

# Null Hypothesis

Sophie Wulfinf

2/12/2022

## Note

This is what I've come up with regarding how to tackle the null hypothesis. I know we've taken NLA data out of circulation but I'm going to continue to use it here only because I was using it earlier as an example set and I wanted to get feedback from you all about the methods and learn a little more about the changes to data before I applied it to other datasets.

```
model_2013 <- lm(NITRATE_N_RESULT~DOC_RESULT, data = test_data)
summary(model_2013)
```

```
##
## Call:
## lm(formula = NITRATE_N_RESULT ~ DOC_RESULT, data = test_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.8392 -0.7624 -0.6329 -0.2571  24.3065
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.844810   0.061624  13.709  <2e-16 ***
## DOC_RESULT  -0.015579   0.007813  -1.994   0.0463 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.224 on 2238 degrees of freedom
## (21 observations deleted due to missingness)
## Multiple R-squared:  0.001773, Adjusted R-squared:  0.001327
## F-statistic: 3.976 on 1 and 2238 DF, p-value: 0.04628
```

```
test_dataLOGS <- test_data %>%
  mutate(log_Nitrate = log(NITRATE_N_RESULT)) %>%
  mutate(log_TP = log(PTL_RESULT)) %>%
  mutate(log_DOC = log(DOC_RESULT))
# Duplicate data
test_dataLOGS$log_Nitrate[is.na(test_dataLOGS$log_Nitrate) | test_dataLOGS$log_Nitrate == "-Inf"] <- NA
quantile(test_dataLOGS$DOC_RESULT)
```

```
##      0%    25%    50%    75%   100%
##  0.19  1.81  3.42  6.01 62.76
```

```

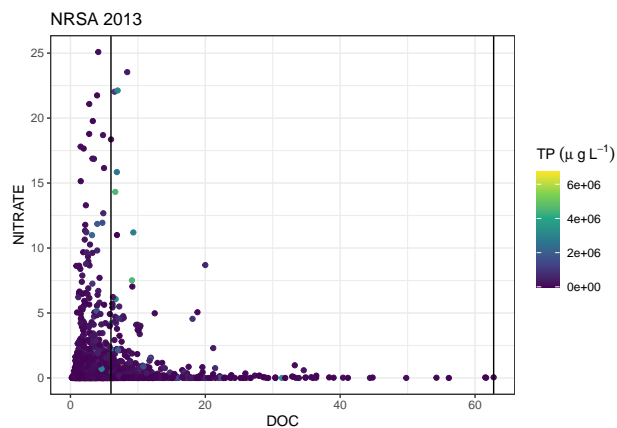
q4 <- quantile(test_data$DOC_RESULT)[4]
q5 <- quantile(test_data$DOC_RESULT)[5]

lowDOC_data <- test_data %>%
  filter(DOC_RESULT < quantile(test_data$DOC_RESULT)[4])

par(mfrow = c(1,2))
p <- ggplot(test_dataLOGS) +
  geom_point(aes(DOC_RESULT, NITRATE_N_RESULT, color = PTL_RESULT * 1000)) +
  scale_color_viridis_c("TP"~(mu~g~L^-1)) +
  theme_bw() +
  labs(title = "NRSA 2013",
        x = "DOC", y = "NITRATE")
p + geom_vline(aes(xintercept = q4)) + geom_vline(aes(xintercept = q5))

```

## Warning: Removed 21 rows containing missing values (geom\_point).

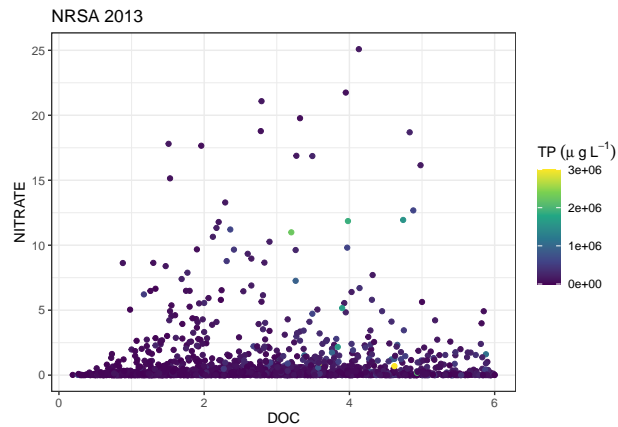


```

ggplot(lowDOC_data) +
  geom_point(aes(DOC_RESULT, NITRATE_N_RESULT, color = PTL_RESULT * 1000)) +
  scale_color_viridis_c("TP"~(mu~g~L^-1)) +
  theme_bw() +
  labs(title = "NRSA 2013",
        x = "DOC", y = "NITRATE")

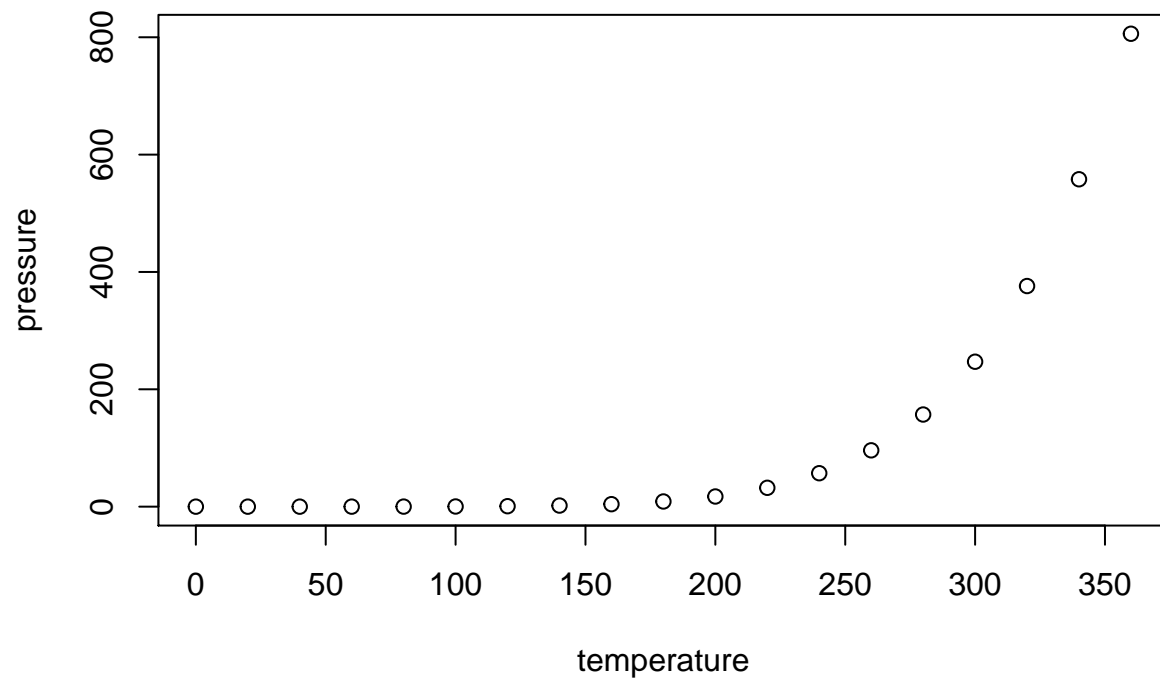
```

## Warning: Removed 4 rows containing missing values (geom\_point).



### Note:

For now I am using the 4th quantile of this specific dataset. I realize that will change both where the quantile is/where we choose to cut off data



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.