fish_feed_PFAS_analysis

2023-03-02

```
library(nlme)

# Read in raw data
feeds <- read.csv('feed_pfas.csv')

# Count occurrences of each unique value in the mfr (manufacturer) column
mfr_counts <- table(feeds$mfr)
print(mfr_counts)
```

```
##
## A B C D E
## 3 2 4 3 1
```

```
# Because manufacturers B and E have occurrences of less than 3, I will omit them from my analyses

# Count occurrences of each unique value in the protein_source column
protein_source_counts <- table(feeds$protein_source)
print(protein_source_counts)
```

```
##
## ap fm pp sb
##  1  6  1  5
```

```
# Likewise, because protein sources 'ap' and 'pp' each only occur once, I will omit them from my analys

# Subset the data to exclude certain protein_source and mfr values
excluded_ps <- c('ap', 'pp')
excluded_mfr <- c('B', 'E')
subset_feeds <- subset(feeds, !(protein_source %in% excluded_ps) & !(mfr %in% excluded_mfr))

# Factor categorical variables
subset_feeds$mfr <- factor(subset_feeds$mfr)
subset_feeds$protein_source <- factor(subset_feeds$protein_source)
```

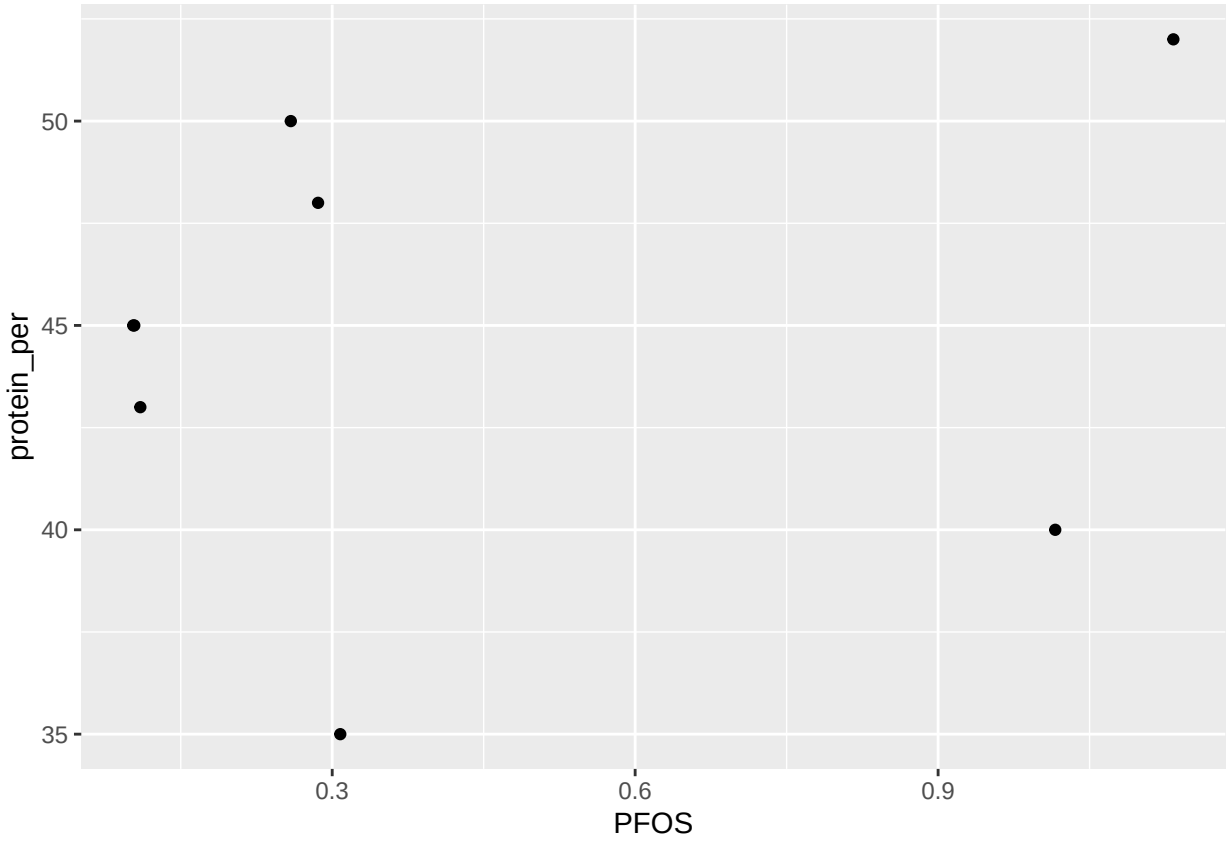
```
### PFOS Model

# Run the multiple linear regression
PFOSmodel <- lm(PFOS ~ protein_per + protein_source + mfr, data = subset_feeds)

# View the summary of the model
summary(PFOSmodel)
```

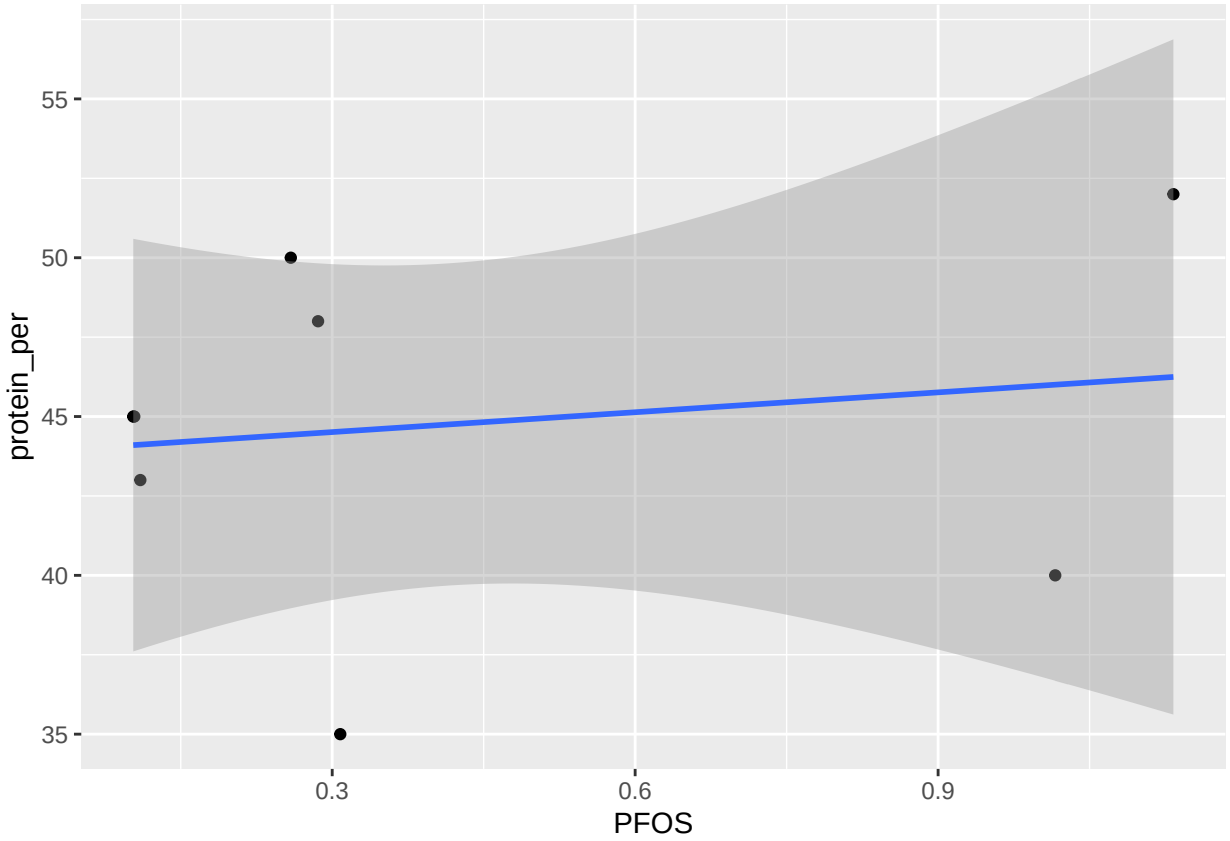
```
##
## Call:
## lm(formula = PFOS ~ protein_per + protein_source + mfr, data = subset_feeds)
##
## Residuals:
##          1          2          4          5          6          8         12
## -5.165e-02  5.165e-02  1.858e-01 -2.429e-01  1.203e-01  1.041e-17 -6.312e-02
##         13
## -3.296e-17
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -4.32296    1.14246  -3.784   0.0324 *
## protein_per       0.12094    0.02585   4.678   0.0185 *
## protein_sourcesb  0.44970    0.30810   1.460   0.2405
## mfrC             -0.95323    0.34796  -2.739   0.0714 .
## mfrD             -1.46505    0.31026  -4.722   0.0180 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1978 on 3 degrees of freedom
## Multiple R-squared:  0.9035, Adjusted R-squared:  0.7749
## F-statistic: 7.025 on 4 and 3 DF,  p-value: 0.07056
```

```
### scatter plot of protein % and PFOS
library(ggplot2)
ggplot(subset_feeds, aes(x = PFOS, y = protein_per)) +
  geom_point()
```

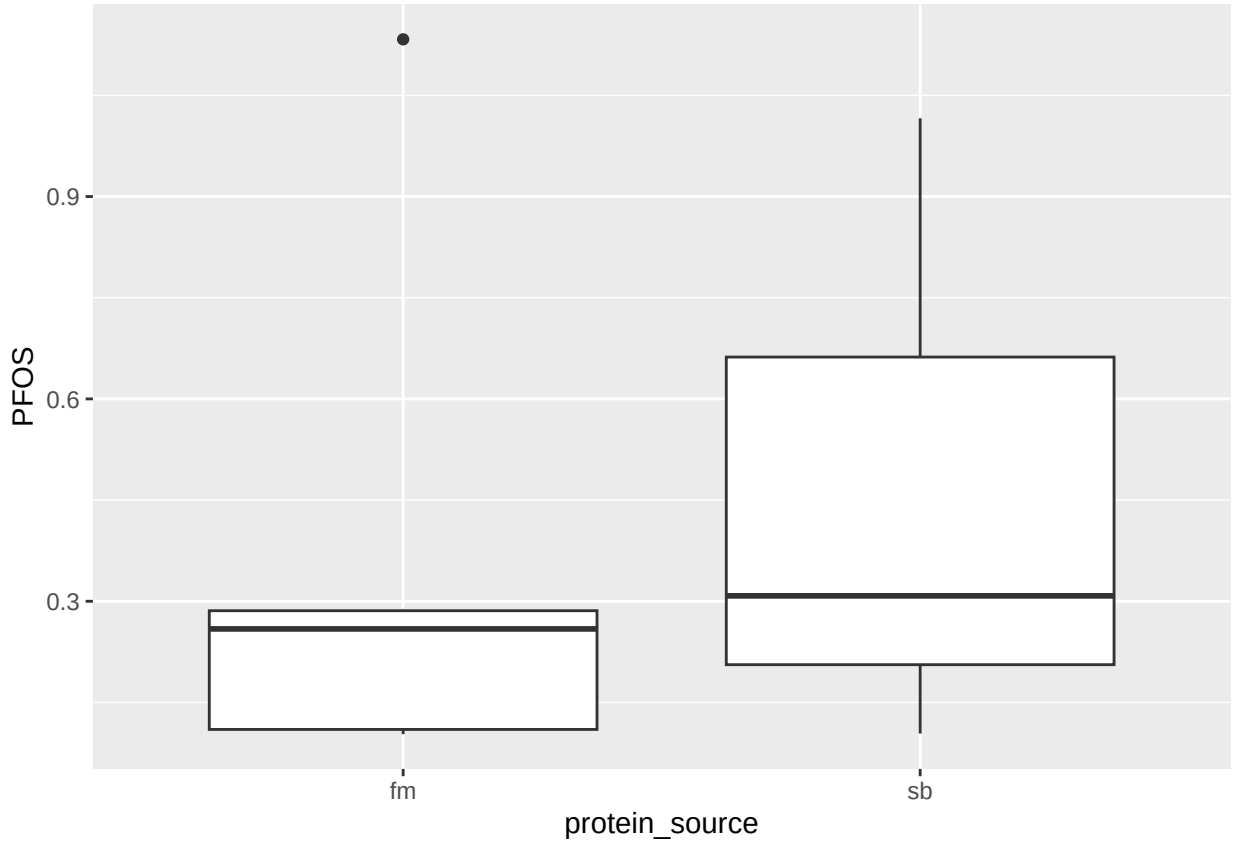



```
ggplot(subset_feeds, aes(x = PFOS, y = protein_per)) +
  geom_point() +
  geom_smooth(method = "lm")
```

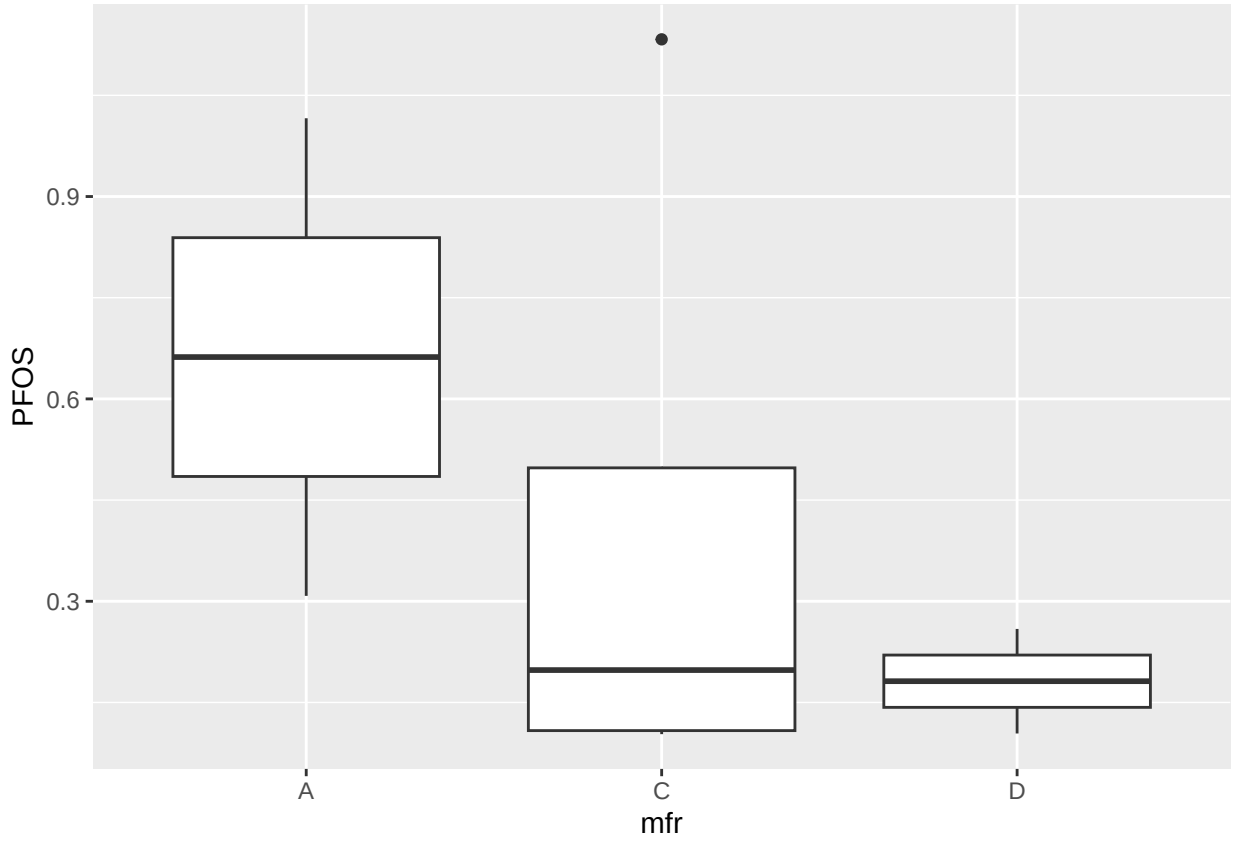
```
## `geom_smooth()` using formula = 'y ~ x'
```




```
# Boxplot of protein source and PFOS
ggplot(subset_feeds, aes(x = protein_source, y = PFOS)) +
  geom_boxplot()
```




```
# Boxplot of manufacturer and PFOS
ggplot(subset_feeds, aes(x = mfr, y = PFOS)) +
  geom_boxplot()
```




```
### PFNA Model

# Run the multiple linear regression
PFNAmodel <- lm(PFNA ~ protein_per + protein_source + mfr, data = subset_feeds)

# View the summary of the model
summary(PFNAmodel)
```

```
##
## Call:
## lm(formula = PFNA ~ protein_per + protein_source + mfr, data = subset_feeds)
##
## Residuals:
##          1          2          4          5          6          8         12
##  3.924e-01 -3.924e-01 -1.557e-01  1.654e-02  2.287e-01 -3.192e-16 -8.959e-02
##         13
##  2.220e-16
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       2.37132    2.09043   1.134    0.339
## protein_per      -0.03504    0.04731  -0.741    0.513
## protein_sourcesb -0.17721    0.56375  -0.314    0.774
## mfrC             -0.60581    0.63669  -0.951    0.412
## mfrD             -0.51718    0.56770  -0.911    0.429
##
```

```
## Residual standard error: 0.3618 on 3 degrees of freedom
## Multiple R-squared:  0.7087, Adjusted R-squared:  0.3202
## F-statistic: 1.824 on 4 and 3 DF,  p-value: 0.3244
```

```
# Nothing significant. This is likely due to low power and low variability. The LOD (limit of detection
```

```
### PFDA Model

# Run the multiple linear regression
PFDAmodel <- lm(PFDA ~ protein_per + protein_source + mfr, data = subset_feeds)

# View the summary of the model
summary(PFDAmodel)
```

```
##
## Call:
## lm(formula = PFDA ~ protein_per + protein_source + mfr, data = subset_feeds)
##
## Residuals:
##          1          2          4          5          6          8         12
##  3.802e-01 -3.802e-01 -1.637e-01  4.104e-02  2.072e-01 -2.498e-16 -8.458e-02
##         13
##  2.637e-16
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       2.56238    2.02225   1.267    0.295
## protein_per      -0.04154    0.04577  -0.908    0.431
## protein_sourcesb -0.20869    0.54536  -0.383    0.727
## mfrC             -0.50958    0.61593  -0.827    0.469
## mfrD             -0.38346    0.54919  -0.698    0.535
##
## Residual standard error: 0.35 on 3 degrees of freedom
## Multiple R-squared:  0.692,  Adjusted R-squared:  0.2813
## F-statistic: 1.685 on 4 and 3 DF,  p-value: 0.3484
```

```
### PFUnA Model

# Run the multiple linear regression
PFUnAmodel <- lm(PFUnA ~ protein_per + protein_source + mfr, data = subset_feeds)

# View the summary of the model
summary(PFUnAmodel)
```

```
##
## Call:
## lm(formula = PFUnA ~ protein_per + protein_source + mfr, data = subset_feeds)
##
## Residuals:
##          1          2          4          5          6          8         12
## -9.290e-02  9.290e-02  5.181e-02 -2.939e-02 -3.896e-02  3.123e-17  1.653e-02
##         13
```

```
## -6.592e-17
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -0.43274    0.50142  -0.863    0.452
## protein_per       0.01564    0.01135   1.378    0.262
## protein_sourcesb  0.07721    0.13522   0.571    0.608
## mfrC             -0.18863    0.15272  -1.235    0.305
## mfrD             -0.24731    0.13617  -1.816    0.167
##
## Residual standard error: 0.08679 on 3 degrees of freedom
## Multiple R-squared:  0.6182, Adjusted R-squared:  0.1092
## F-statistic: 1.214 on 4 and 3 DF,  p-value: 0.4547
```

```
# Overall, it appears that PFOS is the only compound that appears to differ significantly between manuf
# The other three compounds found no significant differences. There are several explanations and conclu
# Firstly, as mentioned above, the limit of detection (LOD) was used as the concentration of PFNA, PFDA
# The LOD is the lowest concentration that can be reliably detected and distinguished from background n
# Therefore, it is likely that for the feeds in which the LOD was used, then that compound was either v
# Because the LOD was used in about half of the feeds for each of the three compounds, there is likely

# Next, we can conclude that there are little amounts of PFAS compounds present in commercial fish feed
# PFOS was the only compound that was present in all of the feeds tested, and there were some significa
# PFOS is the most widespread PFAS compound currently known, so it is not surprising to see it in every
```