GLM model on toxin concetration in Dolphins

# Load the dataset and inspect it  
dolphins=read.csv("dolphins.csv",header = T)  
str(dolphins)

## 'data.frame': 47 obs. of 22 variables:  
## $ ID : Factor w/ 46 levels "CAHA 106","CAHA377",..: 1 3 4 5 6 12 13 14 15 16 ...  
## $ Sex : Factor w/ 2 levels "F","M": 2 2 2 2 1 2 1 2 2 2 ...  
## $ Age\_Class : Factor w/ 2 levels "Adult","Juvenile": 1 2 1 1 2 2 1 1 1 1 ...  
## $ Location : Factor w/ 2 levels "FL","NC": 2 2 2 2 2 1 1 1 1 1 ...  
## $ Morbillivirus: Factor w/ 2 levels "N","Y": 1 1 2 2 2 2 1 1 1 1 ...  
## $ PCB\_1268 : num NA NA NA NA NA NA NA NA NA NA ...  
## $ DEP : num 0.04 NA 46.27 5.22 8.29 ...  
## $ NPE : num 0.01 NA 0.73 0.01 0.01 ...  
## $ Triclosan : num 0.03 NA 7.36 11.87 0.03 ...  
## $ BPA : num 0.01 NA 0.01 49.55 0.01 ...  
## $ Atrazine : num 0.01 NA 0.45 0.01 0.01 NA NA NA 0.06 0.01 ...  
## $ Hg : num 1402.21 9.89 NA 279.18 388.91 ...  
## $ Cu : num 34 27.1 NA 17.1 25.3 ...  
## $ Se : num 448.11 5.12 NA 101.96 149.47 ...  
## $ Cd : num 149.41 0.41 NA 0.36 3.08 ...  
## $ Tl : num 0.05 0.05 NA 0.05 0.05 0 0.01 0.02 NA 0 ...  
## $ Pb : num 0.64 0.05 NA 0.1 0.11 0.03 0.11 0.09 NA 0.26 ...  
## $ Mn : num 10.03 13.91 NA 4.21 15.3 ...  
## $ Co : num 0.03 0.03 NA 0.03 0.03 0.02 0.02 0.04 NA 0.02 ...  
## $ Zn : num 90.2 216 NA 143.8 240.9 ...  
## $ As : num 1.19 1.22 NA 0.5 0.68 0.78 0.3 1.73 NA 0.2 ...  
## $ Fe : num 3675 917 NA 1436 1336 ...

head(dolphins)

## ID Sex Age\_Class Location Morbillivirus PCB\_1268 DEP NPE Triclosan  
## 1 CAHA 106 M Adult NC N NA 0.04 0.01 0.03  
## 2 CALO 13-20 M Juvenile NC N NA NA NA NA  
## 3 CALO 13-22 M Adult NC Y NA 46.27 0.73 7.36  
## 4 CALO 14-05 M Adult NC Y NA 5.22 0.01 11.87  
## 5 CALO 14-13 F Juvenile NC Y NA 8.29 0.01 0.03  
## 6 HBOI-1313 M Juvenile FL Y NA NA NA NA  
## BPA Atrazine Hg Cu Se Cd Tl Pb Mn Co Zn As  
## 1 0.01 0.01 1402.21 34.02 448.11 149.41 0.05 0.64 10.03 0.03 90.22 1.19  
## 2 NA NA 9.89 27.09 5.12 0.41 0.05 0.05 13.91 0.03 216.02 1.22  
## 3 0.01 0.45 NA NA NA NA NA NA NA NA NA NA  
## 4 49.55 0.01 279.18 17.14 101.96 0.36 0.05 0.10 4.21 0.03 143.76 0.50  
## 5 0.01 0.01 388.91 25.31 149.47 3.08 0.05 0.11 15.30 0.03 240.93 0.68  
## 6 NA NA NA 30.99 2.87 0.64 0.00 0.03 12.57 0.02 207.06 0.78  
## Fe  
## 1 3674.68  
## 2 916.81  
## 3 NA  
## 4 1435.62  
## 5 1336.10  
## 6 NA

# Perform the summary statistics  
# Categorical variables: create a function that performs the test then call it on the variables  
   
mytable=function(x){  
 result=table(x)  
 return(result)  
}  
lapply(dolphins[,c(2:5)], mytable)

## $Sex  
## x  
## F M   
## 17 30   
##   
## $Age\_Class  
## x  
## Adult Juvenile   
## 31 16   
##   
## $Location  
## x  
## FL NC   
## 17 30   
##   
## $Morbillivirus  
## x  
## N Y   
## 35 12

# Continuos variables   
summary(dolphins[,c(6:22)])

## PCB\_1268 DEP NPE Triclosan   
## Min. : 0.030 Min. : 0.020 Min. : 0.010 Min. : 0.01   
## 1st Qu.: 0.030 1st Qu.: 0.050 1st Qu.: 0.010 1st Qu.: 0.03   
## Median : 0.030 Median : 0.410 Median : 0.210 Median : 0.23   
## Mean : 3.615 Mean : 7.491 Mean : 4.168 Mean : 14.78   
## 3rd Qu.: 3.810 3rd Qu.: 9.370 3rd Qu.: 1.525 3rd Qu.: 14.87   
## Max. :17.330 Max. :46.270 Max. :31.270 Max. :173.69   
## NA's :28 NA's :28 NA's :28 NA's :28   
## BPA Atrazine Hg Cu   
## Min. : 0.01 Min. :0.01000 Min. : 3.62 Min. : 7.86   
## 1st Qu.: 0.01 1st Qu.:0.01000 1st Qu.: 28.56 1st Qu.: 17.14   
## Median : 0.01 Median :0.01000 Median : 149.98 Median : 25.31   
## Mean : 18.98 Mean :0.03579 Mean : 247.88 Mean : 33.97   
## 3rd Qu.: 0.01 3rd Qu.:0.01000 3rd Qu.: 325.92 3rd Qu.: 33.13   
## Max. :258.29 Max. :0.45000 Max. :1415.50 Max. :213.60   
## NA's :28 NA's :28 NA's :10 NA's :2   
## Se Cd Tl Pb   
## Min. : 2.87 Min. : 0.04 Min. :0.00000 Min. :0.0300   
## 1st Qu.: 13.73 1st Qu.: 0.06 1st Qu.:0.05000 1st Qu.:0.0500   
## Median : 66.10 Median : 0.17 Median :0.05000 Median :0.0600   
## Mean :107.65 Mean : 12.95 Mean :0.04378 Mean :0.1087   
## 3rd Qu.:149.47 3rd Qu.: 0.49 3rd Qu.:0.05000 3rd Qu.:0.1200   
## Max. :585.86 Max. :366.00 Max. :0.07000 Max. :0.6400   
## NA's :2 NA's :2 NA's :2 NA's :2   
## Mn Co Zn As   
## Min. : 2.52 Min. :0.01000 Min. : 42.77 Min. :0.070   
## 1st Qu.: 7.96 1st Qu.:0.03000 1st Qu.:116.64 1st Qu.:0.660   
## Median :11.83 Median :0.03000 Median :181.51 Median :1.070   
## Mean :11.55 Mean :0.02933 Mean :188.28 Mean :1.233   
## 3rd Qu.:13.65 3rd Qu.:0.03000 3rd Qu.:231.06 3rd Qu.:1.640   
## Max. :25.97 Max. :0.08000 Max. :702.67 Max. :5.810   
## NA's :2 NA's :2 NA's :2 NA's :2   
## Fe   
## Min. : 411.1   
## 1st Qu.: 647.2   
## Median : 928.6   
## Mean :1128.0   
## 3rd Qu.:1336.1   
## Max. :3674.7   
## NA's :10

For bottlenose dolphin cases, the Mann-Whitney U-test was used to compare the median toxicant and nutrient concentrations between: adult males versus adult females, juveniles versus adults, Florida-stranded versus North Carolina-stranded,confirmed/suspected Morbillivirus+ versus uninfected.

mydifference=function(y){  
result=wilcox.test(dolphins$Hg~y,data = dolphins)   
 return(result)  
}  
lapply(dolphins[,c(2:5)], mydifference)

## $Sex  
##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: dolphins$Hg by y  
## W = 148, p-value = 0.6954  
## alternative hypothesis: true location shift is not equal to 0  
##   
##   
## $Age\_Class  
##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: dolphins$Hg by y  
## W = 295, p-value = 2.903e-05  
## alternative hypothesis: true location shift is not equal to 0  
##   
##   
## $Location  
##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: dolphins$Hg by y  
## W = 168, p-value = 0.05741  
## alternative hypothesis: true location shift is not equal to 0  
##   
##   
## $Morbillivirus  
##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: dolphins$Hg by y  
## W = 152, p-value = 0.5725  
## alternative hypothesis: true location shift is not equal to 0

#### Difference in Hg based on Sex

* Since the p-value is greater than 0.05, we fail to reject the null hypothesis. We do not have sufficient evidence to say that the amount of Mercury (Hg) in the Male is different from the Female dolphins.

#### Difference in Hg based on Age\_class

* Since the p-value is less than 0.05, we reject the null hypothesis. We conclude that the amount of Mercury (Hg) in the Juvenile is different from the adult dolphins.

#### Difference in Hg based on Location

* Since the p-value is greater than 0.05, we fail to reject the null hypothesis. We do not have sufficient evidence to say that the amount of Mercury (Hg) in the Florida-stranded is different from the North Carolina-stranded dolphins

#### Difference in Hg based on Mobilivirus

* Since the p-value is greater than 0.05, we fail to reject the null hypothesis. We do not have sufficient evidence to say that the amount of Mercury (Hg) in the confirmed/suspected Morbillivirus dolphins is different from the uninfected dolphins.

#### Before we fit a GLM model we handle the missing values first using multiple imputation method

# Check the number of missing  
sum(is.na(dolphins)) #all missing values

## [1] 206

colnames(dolphins)[colSums(is.na(dolphins)) > 0] #columns with missing values

## [1] "PCB\_1268" "DEP" "NPE" "Triclosan" "BPA" "Atrazine"   
## [7] "Hg" "Cu" "Se" "Cd" "Tl" "Pb"   
## [13] "Mn" "Co" "Zn" "As" "Fe"

#### Pattern of Missing Data Exploration

Before moving on to determining the specifics of multiple imputation, we should first explore and see the pattern of missing data in our dataset.

p\_missing <- unlist(lapply(dolphins, function(x) sum(is.na(x))))/nrow(dolphins)  
sort(p\_missing[p\_missing > 0], decreasing = TRUE)

## PCB\_1268 DEP NPE Triclosan BPA Atrazine Hg   
## 0.59574468 0.59574468 0.59574468 0.59574468 0.59574468 0.59574468 0.21276596   
## Fe Cu Se Cd Tl Pb Mn   
## 0.21276596 0.04255319 0.04255319 0.04255319 0.04255319 0.04255319 0.04255319   
## Co Zn As   
## 0.04255319 0.04255319 0.04255319

# Drop the variables that have more than 25% missing values, they will interfere with the model  
  
# Select out variables that could cause problems in the imputation process, include the ID also  
  
library(tidyverse)

## -- Attaching packages ---------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.2.1 v purrr 0.3.3  
## v tibble 2.1.3 v dplyr 0.8.3  
## v tidyr 1.0.0 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.5.0

## -- Conflicts ------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

dolphins <- dolphins %>%   
 dplyr::select(-'ID',-'PCB\_1268',-'DEP', -'NPE',-'Triclosan',-'BPA',-'Atrazine')  
  
sum(is.na(dolphins))

## [1] 38

### Perform the imputation

# We run the mice code with 0 iterations   
  
library(dplyr)  
library(mice)

##   
## Attaching package: 'mice'

## The following objects are masked from 'package:base':  
##   
## cbind, rbind

set.seed(145)  
imp <- mice(dolphins, maxit=0)  
  
# Extract predictorMatrix and methods of imputation   
predM = imp$predictorMatrix  
meth = imp$method  
head(predM)

## Sex Age\_Class Location Morbillivirus Hg Cu Se Cd Tl Pb Mn Co Zn  
## Sex 0 1 1 1 1 1 1 1 1 1 1 1 1  
## Age\_Class 1 0 1 1 1 1 1 1 1 1 1 1 1  
## Location 1 1 0 1 1 1 1 1 1 1 1 1 1  
## Morbillivirus 1 1 1 0 1 1 1 1 1 1 1 1 1  
## Hg 1 1 1 1 0 1 1 1 1 1 1 1 1  
## Cu 1 1 1 1 1 0 1 1 1 1 1 1 1  
## As Fe  
## Sex 1 1  
## Age\_Class 1 1  
## Location 1 1  
## Morbillivirus 1 1  
## Hg 1 1  
## Cu 1 1

With this command, we tell mice to impute the dolphins data, create 5 datasets, use predM as the predictor matrix and don’t print the imputation process. If you would like to see the process, set print as TRUE

imp2 <- mice(dolphins, maxit = 5,   
 predictorMatrix = predM,   
 method = meth, print = FALSE)  
  
  
# The imputation created 5 datasets with different plausible values for missing values. You can look at imputed datasets and values with the following commands:  
   
# Look at head and tail of imputed values for china\_econ variable   
head(imp2$imp$Hg)

## 1 2 3 4 5  
## 3 10.84 169.91 388.91 550.38 310.65  
## 6 194.65 209.96 377.12 172.33 550.38  
## 7 398.32 377.12 169.91 550.38 378.74  
## 8 398.32 550.38 639.95 639.95 639.95  
## 9 209.96 138.54 550.38 124.36 398.32  
## 10 464.85 138.54 550.38 169.91 464.85

tail(imp2$imp$Hg)

## 1 2 3 4 5  
## 9 209.96 138.54 550.38 124.36 398.32  
## 10 464.85 138.54 550.38 169.91 464.85  
## 11 149.98 138.54 279.18 25.96 209.96  
## 22 1402.21 1402.21 398.32 398.32 1402.21  
## 23 639.95 169.91 464.85 1415.50 378.74  
## 24 550.38 169.91 1415.50 378.74 1415.50

# Can also extract the first imputed, complete dataset and look at the first rows using the complete function  
dolphincomp <- mice::complete(imp2, 1)  
head(dolphincomp)

## Sex Age\_Class Location Morbillivirus Hg Cu Se Cd Tl Pb  
## 1 M Adult NC N 1402.21 34.02 448.11 149.41 0.05 0.64  
## 2 M Juvenile NC N 9.89 27.09 5.12 0.41 0.05 0.05  
## 3 M Adult NC Y 10.84 30.99 3.93 0.05 0.05 0.03  
## 4 M Adult NC Y 279.18 17.14 101.96 0.36 0.05 0.10  
## 5 F Juvenile NC Y 388.91 25.31 149.47 3.08 0.05 0.11  
## 6 M Juvenile FL Y 194.65 30.99 2.87 0.64 0.00 0.03  
## Mn Co Zn As Fe  
## 1 10.03 0.03 90.22 1.19 3674.68  
## 2 13.91 0.03 216.02 1.22 916.81  
## 3 9.00 0.02 218.81 0.46 1762.64  
## 4 4.21 0.03 143.76 0.50 1435.62  
## 5 15.30 0.03 240.93 0.68 1336.10  
## 6 12.57 0.02 207.06 0.78 779.88

# First, turn the datasets into long format  
dolimp\_long <- mice::complete(imp2, action="long", include = TRUE)  
dim(dolimp\_long)

## [1] 282 17

head(dolimp\_long)

## .imp .id Sex Age\_Class Location Morbillivirus Hg Cu Se Cd  
## 1 0 1 M Adult NC N 1402.21 34.02 448.11 149.41  
## 2 0 2 M Juvenile NC N 9.89 27.09 5.12 0.41  
## 3 0 3 M Adult NC Y NA NA NA NA  
## 4 0 4 M Adult NC Y 279.18 17.14 101.96 0.36  
## 5 0 5 F Juvenile NC Y 388.91 25.31 149.47 3.08  
## 6 0 6 M Juvenile FL Y NA 30.99 2.87 0.64  
## Tl Pb Mn Co Zn As Fe  
## 1 0.05 0.64 10.03 0.03 90.22 1.19 3674.68  
## 2 0.05 0.05 13.91 0.03 216.02 1.22 916.81  
## 3 NA NA NA NA NA NA NA  
## 4 0.05 0.10 4.21 0.03 143.76 0.50 1435.62  
## 5 0.05 0.11 15.30 0.03 240.93 0.68 1336.10  
## 6 0.00 0.03 12.57 0.02 207.06 0.78 NA

# Convert back to mids type - mice can work with this type  
dolimp\_long\_mids<-as.mids(dolimp\_long)

#### Meeting the Assumptions of Logistic regression model

* The dependent variable is binnary (all the demographic groups are binary)

str(dolphincomp[,1:4]) #Yes they are binary; all are factors with 2 levels

## 'data.frame': 47 obs. of 4 variables:  
## $ Sex : Factor w/ 2 levels "F","M": 2 2 2 2 1 2 1 2 2 2 ...  
## $ Age\_Class : Factor w/ 2 levels "Adult","Juvenile": 1 2 1 1 2 2 1 1 1 1 ...  
## $ Location : Factor w/ 2 levels "FL","NC": 2 2 2 2 2 1 1 1 1 1 ...  
## $ Morbillivirus: Factor w/ 2 levels "N","Y": 1 1 2 2 2 2 1 1 1 1 ...

* Observations to be independent of each other: #check for multicollinearity and drop indpependent variables with high correlation

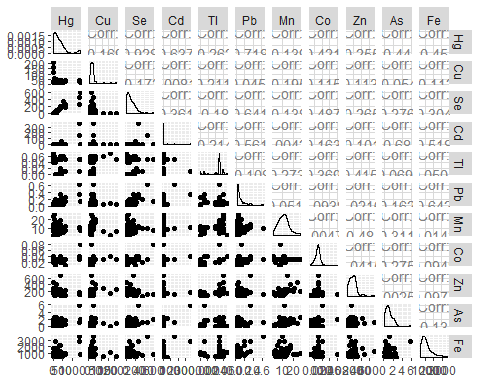
x=dolphincomp[,c(5:15)]   
library(GGally)

## Registered S3 method overwritten by 'GGally':  
## method from   
## +.gg ggplot2

##   
## Attaching package: 'GGally'

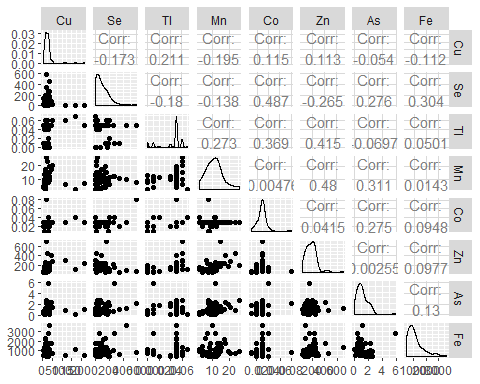
## The following object is masked from 'package:dplyr':  
##   
## nasa

ggpairs(x) #plot the correlation plot the identify the highly correlated variables



Above Hg, Cd, and Pb were correlated with atleast one other variable they will be excluded in the model

# plot the corrplot and check again  
y=dolphincomp[,c(6,7,9,11,12,13,14,15)]  
ggpairs(y)



* The sample size should be large enough N=(10k/p) k=10 independent variables p=smallest proportion of the demographic variables: p=0.25. Therefore this assumption is not met since N=10\*10/0.25 = 400 and our sample size is 47.

dim(dolphincomp)

## [1] 47 15

## GLM model with logistic link function

# Demographic group: Morbillivirus  
fitimp <- with(dolimp\_long\_mids, glm(Morbillivirus~Cu+Se+Tl+Mn+Co+Zn+As+Fe,family = 'binomial'))  
sum\_model=summary(pool(fitimp))  
sum\_model

## term estimate std.error statistic df p.value  
## 1 (Intercept) -5.151495e-01 2.421171e+00 -0.21276875 26.88271 0.8331125  
## 2 Cu -2.548677e-02 2.429830e-02 -1.04891162 34.47689 0.3015164  
## 3 Se -1.150604e-02 7.124442e-03 -1.61500906 34.55686 0.1154038  
## 4 Tl 2.490852e+01 4.444032e+01 0.56049367 24.02014 0.5803336  
## 5 Mn -8.168986e-03 1.013542e-01 -0.08059839 34.18735 0.9362313  
## 6 Co -2.959022e+01 1.247564e+02 -0.23718393 24.91817 0.8144533  
## 7 Zn -6.311534e-05 4.260925e-03 -0.01481259 34.63539 0.9882667  
## 8 As 4.506926e-02 6.261491e-01 0.07197848 34.93355 0.9430296  
## 9 Fe 7.053183e-04 7.633799e-04 0.92394134 22.90998 0.3651466

# Demographic group: Sex  
fitimp2 <- with(dolimp\_long\_mids, glm(Sex~Cu+Se+Tl+Mn+Co+Zn+As+Fe,family = 'binomial'))  
sum\_model2=summary(pool(fitimp2))  
sum\_model2

## term estimate std.error statistic df p.value  
## 1 (Intercept) 1.027367e+00 1.795790e+00 0.5720976 29.88718 0.5715343  
## 2 Cu -1.635651e-03 1.012922e-02 -0.1614785 34.84745 0.8726491  
## 3 Se 4.346377e-03 4.948117e-03 0.8783901 34.64085 0.3857843  
## 4 Tl 1.357312e+01 3.132589e+01 0.4332875 22.04322 0.6690164  
## 5 Mn 4.138121e-02 9.184749e-02 0.4505426 34.80216 0.6551137  
## 6 Co -1.045658e+02 7.370910e+01 -1.4186280 32.84778 0.1654266  
## 7 Zn 1.172520e-03 4.655763e-03 0.2518426 32.90929 0.8027289  
## 8 As -1.532807e-01 4.480095e-01 -0.3421371 30.48301 0.7345955  
## 9 Fe 1.096260e-03 8.552182e-04 1.2818485 23.20536 0.2125556

# Demographic group: Morbillivirus  
fitimp3 <- with(dolimp\_long\_mids, glm(Age\_Class~Cu+Se+Tl+Mn+Co+Zn+As+Fe,family = 'binomial'))  
sum\_model3=summary(pool(fitimp3))  
sum\_model3

## term estimate std.error statistic df p.value  
## 1 (Intercept) 4.095985e-01 3.386682e+00 0.1209439 32.96184 0.90446974  
## 2 Cu 1.777447e-01 1.313310e-01 1.3534100 23.13882 0.18899956  
## 3 Se -3.849467e-02 1.611909e-02 -2.3881414 33.75580 0.02267573  
## 4 Tl 2.017001e+01 1.008239e+02 0.2000520 16.52531 0.84388107  
## 5 Mn 4.583040e-01 2.907289e-01 1.5763965 34.66976 0.12401668  
## 6 Co -2.059320e+02 2.813233e+02 -0.7320118 21.13868 0.47219983  
## 7 Zn -1.093746e-02 1.411441e-02 -0.7749146 26.51447 0.44524636  
## 8 As -1.362586e+00 1.650921e+00 -0.8253493 22.74068 0.41774787  
## 9 Fe -9.078241e-04 1.821771e-03 -0.4983195 20.55082 0.62355009

# Demographic group: Location  
fitimp4 <- with(dolimp\_long\_mids, glm(Location~Cu+Se+Tl+Mn+Co+Zn+As+Fe,family = 'binomial'))  
sum\_model4=summary(pool(fitimp4))  
sum\_model4

## term estimate std.error statistic df p.value  
## 1 (Intercept) -2.697660e+00 3.608502e+00 -0.7475845 14.75886 0.46645604  
## 2 Cu -9.728354e-03 1.185616e-02 -0.8205314 35.25540 0.41742687  
## 3 Se -2.346611e-02 1.077065e-02 -2.1787084 22.53486 0.04007823  
## 4 Tl 1.547943e+02 7.547896e+01 2.0508269 26.57106 0.05026571  
## 5 Mn 3.367761e-02 1.486356e-01 0.2265784 32.39030 0.82217548  
## 6 Co -1.813546e+02 1.974381e+02 -0.9185389 20.56835 0.36898347  
## 7 Zn -7.112181e-03 4.743818e-03 -1.4992526 34.02299 0.14303075  
## 8 As 1.671725e+00 1.149756e+00 1.4539829 23.00820 0.15945708  
## 9 Fe 3.190849e-03 1.875978e-03 1.7008990 15.43105 0.10902132

### ODDS Ratios: Coefficient estimates for the regression equation

# Create a dataframe of the odds ratios  
odds\_ratios=exp(sum\_model$estimate) #convert the log odds to odds ratios  
df\_odd=as.data.frame(odds\_ratios) #create the dataframe  
df\_odd$independent\_variable=sum\_model$term #add the names of the independent variables  
df\_odd

## odds\_ratios independent\_variable  
## 1 5.974113e-01 (Intercept)  
## 2 9.748353e-01 Cu  
## 3 9.885599e-01 Se  
## 4 6.571012e+10 Tl  
## 5 9.918643e-01 Mn  
## 6 1.409719e-13 Co  
## 7 9.999369e-01 Zn  
## 8 1.046100e+00 As  
## 9 1.000706e+00 Fe

odds\_ratios2=exp(sum\_model2$estimate) #convert the log odds to odds ratios  
df\_odd2=as.data.frame(odds\_ratios2) #create the dataframe  
df\_odd2$independent\_variable=sum\_model2$term #add the names of the independent variables  
df\_odd2

## odds\_ratios2 independent\_variable  
## 1 2.793700e+00 (Intercept)  
## 2 9.983657e-01 Cu  
## 3 1.004356e+00 Se  
## 4 7.847469e+05 Tl  
## 5 1.042249e+00 Mn  
## 6 3.869498e-46 Co  
## 7 1.001173e+00 Zn  
## 8 8.578889e-01 As  
## 9 1.001097e+00 Fe

odds\_ratios3=exp(sum\_model3$estimate) #convert the log odds to odds ratios  
df\_odd3=as.data.frame(odds\_ratios3) #create the dataframe  
df\_odd3$independent\_variable=sum\_model3$term #add the names of the independent variables  
df\_odd3

## odds\_ratios3 independent\_variable  
## 1 1.506213e+00 (Intercept)  
## 2 1.194520e+00 Cu  
## 3 9.622368e-01 Se  
## 4 5.750769e+08 Tl  
## 5 1.581390e+00 Mn  
## 6 3.671666e-90 Co  
## 7 9.891221e-01 Zn  
## 8 2.559978e-01 As  
## 9 9.990926e-01 Fe

odds\_ratios4=exp(sum\_model4$estimate) #convert the log odds to odds ratios  
df\_odd4=as.data.frame(odds\_ratios4) #create the dataframe  
df\_odd4$independent\_variable=sum\_model4$term #add the names of the independent variables  
df\_odd4

## odds\_ratios4 independent\_variable  
## 1 6.736296e-02 (Intercept)  
## 2 9.903188e-01 Cu  
## 3 9.768071e-01 Se  
## 4 1.683850e+67 Tl  
## 5 1.034251e+00 Mn  
## 6 1.732631e-79 Co  
## 7 9.929131e-01 Zn  
## 8 5.321341e+00 As  
## 9 1.003196e+00 Fe

### Comparison of toxin concentration in each demographic group

Doll\_loc <- dolphincomp %>%  
 select(Location,Hg,Cu,Se,Cd,Tl,Pb,Mn,Co,Zn,As) %>%  
 group\_by(Location) %>%  
 summarise(Hg = mean(Hg), Cu = mean(Cu),Se = mean(Se),  
 Cd = mean(Cd), Tl = mean(Tl),Pb = mean(Pb),  
 Mn = mean(Mn), Co = mean(Co),Zn = mean(Zn),  
 As = mean(As))  
  
Doll\_loc

## # A tibble: 2 x 11  
## Location Hg Cu Se Cd Tl Pb Mn Co Zn As  
## <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 FL 475. 34.3 168. 24.9 0.0294 0.138 9.50 0.0294 161. 1.23  
## 2 NC 184. 33.1 68.4 5.29 0.0507 0.0867 12.3 0.0290 203. 1.18

Doll\_sex <- dolphincomp %>%  
 select(Sex,Hg,Cu,Se,Cd,Tl,Pb,Mn,Co,Zn,As) %>%  
 group\_by(Sex) %>%  
 summarise(Hg = mean(Hg), Cu = mean(Cu),Se = mean(Se),  
 Cd = mean(Cd), Tl = mean(Tl),Pb = mean(Pb),  
 Mn = mean(Mn), Co = mean(Co),Zn = mean(Zn),  
 As = mean(As))  
  
Doll\_sex

## # A tibble: 2 x 11  
## Sex Hg Cu Se Cd Tl Pb Mn Co Zn As  
## <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 F 323. 41.3 108. 25.0 0.0447 0.0853 10.7 0.0324 172. 1.40  
## 2 M 270. 29.1 103. 5.24 0.042 0.117 11.6 0.0273 197. 1.09

Doll\_age <- dolphincomp %>%  
 select(Age\_Class,Hg,Cu,Se,Cd,Tl,Pb,Mn,Co,Zn,As) %>%  
 group\_by(Age\_Class) %>%  
 summarise(Hg = mean(Hg), Cu = mean(Cu),Se = mean(Se),  
 Cd = mean(Cd), Tl = mean(Tl),Pb = mean(Pb),  
 Mn = mean(Mn), Co = mean(Co),Zn = mean(Zn),  
 As = mean(As))  
Doll\_age

## # A tibble: 2 x 11  
## Age\_Class Hg Cu Se Cd Tl Pb Mn Co Zn As  
## <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Adult 400. 23.1 147. 18.6 0.0410 0.129 10.3 0.03 176. 1.24  
## 2 Juvenile 73.6 53.8 23.0 0.366 0.0469 0.0594 13.3 0.0275 213. 1.12

Doll\_morbillivirus <- dolphincomp %>%  
 select(Morbillivirus,Hg,Cu,Se,Cd,Tl,Pb,Mn,Co,Zn,As) %>%  
 group\_by(Morbillivirus) %>%  
 summarise(Hg = mean(Hg), Cu = mean(Cu),Se = mean(Se),  
 Cd = mean(Cd), Tl = mean(Tl),Pb = mean(Pb),  
 Mn = mean(Mn), Co = mean(Co),Zn = mean(Zn),  
 As = mean(As))  
Doll\_morbillivirus

## # A tibble: 2 x 11  
## Morbillivirus Hg Cu Se Cd Tl Pb Mn Co Zn As  
## <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 N 338. 36.7 123. 16.5 0.0417 0.12 11.0 0.0297 182. 1.26  
## 2 Y 148. 24.3 51.0 0.472 0.0467 0.0625 12.3 0.0275 206. 1.02