Assignment 2 - Report

Eleni Liarou, Zoë Azra Blei, Frederieke Loth, group 20

2025-03-01

Exercise 1: Titanic

Section a

Section b

Section c

Section d

Section e

Exercise 2: Military Coups

Section a

Section b

Section c

Exercise 3: Stormer viscometer

Section a The nonlinear regression model estimates θ_1 as -3.219 (p = 0.638) and θ_2 as 55.045 (p < 0.001), indicating that only θ_2 is statistically significant. The high residual standard error (95.17) suggests a poor fit, and the nonlinear model may not effectively capture the relationship between Time, Viscosity, and Weight. The linear regression model achieves a much higher R-squared = 0.9938, indicating a strong fit to the data. Both Viscosity (p < 2e-16) and Time (p = 0.00987) are statistically significant in the linear model. However, a high R-squared could also suggest possible overfitting. The large residual standard error (207.3) and some large residuals suggest that the model may be capturing noise rather than true patterns in the data. The plot below shows that the nonlinear fit does not align well with the data, while the linear fit provides a reasonable approximation, which could be because of overfitting, as mentioned. Given that the problem is based on a nonlinear theoretical model, forcing a linear approximation does not make sense, as it ignores the true structure of the data. While the nonlinear model may not fit the data as well numerically, it is more theoretically sound and better reflects the expected relationship between the variables. Therefore, relying on the linear model would be misleading, and refining the nonlinear model (e.g., with better parameter estimation or transformations) would be a more appropriate approach.

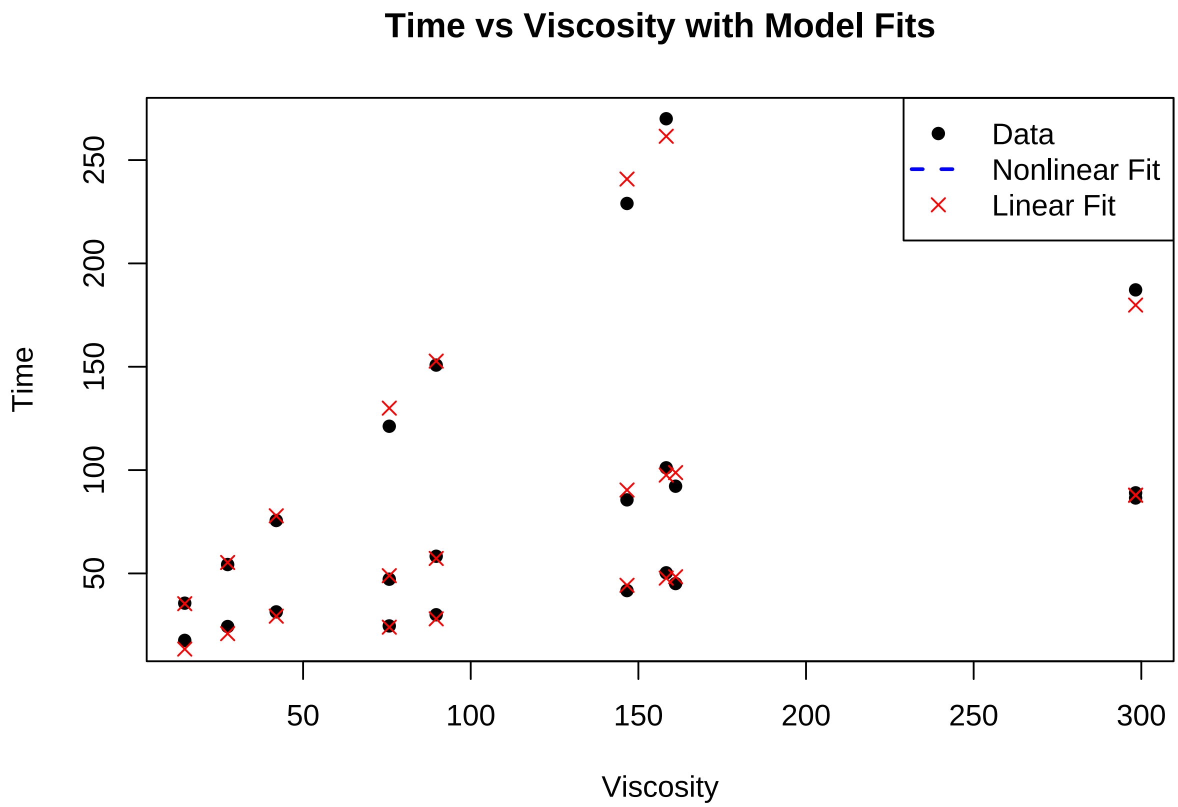
```r
theta1_init = 1
theta2_init = mean(stormer$Wt)
nls_model = nls(Time ~ (theta1 * Viscosity) / (Wt - theta2),
            data = stormer,start = list(theta1 = theta1_init, theta2 =
            theta2_init))
summary(nls_model)
```

```
##
## Formula: Time ~ (theta1 * Viscosity)/(Wt - theta2)
##
## Parameters:
##        Estimate Std. Error t value Pr(>|t|)
## theta1   -3.219      6.740  -0.478    0.638
## theta2   55.045     10.699   5.145 4.26e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 95.17 on 21 degrees of freedom
##
## Number of iterations to convergence: 15
## Achieved convergence tolerance: 8.759e-06
```

```r
lm_model = lm(Wt * Time ~ Viscosity + Time, data = stormer)
summary(lm_model)
```

```
##
## Call:
## lm(formula = Wt * Time ~ Viscosity + Time, data = stormer)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -330.7 -153.8    4.7  170.7  368.3
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 220.1381    82.0080   2.684  0.01426 *
## Viscosity    28.0987     0.5663  49.620  < 2e-16 ***
```

```
## Time          2.0818     0.7302   2.851  0.00987 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 207.3 on 20 degrees of freedom
## Multiple R-squared:  0.9938, Adjusted R-squared:  0.9932
## F-statistic:  1600 on 2 and 20 DF,  p-value: < 2.2e-16
```



Section b A two-tailed t-test is conducted, as we have no expectation about whether θ_1 will be greater or smaller than 25; it could differ in either direction. We consider the null hypothesis H0: $\theta_1 = 25$. For the test, we use the estimated θ_1 and its standard error obtained from question a). The test resulted in a t-statistic of 4.81 and a p-value of 9.45e-05, which is far below the typical significance level of 0.05. This means we reject H0 and conclude that θ_1 is significantly different from 25, further supporting the nonlinear model' results.

```r
theta1_hat = 29.4013   # Estimated parameter
theta1_se = 0.9155     # Standard error
theta1_h0 = 25         # Hypothesized value under H0
df = 21                # Degrees of freedom from nls summary

t_stat = (theta1_hat - theta1_h0) / theta1_se
```

```r
p_value = 2 * pt(-abs(t_stat), df)

cat("Test Statistic (t):", t_stat, "\n")
```

```
## Test Statistic (t): 4.807537
```

```r
cat("P-value:", p_value, "\n")
```

```
## P-value: 9.453736e-05
```

Section c For computing the 92% confidence interval for θ_1 and θ_2, we consider the following formula to calculate the z-value:

$$\hat{\theta} \pm z_{\alpha/2} \cdot SE(\theta)$$

where $z_{\alpha/2}$ is the critical value from the standard normal distribution. For a 92% confidence level, the significance level is $\alpha = 0.08$, thus:

$$z_{0.04/2} = z_{0.02} \approx 1.75$$

This gave a 92% CI for θ_1 of [27.80, 31.00] and for θ_2 of [1.05, 3.38], meaning we are 92% confident that the true values lie within these intervals. Since the confidence interval for θ_1 does not include 25, it further supports rejecting H0 from question b).

```r
theta1_hat <- 29.4013  # Estimated  1
theta1_se <- 0.9155    # Standard error of  1
theta2_hat <- 2.2183   # Estimated  2
theta2_se <- 0.6655    # Standard error of  2

z_value = qnorm(0.96)  # 1.75

theta1_CI = c(theta1_hat - z_value * theta1_se, theta1_hat + z_value * theta1_se)
theta2_CI = c(theta2_hat - z_value * theta2_se, theta2_hat + z_value * theta2_se)
```

92% CI for 1: [27.79855 31.00405]

92% CI for 2: [1.053218 3.383382]

Section d The expected values are computed using the nonlinear model with $w = 50$, and viscosity values ranging from 10 to 300. The 94% confidence intervals were derived using asymptotic normality, where the standard error of T was estimated through error propagation. The confidence bounds were calculated as:

$$T(v) \pm z_{\alpha/2} \cdot SE(T)$$

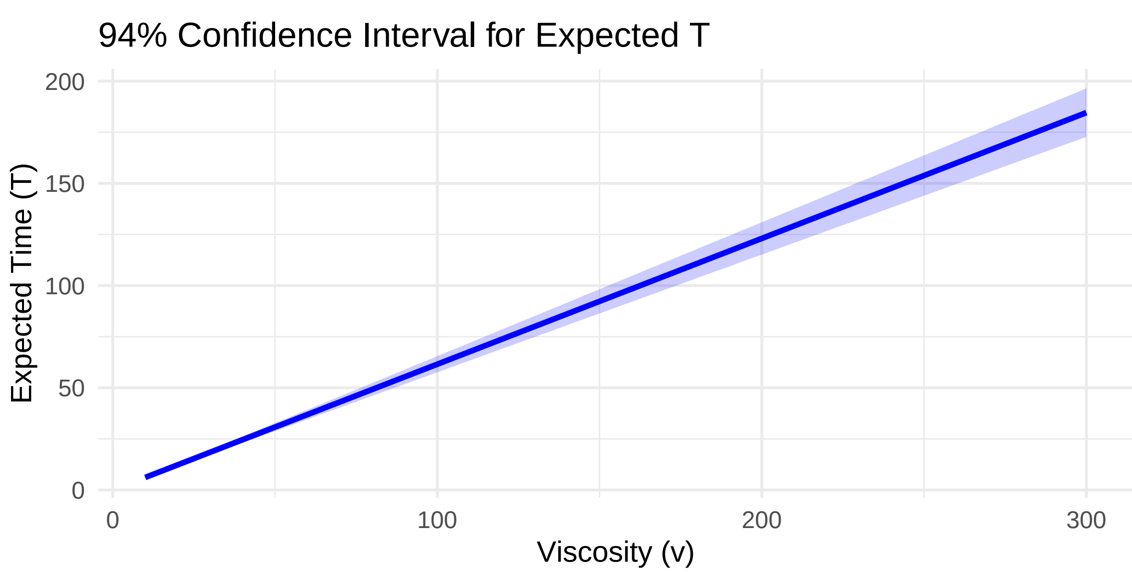
where $z_{0.03} = 1.88$ is the critical value for a 94% confidence level. The plot shows the expected T along with a shaded confidence band, indicating the uncertainty in our estimates. The confidence interval widens as viscosity increases, reflecting greater uncertainty for larger v. The linear trend suggests a strong relationship between viscosity and time, but further diagnostics are needed to confirm model assumptions. Overall, the plot aligns well with the theoretical nonlinear model, supporting its validity over a simple linear approximation.

```
theta1_hat = 29.4013  # Estimate of theta1
theta2_hat = 2.2183   # Estimate of theta2
theta1_se = 0.9155    # Standard error of theta1
theta2_se = 0.6655    # Standard error of theta2
w_fixed = 50
v_values = seq(10, 300, length.out = 100)

# Compute expected T values using the nonlinear model: T = (theta1 * v) / (w - theta2)
T_hat <- (theta1_hat * v_values) / (w_fixed - theta2_hat)

# Compute standard error propagation for T
T_se <- sqrt(
  (v_values / (w_fixed - theta2_hat))^2 * theta1_se^2 +
  (theta1_hat * v_values / (w_fixed - theta2_hat)^2)^2 * theta2_se^2
)

# Compute 94% confidence intervals
z_value <- qnorm(0.97)  # z-value for 94% confidence interval (alpha = 0.06, so z(0.03) = 1.88
T_lower <- T_hat - z_value * T_se
T_upper <- T_hat + z_value * T_se
```



Section e