
Mixed Effect Model in R

Ayobami Emmanuel Omolusi
School of Collective Intelligence
Rabat, Morocco
emmanuel.omolusi@um6p.ma

Abstract

This study employs a mixed-effects modeling approach to analyze the sleepstudy dataset, investigating the impact of sleep deprivation on psychomotor performance. The dataset features repeated measurements on individual subjects, making it inherently suited for mixed-effects models. My analysis examines the interplay between reaction times and the duration of sleep deprivation, considering both fixed and random effects. Diagnostic assessments validate the model's appropriateness, affirming the normality and homoscedasticity assumptions. The results offer great insights into subject-specific variations, providing a comprehensive understanding of psychomotor performance dynamics over time. This approach ensures robust statistical inferences in exploring the effects of sleep deprivation on cognitive function.

1 Data Exploration

1.1 Loading the “Sleepstudy” dataset

The dataset was loaded from the R library using the command `"data('sleepstudy')." To gain insight into the dataset's contents, the first six rows, along with the details of the columns, were displayed using "head(sleepstudy)".`

1.2 Structure of the dataset

The command `"str(sleepstudy)"` was used to examine the structure of the dataset. The structure reveals that the dataset is a data frame containing 180 rows and 3 columns. The dataset has 3 columns: • Reaction: A numerical Variable • Days: A numerical variable • Subject: A 18 levels factored variable

```
'data.frame':   180 obs. of  3 variables:
 $ Reaction: num  250 259 251 321 357 ...
 $ Days    : num   0  1  2  3  4  5  6  7  8  9 ...
 $ Subject : Factor w/ 18 levels "308","309","310",...: 1 1 1 1 1 1 1 1 1 1 ...
```

1.3 Visualization

Some Visualizations were done to understand the distribution and relationship between the variables. Scatterplot, Boxplot and Density plot were the basic visualizations done.

2 Descriptive Statistics

The summary statistics was two in two groups- by Days and by Subject.

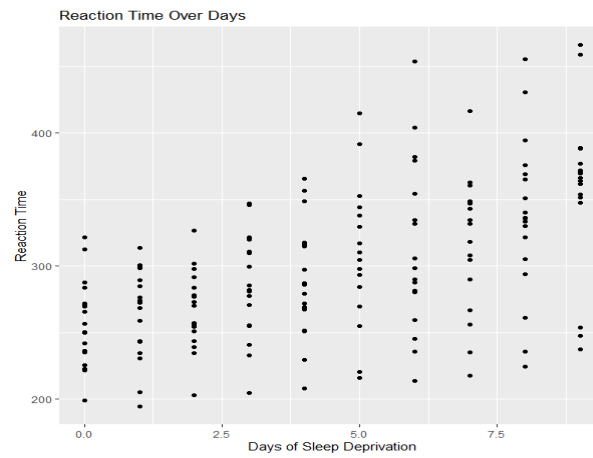


Figure 1: Scatterplot of reaction time over days

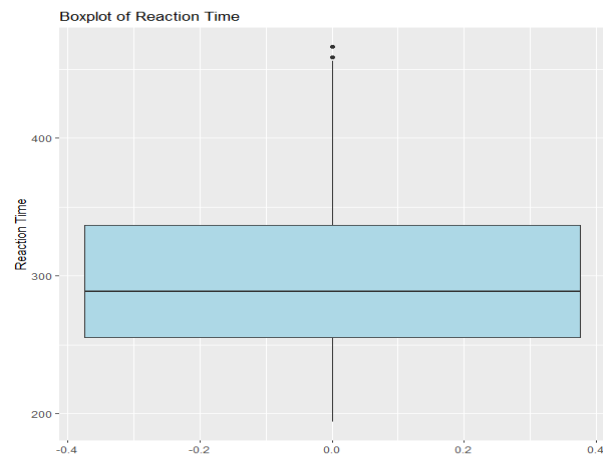


Figure 2: boxplot of reaction time

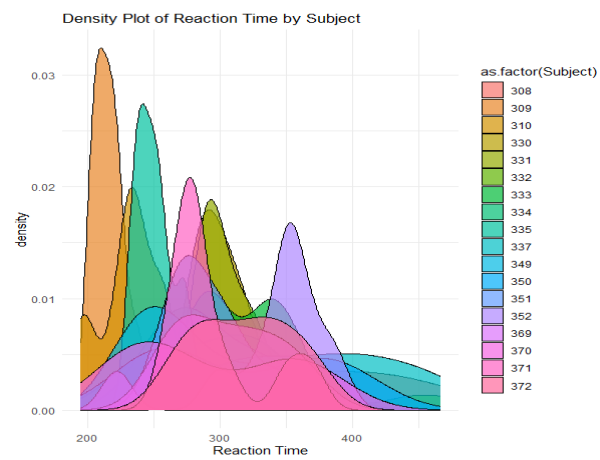


Figure 3: densityplot of reaction time by subject

Table 1: Summary Statistics of Reaction Time Grouped by Days

Days	Mean Reaction	Median Reaction	SD Reaction	Min Reaction	Max Reaction
0	256.7	253.4	32.1	199.1	321.5
1	264.5	273.0	33.4	194.3	313.8
2	265.4	263.6	29.5	203.0	326.9
3	283.0	281.2	38.9	204.7	346.9
4	288.6	282.5	42.5	207.7	365.7
5	308.5	307.3	51.8	216.0	414.7
6	312.2	294.1	63.2	213.6	454.2
7	318.8	333.2	50.1	217.7	416.7
8	336.6	336.0	60.2	224.3	455.9
9	350.9	365.3	67.0	237.2	466.4

Table 2: Summary Statistics of Reaction Time by Subject

Subject	Mean Reaction	Median Reaction	SD Reaction	Min Reaction	Max Reaction
308	342.13383	339.14585	79.82176	249.56	466.3535
309	215.23298	214.79605	10.81219	202.9778	237.3142
310	231.00127	233.5808	21.85600	194.3322	261.0125
330	303.22142	298.99285	22.90920	280.2396	354.0487
331	309.43605	297.78375	27.24261	285.0	371.5811
332	307.30207	309.8832	64.30613	234.8606	454.1619
333	316.15831	315.9181	30.06821	276.7693	362.0428
334	295.30205	281.6078	41.85561	243.3647	377.299
335	250.07004	248.4521	13.83385	235.311	273.9472
337	375.72101	378.78545	59.62379	291.6112	458.9167
349	275.83447	262.3482	42.93794	230.3167	351.6451
350	313.60268	299.3206	63.36056	243.4543	394.4872
351	290.09775	284.16785	28.97881	250.5265	347.5655
352	337.42154	350.78445	47.60238	221.6771	388.5417
369	306.03464	306.47875	37.46043	257.2424	366.5131
370	291.70177	274.3427	59.20821	225.264	372.2288
371	294.98404	280.48	36.50599	259.2658	369.4692
372	317.88613	320.1196	35.82280	269.4117	369.1417

2.1 summary statistics for the key variables

Grouped by Days

Grouped by Subject

2.2 visualizations to better understand the distribution of reaction times over different days

3 Fitting adequate Model(s).

The dataset involves repeated measurements on the same subjects over multiple days, introducing a level of dependency in the data. A mixed-effects model is well-suited for handling this type of correlated observations within subjects. Reaction times in the dataset show subject-specific variability that a simple fixed-effects model would not be able to adequately represent. The model recognizes and accounts for individual differences among participants by allowing for subject-specific intercepts through the incorporation of a random effect term (1|Subject).

```
model <- lmer(Reaction ~ Days + (1 | Subject), data = sleepstudy)
> model
Linear mixed model fit by REML ['lmerMod']
Formula: Reaction ~ Days + (1 | Subject)
Data: sleepstudy
REML criterion at convergence: 1786.465
```

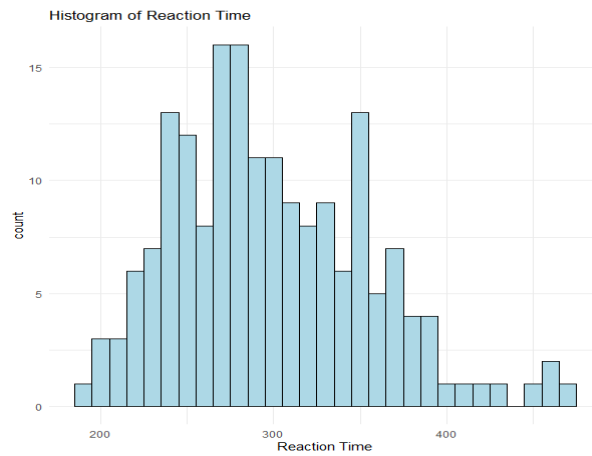


Figure 4: count of reaction time

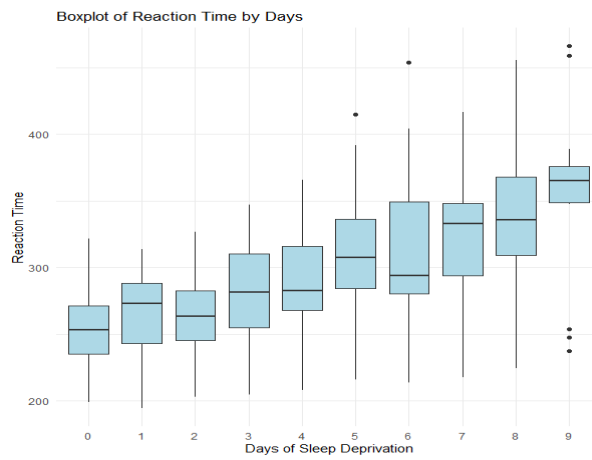


Figure 5: boxplot of reaction time over days

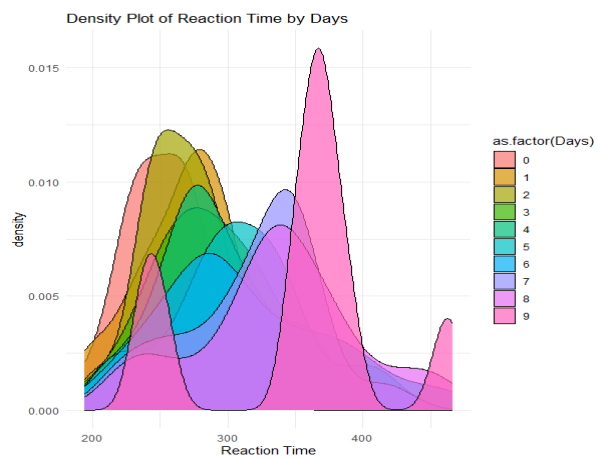


Figure 6: densityplot of reaction time over days

```

Random effects:
  Groups   Name      Std.Dev.
  Subject  (Intercept) 37.12
  Residual                    30.99
Number of obs: 180, groups: Subject, 18
Fixed Effects:
      (Intercept)      Days
      251.41      10.47

```

3.1 Model Interpretation

The dataset involves repeated measurements on the same subjects over multiple days, introducing a level of dependency in the data. Hence, the need for a mixed effect model. A mixed-effects model is well-suited for handling this type of correlated observations within subjects.

Fixed Effects: When the number of days (Days) is zero, the projected mean reaction time is found at the intercept (251.41). It stands for the average response time at the baseline (Day 0) in this context. Days (10.47): The expected variation in response time for every day that is increased by one unit. Stated differently, the average daily increase in reaction time is 10.47 milliseconds.

Random Effects: Subject (Intercept, Std.Dev. = 37.12): Explains the variation in reaction times among subjects. Subject-specific intercepts have an estimated standard deviation of 37.12 milliseconds. This illustrates how various respondents' baseline reaction times vary from one another. Stubs (average development = 30.99):

the estimated residual standard deviation, which, after fixed and random influences are taken into account, represents the unexplained variance in reaction times.

According to the model, the reaction time increases by roughly 10.47 milliseconds for every extra day, with an average of 251.41 milliseconds at Day 0 (Intercept).

The random effect for Subject's Intercept (Std.Dev. = 37.12) shows significant variation in subjects' baseline reaction times.

The REML criteria indicates that the total model fit is 1786.465, which gives an indication of how well the model accounts for the observed data.

4 Residual Analysis

```

Residual Analysis
> res <- residuals(model)
> #set graphical parameters to generate plot matrix
> par(mfrow =c(1,3))
> # Plot 1 histogram
> hist(res)
> # Plot 2 Q-Q Plot
> qqnorm(res)
> qqline(res)
> plot(fitted(model), res)

```

From the plot distributions, the following can be observed:

Normal Distribution : (a) The residuals' distribution appears to be bell-shaped and not skewed. The residuals histogram shows that the distribution is roughly bell-shaped and symmetric. This implies that there is no apparent skewness in the residuals, which are instead centered around zero. A distribution that is symmetric suggests that the normality assumption is fitted.

(b) The residuals in this plot conform to the normal distribution when the sample quantiles assemble into a straight line: The sample quantiles closely coincide with the theoretical quantiles of a normal distribution, as shown by the normal quantile-quantile (Q-Q) figure. The Q-Q plot's straight line suggests that the residuals have a normal distribution. Deviations from normality would be shown by deviations from the diagonal line.

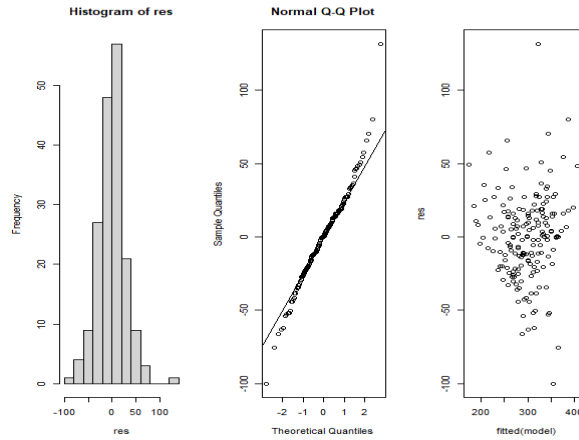


Figure 7: densityplot of reaction time over days

Homoscedasticity Assumption The residuals are consistently distributed throughout the range of fitted values, creating a relatively uniform "blob" shape on the residuals vs. fitted values diagram. This implies that the homoscedasticity assumption—that is, the residuals' constant variance—is satisfied. It is a sign that the response variable's variability is constant at all predictor variable levels when the residual spread is nearly identical.