

## 010- p450 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()  
#pdata<- read.csv("/Users/cmantegna/Documents/WDFWmussels/data/p450data.csv")  
apdata<- read.csv("/Users/cmantegna/Documents/WDFWmussels/data/p450_analytes.csv")  
  
#metal<- read.csv("/Users/cmantegna/Documents/WDFWmussels/data/metals.csv")
```

### Check data

```
summary(apdata)
```

##	site_name	latitude	longitude	reporting_area
##	Length:299	Min. :47.05	Min. :-123.5	Min. : 6.000
##	Class :character	1st Qu.:47.34	1st Qu.: -122.7	1st Qu.: 8.200
##	Mode :character	Median :47.62	Median : -122.6	Median :10.000
##		Mean :47.72	Mean : -122.6	Mean : 9.978
##		3rd Qu.:48.02	3rd Qu.: -122.4	3rd Qu.:11.000
##		Max. :48.82	Max. : -122.2	Max. :13.000
##	site_number	sample_id	p450	mercury
##	Min. : 1.00	Min. : 1.0	Min. : 291625	Min. : -0.02570
##	1st Qu.:20.00	1st Qu.: 79.5	1st Qu.: 2351930	1st Qu.: 0.03128
##	Median :39.00	Median :161.0	Median : 4081903	Median : 0.03396

```
## Mean :39.55 Mean :158.6 Mean : 5469202 Mean : 0.03370
## 3rd Qu.:59.00 3rd Qu.:236.5 3rd Qu.: 6629056 3rd Qu.: 0.03925
## Max. :77.00 Max. :312.0 Max. :52691211 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.:1.741 1st Qu.: 5.812 1st Qu.:0.20122
## Median : 8.466 Median :1.862 Median : 6.550 Median :0.24936
## Mean : 8.517 Mean :1.866 Mean : 7.099 Mean :0.28191
## 3rd Qu.: 8.820 3rd Qu.:1.956 3rd Qu.: 7.547 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.: 79.30 1st Qu.: 112.89 1st Qu.: 71.00 1st Qu.: 36.55
## Median : 86.21 Median : 188.18 Median : 150.63 Median : 49.07
## Mean : 85.84 Mean : 416.70 Mean : 560.26 Mean : 58.41
## 3rd Qu.: 91.81 3rd Qu.: 364.36 3rd Qu.: 292.07 3rd Qu.: 68.07
## 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.: 75.07 1st Qu.: 182.9 1st Qu.:1.000 1st Qu.:2.000
## Median : 156.65 Median : 356.9 Median :2.000 Median :2.000
## Mean : 593.43 Mean : 1006.7 Mean :1.977 Mean :2.482
## 3rd Qu.: 282.43 3rd Qu.: 599.0 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 :2.003 Mean :2.528
## 3rd Qu.:3.000 3rd Qu.:4.000
## Max. :4.000 Max. :5.000
```

## Shapiro-Wilkes

```
#test for normality. No data is normally distributed.
```

```
shapiro.test(apdata$p450)
```

```
##
## Shapiro-Wilk normality test
##
## data: apdata$p450
## W = 0.66138, p-value < 2.2e-16
```

```
shapiro.test(apdata$sumPAH)
```

```
##
## Shapiro-Wilk normality test
##
## data: apdata$sumPAH
## W = 0.36439, p-value < 2.2e-16
```

```
shapiro.test(apdata$lmwPAH)
```

```
##
```

```

## Shapiro-Wilk normality test
##
## data:  apdata$lmwPAH
## W = 0.37833, p-value < 2.2e-16
shapiro.test(apdata$lmwPAH)

##
## Shapiro-Wilk normality test
##
## data:  apdata$hmwPAH
## W = 0.34123, p-value < 2.2e-16
shapiro.test(apdata$hmwPAH)

##
## Shapiro-Wilk normality test
##
## data:  apdata$PAH16
## W = 0.32484, p-value < 2.2e-16
shapiro.test(apdata$PAH16)

##
## Shapiro-Wilk normality test
##
## data:  apdata$sumPCB
## W = 0.82775, p-value < 2.2e-16
shapiro.test(apdata$sumPCB)

##
## Shapiro-Wilk normality test
##
## data:  apdata$mercury
## W = 0.62037, p-value < 2.2e-16
shapiro.test(apdata$mercury)

##
## Shapiro-Wilk normality test
##
## data:  apdata$arsenic
## W = 0.97529, p-value = 4.983e-05
shapiro.test(apdata$arsenic)

##
## Shapiro-Wilk normality test
##
## data:  apdata$cadmium
## W = 0.96871, p-value = 4.385e-06
shapiro.test(apdata$cadmium)

##
## Shapiro-Wilk normality test
##
## data:  apdata$copper

```

```
## W = 0.43289, p-value < 2.2e-16
```

```
shapiro.test(apdata$lead)
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
```

```
## data: apdata$lead
```

```
## W = 0.83901, p-value < 2.2e-16
```

```
shapiro.test(apdata$zinc)
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
```

```
## data: apdata$zinc
```

```
## W = 0.98467, p-value = 0.002799
```

## 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(p450 ~ site_name, data = apdata)
```

```
##
```

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

```
##
```

```
## data: p450 by site_name
```

```
## Kruskal-Wallis chi-squared = 138.13, df = 73, p-value = 6.442e-06
```

```
kruskal.test(p450 ~ reporting_area, data = apdata)
```

```
##
```

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

```
##
```

```
## data: p450 by reporting_area
```

```
## Kruskal-Wallis chi-squared = 30.464, df = 8, p-value = 0.0001749
```

### PAH groups

```
kruskal.test(p450 ~ PAHgroup3, data = apdata)
```

```
##
```

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

```
##
```

```
## data: p450 by PAHgroup3
```

```
## Kruskal-Wallis chi-squared = 1.8934, df = 2, p-value = 0.388
```

```
kruskal.test(p450 ~ PAHgroup4, data = apdata)

##
## Kruskal-Wallis rank sum test
##
## data:  p450 by PAHgroup4
## Kruskal-Wallis chi-squared = 3.2105, df = 3, p-value = 0.3603
kruskal.test(p450 ~ PAHgroup5, data = apdata)

##
## Kruskal-Wallis rank sum test
##
## data:  p450 by PAHgroup5
## Kruskal-Wallis chi-squared = 2.7208, df = 4, p-value = 0.6056
kruskal.test(p450 ~ PAHgroup6, data = apdata)

##
## Kruskal-Wallis rank sum test
##
## data:  p450 by PAHgroup6
## Kruskal-Wallis chi-squared = 6.1434, df = 5, p-value = 0.2925
```

## 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

```
library(pgirmess)

# no significance confirmed between sites
# significant differences across reporting area pairwise matches: 8.2-12, 10-13, 11-12, and 12-13

mc_site<- as.data.frame(kruskalmc(p450 ~ site_name, data = apdata, method = "bonferroni"))
mc_reporting<- as.data.frame(kruskalmc(p450 ~ reporting_area, data = apdata, method = "bonferroni"))

#head(mc_site)
#head(mc_reporting)
```

## 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(apdata$p450, apdata$PAHgroup3, method = "pearson")  
print(correlation_result)
```

```
##  
## Pearson's product-moment correlation  
##  
## data: apdata$p450 and apdata$PAHgroup3  
## t = -0.22448, df = 297, p-value = 0.8225  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.1262684 0.1005543  
## sample estimates:  
## cor  
## -0.01302458
```

```
#no correlation  
correlation_result <- cor.test(apdata$p450, apdata$PAHgroup4, method = "pearson")  
print(correlation_result)
```

```
##  
## Pearson's product-moment correlation  
##  
## data: apdata$p450 and apdata$PAHgroup4  
## t = -0.62396, df = 297, p-value = 0.5331  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.14900081 0.07756666  
## sample estimates:  
## cor  
## -0.03618201
```

```
#no correlation  
correlation_result <- cor.test(apdata$p450, apdata$PAHgroup5, method = "pearson")  
print(correlation_result)
```

```
##  
## Pearson's product-moment correlation  
##  
## data: apdata$p450 and apdata$PAHgroup5  
## t = -0.36995, df = 297, p-value = 0.7117  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.1345644 0.0921931  
## sample estimates:  
## cor  
## -0.02146166
```

```
#no correlation  
correlation_result <- cor.test(apdata$p450, apdata$PAHgroup6, method = "pearson")  
print(correlation_result)
```

```
##
```

```
## Pearson's product-moment correlation
##
## data:  apdata$p450 and apdata$PAHgroup6
## t = -0.30211, df = 297, p-value = 0.7628
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.13069813  0.09609371
## sample estimates:
##          cor
## -0.01752766
```

## Correlation- sumPAH, lmwPAH, hmwPAH and PAH16

```
# no correlation
correlation_result <- cor.test(apdata$p450, apdata$sumPAH, method = "pearson")
print(correlation_result)
```

```
##
## Pearson's product-moment correlation
##
## data:  apdata$p450 and apdata$sumPAH
## t = -0.13317, df = 297, p-value = 0.8941
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.1210515  0.1057957
## sample estimates:
##          cor
## -0.007727312
```

```
# no correlation
correlation_result <- cor.test(apdata$p450, apdata$hmwPAH, method = "pearson")
print(correlation_result)
```

```
##
## Pearson's product-moment correlation
##
## data:  apdata$p450 and apdata$hmwPAH
## t = -0.36791, df = 297, p-value = 0.7132
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.13444804  0.09231059
## sample estimates:
##          cor
## -0.02134321
```

```
# no correlation
correlation_result <- cor.test(apdata$p450, apdata$lmwPAH, method = "pearson")
print(correlation_result)
```

```
##
## Pearson's product-moment correlation
##
## data:  apdata$p450 and apdata$lmwPAH
## t = 0.24705, df = 297, p-value = 0.805
## alternative hypothesis: true correlation is not equal to 0
```

```
## 95 percent confidence interval:
## -0.09925789 0.12755674
## sample estimates:
##      cor
## 0.01433382

# no correlation
correlation_result <- cor.test(apdata$p450, apdata$PAH16, method = "pearson")
print(correlation_result)

##
## Pearson's product-moment correlation
##
## data: apdata$p450 and apdata$PAH16
## t = -0.073675, df = 297, p-value = 0.9413
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.1176483 0.1092083
## sample estimates:
##      cor
## -0.004275003
```

## Correlation test - sumPCB

```
# no correlation
correlation_result <- cor.test(apdata$p450, apdata$sumPCB, method = "pearson")
print(correlation_result)

##
## Pearson's product-moment correlation
##
## data: apdata$p450 and apdata$sumPCB
## t = -1.0941, df = 297, p-value = 0.2748
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.1755268 0.0504348
## sample estimates:
##      cor
## -0.06335798
```

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

```
# Correlation of .1505; p-value= .0092
correlation_result <- cor.test(apdata$p450, apdata$mercury, method = "pearson")
print(correlation_result)

##
## Pearson's product-moment correlation
##
## data: apdata$p450 and apdata$mercury
## t = 2.6229, df = 297, p-value = 0.009167
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
```



```
## 0.03767878 0.25946790
## sample estimates:
##      cor
## 0.150466

# no correlation
correlation_result <- cor.test(apdata$p450, apdata$arsenic, method = "pearson")
print(correlation_result)

##
## Pearson's product-moment correlation
##
## data: apdata$p450 and apdata$arsenic
## t = 0.88157, df = 297, p-value = 0.3787
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.06270669 0.16356947
## sample estimates:
##      cor
## 0.051087

# no correlation
correlation_result <- cor.test(apdata$p450, apdata$cadmium, method = "pearson")
print(correlation_result)

##
## Pearson's product-moment correlation
##
## data: apdata$p450 and apdata$cadmium
## t = 0.42034, df = 297, p-value = 0.6745
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.08929385 0.13743364
## sample estimates:
##      cor
## 0.02438344

# no correlation
correlation_result <- cor.test(apdata$p450, apdata$copper, method = "pearson")
print(correlation_result)

##
## Pearson's product-moment correlation
##
## data: apdata$p450 and apdata$copper
## t = -1.247, df = 297, p-value = 0.2134
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.18409302 0.04160115
## sample estimates:
##      cor
## -0.07216973

# no correlation
correlation_result <- cor.test(apdata$p450, apdata$lead, method = "pearson")
print(correlation_result)
```

```
##
## Pearson's product-moment correlation
##
## data:  apdata$p450 and apdata$lead
## t = -1.1043, df = 297, p-value = 0.2704
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.17609769  0.04984723
## sample estimates:
##          cor
## -0.06394465

# no correlation
correlation_result <- cor.test(apdata$p450, apdata$zinc, method = "pearson")
print(correlation_result)

##
## Pearson's product-moment correlation
##
## data:  apdata$p450 and apdata$zinc
## t = 0.72045, df = 297, p-value = 0.4718
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.07200302  0.15446692
## sample estimates:
##          cor
## 0.04176844
```