Statistical Report Nitrate-Ph-Permeability Of Region Orana And Riverina

Niyati

INTRODUCTION

The following statistical report analyzes a simulated dataset containing data on 247 samples from the ORAN and RIVERINA regions. The dataset contains variables like Subject ID, region, pH level, nitrate content and permeability. The main objectives of this analysis are to answer the following questions: 1. Is there any difference in the average nitrate of Orana and Riverina specimens? 2. What is the relation between the pH of specimens and the permeability?

ORANA AND RIVERINA DATASET

ID	region	pH	nitrate	permeability
subj1	Orana	5.0	8.40	58.8
subj2	Riverina	6.5	8.56	89.1
subj3	Riverina	6.3	9.55	85.0
subj4	Orana	4.6	8.44	55.5
subj5	Riverina	5.6	6.39	84.6
subj6	Riverina	6.8	9.01	89.2
subj7	Riverina	6.8	7.73	92.5

1) To check is there any difference in the average nitrate of Orana and riverina specimens we have to perform two-sample t-test with assuming equal variance as their both the samples have different no. of observations.

Orana	Riverina
8.40	8.56
8.44	9.55
9.12	6.39
7.41	9.01
9.24	7.73
7.20	5.95
8.11	9.30

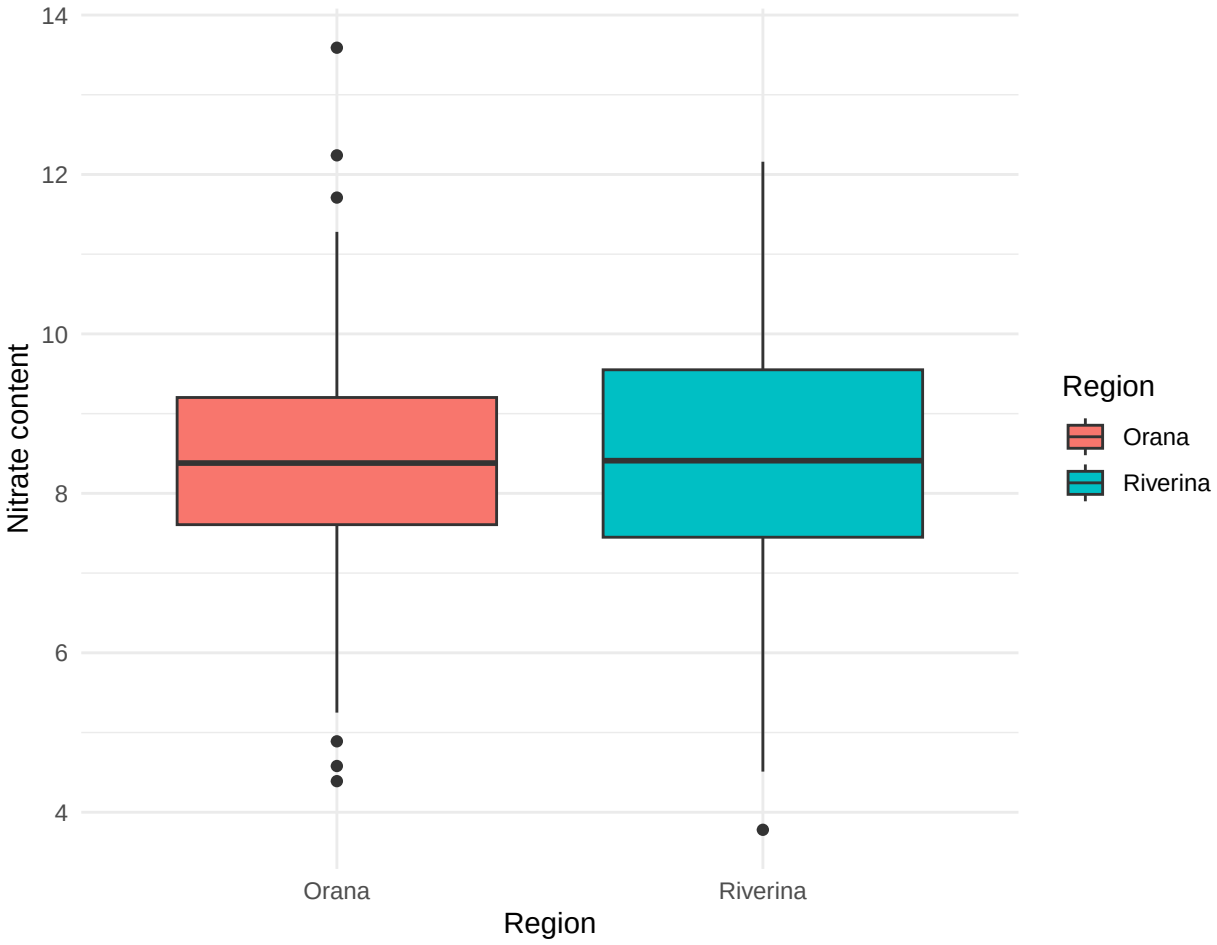


Fig1.Box Plot Comparison of Regions.

Box Plot visualise that mean and spread of both region are almost same though we can assume to be equal to check the t stat value and moreover,Orana contains some outliers too that may effect some of statistical analysis.

Statistical Analysis

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

```
Two Sample t-test
```

data: Orana and Riverina t = -0.19351, df = 245, p-value = 0.8467 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -0.4231283 0.3474273 sample estimates: mean of x mean of y 8.383106 8.420957

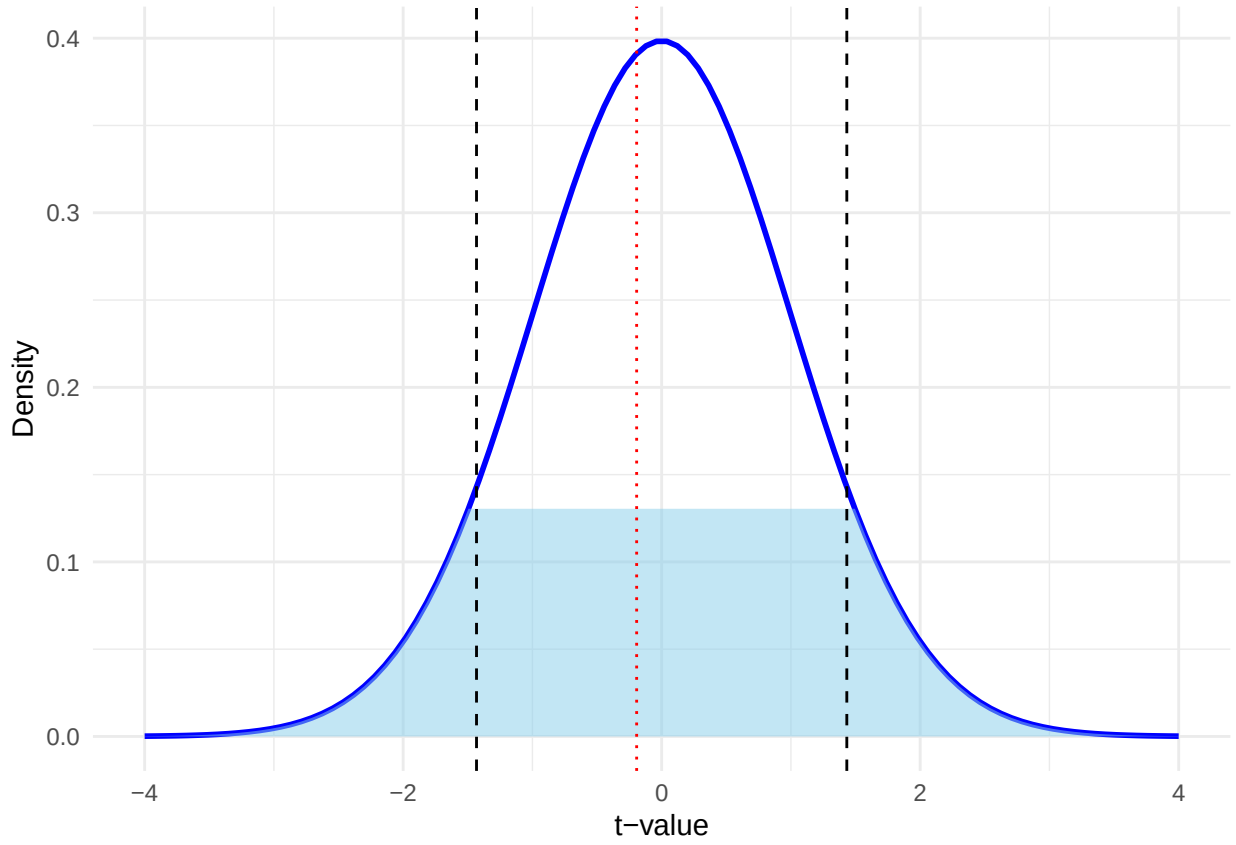


Fig2.t-test distribution of region

T-Stat value is -0.19351 and p-value is 0.8467 which is greater than >0.05 therefore we will donot reject H_o and conclude that their is insufficent evidence for difference in the average nitrate of Orana and Riverina specimens.

95% confidence interval is : (-0.4231283 0.3474273)

The values -0.4231283 and 0.3474273 represent the lower and upper bounds, respectively, of the 95% confidence interval for a particular statistic. That ensures that the true value of the parameters lie between this interval and we are 95% sure about this.

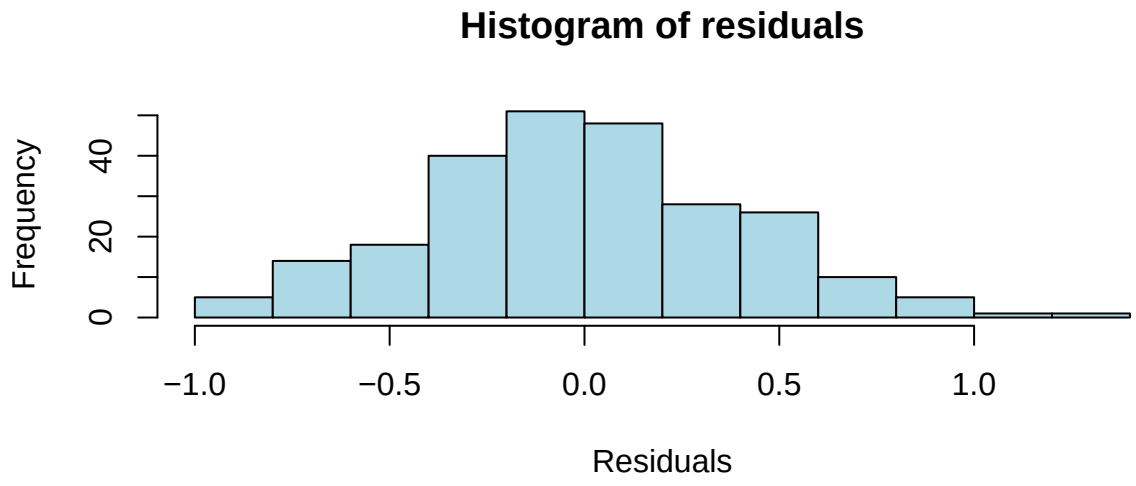
2)To check the relation between the pH of specimens and the permeability we will conduct a simple linear regression

Simple Linear Regression

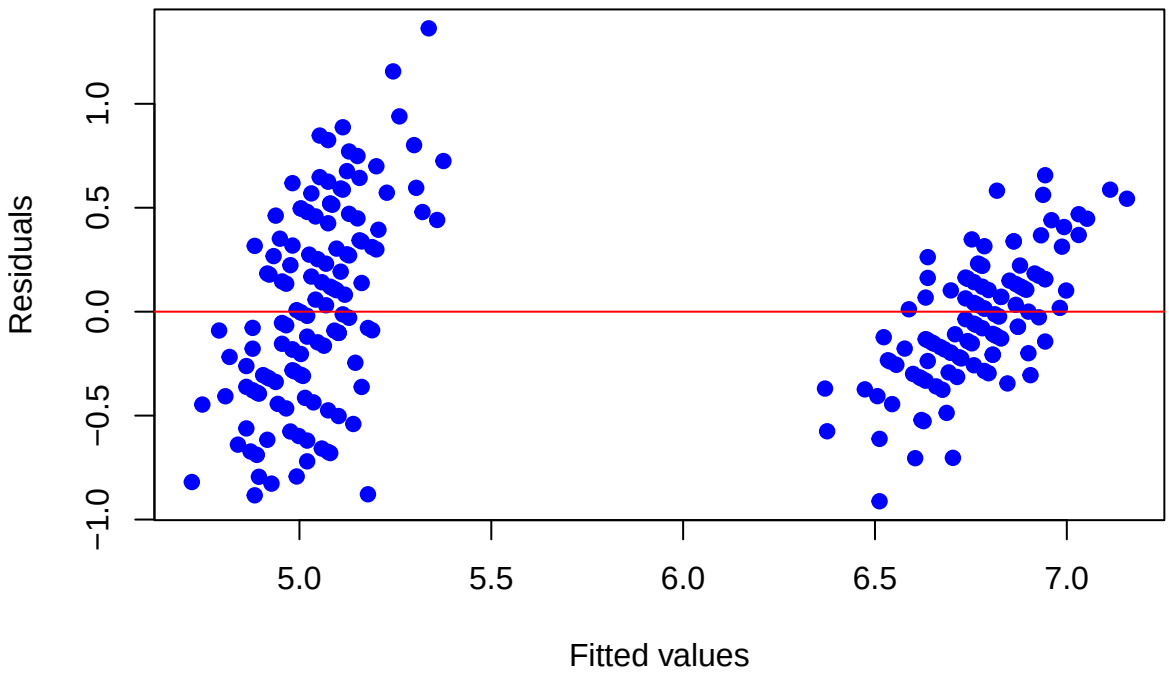
```
##
## Call:
## lm(formula = pH ~ permeability, data = dataset)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9120 -0.2862 -0.0201  0.2690  1.3630
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.888849   0.120443   15.68   <2e-16 ***
## permeability 0.054647   0.001625   33.63   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4076 on 245 degrees of freedom
## Multiple R-squared:  0.8219, Adjusted R-squared:  0.8212
## F-statistic:  1131 on 1 and 245 DF,  p-value: < 2.2e-16
```

Coefficient Interpretation: The Intercept (1.888849) is an estimate of the pH value at the point where the permeability reaches 0. In practice, this may not be directly physical, especially if the data cannot be zero. The Permeability Coefficient (0.054647) indicates that for each unit increase in the permeability value, the pH value increases by 0.05457 units, assuming that all other variables are constant.

$pH = 1.888849 + 0.054647 \times \text{permeability}$




```
## Fig3.Histogram of Residuals
```




```
## Fig4. Residual plot
```

Goodness of fit:This study found that the linear relationship between pH values and permeability was responsible for approximately 82.19 percent of the variance in pH values. This indicates that the model is reasonably well-fitted to the data.Additionally, the adjusted RSQ (0.8212) was close to the multiple RSquare, indicating that the model was not oversimplifying the data.

Residual Standard Error:The residual standard error (RSD) of the study was 0.4076, indicating the average amount of deviation from the predicted values of the model.

CONCLUSION

1) Based on statistical analysis, the results do not support a statistically significant difference in average nitrate concentration between the ORAN and RIVERINA samples.

Therefore, we do not accept the null hypothesis and conclude there is not enough evidence to support a meaningful difference in nitrate concentration between these two regions.

It is important to note the limitations of this study, including the size of the sample and any possible confounders that may have affected the results. A larger and more varied sample size, as well as an investigation of additional variables, may provide a better understanding of nitrate content differences between the ORANA and RIVERINA regions. This research adds to the existing evidence base and

emphasizes the necessity of ongoing research and surveillance of nitrate concentrations in various geographical areas in order to ensure successful environmental management and agricultural operations.

2) The statistical analysis conducted revealed a strong positive linear relationship between the pH levels of the samples and the permeability of the water.

It can be concluded that pH plays a major role in the permeability of samples, thus emphasizing the significance of monitoring and comprehending these relationships for the successful management of the environment and agricultural practices.

Nevertheless, it is important to be aware of any limitations, such as the size of the sample or other variables, that may limit the applicability of the findings. Further investigation with a larger and varied sample could provide further insight into this fundamental ecological relationship.

Appendix

ASSUMPTION IN FIRST TEST AND NORMALITY CHECK

1. Both samples have equal spread $\sigma_1 = \sigma_2$

To check normality

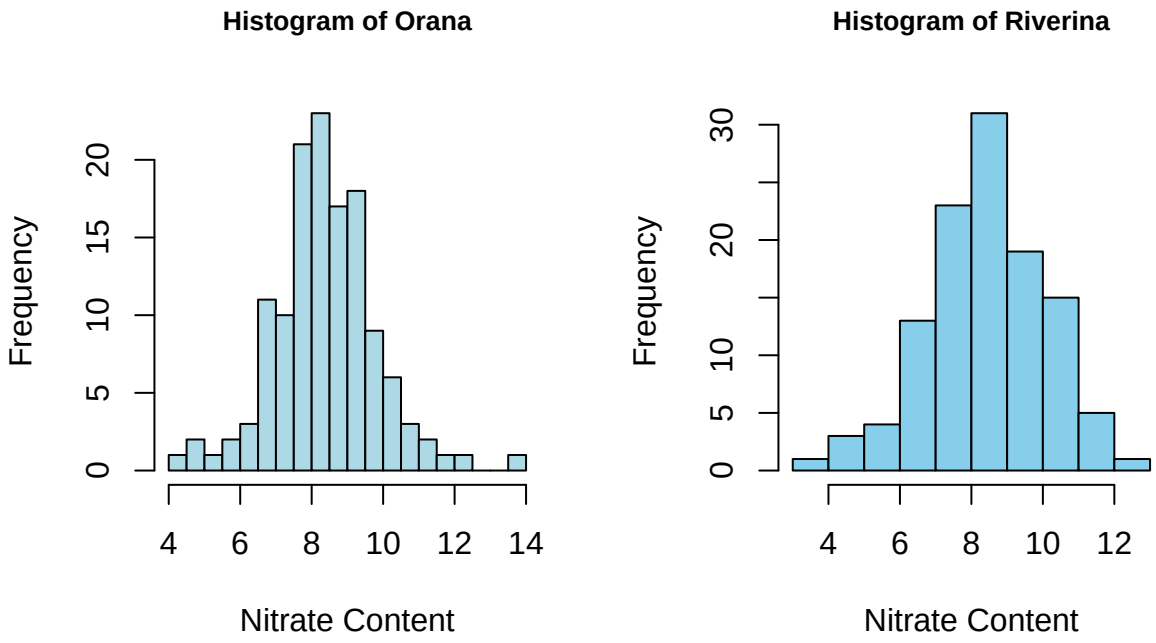


Fig a.1.Normality check of Orana and Riverna

Above histogram depicts that the specimen collected from Orana and Riverina region are**normally distributed** in terms of nitrate content.Hence two sample t-test conditions are being satisfied

FORMULA AND METHOD OF T-TEST

The results of the two-sample t-test are as follows:

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

1. $\bar{X}_1$ and $\bar{X}_2$ represent the sample means of the two groups,
2. s_1 and s_2 are the sample standard deviations of the two groups,
3. n_1 and n_2 are the sizes of the two groups.

The null hypothesis is rejected if the p-value is less than the chosen significance level, typically 0.05.

CORRELATION OF FEATURES INFLUENCING SIMPLE LINEAR REGRESSION

Correlation between pH of the region and permeability is : 0.906603 Correlation in Orana of pH and permeability: 0.7101706Correlation in Riverina of pH and permeability: 0.8593329

A correlation coefficient (CoR) of 0.906603 shows a positive linear correlation between pH levels and sample permeability. This indicates that the higher the pH, the higher the permeability, and vice-versa.

When examining the correlations between the two regions on their own, it is clear that the Orana region and Riverina region demonstrate a strong positive correlation between the pH level and the permeability level. The correlation coefficient for the Orana region is estimated to be 0.734862, while the correlation coefficient for Riverina is slightly higher at 0.8093329. This indicates that the pH-permeability relationship is consistent across both regions, suggesting a similar pattern across both geographic areas.

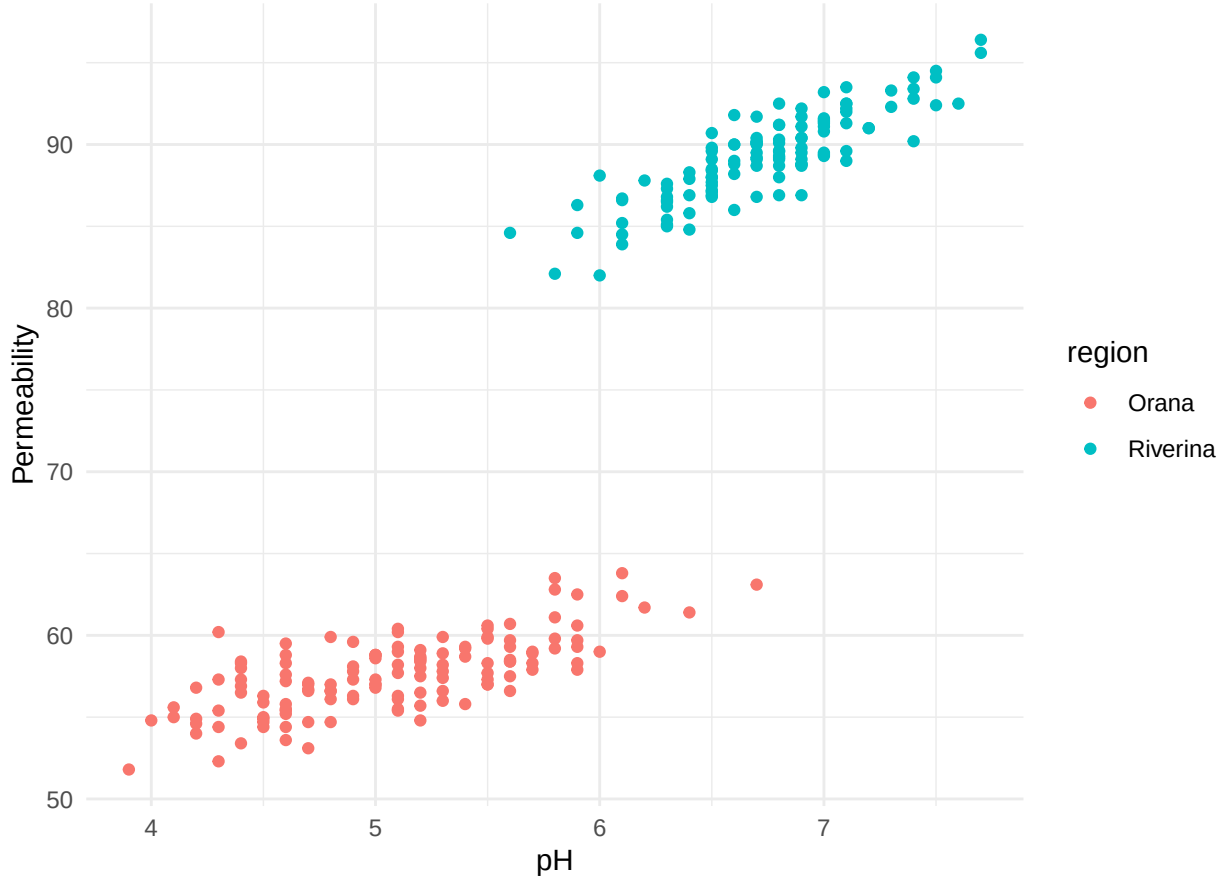


Fig a.2.Scatter Plot of pH vs. Permeability

FORMULA AND METHOD OF SIMPLE LINEAR REGRESSION The simple linear regression model can be represented as:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

where:

- y is the dependent variable,
- x is the independent variable,
- β_0 is the intercept,
- β_1 is the coefficient for the independent variable
- ε is the error term.