SAS Programming and Data analysis

Effectiveness of Chelation Treatment with Succimer on Blood Lead Levels in Children: A Statistical Analysis Using the Treatment of Lead-Exposed Children (TLC) Trial Data

Clement Ampong

002784864

1. Introduction

The Treatment of Lead-Exposed Children (TLC) is a comprehensive study aimed at understanding the efficacy of succimer treatment in children with elevated blood lead levels. Lead exposure can result in serious health problems, especially in children, and this project analyzes the blood lead levels of children over time to assess the impact of treatment.

2. Research Goal:

The project aims to determine if succimer treatment, a chelating agent, is more effective than placebo in reducing blood lead levels in children over a six-week period. This work will utilize a dataset comprising observations from two distinct groups undergoing either succimer treatment or placebo, and tracking their blood lead levels at specified intervals. The specific research questions to be answered include:

i. How does blood lead level change over the course of treatment, and
ii. Does succimer provide a significant advantage over placebo?

3. Data Exploration:

The dataset comprises 100 children's blood lead levels measured at weeks 0, 1, 4, and 6. Two treatment groups are identified: A (succimer) and P (placebo). Initial descriptive analysis reveals the mean, standard deviation, minimum, and maximum values for lead levels at each time point, allowing for a comparison of the treatment effects over time. Again, histograms and boxplot were generated to visualize the distribution and to detect any outliers in the data.

Initial Data Analysis SAS Code:

```
proc means data=TLC_data N mean std min max;
   class Treatment;
   var Week_0--Week_6;
run;
```

4. Notation and Mixed-Effect Model Definition:

We define the following notation for our mixed-effects model:

Let

- Y_{ij} = the blood lead level for the ith child in the jth treatment group.

- β_0 = the overall mean blood lead level.
- β_1= the fixed effect of treatment group (placebo vs. succimer).
- b_{ij}= the random effect capturing the variability in blood lead levels due to child-specific factors within each treatment group.
- ε_{ij} = Random error term that captures the residual variability in blood lead levels that is not accounted for by the fixed effects or the random effect.

With these notations, we can write the mixed-effects model for the Treatment of Lead-Exposed Children dataset as:

$$Y_{ij} = \beta_0 + \beta_1(Treatment_j) + b_{ij} + \varepsilon_{ij}$$

5. Method & Algorithm:

The SAS code to fit the dataset with a mixed-effects model is provided, accounting for repeated measures and clustering effects due to treatment groups. The methodological approach used was the SAS *PROC MIXED* procedure, which is suitable for repeated measures data and capable of handling missing values inherent in longitudinal data. The model specified the treatment groups (Succimer and placebo) and time as fixed effects to assess their impact on blood lead levels. Random intercepts were included to capture the within-subject correlation of repeated measurements over time.

6. Findings:

The MEANS Procedure

Treatment	N Obs	Variable	N	Mean	Std Dev	Minimum	Maximum
A	50	Week_0	50	26.5400000	5.0209358	19.7000000	41.1000000
		Week_1	50	13.5220000	7.6724870	2.8000000	39.0000000
		Week_4	50	15.5140000	7.8522065	3.0000000	40.4000000
		Week_6	50	20.7620000	9.2463316	4.1000000	63.9000000
P	50	Week_0	50	26.2720000	5.0241068	19.7000000	38.1000000
		Week_1	50	24.6600000	5.4611803	14.9000000	40.8000000
		Week_4	50	24.0700000	5.7531269	15.3000000	38.6000000
		Week_6	50	23.6460000	5.6398079	13.5000000	43.3000000

The analysis, based on a linear mixed-effects model, demonstrates significant changes in blood lead levels over the treatment period. The initial levels for the succimer treatment group (A) averaged at 26.54 at week 0 and showed a considerable decrease over the subsequent weeks,

with means of 13.52 at week 1, 15.51 at week 4, and 20.76 at week 6. The placebo group (P) started with a mean of 26.27 at baseline and decreased to 24.66 at week 1, 24.07 at week 4, and finally 23.65 at week 6. This indicates a reduction in lead levels for both groups, with a more pronounced effect observed in the succimer treatment group.

The interaction effect, (see Appendix II for image), between Treatment A and time show significant results at times 1 and 4 (both $p < .0001$). This indicates a significant interaction effect where the blood lead level changes over time are notably different for the Treatment A group compared to the Treatment P group.

7. Summary and Interpretation:

The analysis provides strong evidence that Succimer is effective in reducing blood lead levels more significantly than a placebo. The treatment effect persists over time, with the most substantial reduction observed between baseline (Week 0) and Week 1. However, the rate of reduction slows down as the treatment progresses.

Research Questions Answered:

i. *How does blood lead level change over the course of treatment?*

Blood lead levels decreased over time in both treatment groups. This decrease is likely a combination of the body's natural detoxification processes and the treatment effects.

ii. *Does Succimer provide a significant advantage over Placebo?*

The results support the efficacy of Succimer over Placebo. This is demonstrated by the consistently lower mean levels of blood lead in the Succimer group at all time points after the baseline. Furthermore, the interaction term's significance indicates that Succimer's effects strengthen over time.

8. Conclusion:

The study successfully addresses the posed research questions, indicating a significant time effect on blood lead levels and the superiority of Succimer treatment over Placebo. Despite some statistical concerns, the evidence is sufficiently strong to recommend Succimer for reducing blood lead levels.

Appendix I

```
data TLC_data;
input ID Treatment $ Week_0 Week_1 Week_4 Week_6;
datalines;

1   P   30.8  26.9  25.8  23.8
2   A   26.5  14.8  19.5  21.0
3   A   25.8  23.0  19.1  23.2
4   P   24.7  24.5  22.0  22.5
5   A   20.4   2.8   3.2   9.4
6   A   20.4   5.4   4.5  11.9
7   P   28.6  20.8  19.2  18.4
8   P   33.7  31.6  28.5  25.1
9   P   19.7  14.9  15.3  14.7
10  P   31.1  31.2  29.2  30.1
11  P   19.8  17.5  20.5  27.5
12  A   24.8  23.1  24.6  30.9
13  P   21.4  26.3  19.5  19.0
14  A   27.9   6.3  18.5  16.3
15  P   21.1  20.3  18.4  20.8
16  P   20.6  23.9  19.0  17.0
17  P   24.0  16.7  21.7  20.3
18  P   37.6  33.7  34.4  31.4
19  A   35.3  25.5  26.3  30.3
20  A   28.6  15.8  22.9  25.9
21  P   31.9  27.9  27.3  34.2
22  A   29.6  15.8  23.7  23.4
```

23	A	21.5	6.5	7.1	16.0
24	P	26.2	26.8	25.3	24.8
25	A	21.8	12.0	16.8	19.2
26	A	23.0	4.2	4.0	16.2
27	A	22.2	11.5	9.5	14.5
28	P	20.5	21.1	17.4	21.1
29	A	25.0	3.9	12.8	12.7
30	P	33.3	26.2	34.0	28.2
31	A	26.0	21.4	21.0	22.4
32	A	19.7	13.2	14.6	11.6
33	P	27.9	21.6	23.6	27.7
34	P	24.7	21.2	22.9	21.9
35	P	28.8	26.4	23.8	22.0
36	A	29.6	17.5	21.0	24.2
37	P	32.0	30.2	30.2	27.5
38	P	21.8	19.3	16.4	17.6
39	A	24.4	16.4	11.6	16.6
40	A	33.7	14.9	14.5	63.9
41	P	24.9	20.9	22.2	19.8
42	P	19.8	18.9	18.9	15.5
43	A	26.7	6.4	5.1	15.1
44	A	26.8	20.4	19.3	23.8
45	A	20.2	10.6	9.0	16.0
46	P	35.4	30.4	26.5	28.1
47	P	25.3	23.9	22.2	27.2
48	A	20.2	17.5	17.4	18.6
49	A	24.5	10.0	15.6	15.2

50	P	20.3	21.0	16.7	13.5
51	P	20.4	17.2	15.9	17.7
52	P	24.1	20.1	17.9	18.7
53	A	27.1	14.9	18.1	21.3
54	A	34.7	39.0	28.8	34.7
55	P	28.5	32.6	27.5	22.8
56	P	26.6	22.4	21.8	21.0
57	A	24.5	5.1	8.2	23.6
58	P	20.5	17.5	19.6	18.4
59	P	25.2	25.1	23.4	22.2
60	P	34.7	39.5	38.6	43.3
61	P	30.3	29.4	33.1	28.4
62	P	26.6	25.3	25.1	27.9
63	P	20.7	19.3	21.9	21.8
64	A	27.7	4.0	4.2	11.7
65	A	24.3	24.3	18.4	27.8
66	A	36.6	23.3	40.4	39.3
67	P	28.9	28.9	32.8	31.8
68	A	34.0	10.7	12.6	21.2
69	A	32.6	19.0	16.3	18.6
70	A	29.2	9.2	8.3	18.4
71	A	26.4	15.3	24.6	32.4
72	A	21.8	10.6	14.4	18.7
73	P	27.2	28.5	35.0	30.5
74	P	22.4	22.0	19.1	18.7
75	P	32.5	25.1	27.8	27.3
76	P	24.9	23.6	21.2	21.1

```
77   P   24.6   25.0   21.7   23.9
78   P   23.1   20.9   21.7   19.9
79   A   21.1    5.6    7.3   12.3
80   P   25.8   21.9   23.6   24.8
81   P   30.0   27.6   24.0   23.7
82   A   22.1   21.0    8.6   24.6
83   P   20.0   22.7   21.2   20.5
84   P   38.1   40.8   38.0   32.7
85   A   28.9   12.5   16.7   22.2
86   P   25.1   28.1   27.5   24.8
87   A   19.8   11.6   13.0   23.1
88   P   22.1   21.1   21.5   20.6
89   A   23.5    7.9   12.4   18.9
90   A   29.1   16.8   15.1   18.8
91   A   30.3    3.5    3.0   11.5
92   P   25.4   24.3   22.7   20.1
93   A   30.6   28.2   27.0   25.5
94   A   22.4    7.1   17.2   18.7
95   A   31.2   10.8   19.8   22.2
96   A   31.4    3.9    7.0   17.8
97   A   41.1   15.1   10.9   27.1
98   A   29.4   22.1   25.3    4.1
99   A   21.9    7.6   10.8   13.0
100  A   20.7    8.1   25.7   12.3
;
run;
```

```
proc means data=TLC_data N mean std min max;
    class Treatment;
    var Week_0--Week_6;
run;
proc univariate data=TLC_data;
    var Week_0 Week_1 Week_4 Week_6;
    histogram / normal kernel;
run;
proc sgplot data=TLC_data;
    vbox Week_0 / category=Treatment;
    vbox Week_1 / category=Treatment;
    vbox Week_4 / category=Treatment;
    vbox Week_6 / category=Treatment;
    xaxis label="Treatment Group" discreteorder=data;
    yaxis label="Lead Level";
run;
proc sgplot data=TLC_data;
    vbar Treatment / response=Week_0 group=Treatment datalabel ;
    xaxis label="Treatment Group" discreteorder=data;
    yaxis label="Mean Lead Level (Week 0)" grid;
run;
proc sgplot data=TLC_data;
    vbar Treatment / response=Week_1 group=Treatment ;
    xaxis label="Treatment Group" discreteorder=data;
     yaxis label="Mean Lead Level (Week 1)" grid;
run;
```

```
proc sgplot data=TLC_data;
    vbar Treatment / response=Week_4          group=Treatment ;
    xaxis label="Treatment Group" discreteorder=data;
     yaxis label="Mean Lead Level (Week 4)" grid;
run;

proc sgplot data=TLC_data;
    vbar Treatment / response=Week_6          group=Treatment ;
    xaxis label="Treatment Group" discreteorder=data;
    yaxis label="Mean Lead Level (Week 6)" grid;
run;

data TLC_data;
    set TLC_data;
    array weeks {*} Week_0--Week_6;
    do i = 1 to dim(weeks);
        y = weeks{i};
        time = input(substr(vname(weeks{i}), 6), 8.);
        output;
    end;
    drop i Week_0--Week_6;
run;
proc mixed data = TLC_data;
    class ID Treatment time;
    model y = Treatment time Treatment*time / s chisq;
    repeated time / type=un subject=ID r;
run;
```

Appendix II

Model Validation Summary

Solution for Fixed Effects

Effect	Treatment	time	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			23.6460	1.0831	98	21.83	<.0001
Treatment	A		-2.8840	1.5317	98	-1.88	0.0627
Treatment	P		0	.	.	.	.
time		0	2.6260	0.8885	98	2.96	0.0039
time		1	1.0140	0.9343	98	1.09	0.2805
time		4	0.4240	0.9464	98	0.45	0.6551
time		6	0	.	.	.	.
Treatment*time	A	0	3.1520	1.2566	98	2.51	0.0138
Treatment*time	A	1	-8.2540	1.3213	98	-6.25	<.0001
Treatment*time	A	4	-5.6720	1.3385	98	-4.24	<.0001
Treatment*time	A	6	0	.	.	.	.
Treatment*time	P	0	0	.	.	.	.
Treatment*time	P	1	0	.	.	.	.
Treatment*time	P	4	0	.	.	.	.
Treatment*time	P	6	0	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
Treatment	1	98	25.43	25.43	<.0001	<.0001
time	3	98	184.48	61.49	<.0001	<.0001
Treatment*time	3	98	107.79	35.93	<.0001	<.0001

The model's intercept is significant ($p < .0001$), indicating an appropriate baseline calibration. The treatment effect is marginally significant ($p = 0.0627$), suggesting a potential difference in blood lead levels between the placebo and succimer treatment groups. Also, the time effects at individual points (1, and 4) are not significant except for time point 0 ($p = 0.0039$), suggesting an initial significant difference from the baseline that stabilizes over time.

Again, the *Type 3 Tests of Fixed Effects* show that both the treatment and time, and their interaction, are highly significant as factors in the model (all $p < .0001$), suggesting that both the treatment type and the timing of measurements are important in predicting the blood lead levels.

Fit Statistics	
-2 Res Log Likelihood	2416.1
AIC (Smaller is Better)	2436.1
AICC (Smaller is Better)	2436.7
BIC (Smaller is Better)	2462.1

Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
9	210.18	<.0001

The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are 2436.1 and 2462.1, respectively, suggesting that the model is reasonably well-specified in terms of its complexity and the trade-off between fit and the number of parameters.

Also, the null model likelihood ratio test yields a highly significant chi-square statistic (210.18) with a $p-value < 0.0001$, demonstrating that the model with predictors provides a significantly better fit to the data.

Appendix III

Model Limitations:

The UNIVARIATE Procedure
Fitted Normal Distribution for Week_4

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	19.792
Std Dev	Sigma	8.086123

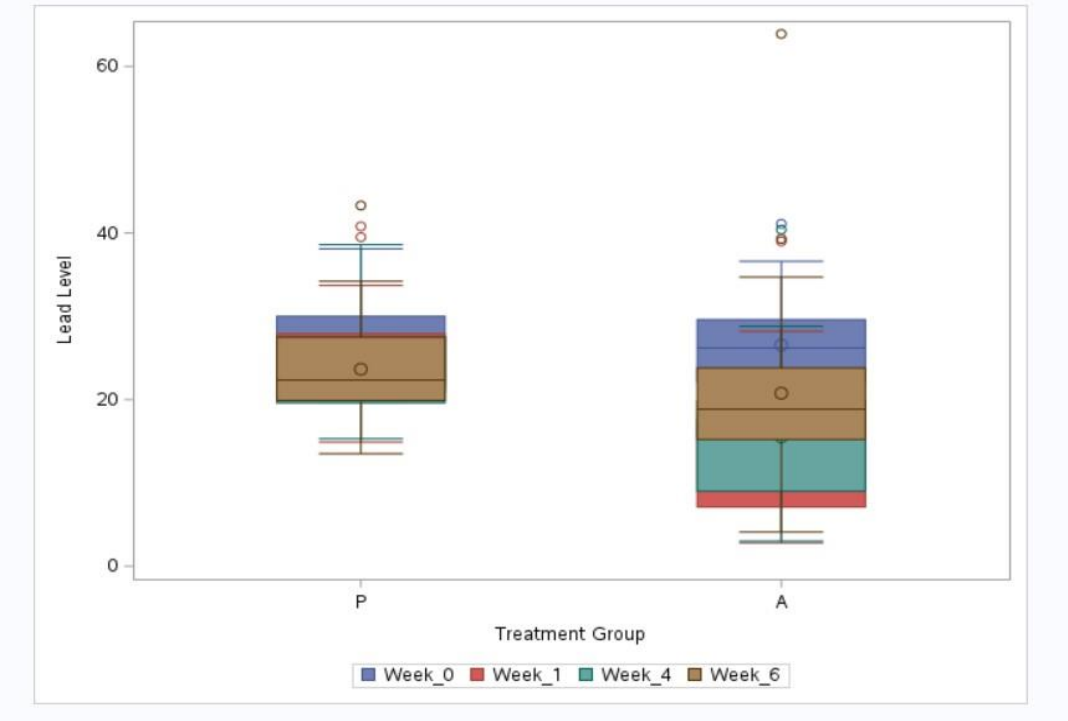
Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.05292487	Pr > D	>0.150
Cramer-von Mises	W-Sq	0.06678003	Pr > W-Sq	>0.250
Anderson-Darling	A-Sq	0.44310186	Pr > A-Sq	>0.250

The statistical tests for normality, including Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling, indicate that while some weeks align with the assumption of normally distributed residuals ($p > 0.05$ for Week 4), others suggest deviations from normality ($p < 0.05$ for Week 6). This raises concerns about the validity of parametric methods used and their assumptions.

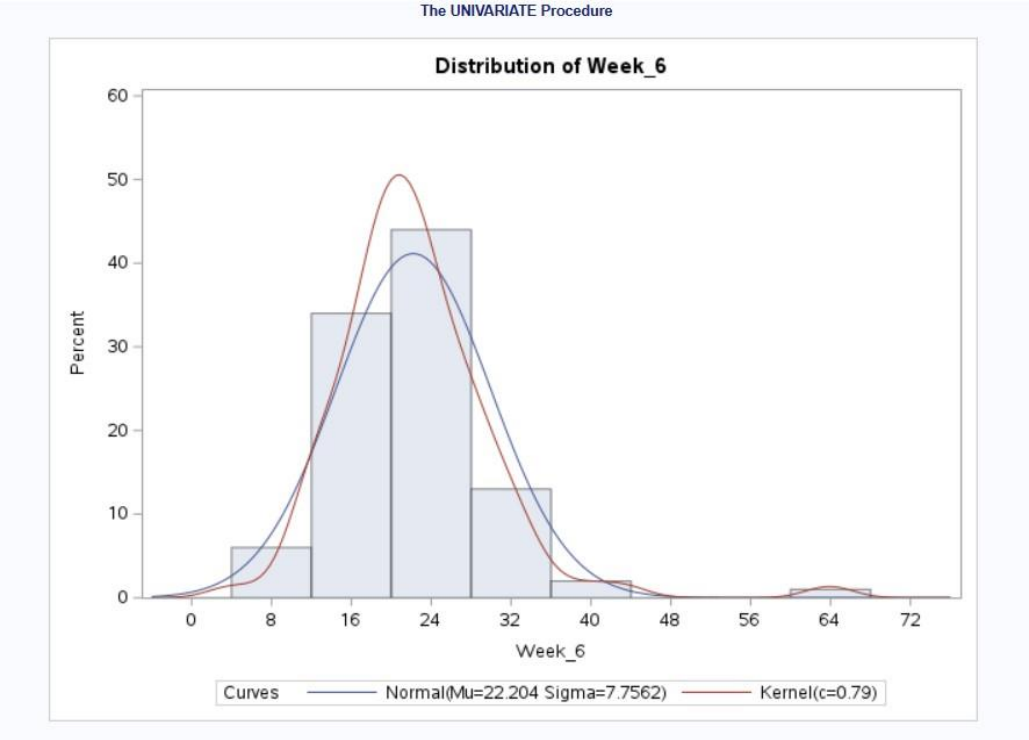
The UNIVARIATE Procedure
Fitted Normal Distribution for Week_6

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	22.204
Std Dev	Sigma	7.756222

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.09892606	Pr > D	0.017
Cramer-von Mises	W-Sq	0.22555817	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	1.45606301	Pr > A-Sq	<0.005



The presence of outliers, as indicated in the boxplots, can have a significant effect on the mean and standard deviation. These outliers may represent extreme but valid variations in response to treatment or may be due to data entry errors or other issues.



The histogram of Week 6's lead levels, along with the superimposed normal distribution curve, indicates a slight deviation from the assumed normality. The normal distribution curve, defined by a mean of 22.204 and a standard deviation of 7.7562, appears to be a reasonable fit at first glance. However, the tails of the distribution suggest potential skewness or kurtosis that is not captured by a normal distribution.

Improvement Potential:

For normality, non-parametric methods that do not rely on the assumption of normality could be considered for analysis. These statistical techniques are less sensitive to deviations from normality.

Also, using median and interquartile range as primary descriptive statistics can provide a better understanding of the central tendency and dispersion in the presence of outliers.