Analysis of Cognitive Score of Infants

PROBABILITY MODELS

BANA 7031 - 002

Presented by

Pati, Pruthvi Ranjan Reddy – M12908411

Tammineedi, Mahitha Sree – M12789619

# DATA

The dataset used in this project contains data on cognitive scores of infants in a study of early childhood intervention. There are 58 infants each in treatment and control group whose cognitive scores are recorded at ages 1,1.5, and 2 years. There are 309 rows and 4 columns in the dataset.

The columns in the data are as follows:

|  |  |
| --- | --- |
| **Name** | **Description** |
| Id | An ordered factor of the id number for each infant |
| Cog | A numeric cognitive score |
| Age | The age of the infant at the measurement |
| Trt | A factor with two levels, "N" and "Y", indicating if the infant is in the early childhood intervention program |

A new column ‘High’ is defined which has two levels – 1 or 0 depending whether the cognitive score is greater than or equal to 110.

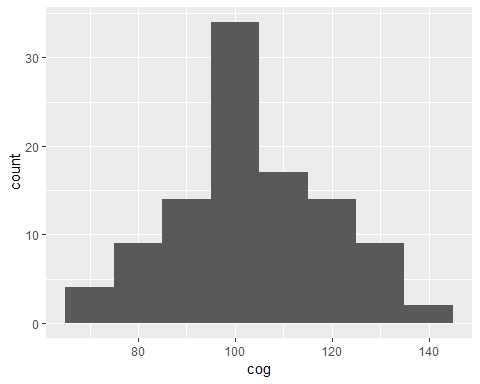
# PROJECT SUMMARY

The project aims at understanding the cognitive scores of infants using various statistical methods. Cognitive scores are analyzed and the distribution is found to be close to normal as inferred from the qq-plot. Cumulative distribution function is plotted using Empirical CDF with a 95% confidence band. We performed Wald test to test whether the average cognitive scores of treatment group is equal to average cognitive score of control group. A 95% confidence interval of difference in average cognitive scores of treatment and control groups is calculated using MLE and bootstrap. Bayesian analysis is done to obtain the posterior distribution of proportion of infants with High cognitive scores considering a beta prior distribution. All the above analysis is done for infants aged 1.5 years.

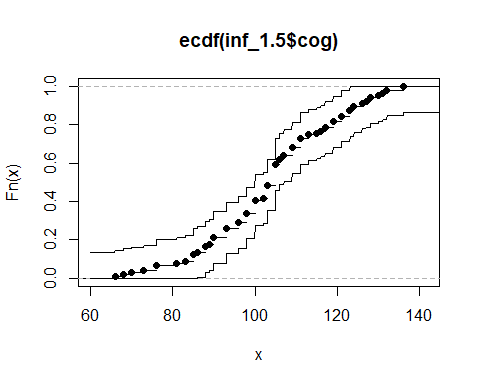
# Analysis

## DISTRIBUTION OF COGNITIVE SCOREs

To understand the distribution of cognitive scores of infants aged 1.5 years, a histogram and qq-plot is plotted. From the below plots we infer that the distribution is close to normal.

Empirical cumulative distribution and the 95% confidence band of the cognitive scores of infants (Age 1.5) is as shown below.



## MLE of difference between treatment and control

The aim of **3.2** and **3.3** is to provide a 95% confidence interval of difference in average cognitive scores of treatment and control groups. The variables “TRT” has two categories “Treatment” and “Control”. This enables us to compute the average cognitive scores of infants under these two categories. In this section MLE of the difference is calculated as follows:

MLE of μ2 - μ1 =  x̅treatment - x̅control = 14.40421

## Standard error of MLE using bootstrap AND CONFIDENCE INTERVALS

In this section, standard error for the above MLE is calculated analytically, with parametric and non-parametric bootstrap. The 95% confidence intervals are built using these standard errors.

|  |  |
| --- | --- |
| Analytical | (9.192859,19.615570) |
| Parametric Bootstrap | (9.172427,19.636002) |

Non-Parametric Bootstrap

|  |  |
| --- | --- |
| Normal | (9.007697,19.800732) |
| Pivotal | (9.169636,19.842912) |
| Quantile | (8.965517,19.638793) |

## Hypothesis test for difference between treatment and control

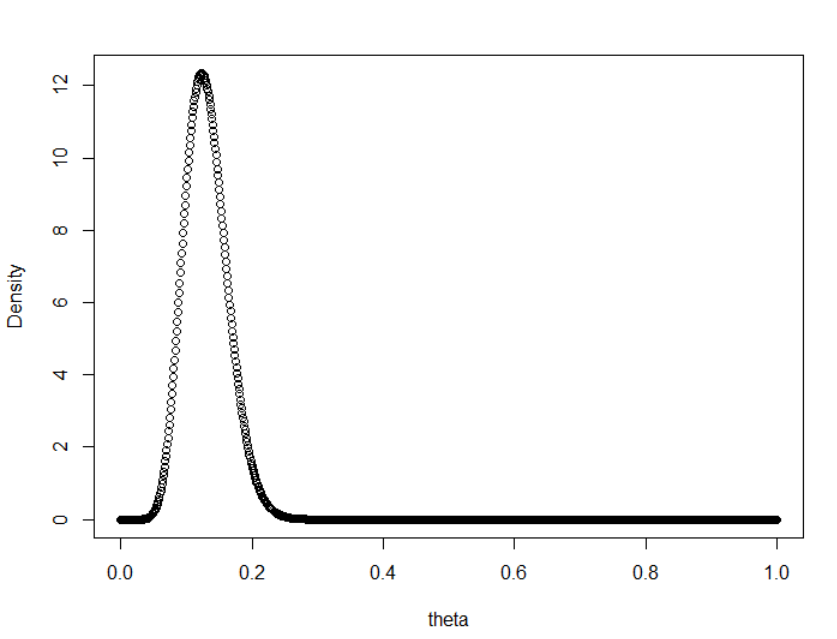
Let’s check if the cognitive scores of infants aged 1.5 years significantly differ for treatment and control group. Our hypotheses are:

Wald test is carried out to test the hypothesis. P-value obtained is 3.238825e-08. At a significance level of 0.05, we reject the null hypothesis that mean of cognitive scores for treatment group and control group are same and conclude that there is significant difference in cognitive scores of treatment group and control group.

## Bayesian Analysis on proportion of high cognitive score infants

The ‘High’ column in the data provides information whether the cognitive score of an infant is greater than 110. This will be used to calculate the proportion of high cognitive infants with age 1.5 years and to find the posterior distribution of proportion of high cognitive infants (Age – 1.5 yrs.) using Bayesian Analysis.

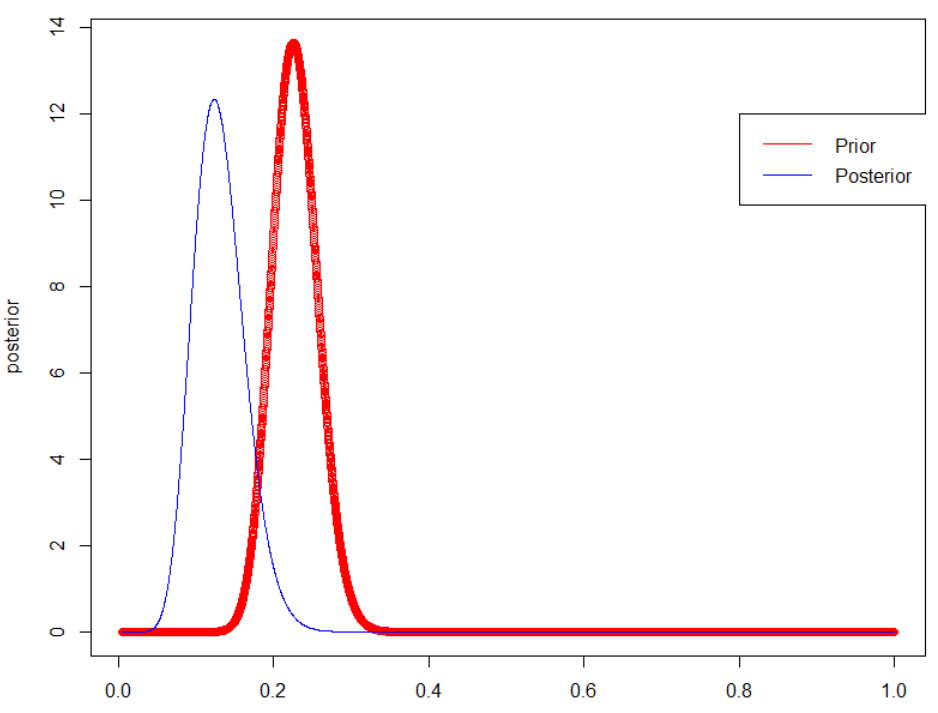
For calculating posterior distribution, we need a prior distribution. In order to estimate this, we have considered that distribution of prior is beta, median of prior is 0.25, 99th percentile of prior is 0.4, and 0.1th percentile is 0.05. Using these values and LearnBayes library, estimated prior distribution is Beta(13.710, 91.218)



For estimating likelihood function, a random sample of 100 is chosen and it is observe that there are 33 successes. The posterior distribution is given by

*Beta( (alpha of prior) + successes, (beta of prior) + total observations – success )*

The obtained beta distribution is Beta ( 46.71, 158.219)



Statistics of posterior distribution:

Mean of posterior distribution = 0.228

Variance of posterior distribution = 36.24

# CONCLUSION

The cognitive scores of 1.5 year old infants is analyzed and it is observed that the distribution of cognitive scores is close to normal. An empirical CDF is also calculated with 95% confidence bands. Then we calculated the MLE of difference between the means of cognitive scores of treatment group and control group and 95% confidence interval for the MLE is calculated using various methods – non-parametric bootstrap, parametric bootstrap (assuming the scores are normally distributed), and analytical method. It is observed that the confidence limits obtained are almost same in all the methods. From the hypothesis test on difference of means of cognitive scores of treatment and control group, we concluded that there is significant difference between the two groups. A posterior distribution of proportion of infants with High cognitive scores is also calculated assuming a beta prior distribution.

# Bibiliography

1. All of Statistics: A Concise Course in Statistical Inference by Larry Wasserman
2. <http://a-little-book-of-r-for-bayesian-statistics.readthedocs.io/en/latest/src/bayesianstats.html>
3. <https://cran.r-project.org/web/packages/mlmRev/mlmRev.pdf>