Stastitical Study of Narcotic Abstinence Syndrome

Liang Chen

1 Introduction

Infants prenatally exposed to opiates were observed for the development of narcotic abstinence syndrome (NAS) and given supportive care. NAS is a group of conditions caused when a baby withdraws from certain drugs (most often opioids) that this baby is exposed to in the womb before birth. Besides opioids, NAS can also be caused by antidepressants (used to treat depression), barbiturates or benzodiazepines (sleeping pills). When mothers take these drugs during their pregnancies, those drugs can pass through the placenta and cause serious problems for their babies. Babies with NAS may have health conditions that need treatment in the newborn intensive care unit (NICU).

Auricular acupuncture facilitates narcotic withdrawal in opiate-dependent adults. Effects of auricular acupuncture in neonates with NAS are unknown. The objective of this study was to evaluate the potential efficacy of auricular acupressure augmentation of standard medical management of babies with NAS by comparing the length of hospital stay between babies randomized to receive standard medical care alone, or standard care plus auricular acupressure. We are also interested in identifying which factor(s) are related with the decreased length of hospital stay.

Auricular Acupressure in this study has a detailed procedure. A small herbal seed was taped to 1-3 left or right ear acupoints in babies assigned to receive auricular acupressure in addition to standard medical management. These points are recommended for use by the National Acupuncture Detoxification Association (NADA). Points selected for treatment were based on the NAS symptoms present in the infant at the time of seed application.

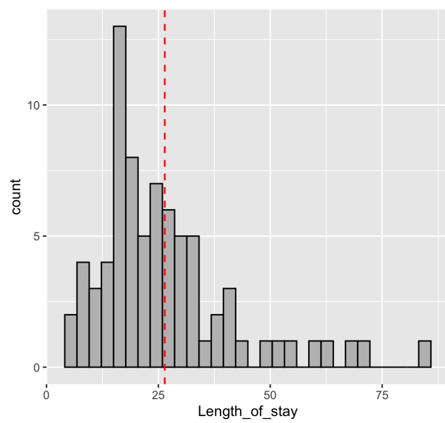
2 Data Description

The information of the dataset of interest is introduced as follows. This dataset includes 77 NAS babies with 37 receiving standard care and 40 receiving standard care plus acupressure. This dataset has 18 features. They can be categorized into two classes based on the subjects: Maternal data and Neonatal data.

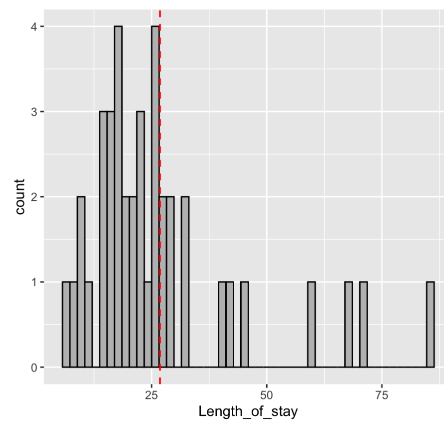
The primary outcome in this study that we are interested in is the length of hospital stay, which is a continuous variable. In terms of this variable, we would like to investigate the potential efficacy of auricular acupressure augmentation of standard medical management of babies with NAS by comparing the length of hospital stay between babies randomized to receive standard medical care alone, or standard care plus auricular acupressure. We would also like to investigate which factor(s) are related with the decreased length of hospital stay by applying multiple linear regression model. The second outcomes in this study are race, use of heroin, use of methadone, use of alcohol, state of homelessness, state of incarceratedness, etc. Specifically, we would like to investigate whether they are related to each other by doing categorical data analysis.

Length of Hospital Stay			
Statistics	Total (days)	W/ Acupressure (days)	W/O Acupressure (days)
Mean	26.43	26.82	26.00
Median	22.63	22.92	22.38
Variance	239.86	303.72	176.99
Std	15.49	17.43	13.30
Q1	16.50	16.48	17.58
Q3	31.56	28.85	32.25

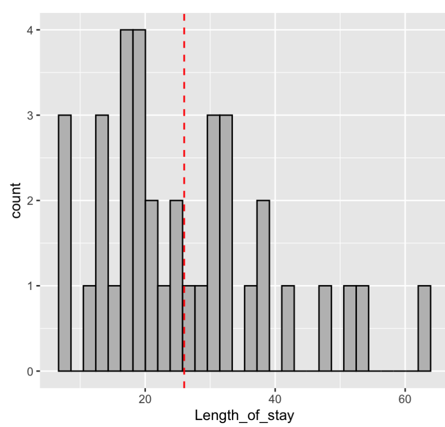
Table 1: summary statistics on length of hospital stay



(a) with and without acupressure



(b) with acupressure

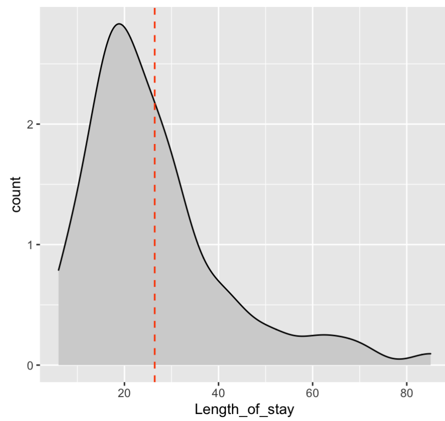


(c) without acupressure

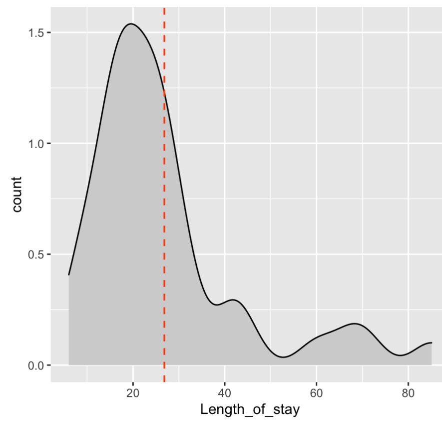
Figure 1: histograms of length of stay in terms of with or without acupressure

Table 1 illustrates the summary statistics for length of hospital stay for three cases. From the table, we can see that the mean and median values for length of hospital stay with acupressure and without acupressure are roughly the same. The interquartile range of length of hospital stay without acupressure is slightly larger than length of hospital stay with acupressure. Nevertheless, we cannot directly conclude that the procedure of acupressure effectively reduces the length of hospital stay. Hypothesis testing needs to be done further.

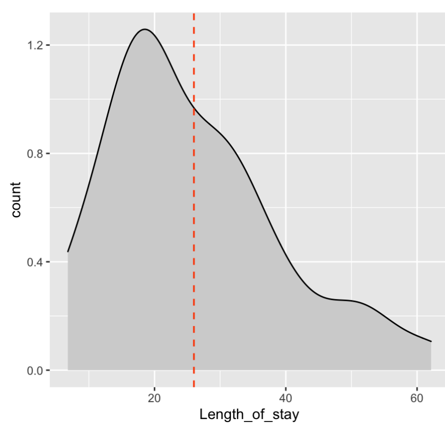
As shown in Figure 1, the red lines indicate the mean values of length of hospital stay for each scenario: (a) including both with and without acupressure (b) including only acupressure (c) including only without acupressure.



(a) with and without acupressure



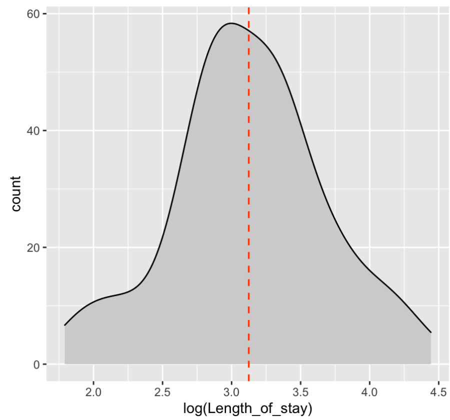
(b) with acupressure



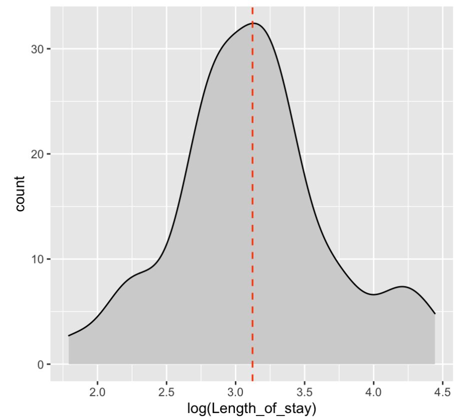
(c) without acupressure

Figure 2: density plots of length of stay in terms of with or without acupressure

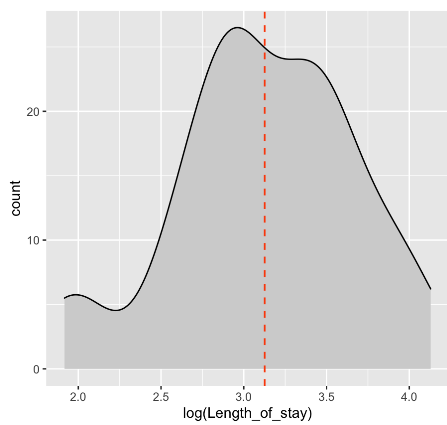
Figure 2 shows the density plots of length of hospital stay for three cases. We can clearly see that the distributions for length of stay for all three cases are right-skewed rather than not normal. Therefore, the normality assumption for length of hospital stay may not exactly hold, but we could normalize the data by doing log transformation. The red lines indicate the mean values.



(a) with and without acupressure



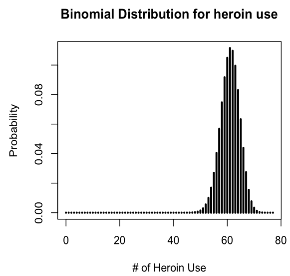
(b) with acupressure



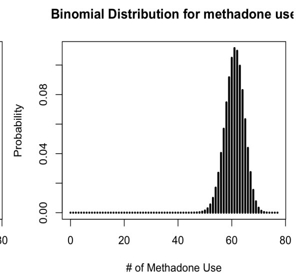
(c) without acupressure

Figure 3: density plots of length of stay after log transformation

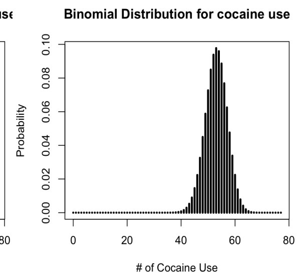
After log transformation, we can see that the assumption holds that the length of hospital stay is normal distribution for all three cases.



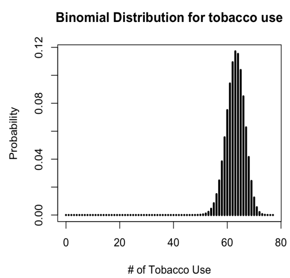
(a) heroin use



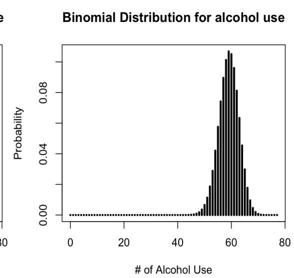
(b) methadone use



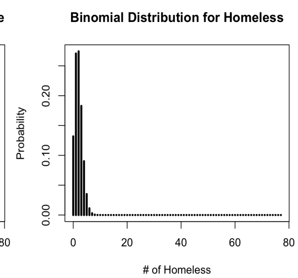
(c) cocaine use



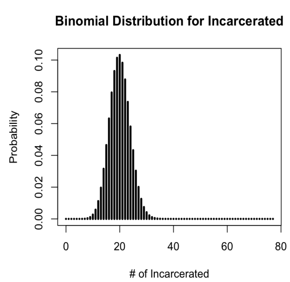
(d) tabacco use



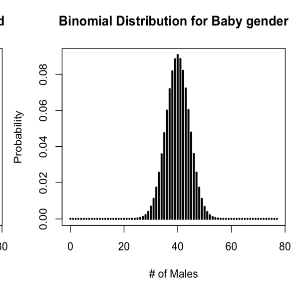
(e) alcohol use



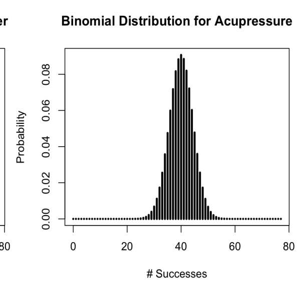
(f) homeless people



(g) incarcerated people



(h) baby gender



(i) acupressure

Figure 4: Binomial distributions of various categorical variables

As shown in Figure 4, all categorical variables in this dataset follows binomial distribution. We can also see that the probability of value being 1 for heroin use, methadone use, cocaine use, tobacco use, and alcohol use are roughly the same, which is around $\hat{p} = 65\%$. The expected value of samples is $n\hat{p}$, which is an unbiased estimate. For binomial distribution, if it satisfies 'Rule of Five', then the normal approximation (assumption) is valid. In our dataset, all binary variables except for 'Homeless' satisfy 'Rule of Five'. We can consider normal approximation for all binary variable in the dataset except for 'Homeless'. This is important when computing confidence intervals for those variables later.

3 Point Estimates and Confidence Intervals

First, we consider the point estimate for the continuous variable length of hospital stay. The maximum likelihood estimate for the mean of length of hospital stay with and without acupressure is 26.42, the MLE for variance is 236.75; the MLE for the mean of length of hospital stay with acupressure is 26.82, the MLE for variance is 296.13; the MLE for the mean of length of hospital stay without acupressure is 26.00, the MLE for variance is 172.21. The maximum likelihood estimate assumes the data are normal distributed. Next, we consider the point estimate for categorical variables in this dataset. Since the categorical variables in this dataset are Bernoulli random variables, the point estimate is $\hat{p}$. Specifically, for heroin use $\hat{p} = 0.79$; for methadone use $\hat{p} = 0.79$; for cocaine use $\hat{p} = 0.69$; for tobacco use $\hat{p} = 0.82$; for alcohol use $\hat{p} = 0.77$; for homeless $\hat{p} = 0.026$; for incarcerated $\hat{p} = 0.260$; for baby gender $\hat{p} = 0.52$; for acupressure $\hat{p} = 0.52$.

Assume the length of hospital stay is normally distributed, and that the variance is unknown. Then we can use the student's t-test to compute the confidence interval for length of hospital stay. The 95% CI for length of hospital stay is $[22.913, 29.944]$. The 95% CI for length of hospital stay with acupressure is $[21.249, 32.396]$. The 95% CI for length of hospital stay without acupressure is $[21.567, 30.438]$. For binomial distribution, if 'Rule of Five' is satisfied, then approximate confidence interval is used. In our dataset, all binary variables except for 'Homeless' satisfy 'Rule of Five'. The approximate confidence intervals for those variables are shown in the following table.

95% Approximate Confidence Intervals		
Binomial Variables	Lower Bound	Upper Bound
heroin use	0.689	0.868
methadone use	0.689	0.868
cocaine use	0.578	0.781
tobacco use	0.718	0.888
alcohol use	0.660	0.847
Incarcerated	0.175	0.367
Baby gender	0.410	0.627
Acupressure	0.410	0.627

Table 2: 95% approximate confidence intervals on binomial variables

For binomial variable 'Homeless', since it does not satisfy 'Rule of Five', we compute its exact confidence interval. The R function for testing exact confidence interval for binomial variable is binom.test. Therefore, the exact confidence interval for 'Homeless' variable is $[0.003, 0.09]$.

4 Hypothesis Testing

One of the scientific questions that we would like to investigate for this study is the potential efficacy of auricular acupressure augmentation of standard medical management of babies with NAS by comparing the length of hospital stay between babies randomized to receive standard medical care alone, or standard care plus auricular acupressure. For this topic, hypothesis testing can be used to investigate whether there is significant difference over length of hospital stay between standard medical procedure and standard medical procedure plus auricular acupressure. Therefore, we can categorize the samples into two groups: one with standard medical procedure alone, the other one with standard medical procedure plus auricular acupressure. And accordingly two-sample hypothesis testing is performed. Specifically, we perform two-sample hypothesis testing with independent samples since two groups of samples are independent.

First, we use F test to roughly test the equality of variances of two groups. The p-value is 0.105. We fail to reject the null hypothesis H_0. Therefore, there is significant evidence that variances of two groups are equal.

Since the variances of two groups are equal, we test the difference of the mean values of length of hospital stay between two groups with unknown and equal variances. The null hypothesis H_0 is the mean values between two groups are the same; the alternative hypothesis H_1 is the mean values between two groups are not the same, which is a two-sided test. The assumption made for this hypothesis testing is that the data is normally distributed and the variances of two groups are unknown and equal. The confidence level is 95%. The p-value for this hypothesis testing is 0.8165. Therefore, we can conclude that there is no significant difference between the mean values of length of hospital stay between two groups.

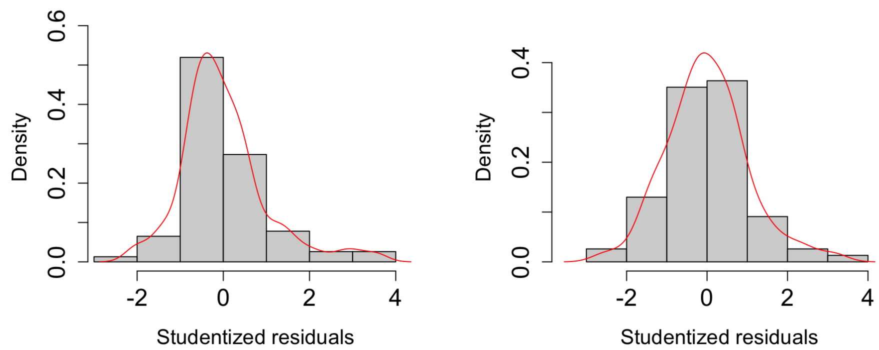
Bootstrap method is used to further perform the hypothesis testing. bootstrap subset sample size n = 25, 2000 trials are performed. We still assume the data is normally distributed as well as the variances of two groups are equal and unknown. After performing bootstrap method, the p-value is 0.489, the confidence interval is $[-7.74, 9.43]$. If we assume the sampling distribution is normal rather than the data is normally distributed (no assumption made for data itself in this case), then the confidence interval is $[-40.79, 42.47]$.

Hypothesis testing for difference of hospital stay between two groups performed above is under the assumption that the data is normally distributed. However, as we have seen in the summary statistics, length of hospital stay does not exactly follow a normal distribution. Instead, the distribution is right-skewed. Therefore, nonparametric method for hypothesis testing is necessary in this case. Since two groups of samples are independent with each other, Wilcoxon rank-sum test procedures with correction are performed on this dataset. Since the sample size is larger than 16, we only need the approximate distribution of R_1. After performing Wilcoxon rank-sum test procedures with correction, p-value is 0.7831. Bootstrap method is also performed in this case with bootstrap subset sample size n = 25, number of trials being 2000. The p-value is 0.7830. Therefore, for nonparametric method, with no assumption about the distribution of data, we conclude that there is no significant difference over length of hospital stay between two groups of samples.

5 Linear Model

Another scientific question that we are interested in is to identify which factor(s) are related with the decreased length of hospital stay. We can investigate this question by performing linear regression on this dataset. First of all, we build a model in which the dependent variable (outcome variable) is length of hospital stay, the independent variables (predictors) are all other variables in this dataset except for length of hospital stay. F test in regression compares the fits of different linear models. In other words, F test can assess multiple coefficients simultaneously.The F test for this linear model yields a p-value = 0.0271. Therefore, we can conclude that there is significant evidence that the model with predictors has a better fit on the data than the model with intercept only. We can also see that the p-value for Av_Nas_score is 0.000813, from which we can conclude that there is significant evidence that the coefficient of Av_Nas_score is not zero. The coefficient of Av_Nas_score is 6.59, which makes an increase contribution to length of hospital stay. In this model, for all other predictors except for Av_Nas_score, there is no significant evidence that the coefficients of those predictors are not zero. Among those predictors, if the coefficients are negative, then it means the predictors are related with the decreased length of hospital stay. Therefore, maternal_age, maternal_cocaine_use, maternal_tobacco_use, homeless, and birth weight are related with the decreased length of hospital stay in this model. The adjusted R-squared value is 0.1867.

Linear model assumes the data is normal distributed. Since length of hospital stay is not exactly normal, log transformation is performed. As we can see in Figure 5, the data becomes more normal after log transformation on studentized residuals of length of hospital stay. And accordingly, we redo a linear model on this log-transformed dataset.

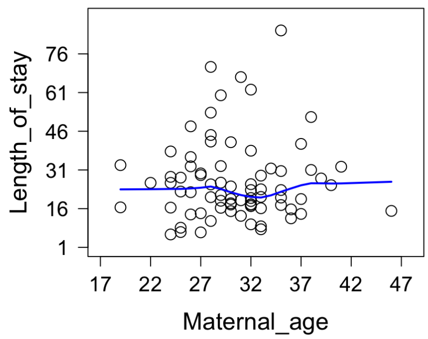


(a) before log transformation

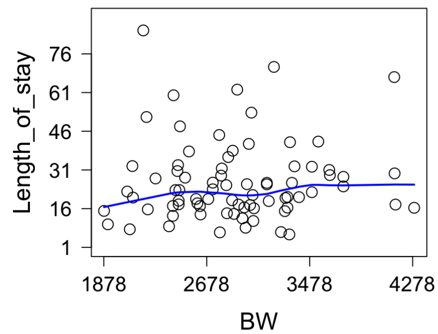
(b) after log transformation

Figure 5: distribution of studentized residuals before and after log transformation

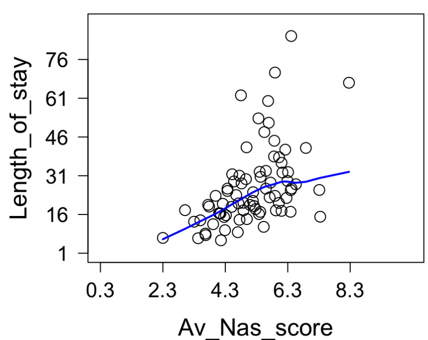
Next, we look closer on the linearity of this linear model. For investigation of linearity, only continuous variables are considered. Specifically, variables such as maternal age, birth weight, av_nas_score, gestational age, and time in hours from birth to first medical treatment for NAS are investigated. As shown in Figure 6, the linearity for all investigated variables is not perfect. Among these variables, the linearity of av_nas_score is the best.



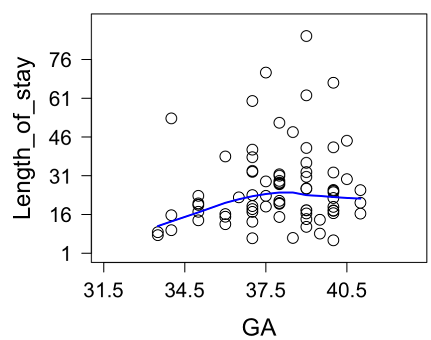
(a) linearity of maternal age



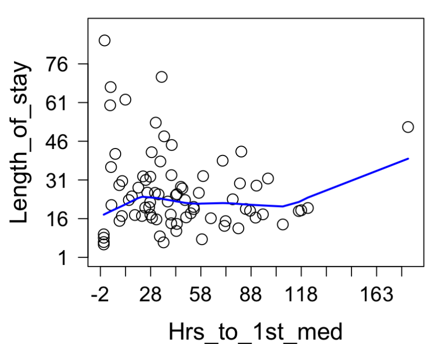
(b) linearity of birth weight



(c) linearity of av_nas_score



(d) linearity of gestational age



(e) linearity of time in hours

Figure 6: investigation of linearity for some continuous variables

Therefore, we choose to add quadratic term for each variable. We also compare quadratic model with cubic model in which we add both quadratic and cubic terms for each variable.

From the summary results of the quadratic model, we can see that the adjusted R-squared value is 0.1939, which increases compared to the model without quadratic term. From the summary results of the cubic model, we can see that the adjusted R-squared value is 0.1804, not as good as the original model. The R-squared value is 0.1843 for the model with both quadratic and cubic terms, not as good as the original model. Therefore, we can conclude that the quadratic model improves the linearity of the original model and accordingly yields better results. More comparison among different models is shown in Table 3 as follows.

Comparison among different linear models		
Models	Ad R-sq value	P-value
Original	0.1867	0.027
Original with log	0.3505	0.0003
Quadratic	0.1939	0.033
Quadratic with log	0.3776	0.0002
Cubic	0.1804	0.847
Cubic with log	0.3694	0.0004
Quad_Cub	0.1843	0.06
Quad_Cub with log	0.3557	0.0014

Table 3: Comparison among different linear models

From Table 3, we can see that the model with the highest adjusted R-squared value is quadratic model after log transformation. The adjusted R-squared value for this model is 0.3776, the p-value for this model is 0.0002. The p-value of F statistic of this model tells that there is significant evidence that the model with predictors is better than the model with intercept only. In this model, for each predictor, 'Av_Nas_score' and 'RaceH' have p-values lower than 0.05, which means that we are 95% confident that the coefficients of these two variables are not zero. Finally, we look for influential points by doing diagnostic

statistics analysis. After examination, the indexes for influential points are $5, 8, 13, 16, 17, 21, 23, 28, 29, 31, 45, 62, 63, 67, 71$.
Above all, the quadratic model with log transformation is used because it yields the highest adjusted R-squared value. The statistically significant predictors in this model are 'Av Nas score' and 'RaceH'. From this model, we can see that 'Maternal methadone use', 'Maternal cocaine use','Maternal tobacco use','Homeless','Baby gender', 'squared birth weight', 'squared gestational age','Hrs to 1st med',and 'squared Av Nas score' are the factors that are related with the decreased length of hospital stay.

6 Categorical Data Analysis

In addition to investigation of the potential efficacy of auricular acupressure augmentation of standard medical management of babies with NAS as well as investigation of factors that are related with the decreased length of hospital stay, we would also like to do some association analysis between categorical variables in this dataset. Particularly, we are interested in investigating whether the race of the mothers are associated with their behaviors. Specifically, we would like to investigate by doing association test between race and the maternal use of heroin, the maternal use of methadone, the maternal use of alcohol, status of homelessness, status of incarceratedness. $N\ timesC$ contingency tables are established for each association test. The chi-square test and the Fisher's exact test are used depending on the following criteria. The chi-square test is valid if 1) No more than 20% of the cells have expected values which are smaller than 5 2) No cell has an expected value which is smaller than 1.

After computing expected values for each association test. For the association tests of race and status of homelessness, and of race and use of heroin, and of race and use of methadone, we use Fisher's exact test; For all other association tests, since they satisfy the rules above, we use the chi-square test. The test results are shown in Table 4 as follows.

P-values for different association tests		
Tests	Chi-square	Fisher's Exact
Race/Incarcerated	0.7756	N/A
Race/Heroin	N/A	0.7099
Race/Methadone	N/A	0.8732
Race/Alcohol	0.5884	N/A
Race/Homeless	N/A	1

Table 4: Association tests for different categorical variables

From the results above, we can conclude that there is no significant association between race of the mothers and variables such as the maternal use of heroin, the maternal use of methadone, the maternal use of alcohol, the status of homelessness, and the status of incarceratedness. Table 5 shows an exemplary 3×2contingency table for association test of race and use of alcohol.The values within parentheses in the following table are the expected values. Other results of contingency tables for association tests can be referred to R codes.

3×2 contingency table of race and use of alcohol			
Race	Alcohol	Non-alcohol	Total
A	17(18.390)	7(5.610)	24
C	31(29.117)	7(8.883)	38
H	11(11.494)	4(3.506)	15
Total	59	18	77

Table 5: 3×2 contingency table of race and use of alcohol

7 Study Plan

In this study, we are interested in the potential efficacy of auricular acupressure augmentation of standard medical management of babies with NAS by comparing the length of hospital stay between babies randomized to receive standard medical care alone, or standard care plus auricular acupressure. We can also use other criteria besides the length of hospital stay to determine the efficacy of auricular acupressure such as monitoring some factors associated with the recovery of NAS. This can be treated as an extension of the current study. We can also collect more features about the maternal data as well as the neonatal data because we don't know for sure which feature is important for the potential efficacy of auricular acupressure unless we test it. For example, besides collecting the behavioral data about mothers such as their use of drugs, we can also collect some biological data about the mothers, such as their weights, heights, blood pressure, etc. We should also keep in mind about the correlation between features. We don't want too many high-correlated features as predictors when performing linear regression since predictors should be as independent as possible. Meanwhile, the sample size can be further enlarged to more than 200, we want to collect as many samples as possible since current dataset is relatively small. If the dataset is large enough, we could be more confident to make the assumption that the data is normally distributed by applying the central limit theorem.

8 Conclusion and Discussion

The exploratory data analysis has been done on this dataset by doing summary statistics. The MLE point estimates has been computed for some variables as well as their confidence intervals. The nonparametric method such as bootstrap has been used. Hypothesis testing has been performed for the difference of length of hospital stay between two groups of patients, one receiving standard medical procedure, the other receiving auricular acupressure. Bootstrap method has been used for computing confidence intervals as well asp-values given different subset samples. After doing hypothesis testing, it is concluded that there is no significant difference over length of hospital stay between two groups of patients. Moreover, linear regression models have been performed to identify which factors are related with the decreased length of hospital stay. After assessing the linearity and normality of the models, we chose the quadratic model with log transformation on length of hospital stay because this model yields the highest adjusted R-squared value among the others. After performing linear regression model, we concluded that the model with predictors better fits the data than the model with intercept only. We also concluded that two variables 'Av

Nas score' and 'RaceH' are statistically significant. Moreover, from the model, we can clearly know that variables such as 'Maternal methadone use', 'Maternal cocaine use','Maternal tobacco use','Homeless','Baby gender', 'squared birth weight', 'squared gestational age','Hrs to 1st med',and 'squared Av Nas score' are the factors that are related with the decreased length of hospital stay. Then we performed categorical data analysis by applying association tests on variables of interest. Particularly, we performed association tests between race and the maternal use of heroin, the maternal use of methadone, the maternal use of alcohol, the status of homelessness, and the status of incarceratedness. Various $N \times C$ contingency tables are established. Chi-square test as well as Fisher's Exact test were performed by following the criteria. From the tests, we can concluded that there is no significant association between the race of the mothers and variables listed above.

The strength of this study is that we fully use different features of this dataset, we've not only explored the continuous variables and built linear regression models upon it, we've also explored categorical variables and done categorical data analysis such as association tests. The weakness of this study is that the sample size is not large enough, it could be better if we have a larger sample size. Meanwhile, the data is not strictly normally distributed. Although, the normality of the data has been evaluated as well as improved by doing log transformation, it is still not perfect. We could potentially enlarge the sample size so that the normal approximation could be more accurate due to the central limit theorem.

For further study, we can use some other factors besides the length of hospital stay to determine the efficacy of auricular acupressure. For example, we can measure some factors associated with the disease and monitor the data as patients are on their way to recovery. And we can use these recorded data to determine the efficacy of added auricular acupressure. This can be treated as an extension of the current study.