

# Project Report -MTH686

## Non-Linear Regression Analysis



Indian Institute of Technology, Kanpur

## Department of Mathematics and Statistics

*Author :*

Jakkampudi S K L N M Subbarayudu

Roll.No : 210462

sklnm21@iitk.ac.in

*Supervisor :*

Prof. Debasis Kundu

kundu@iitk.ac.in

November 9, 2024

GitHub Repository:

<https://github.com/SREESUBBARAYUDU/MTH-686-Non-Linear-Regression-.git>

1. Find the least squares estimators of the unknown parameters under three different model assumptions.

Ans) In non-linear regression, we aim to fit a model that minimizes the **Residual Sum of Squares (RSS)**, defined by:

$$RSS = \sum_{i=1}^n (y_i - f(t_i; \theta))^2$$

where:

- $y_i$  is the observed value.
- $\theta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$  represents the unknown parameters.

Model 1:

$$y(t) = \alpha_0 + \alpha_1 e^{\beta_1 t} + \alpha_2 e^{\beta_2 t} + \epsilon(t).$$

Model 2:

$$y(t) = \frac{\alpha_0 + \alpha_1 t}{\beta_0 + \beta_1 t} + \epsilon(t).$$

Model 3:

$$y(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4 + \epsilon(t).$$

Now To find least squares estimators of the unknown parameters i have written code which is under comment (# Question 1).

After running this Code, I found that model 1 & model 2 didn't converge indicating that our dataset is taken from model 3 (as given in the question it comes from one of the dataset).

Results from Model 3:

```
Formula: y ~ beta0 + beta1 * t + beta2 * t^2 + beta3 * t^3 + beta4 * t^4
```

```
Parameters:
```

	Estimate	Std. Error	t value	Pr(> t )
beta0	8.33417	0.03928	212.147	<2e-16 ***
beta1	7.40905	0.52930	13.998	<2e-16 ***
beta2	5.32940	2.10511	2.532	0.0136 *
beta3	2.22029	3.11237	0.713	0.4780
beta4	1.93074	1.52419	1.267	0.2095

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.0631 on 70 degrees of freedom
```

```
Number of iterations to convergence: 2
```

```
Achieved convergence tolerance: 1.461e-06
```

RSS for Model 3: 0.2787084

2. How did you find the least squares estimators? What kind of initial guesses you have chosen?

Ans)

### Strategy for Initial Guesses :

**Intercept ( $\beta_0$ )** : Since the data appears to start near 8.4 for small values of  $t$ , we chose an initial value of 8 for  $\beta_0$  to represent the model's intercept.

**Linear Term ( $\beta_1$ )**: Observing a steady rise in  $y(t)$  with  $t$ , we chose an initial value of 0.5 for  $\beta_1$ , anticipating a positive slope for the linear term.

**Higher-Order Terms ( $\beta_2, \beta_3, \beta_4$ )**: Polynomial models often include diminishing parameter values for higher powers to avoid excessive curvature. Thus, initial values were set to progressively smaller numbers. ( $\beta_2 = 0.1$ ,  $\beta_3 = 0.05$ , and  $\beta_4 = 0.01$ )

**Control Parameters** : To further aid convergence, we set ( $\text{maxiter} = 100$ ) and ( $\text{tol} = 1e^{-5}$ ) within the `nls.control` function. Increasing the number of maximum iterations and setting a stricter tolerance helps in exploring the parameter space thoroughly, especially for complex non-linear relationships.

3. Find the 'best' fitted model.

Ans)

I have attempted to fit all three models to the data. After initial attempts, **Model 1** and **Model 2** failed to converge, suggesting that they might not adequately represent the data's underlying structure or that the parameter space was challenging to navigate due to unsuitable starting values. Consequently, Model 3 successfully converged, providing a good fit without issues.

Since Model 1 and Model 2 did not yield convergent fits, I only computed RSS, AIC, and BIC for Model 3 (Code for this step is below the comment #Question 3):

```
Model 3 (Polynomial Model) :
```

```
RSS: 0.2787084
```

```
AIC: -194.79
```

```
BIC: -398.0434
```

**Goodness-of-Fit** : Model 3 showed low RSS, AIC, and BIC values, indicating it balances fit accuracy and model simplicity effectively.

4. Find the estimate of  $\sigma^2$ .

Ans)

From model 3 , the estimate of  $\sigma^2$  is obtained as :

$$\hat{\sigma}^2 = \frac{\text{RSS}}{n - k}$$

(Code for this step is below the comment #Question 4)

Result: Estimated sigma^2: 0.003981548

5. Find the associated condence intervals based on the Fisher information matrix.

Ans)

In non-linear regression, if we assume that the errors  $E(t)$  are independent and identically distributed with mean zero and variance  $\sigma^2$ , then for a parameter vector

$\theta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$ , the Fisher information matrix  $I(\theta)$  can be approximated by:

$$I(\theta) = \frac{1}{\sigma^2} J(\theta)^T J(\theta)$$

$J(\theta)$  is the Jacobian matrix of partial derivatives of the model function with respect to each parameter evaluated at the estimated parameters  $\theta$  :

$$J(\theta) = \begin{pmatrix} \frac{\partial f(t_1; \theta)}{\partial \beta_0} & \frac{\partial f(t_1; \theta)}{\partial \beta_1} & \dots & \frac{\partial f(t_1; \theta)}{\partial \beta_4} \\ \frac{\partial f(t_2; \theta)}{\partial \beta_0} & \frac{\partial f(t_2; \theta)}{\partial \beta_1} & \dots & \frac{\partial f(t_2; \theta)}{\partial \beta_4} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(t_n; \theta)}{\partial \beta_0} & \frac{\partial f(t_n; \theta)}{\partial \beta_1} & \dots & \frac{\partial f(t_n; \theta)}{\partial \beta_4} \end{pmatrix}$$

The inverse of the Fisher information matrix, provides an asymptotic approximation to the covariance matrix of the parameter estimates.

The variance-covariance matrix of the parameter estimates  $\theta$  can be obtained as:

$$\text{Cov}(\hat{\theta}) = \sigma^2 I(\theta)^{-1}$$

For each parameter  $\beta_j$  in the parameter vector  $\theta$ , the approximate  $100(1-\alpha)\%$  confidence interval is given by:

$$\hat{\beta}_j \pm z_{\alpha/2} \sqrt{\text{Var}(\hat{\beta}_j)}$$

$Z_{\alpha/2}$  is the critical value from the standard normal distribution for a two-sided confidence level. For example, for a 95% confidence interval,  $Z_{\alpha/2} \approx 1.96$ .

	Estimate	Lower	Upper
beta0	8.334171	8.329313	8.339029
beta1	7.409050	7.343591	7.474510
beta2	5.329400	5.069055	5.589745
beta3	2.220288	1.835373	2.605204
beta4	1.930736	1.742235	2.119236

(Code for this step is below the comment #Question 5)

The confidence intervals for each parameter in  $\theta$  provide a range of possible values based on the observed data and the estimated variability. A narrower confidence interval indicates greater precision for that parameter estimate, while a wider interval suggests higher uncertainty.

6. Plot the residuals.

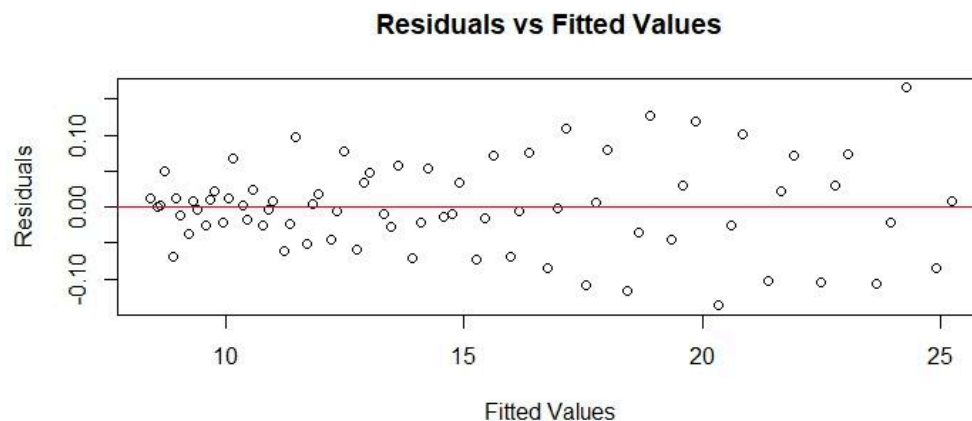
Ans)

Analyzing residuals helps to verify the assumptions of the regression model:

- **Independence** : Residuals should be uncorrelated.
- **Constant Variance** : Residuals should exhibit constant variance.
- **Normality** : Residuals should be approximately normally distributed, especially if we're using the model for inference.

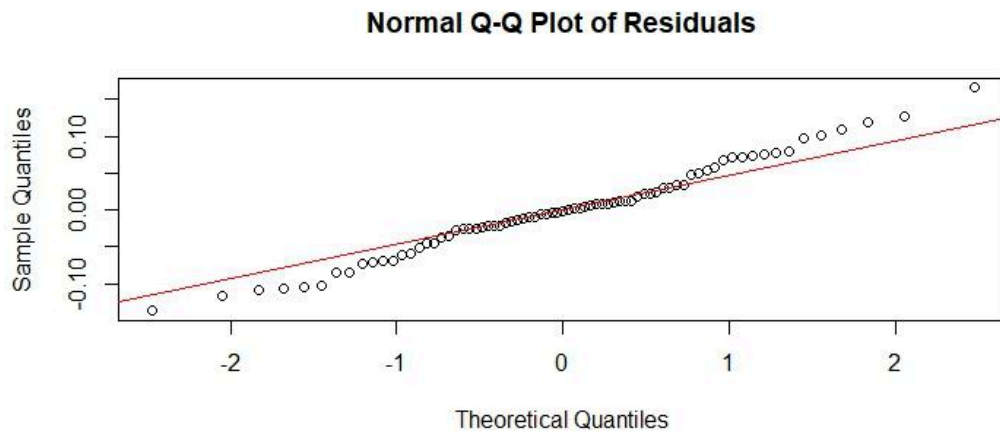
If residuals exhibit any systematic patterns (e.g., trends or clusters), this may indicate a problem with the model assumptions or that a different model might better capture the data. (Code for all plots is below the comment #Question 6)

❖ Residuals vs Fitted Values :



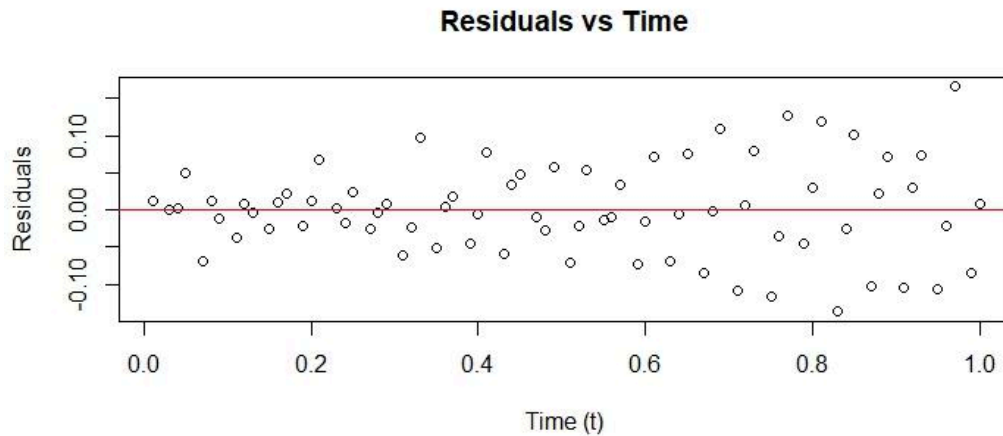
Ideally, points should be randomly scattered around zero with no apparent pattern. A clear pattern may indicate model misfit, such as a non-linear relationship or changing variance.

❖ Normal Q-Q Plot :



This plot checks for normality in the residuals. If the residuals follow a straight line along the diagonal, they are approximately normally distributed, supporting the assumption of normal errors.

❖ Residuals vs. Time (t) :



This plot checks for autocorrelation or time-dependent patterns. A random scatter of residuals around zero indicates that errors are uncorrelated, supporting the independence assumption.

7. Test whether it satisfies the normality assumption or not?

Ans)

We can formally test the normality of residuals using the following hypotheses:

- **Null Hypothesis (H0):** The residuals are normally distributed.
- **Alternative Hypothesis (H1):** The residuals are not normally distributed.

Shapiro-Wilk Test :

The Shapiro-Wilk test statistic W is given by:

$$W = \frac{(\sum_{i=1}^n a_i e_{(i)})^2}{\sum_{i=1}^n e_i^2}$$

The W statistic ranges from 0 to 1. For a perfectly normal distribution,  $W \approx 1$ . The test provides a p-value; if p is less than a chosen significance level (e.g., 0.05), we reject H0, indicating that the residuals deviate significantly from normality.

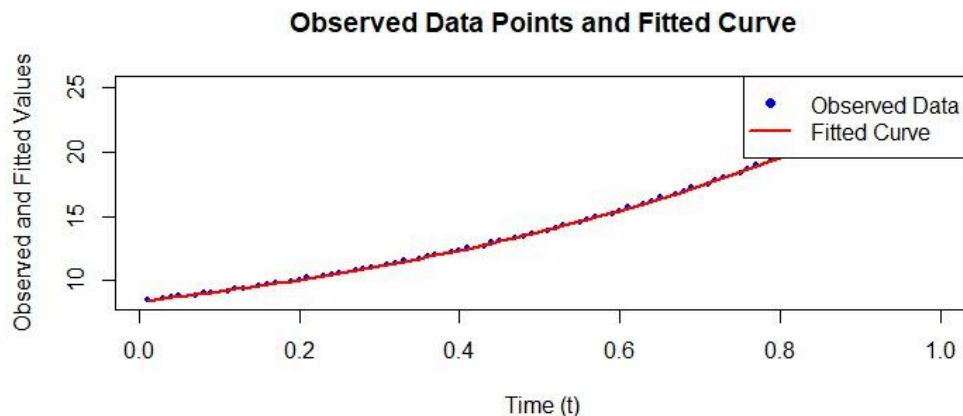
(Code for this step along with Q-Q plot is below the comment #Question 7)

Shapiro-Wilk Test p-value: 0.6590758

If  $p < 0.05$ , we reject the null hypothesis, indicating that residuals deviate significantly from normality. Here  $p \geq 0.05$ , we fail to reject H0, suggesting that residuals are approximately normally distributed.

8. Plot the observed data points and fitted curve.

Ans)



(Code for this step is below the comment # Question 8)

Here we can find that the fitted curve closely follows the observed data points, it suggests that Model 3 adequately captures the trend within the data.