Group 3: Storm Events and Water Impairments

Marc Los Huertos

2025-04-30

Correlation Tests

A correlation test in R is used to evaluate the strength and direction of a relationship between two numeric variables. The most common correlation test is Pearson's correlation, but others include Spearman's and Kendall's if the data isn't normally distributed.

Here's a simple example using Pearson's correlation with the built-in cor.test() function in R.

See Correlation Test Tutorial

An example

```
# Sample data
height <- c(160, 165, 170, 175, 180)
weight <- c(55, 60, 65, 70, 75)

# Perform Pearson correlation test
cor.test(height, weight, method = "pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  height and weight
## t = Inf, df = 3, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  1 1
## sample estimates:
## cor
##   1
```

cor: 0.9984 → Strong positive correlation

p-value: 0.000635 → Very small, so the result is statistically significant

95% CI: [0.9771, 0.9999] → We're 95% confident the true correlation lies in this interval.

Hypotheses

Data Set

```
group3.csv = "/home/mwl04747/RTricks/00_Project_Group_Demos/Group3_FakData.csv"
riohonda = read.csv(group3.csv)
head(riohonda)
```

```
##   Sample.ID       Date Phosphate..mg.L. Ammonia..mg.L.
## 1        S1 2025-04-01              0.8            0.2
## 2        S2 2025-04-01              1.0            0.4
## 3        S3 2025-04-01              1.5            0.6
## 4        S4 2025-04-02              2.0            1.2
## 5        S5 2025-04-02              0.5            0.1
## 6        S6 2025-04-03              0.4            0.1
```

```
names(riohonda) = c("ID", "Date", "Phosphate", "Ammonia")

data2.csv = "/home/mwl04747/RTricks/00_Project_Group_Demos/Group3_FakeData2.csv"
data2 = read.csv(data2.csv);

names(data2) = c("ID", "Distance", "Date", "Phosphate", "Ammonia",  "Nitrate")
```

```
#install.packages("ggpubr")

library(ggpubr)
```

```
## Loading required package: ggplot2
```

```
library(ggplot2)
```

Summary Stats

```
summary(riohonda)
```

```
##       ID                Date             Phosphate        Ammonia
##  Length:6           Length:6           Min.   :0.400   Min.   :0.1000
##  Class :character   Class :character   1st Qu.:0.575   1st Qu.:0.1250
##  Mode  :character   Mode  :character   Median :0.900   Median :0.3000
##                                        Mean   :1.033   Mean   :0.4333
##                                        3rd Qu.:1.375   3rd Qu.:0.5500
##                                        Max.   :2.000   Max.   :1.2000
```

```
summary(data2)
```

```
##       ID                Distance         Date             Phosphate
##  Length:6           Min.   :0.500   Length:6           Min.   :0.400
##  Class :character   1st Qu.:1.125   Class :character   1st Qu.:0.575
##  Mode  :character   Median :1.750   Mode  :character   Median :0.900
##                     Mean   :1.750                      Mean   :1.033
##                     3rd Qu.:2.375                      3rd Qu.:1.375
##                     Max.   :3.000                      Max.   :2.000
##     Ammonia          Nitrate
##  Min.   :0.1000   Min.   :0.0900
```

```
##  1st Qu.:0.1250   1st Qu.:0.5650
##  Median :0.3000   Median :0.9850
##  Mean   :0.4333   Mean   :0.8567
##  3rd Qu.:0.5500   3rd Qu.:1.2025
##  Max.   :1.2000   Max.   :1.3800
```

Hypothesis Tests

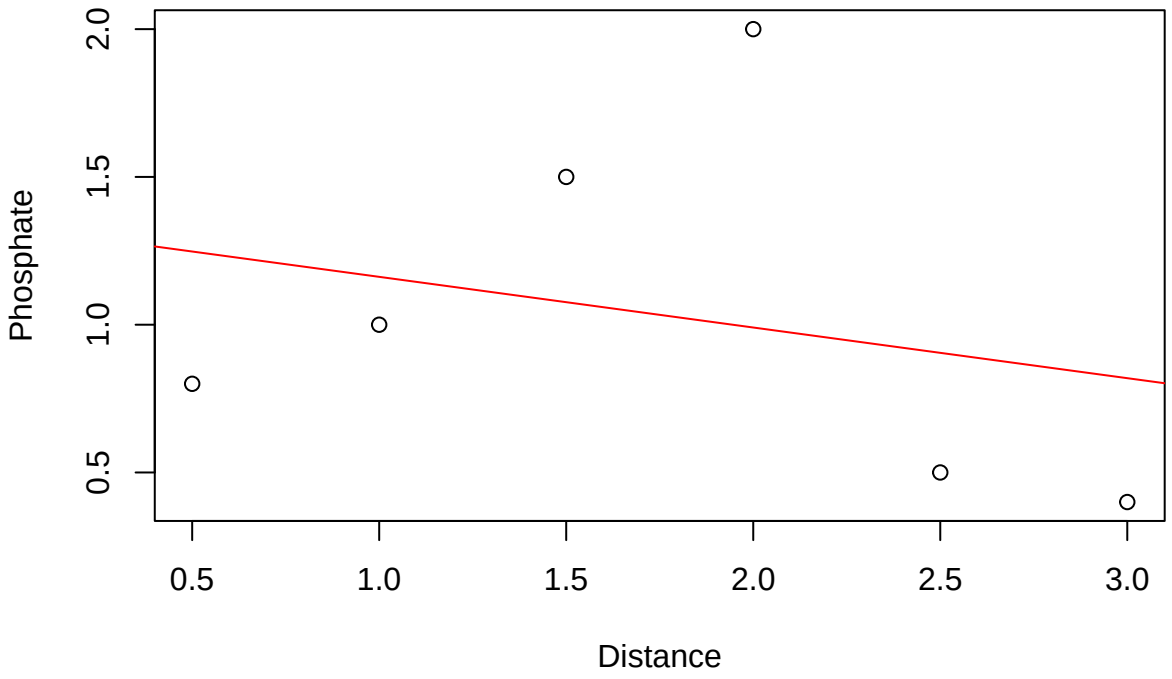
```
cor(riohonda$Ammonia, riohonda$Phosphate,  method = "pearson")
```

```
## [1] 0.971399
```

```
# Create correlation matrix
```

Plots

```
plot(Phosphate ~ Distance, data=data2)
abline(coef(lm(Phosphate ~ Distance, data=data2)), col="red")
```




```
summary(lm(Phosphate ~ Distance, data=data2))
```

```
##
## Call:
## lm(formula = Phosphate ~ Distance, data = data2)
##
## Residuals:
##       1       2       3       4       5       6
## -0.4476 -0.1619  0.4238  1.0095 -0.4048 -0.4190
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)   1.3333     0.6184   2.156   0.0973 .
## Distance     -0.1714     0.3176  -0.540   0.6180
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6642 on 4 degrees of freedom
## Multiple R-squared:  0.06791,    Adjusted R-squared:  -0.1651
## F-statistic: 0.2914 on 1 and 4 DF,  p-value: 0.618
```

k