chi-squared = 7.8008, df = 4, p-value = 0.09915
```

```
kruskal.test(sod ~ PAHgroup6, data = asdata)
```

```
##
##  Kruskal-Wallis rank sum test
```

```
##
## data:  sod by PAHgroup6
## Kruskal-Wallis chi-squared = 5.8146, df = 5, p-value = 0.3247
```

Kruskal-Wallac Multiple Comparisons (post hoc)

Reporting Areas Are:
6 - East Juan de Fuca Strait
7 - San Juan Islands
8.1 - Deception Pass, Hope Island, and Skagit Bay
8.2 - Port Susan and Port Gardner
9 - Admiralty Inlet
10 - Seattle-Bremerton
11 - Tacoma-Vashon
12 - Hood Canal
13 - South Puget Sound

```r
library(pgirmess)

# no significance confirmed between sites
# significance confirmed between low(1) and very high(4)

mc_site<- as.data.frame(kruskalmc(sod ~ site_name, data = asdata, method = "bonferroni"))
mc_group4<- as.data.frame(kruskalmc(sod ~ PAHgroup4, data = asdata, method = "bonferroni"))

head(mc_site)
```

```
##                                                                                 statistic
## Aiston Preserve-Arroyo Beach                    Multiple comparison test after Kruskal-Wallis
## Aiston Preserve-Blair Waterway                  Multiple comparison test after Kruskal-Wallis
## Aiston Preserve-Blair Waterway #2               Multiple comparison test after Kruskal-Wallis
## Aiston Preserve-Brackenwood Ln                  Multiple comparison test after Kruskal-Wallis
## Aiston Preserve-Broad Spit (Fisherman's Point) Multiple comparison test after Kruskal-Wallis
## Aiston Preserve-Browns Point Lighthouse         Multiple comparison test after Kruskal-Wallis
##                                                alpha dif.com.obs.dif
## Aiston Preserve-Arroyo Beach                    0.05          81.525
## Aiston Preserve-Blair Waterway                  0.05          63.875
## Aiston Preserve-Blair Waterway #2               0.05          81.125
## Aiston Preserve-Brackenwood Ln                  0.05          63.000
## Aiston Preserve-Broad Spit (Fisherman's Point)  0.05          72.625
## Aiston Preserve-Browns Point Lighthouse         0.05          11.625
##                                                dif.com.critical.dif
## Aiston Preserve-Arroyo Beach                               258.3048
## Aiston Preserve-Blair Waterway                             272.2772
## Aiston Preserve-Blair Waterway #2                          272.2772
## Aiston Preserve-Brackenwood Ln                             272.2772
## Aiston Preserve-Broad Spit (Fisherman's Point)             272.2772
## Aiston Preserve-Browns Point Lighthouse                    272.2772
##                                                dif.com.stat.signif
## Aiston Preserve-Arroyo Beach                                 FALSE
## Aiston Preserve-Blair Waterway                               FALSE
## Aiston Preserve-Blair Waterway #2                            FALSE
## Aiston Preserve-Brackenwood Ln                               FALSE
## Aiston Preserve-Broad Spit (Fisherman's Point)               FALSE
```

```
## Aiston Preserve-Browns Point Lighthouse                      FALSE
```

```
head(mc_group4)
```

```
##                                       statistic alpha dif.com.obs.dif
## 1-2 Multiple comparison test after Kruskal-Wallis  0.05        30.31410
## 1-3 Multiple comparison test after Kruskal-Wallis  0.05        17.61311
## 1-4 Multiple comparison test after Kruskal-Wallis  0.05        48.52800
## 2-3 Multiple comparison test after Kruskal-Wallis  0.05        12.70099
## 2-4 Multiple comparison test after Kruskal-Wallis  0.05        18.21390
## 3-4 Multiple comparison test after Kruskal-Wallis  0.05        30.91489
##     dif.com.critical.dif dif.com.stat.signif
## 1-2             37.98846               FALSE
## 1-3             37.86805               FALSE
## 1-4             38.23757                TRUE
## 2-3             37.86805               FALSE
## 2-4             38.23757               FALSE
## 3-4             38.11795               FALSE
```

Correlation- PAHgroup3, PAHgroup4, PAHgroup5 and PAHgroup6

0= VLow
1= Low
2= Mid
3= High
4= VHigh
5= EHigh PAH groups determined by set quantiles found in grouping.rmd.
Group3= 1-3 | Group4= 1-4 | Group5= 0-4 | Group6= 0-5

```
# no correlation
correlation_result <- cor.test(asdata$sod, asdata$PAHgroup3, method = "pearson")
print(correlation_result)
```

```
##
##  Pearson's product-moment correlation
##
## data:  asdata$sod and asdata$PAHgroup3
## t = 1.8166, df = 309, p-value = 0.07025
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.00852043  0.21159258
## sample estimates:
##       cor
## 0.1027943
```

```
# correlation= 0.1774695 with a p-value= 0.001677 indicates a weak positive correlation.
correlation_result <- cor.test(asdata$sod, asdata$PAHgroup4, method = "pearson")
print(correlation_result)
```

```
##
##  Pearson's product-moment correlation
##
## data:  asdata$sod and asdata$PAHgroup4
## t = 3.1699, df = 309, p-value = 0.001677
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
```

```
##  0.06758614 0.28309907
## sample estimates:
##       cor
## 0.1774695
```

```
# correlation= 0.1351344 with a p-value= 0.0171 indicates a weak positive correlation.
correlation_result <- cor.test(asdata$sod, asdata$PAHgroup5, method = "pearson")
print(correlation_result)
```

```
##
##  Pearson's product-moment correlation
##
## data:  asdata$sod and asdata$PAHgroup5
## t = 2.3974, df = 309, p-value = 0.0171
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.02428207 0.24270410
## sample estimates:
##       cor
## 0.1351344
```

```
# correlation= 0.132161 with a p-value= 0.01973 indicates a weak positive correlation.
correlation_result <- cor.test(asdata$sod, asdata$PAHgroup6, method = "pearson")
print(correlation_result)
```

```
##
##  Pearson's product-moment correlation
##
## data:  asdata$sod and asdata$PAHgroup6
## t = 2.3437, df = 309, p-value = 0.01973
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.02125615 0.23985285
## sample estimates:
##      cor
## 0.132161
```

Correlation test - sumPAH, lmwPAH, hmwPAH and PAH16

```
# no correlation
correlation_result <- cor.test(asdata$sod, asdata$sumPAH, method = "pearson")
print(correlation_result)
```

```
##
##  Pearson's product-moment correlation
##
## data:  asdata$sod and asdata$sumPAH
## t = 1.3239, df = 309, p-value = 0.1865
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.03641879  0.18477666
## sample estimates:
##        cor
## 0.07510273
```

```r
# no correlation
correlation_result <- cor.test(asdata$sod, asdata$lmwPAH, method = "pearson")
print(correlation_result)
```

```
##
##  Pearson's product-moment correlation
##
## data:  asdata$sod and asdata$lmwPAH
## t = 1.1861, df = 309, p-value = 0.2365
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.0442258  0.1772131
## sample estimates:
##        cor
## 0.06732266
```

```r
# no correlation
correlation_result <- cor.test(asdata$sod, asdata$hmwPAH, method = "pearson")
print(correlation_result)
```

```
##
##  Pearson's product-moment correlation
##
## data:  asdata$sod and asdata$hmwPAH
## t = 1.3837, df = 309, p-value = 0.1674
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.03303055  0.18805110
## sample estimates:
##        cor
## 0.07847505
```

```r
# no correlation
correlation_result <- cor.test(asdata$sod, asdata$PAH16, method = "pearson")
print(correlation_result)
```

```
##
##  Pearson's product-moment correlation
##
## data:  asdata$sod and asdata$PAH16
## t = 1.2439, df = 309, p-value = 0.2145
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.04095302  0.18038701
## sample estimates:
##        cor
## 0.07058579
```

Correlation test - sumPCB

```r
# correlation= 0.1557987 with a p-value= .0059. This is a significant weak correlation.
correlation_result <- cor.test(asdata$sod, asdata$sumPCB, method = "pearson")
print(correlation_result)
```

```
##
```

```
##  Pearson's product-moment correlation
##
## data:  asdata$sod and asdata$sumPCB
## t = 2.7725, df = 309, p-value = 0.0059
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.04536753 0.26246814
## sample estimates:
##       cor
## 0.1557987
```

Correlation test - mercury, arsenic, cadmium, copper, lead and zinc

```
# no correlation
correlation_result <- cor.test(asdata$sod, asdata$mercury, method = "pearson")
print(correlation_result)
```

```
##
##  Pearson's product-moment correlation
##
## data:  asdata$sod and asdata$mercury
## t = 0.032392, df = 309, p-value = 0.9742
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.1093970  0.1130368
## sample estimates:
##         cor
## 0.001842693
```

```
# no correlation
correlation_result <- cor.test(asdata$sod, asdata$arsenic, method = "pearson")
print(correlation_result)
```

```
##
##  Pearson's product-moment correlation
##
## data:  asdata$sod and asdata$arsenic
## t = -1.2004, df = 309, p-value = 0.2309
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.17799885  0.04341602
## sample estimates:
##         cor
## -0.06813027
```

```
# no correlation
correlation_result <- cor.test(asdata$sod, asdata$cadmium, method = "pearson")
print(correlation_result)
```

```
##
##  Pearson's product-moment correlation
##
## data:  asdata$sod and asdata$cadmium
## t = -1.1225, df = 309, p-value = 0.2625
## alternative hypothesis: true correlation is not equal to 0
```

```
## 95 percent confidence interval:
##  -0.17371112  0.04783137
## sample estimates:
##         cor
## -0.06372494
```

```
# no correlation
correlation_result <- cor.test(asdata$sod, asdata$copper, method = "pearson")
print(correlation_result)
```

```
##
##  Pearson's product-moment correlation
##
## data:  asdata$sod and asdata$copper
## t = -0.53585, df = 309, p-value = 0.5924
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.14120814  0.08102255
## sample estimates:
##         cor
## -0.03046933
```

```
# no correlation
correlation_result <- cor.test(asdata$sod, asdata$lead, method = "pearson")
print(correlation_result)
```

```
##
##  Pearson's product-moment correlation
##
## data:  asdata$sod and asdata$lead
## t = 1.6447, df = 309, p-value = 0.101
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.0182491  0.2022789
## sample estimates:
##       cor
## 0.09315729
```

```
# no correlation
correlation_result <- cor.test(asdata$sod, asdata$zinc, method = "pearson")
print(correlation_result)
```

```
##
##  Pearson's product-moment correlation
##
## data:  asdata$sod and asdata$zinc
## t = -0.71362, df = 309, p-value = 0.476
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.15109888  0.07097429
## sample estimates:
##         cor
## -0.04056322
```