Durham University
MATH1541 Statistics
Exercise Sheet 7

Kamil Hepak
Tutorial Group 4

Nov 2018

1 Estimate a and b.

We are given that the relationship between y and t is thought to be of a hyperbolic form, namely following the equation $y = \frac{at}{b+t}$. Initially, I tried rearranging the formula thus:

$$\frac{1}{y} = \frac{b+t}{at}$$

$$\frac{1}{y} = \frac{b}{at} + \frac{t}{at}$$

$$\frac{1}{y} = \frac{b}{a}\frac{1}{t} + \frac{1}{a}$$

Plotting the reciprocal of y against the reciprocal of t produces a linear-seeming plot, but the residual plot for this transformation does not do huge amounts to suggest homoscedasticity. A different arrangement of the equation produces an even better result:

$$\frac{1}{y} = \frac{b+t}{at}$$

$$\frac{t}{y} = \frac{b}{a} + \frac{t}{a}$$

$$\frac{t}{y} = \frac{1}{a}t + \frac{b}{a}$$

As visible on the attached plots, the relationship between $\frac{t}{y}$ and t is very linear and homoscedastic. In this form, we have an equation that resembles one of linear regression - estimating $\frac{t}{y}$, the intercept is $\frac{b}{a}$ and the coefficient is $\frac{1}{a}$. From the R calculations, we can see that the linear model function returns a value of 0.3451 for $\frac{1}{a}$, and a value of 1.8800 for $\frac{b}{a}$. Rearranging, we estimate a to be 2.8977 and b to be 5.4477.

2 Predict water uptake at time $t = 17$.

Let $q = \frac{t}{y}$. Thus, $\hat{q} = 1.88 + 0.3451t$. Substituting $t = 17$, we find $\hat{q} = 7.7467$. Rearranging to find the water uptake estimate $\hat{y}$, we take $\frac{17}{\hat{q}}$ to obtain a result of $\hat{y} = 2.1945$.

3 Find an interval which should contain 95% of the values of the response variable at time $t = 17$.

First, we find our rmsr s_ϵ by using the formula $s_\epsilon = s_y\sqrt{1 - r^2}$. With $s_y = 2.522$ and $r = 0.9866$, we find $s_\epsilon = 0.4109$.
Let Q be the random variable representing the transformed values of the uptakes at time $t = 17$. Assuming $Q \sim N(7.7467, 0.4109^2)$ we can find the 95% confidence interval for Q by using the equation $7.7467 \pm 1.96(0.4109)$. This results in an interval in terms of q of (6.9413, 8.5521) - rearranging for y, we get an interval of (1.9878, 2.4491).

4 Check any assumptions necessary in making your predictions.

4.1 Linearity of Model - Accuracy of Hyperbolic Law Prediction

A linear, homoscedastic relationship is crucial for making predictions using regression. As visible on the plot of $\frac{t}{y}$ against t, the transformed data follows an almost too-good-to-be-true linear relationship. An r of 0.9866 suggests an incredibly high positive correlation, and the regression line fits neatly through the data points. This suggests that a hyperbolic relationship between the response and explanatory variable is fitting.

4.2 Homoscedasticity

As visible on the plot of the residuals of $\frac{t}{y}$ against t, there appears to be no correlation whatsoever between the two variables, strongly implying homoscedasticity.

4.3 Normal Distribution of Residuals

To accurately predict a confidence interval for a certain value of the explanatory variable, one must assume the residuals are normally distributed. A normal quantile plot of the residuals of $\frac{t}{y}$ confirms this - the data is very linear, implying normal distribution.

4.4 Assumptions that I cannot check

Of course, all of these predictions rely on the assumption that change in t causes change in y, and that there isn't a confounding variable linking the two. We also assume that the sample we are provided with is representative of the population enough to predict from.