

4 Bivariate Statistics

4.1 Introduction

4.2 Correlation Coefficients

The null hypothesis of no correlation is rejected if the calculated t is greater than the critical t ($n-2$ degrees of freedom, $\alpha=0.05$). This test is also only valid if both variables are Gaussian distributed.

$$r_{xy} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

According to this test, we can reject the null hypothesis, that is, that the correlation coefficient is significant, if the calculated t is larger than the critical t ($n-2$ degrees of freedom, $\alpha=0.05$).

This result indeed indicates that we can reject the null hypothesis, and therefore the correlation coefficient is significant.

4.3 Classical Linear Regression Analysis

Classical linear regression offers another way of describing the relationship between the two variables x and y .

Classical linear regression assumes that y responds to x and that the entire dispersion in the data set is contained within the y -value (Fig. 4.4).

The dependent variable y (also known as response variable or regressand) contains errors because its magnitude cannot be determined accurately.

4.4 Analyzing Residuals

The p -value of 0.1481 indicates that we cannot reject the null hypothesis that the residuals follow a Gaussian distribution.

4.5 Bootstrap Estimates of Regression Coefficients

4.6 Jackknife Estimates of Regression Coefficients

The median of the i subsamples provides an improved estimate of the regression coefficients.

4.7 Cross-Validation

In the leave-one-out cross-validation method used here, the regression line

is **calculated** using $n-1$ data points.

The standard deviation of the deviations of the data points from the predicted line **is calculated using**

4.8 Reduced Major Axis Regression

4.9 Curvilinear Regression

4.10 Nonlinear and Weighted Regression

The function `p=nlfit(data(:,1),data(:,2),fun,p0)` returns a vector `p` of coefficient estimates for the nonlinear regression of the responses in `data(:,2)` to the **regressors** in `data(:,1)` using the model specified by `fun`.

```
model = @(phi,x)(phi(1)*exp(x) + phi(2));
```

Recommended Reading

Figure Captions

Fig. 4.8 Curvilinear regression from measurements of barium content. The plot shows the original data points (circles), the regression line for a **second-degree** polynomial (solid line), and the prediction error (dashed lines) of the regression.

Fig. 4.9 Comparison of unweighted (dashed line) and weighted (solid line) regression results from synthetic data. The plot shows the original data points (circles), the error bars for all data points, and the regression line for an exponential model function. In the unweighted regression, the fitted curve is moved toward the first two data points with a large error, while in the weighted regression, it is moved toward the third data point with a small error.