

Double Trouble? A statistical analysis of child and parental outcomes comparing singletons to non-singletons

Final Year Project

By Cathaoir Agnew

Department of Mathematics and Statistics

University of Limerick

Supervisor

Dr Helen Purtill

Table of Contents

[1 Introduction 9](#_Toc37080556)

[1.1 Objective of the Project 10](#_Toc37080557)

[1.2 Growing Up in Ireland Study and Infant Cohort 11](#_Toc37080558)

[1.2.1 Growing Up in Ireland Sample Frame Design and Sample Selection 11](#_Toc37080559)

[2 Literature review 12](#_Toc37080560)

[3 Data Measures and Data Management 14](#_Toc37080561)

[4 Methodology 17](#_Toc37080562)

[4.1 Descriptive statistics 17](#_Toc37080566)

[4.2 Between Group tests 17](#_Toc37080567)

[4.3 Multiple Linear Regression 17](#_Toc37080568)

[4.4 Binary Logistic Regression 18](#_Toc37080569)

[Interpreting Odds Ratio 18](#_Toc37080570)

[Relative Risk 19](#_Toc37080571)

[4.5 Poisson Regression 20](#_Toc37080578)

[4.6 Negative Binomial Regression 21](#_Toc37080579)

[4.7 Cluster Analysis 21](#_Toc37080580)

[4.7.1 Methods of Distance Measures for Cluster Analysis 23](#_Toc37080581)

[4.7.2 Two Step Cluster Analysis 26](#_Toc37080582)

[4.7.3 Hierarchical Cluster Analysis 27](#_Toc37080583)

[4.7.4 K-Means Clustering 29](#_Toc37080584)

[4.8 Quantile Regression 30](#_Toc37080585)

[5 Exploratory Data Analysis and Between Group Tests 32](#_Toc37080586)

[5.1 Descriptive Statistics for Wave 1 Cohort Data 32](#_Toc37080587)

[5.2 Descriptive Statistics for Wave 3 Cohort Data 38](#_Toc37080588)

[6 Analysis of Wave 1 data 43](#_Toc37080589)

[6.1 Length of Pregnancy with key predictors non-singleton and singleton pregnancy 43](#_Toc37080590)

[6.1.1 Results/Conclusion of Length of Pregnancy 44](#_Toc37080591)

[6.2 Depression of Primary carer with key predictor non-singleton and singleton pregnancy 45](#_Toc37080592)

[6.3 Binary Logistic Regression of Primary Carer Depression 46](#_Toc37080593)

[6.3.1 Results/Conclusion of Binary Logistic Regression on Depression of Primary Carers 47](#_Toc37080594)

[6.4 Poisson regression of Primary Carer Depression 48](#_Toc37080597)

[6.5 Negative Binomial Regression of Primary Carer Depression 49](#_Toc37080598)

[6.5.1 Results/Conclusion of Primary Carer Depression Models 50](#_Toc37080599)

[7 Analysis of Wave 3 data 51](#_Toc37080600)

[7.1 Negative Binomial Regression of Primary Carer Parental stress Wave 3 52](#_Toc37080601)

[7.1.1 Results/Conclusion of Negative Binomial Regression of Primary Carer Parental Stress …………………………………………………………………………………………………………………………………53](#_Toc37080602)

[7.2 Hierarchical Cluster Analysis of Wave 3 Academic Variables 53](#_Toc37080603)

[7.2.1 Results/Conclusion of the Cluster Analysis on Academic Variables 56](#_Toc37080604)

[7.3 K-means Cluster Analysis of Strength and Difficulties on Wave 3 Data 57](#_Toc37080605)

[7.3.1 Findings from the K-Means Cluster Analysis 60](#_Toc37080606)

[7.4 Multiple Regression of Wave 3 Naming Vocabulary Scores with key predictors non-singleton and singleton 61](#_Toc37080607)

[7.4.1 Results/Conclusion of Multiple Regression of Naming Vocabulary Scores 62](#_Toc37080608)

[7.5 Quantile Regression on Naming Vocabulary Wave 3 63](#_Toc37080609)

[7.6 Results/Conclusion of Quantile Regression on Naming Vocabulary Scores 66](#_Toc37080610)

[7.7 Marriage Status of Primary Carer from Wave 1 to Wave 3 Crosstabs 66](#_Toc37080611)

[7.7.1 Results/Conclusion of Marital Status from Wave 1 to Wave 3 Crosstabs 68](#_Toc37080612)

[7.8 Multiple Regression of Wave 3 Picture Similarities Scores Squared with key predictors non-singleton and singleton 68](#_Toc37080613)

[7.8.1 Results/Conclusion of Picture Similarity Scores 71](#_Toc37080614)

[8 Summary 71](#_Toc37080615)

[9 Bibliography 73](#_Toc37080616)

[10 Appendix 76](#_Toc37080617)

[10.1 Data Preparation file 76](#_Toc37080618)

[10.2 Multiple Regression for Length of Pregnancy code in Wave 1 in R 78](#_Toc37080619)

[10.3 Poisson Regression for Depression Scores in Wave 1 code in R 78](#_Toc37080620)

[10.4 Negative Binomial Regression for Depression Scores in Wave 1 code in R 78](#_Toc37080621)

[10.5 Quantile Regression for Naming Vocabulary Scores in Wave 3 code in R 78](#_Toc37080622)

[10.6 Negative Binomial Regression of Parental Stress Scores in Wave 3 code in R 79](#_Toc37080623)

[10.7 Multiple Regression of Naming Vocabulary Scores in Wave 3 code in R 79](#_Toc37080624)

[10.8 Multiple Regression for Picture Scores Squared in Wave 3 code in R 80](#_Toc37080625)

[10.9 Further graphs and boxplots for Wave 3 descriptive statistics 80](#_Toc37080626)

List of Tables

[Table 1: Central Statistics Office Data 10](#_Toc37080668)

[Table 2: Odds Ratio Table 19](#_Toc37080669)

[Table 3: Descriptive Statistic 33](#_Toc37080670)

[Table 4: Descriptive Statistics for Wave 3 39](#_Toc37080671)

[Table 5: Multiple Regression of length of pregnancy 44](#_Toc37080672)

[Table 6: Logistic Regression of Depression of Primary Carer 47](#_Toc37080673)

[Table 7: Poisson Regression of Primary Carer Depression 49](#_Toc37080674)

[Table 8: Negative Binomial Regression of Primary Carer Depression 50](#_Toc37080675)

[Table 9: Negative Binomial Regression of Primary Carer Parental Stress 53](#_Toc37080676)

[Table 10: Cluster Analysis of Academic Variables 55](#_Toc37080677)

[Table 11: Cluster Analysis of Academic Variables Controlling for Gender and Nonsingleton 56](#_Toc37080678)

[Table 12: Cluster Analysis of Academic Variables Controlling for Social Factors 56](#_Toc37080679)

[Table 13: Cluster Analysis of SDQ Variables 59](#_Toc37080680)

[Table 14: Cluster Analysis of SDQ Variables Controlling for Gender and Non-Singleton 60](#_Toc37080681)

[Table 15: Cluster Analysis of SDQ Variables Controlling for Social Factors 60](#_Toc37080682)

[Table 16: Multiple Regression of Wave 3 Naming Vocabulary Scores with key predictors non-singleton and singleton 62](#_Toc37080683)

[Table 17: Quantile Regression of Naming Vocabulary Scores 65](#_Toc37080684)

[Table 18: Cross Tab of Marital Status from Wave 1 to Wave 3 67](#_Toc37080685)

[Table 19: Multiple Regression of Picture Similarities Scores Squared 71](#_Toc37080686)

Table of Figures

[Figure 1: Source Wikipedia 20](#_Toc37092635)

[Figure 2: An Example of Clusters in Data 22](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092636)

[Figure 3: Source SlidePlayer Cluster Analysis by Dr. Chayada Bhadrakom 23](#_Toc37092637)

[Figure 4: Example of a Cluster Tree 26](#_Toc37092638)

[Figure 5: Source Towards Data Science 28](#_Toc37092639)

[Figure 6: Example of a Dendrogram formed by Hierarchical Clustering 29](#_Toc37092640)

[Figure 7: Box Plot of length of pregnancy controlling for Non-singleton 34](#_Toc37092641)

[Figure 9: Length of Pregnancy For Non-singleton 35](#_Toc37092642)

[Figure 8: Length of Pregnancy For Singleton 35](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092643)

[Figure 10: Box plot of Depression score of primary carer controlling for non-singleton 35](#_Toc37092644)

[Figure 11: Depression Scores of Primary Carers For singleton 35](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092645)

[Figure 12: Depression Scores of Primary Carers For Non-singleton 35](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092646)

[Figure 13: Box Plot of Age of Primary Carer Controlling for Non-singleton 36](#_Toc37092647)

[Figure 14: Age of Primary Carer for Singleton 36](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092648)

[Figure 15: Age of Primary Carer for Non-singleton 36](#_Toc37092649)

[Figure 16: Box Plot of Annual Income Controlling for Non-singleton 37](#_Toc37092650)

[Figure 17: Income For Singleton Households 37](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092651)

[Figure 18: Income for Non-Singleton Households 37](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092652)

[Figure 19: Box Plot of Parental Stress Scores Controlling for Non-singleton 40](#_Toc37092653)

[Figure 20: Parental Stress Scores of Primary Carer for Singleton 40](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092654)

[Figure 21: Parental Stress Scores of Primary Carer for Non-singleton 40](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092655)

[Figure 22: Box Plots of SDQ Emotional Scores Controlling for Non-singleton 41](#_Toc37092656)

[Figure 24: SDQ Emotional Score for Non-singleton 41](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092657)

[Figure 23: SDQ Emotional Score for Singleton 41](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092658)

[Figure 25: Box Plots of SDQ Conduct Scores Controlling for Non-Singleton 42](#_Toc37092659)

[Figure 26: SDQ Conduct Scores for Singleton 43](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092660)

[Figure 27: SDQ Conduct Scores for Non-singleton 43](#_Toc37092661)

[Figure 28: Dendrogram of Academic Variables 54](#_Toc37092662)

[Figure 29: Dendrogram of SDQ Variables 58](#_Toc37092663)

[Figure 30: Output of Quantile Regression 64](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092664)

[Figure 31: Picture Similarities Scores 69](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092665)

[Figure 32: Picture Similarities Scores Squared 69](#_Toc37092666)

[Figure 33: Residuals of Picture Similarities Scores 70](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092667)

[Figure 34: Residuals of Picture Similarities Scores Squared 70](#_Toc37092668)

[Figure 35: Boxplot of SDQ Hyperactivity Controlling for Non-singleton 81](#_Toc37092669)

[Figure 36: SDQ Hyperactivity Scores for Singleton 82](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092670)

[Figure 37: SDQ Hyperactivity Scores for Non-singleton 82](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092671)

[Figure 38: SDQ Peer Problems Score Controlling for Non-Singletons 82](#_Toc37092672)

[Figure 39: SDQ Peer Problem Scores for Singleton 83](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092673)

[Figure 40: SDQ Peer Problem Scores for Non-singleton 83](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092674)

[Figure 41: SDQ Prosocial Scores Controlling for Non-singleton 83](#_Toc37092675)

[Figure 42: SDQ Prosocial Scores For Singleton 84](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092676)

[Figure 43: SDQ Prosocial Scores For Non-Singleton 84](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092677)

[Figure 44: Naming Vocabulary Scores Controlling for Non-Singleton 84](#_Toc37092678)

[Figure 45: Naming Vocabulary Scores for Singleton 85](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092679)

[Figure 46: Naming Vocabulary Scores for Non-Singleton 85](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092680)

[Figure 47: Picture Similarities Scores Controlling For Non-singleton 85](#_Toc37092681)

[Figure 48: Picture Similarities Scores For Singleton 86](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092682)

[Figure 49: Picture Similarities Scores for Non-singleton 86](#_Toc37092683)

[Figure 50: Attitudes Scores Controlling for Non-singleton 86](#_Toc37092684)

[Figure 51: Attitude Scores for Singleton 87](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092685)

[Figure 52: Attitude Scores for Non-singleton 87](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092686)

[Figure 53: Language Scores Controlling for Non-singletons 87](#_Toc37092687)

[Figure 54: Language Scores for Singleton 88](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092688)

[Figure 55: Language Scores for Non-singleton 88](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092689)

[Figure 56: Linking Scores Controlled for Non-singleton 88](#_Toc37092690)

[Figure 57: Linking Scores for Singleton 89](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092691)

[Figure 58: Linking Scores for Non-singleton 89](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092692)

[Figure 59: Reading Scores Controlling for Non-singleton 89](#_Toc37092693)

[Figure 60: Reading Scores for Singleton 90](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092694)

[Figure 61: Reading Scores for Non-Singleton 90](https://d.docs.live.net/8c047d2e1c8b4a58/Desktop/FYP/Write%20up%20After%20xmas/16171659_Cathaoir_Agnew_1.docx#_Toc37092695)

*Abstract*

The aim of this project was to statistically investigate differences in child and parental outcomes comparing singletons to non-singletons using the Growing Up in Ireland Cohort Study.

Using the Wave 1 data from the Infant Cohort of the Growing Up in Ireland (GUI) study, a multiple regression estimated that the length of pregnancies for a non-singleton were on average 3.7 weeks shorter than a singleton pregnancy. Shorter pregnancies result in lower birth weights of the children which in turn can potentially carry adverse health risks for the children. A binary logistic and negative binomial regression were used on the depression scores to investigate how non-singleton pregnancies affected the mental health of the primary carer compared to their singleton counterparts. The odds of being depressed for a non-singleton primary carer was estimated to be 1.87 times the odds of being depressed for a singleton primary carer. Negative binomial regression analysis estimated the mean depression score of the primary carers of non-singletons to be 2.61 points higher than that of singletons.

Data from Wave 3 of the Infant Cohort of the GUI recorded information about the parents, child and teachers when the study children were 5 years old. Cluster analysis was applied to investigate the social and cognitive ability of the children comparing non-singletons to singletons. A k-means cluster analysis was used on Strength and Difficulties variables. From the k-means cluster analysis, the results suggest that non-singleton females may be more likely to be classified as hyperactive than singleton females. There were no significant differences for non-singleton vs singleton males across the cluster. A hierarchical cluster analysis of academic variables was used to examine the cognitive ability of the children. The findings suggest that non-singletons tended to perform worse in academic tests in school as judged by their teachers, with more non-singleton males being in the lowest performing cluster than their singleton counterparts (18% vs 13.7% respectively). The evidence was similar for females, with more non-singleton females being in the lowest performing cluster than their singleton counterparts (13.1% vs 8.2% respectively).

Primary carer stress scores were recorded at Wave 3. A negative binomial regression analysis of stress, found that primary carer parental stress scores of non-singletons was only slightly higher than that of singletons. When looking at marital status from wave 1 data compared to wave 3 there was no statistically significant evidence to suggest that families with a multiple birth are more likely to separate or divorce than any other families, but more data would be needed to further investigate this. Finally, a quantile regression analysis was undertaken to examine naming vocabulary scores. This analysis did not find any significant differences on naming vocabulary scores between non-singleton and singleton children after controlling for demographic variables.

*Acknowledgement*

First and foremost, I would like to express my sincere appreciation and gratitude to my supervisor, Dr. Helen Purtill, for the continuous support and guidance throughout my final year project, for her patience, motivation and expert knowledge.

Secondly, I would like to acknowledge and thank Dr. Norma Bargary as the second reader of this project, for her insightful comments and encouragement on this project.

# Introduction

The aim of this project is to statistically model differences in child and parental outcomes between singleton and non-singleton births using data from the Infant Cohort of the Growing Up in Ireland study. Due to the advances of medicine and the introduction of assisted reproduction techniques (ART), such as in vitro fertilization (IVF), there is an increasing number of multiple births (Kulkarni, et al., 2013). Previous research suggests that nonsingleton infants and their mothers have increased chances of medical and physiological risks in comparison to singletons (Wimalasundera , et al., 2003).

In Ireland, the Central Statistics Office (CSO) is responsible for the collection of data relating to information of the economic, social and general activities in Ireland. The twinning rate is the proportion of twin deliveries in a given year out of the total number of deliveries, expressed per 1,000 deliveries (Pison, et al., 2019). The percentage of twinning rates has increased from 10.5% to 19% of births from 1986 to 2016, with 2016 being the most up to date year on the Central Statistics Office website (CSO, 2019), .

The average age at which mothers have their first baby in Ireland has increased from 29.3 years old from 1986 to 32.7 in 2016, **Error! Reference source not found.** (CSO, 2019). Some of the outcomes found to be associated with multiple births in international studies are preterm births coupled with low birthweights, mental health and financial problems for parents and child development issues (Beatrice, et al., 2002). Reduced birthweights are a significant risk factor for infant mortality, child development problems and short and long-term disabilities (Yadav, et al., 2011). It is of interest to examine the effects of non-singleton births on child and parental outcomes compared to singleton births within Ireland.

In Ireland, parents of non-singleton babies receive a special grant at the time of birth, and later when the children are 4 and 12 years old. The child benefits payments are increased by 50% in the case of twins and doubled when there are triplets or higher. However, there is no increase in parental leave. In other countries there are such arrangements and benefits to take account of the different circumstances that arise for non-singleton birth children, both in terms of ongoing costs and one-off costs, and reflecting benefits as well as parental leave. Many countries offer extra maternity leave for non-singleton pregnancies such as France, Spain, Belgium, Sweden and the Czech Republic.

In France, paternity leave is raised from 11 days to 18 days with maternity leave benefits paid for twice as long for non-singleton births than singletons. Pre-maternal leave in Belgium starts two weeks earlier for non-singleton births, and in Italy the period for parental leave is longer. Spain gives an additional two weeks maternity leave for each multiple child, and in Sweden, this raises to 6 months. In the Czech Republic maternity benefits lasts an extra 9 weeks in the case of a non-singleton birth (37 weeks).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Year | 1986 | 1991 | 1996 | 2001 | 2006 | 2011 | 2016 |
| Twins (Number) | 642 | 610 | 675 | 891 | 990 | 1307 | 1189 |
| Triplets (Number) | 8 | 6 | 22 | 24 | 17 | 27 | 24 |
| Quadruples (Number) | - | 2 | - | 2 | 1 | 1 | 2 |
| Quintuplets (Number) | - | - | - | 1 | - | - | - |
| Twinning Rate | 10.5 | 11.7 | 13.5 | 15.6 | 15.4 | 18.0 | 19.0 |
| Age Of Mother | 28.9 | 29.6 | 30 | 30.3 | 31.1 | 31.8 | 32.7 |

Table : Central Statistics Office Data

## Objective of the Project

The objective of this project is to examine the differences in the measures of child and parental well-being as well as mental developments such as; cognitive and social involvements between singletons and non-singletons pregnancies in Ireland. Firstly, using the Growing Up in Ireland (GUI) Wave 1 data from the Infant Cohort, when the children of the study were 9 months old. The aim was to investigate the relationship between the mental wellbeing of the primary carer and having a non-singleton pregnancy. It was also of interest to statistical model differences in the length of pregnancy, as having premature pregnancies carries medical risks on both the mother and the child. There is more of an emphasis on the relationship between the mental wellbeing of the primary carer and having a non-singleton pregnancy in wave 1 data, as the topic of mental health is becoming a more important and necessary matter in today’s society, coupled with the fact that we can’t investigate cognitive and social involvements of a 9 month old child.

The second part of the project will use the Wave 3 data from the GUI Infant Cohort, when the children of the study were 5 years old. The aim is to statistically analyze the differences between non-singleton and singleton children’s developments such as; cognitive and social involvements. The aim would be to look at academic results to get an insight into the cognitive developments of both the non-singleton and singleton children as well as using Strength and Difficulties scores to analyze the social development. To use multiple regression and quantile regression on naming vocabulary scores and picture scores to get further insights into the academic ability of the children and compare the non-singletons scores to the singletons. It was also of interest to investigate the relationship between the mental wellbeing of the primary carer and having a non-singleton pregnancy in wave 3 to get an insight into how their mental states vary from wave 1 to wave 3. Finally, to see if potentially having non-singletons has any effect on the parent’s marital status.

## Growing Up in Ireland Study and Infant Cohort

The Growing up in Ireland (GUI) study is an observational longitudinal study undertaken to collect data on factors that influence and contribute to the wellbeing of children living in Ireland. A two-age cohort longitudinal design was adopted, collecting data on 19,702 children. Of the 19,702 children, 11,134 of these were infants aged nine months, born between 1st December 2007 and the 30th June 2008, with the data being collected between September 2008 and April 2009. In the Child Cohort, data was collected on 8,568 nine-years old at the start of the survey. The Growing Up in Ireland was funded by the Irish Government with an estimate of annual costs to be in the region of €2.5 million. This project uses data from the Infant Cohort.

In April 2006, the GUI project began collecting data, led by the Economic and Social Research Institute and Trinity College Dublin. The study has nine main objectives, each important to understanding the physical health, development and wellbeing of the child along with being aware of the future outcomes for the children. The objectives of the GUI study, outlined in the request for tenders by the department of health and children, (Department of Health and Children, 2005), include:

• To describe the lives of Irish children, to establish what is typical and normal as well as what is atypical and problematic.

• To chart the development of Irish children over time, to examine the progress and wellbeing of children at critical periods from birth to adulthood.

• To identify the persistent adverse effects that lead to social disadvantage and exclusion, educational difficulties, ill health and deprivation

• To obtain children’s views and opinions on their lives

### Growing Up in Ireland Sample Frame Design and Sample Selection

A sampling frame is an important part of the design of any study that collects data. The most applicable and ideal sampling frame is one that is up to date and contains a fully comprehensive listing of all elements of the relevant population. Each element of the population should appear once with no omissions or duplications. The Child Benefit Register was identified as an appropriate sampling frame for the GUI Infant Cohort.

The infant’s cohort was defined to be those born between 1st December 2007 and the 30th June 2008 giving a total target population of 41,185. The sample was selected on a payee systematic basis, pre-stratifying by marital status, and nationality of payee as well as number of children in the claim. A simple systematic selection procedure based on a random start and constant sampling fraction was used. The response rate was 58.2%. Giving a sample size of 11,134.

A pilot survey was conducted to identify any problems that could arise while collecting the data. Sampling weights were used to remove bias from the data, thus ensuring the GUI sample was population representative.

# Literature review

International studies have conducted research into twins and multiple births health, speech, language outcomes, behaviour and implications for the parents of the children. Sutcliffe and Derom summarised the various studies and wrote a literature review article on the above topics (Sutcliffe & Derom, 2006). Findings included; twins were estimated to have a stillbirth and neonatal death rate four times higher than singletons with this figure worsening for triplets as they have a death rate six times higher than those of singletons (Wimalasundera , et al., 2003). An explanation for the higher death rates in multiple births than singletons is the preterm birth of multiples which in turn results in lower birth weights in the children (Beatrice, et al., 2002). In this study, Beatrice et al. analysed data from the live birth registrations of Canada, England, Wales, France and the United States from 1980. The study found that twin births had a major impact on the trends of perinatal health indicators, and that preterm new-borns account for about 70% of perinatal mortality.

Maternal mental health is another area of importance in health research. The birth of a child demands dramatic changes and adaptions for women. For many, this means taking leave from work and missing the opportunity to generate an income during that time, which can lead to some financial difficulties as well as sleepless nights and shortened sleeping hours during the week. A recent study estimates that woman, in general, experience major depression at a rate of 1.7 times the reported rate for men (Albert, 2015). This difference is most evident during the life period of caring for infants and young children (Epperson, 1999). There are consistent studies that show evidence that the emotional and mental health of mothers of non-singletons is poorer and are more susceptible to becoming anxious, depressed and clinically exhausted after childbirth than mothers of single children. A study by Hay (Hay, et al., 1990), found that when comparing mothers of 3 month old single infants to mothers of twins of comparable age, 29.7% of the mothers of twins reported depression (5 times higher rate than singleton mothers) and 42% had high anxiety (3 times higher rate than singleton mothers). This figure worsens for mothers of triplets as a study by Robin et al, (Robin, et al., 1991) found that within the first 4 months postpartum, 40% of the mothers of triplets were depressed. A study from (Thorpe, et al., 1991) found that independent of other explanatory factors, mothers of 5-year-old twins were 3 times more likely to be depressed than those mothers of a singleton. From the Scottish Intercollegiate Guidelines research (Scottish Intercollegiate Guidlines Network, 2002), they identified that a poor relationship with a partner is now a major predictor of depression after childbirth in women.

The consequences of multiple births on health is not just limited to the babies. The maternal mortality rate during pregnancy, delivery or within 42 days of delivery, is 3 times higher for multiple births than singleton births, and admission to an intensive care unit is twice as high (Senat & Ancel, 1998). Financial wellbeing plays a major role in mental health, for people in general (Meltzer, et al., 2012). The financial cost of raising multiples may be higher than the cost of raising the same number of singletons, and the expenses are generally immediate or spread out over a shorter time in comparison to singletons (Campbell, et al., 2004). The time taken by mothers to return to work is longer for multiple births compared to singleton births, which can also cause a strain on the mental and financial wellbeing of the mother and family (Campbell, et al., 2004).

Social involvement and cognitive development have also been studied and the results show that multiple births are more likely to be impaired than those of singletons (Haworth, et al., 2013). Some of the developments that were researched consisted of; language and cognitive impairment. A study by Day (Day, 1932), found that twins were less advanced in language development at the ages of 2,3,4 and 5 years of age in comparison to singletons of the same age. Twins also scored on average 10 points lower on an IQ test than a singleton child. A study by Wood (Wood, et al., 1996), found that twins were more likely to develop attention deficit hyperactivity disorder than singletons.

A study by Stephen McKay (McKay, 2010) of English babies identified the financial, marital and mental stress associated with having multiple births vs singletons. Some of the key findings was that twins and multiple births were more common amongst older mothers and families with multiple births were more likely to show and report signs of tiredness and lower levels of confidence in looking after their children. There was also evidence to suggest that families with a multiple birth are more likely to separate or divorce than singleton birth families.

A study by Webbink (Webbink, et al., 2008) which used a cohort study undertaken in the Netherlands that investigated and compared the longitudinal IQ scores for approximately 188,000 singletons and 6000 twins who attended primary school in the Netherlands from 1994 to 2003, found that after controlling for variables of interest such as gender, ethnicity and parent’s education, language scores for twins aged 6 years old scored 16% of a standard deviation lower than non-singletons in the language test, and also scoring a 17% standard deviation lower in arithmetic scores than non-singletons. They also found that language and arithmetic scores for twins aged 8 years old scored 5% of a standard deviation lower than non-singletons in languages, while also scoring a 2% standard deviation lower in arithmetic than non-singletons. However, at ages 10 and 12 the differences were no longer statistically significant. For IQ scores at age 8, the evidence would suggest that twins scored 0.09 points lower than their non-singleton counterparts and this gap increased at age 10, where the IQ scores were 0.83 points lower at age 10.

A study into social development of twins was undertaken by Lisabeth DiLalla, (DiLalla, 2006) , the main aim of this study was to investigate if twins are more or less social than singletons on two basis; that twins would be more social as they grow up with another same-age peer in the home, and conversely that twins are less social than singletons because they have become dependent and more comfortable interacting with a same-age peer who is both genetically and environmentally similar to them, and thus less comfortable interacting with other children that differ from their environment. In this study DiLalla found that twins aged 5 years old were less social when they were placed in a peer play situation with an unfamiliar, same age, same sex peer, but not more aggressive than their singleton counterparts. A further study was conducted when the children were 10 to 15 years old but found no statistically significant difference between social levels of twins and singletons. From this study there is some evidence to suggest that twins in their childhoods appear to show less social behaviour with unfamiliar peers.

While there is extensive international literature on the experiences for non-singleton births, there is a dearth of evidence in the literature on the experiences of Irish mothers of non-singletons.

# Data Measures and Data Management

This section provides definitions of variables used in this project. The selection of the variables was informed by the literature review. A description of each variable of interest used in the study is seen below, along with any recoding that was performed on the original variable.

Non singleton Indicator Variable

There is a total of 398 non-singleton children included in the data, which accounts for 3.6% of the sample. This variable will be the key predictor throughout the project. The variable was coded Singleton = 0 and Non-singleton = 1.

Ethnicity of Primary Carer

Ethnicity of primary carer was coded as Irish = 1, Any other white background = 3, African or any other Black background = 4, Chinese or any other Asian background = 6, Other background = 8.

Irish accounted for 60% of the data, the next largest ethnicity was any other white background around 11%. The variable was recoded to the categories Irish = 1 and Non-Irish = 2, to avoid smaller less frequent categories distorting the statistical models.

Depression Score Primary Carer

Depression scores for the primary carer were measured using the eight-item Centre of Epidemiological Studies Depression Scale (CESD-8). The CESD-8 is a self-report screening survey used to gauge depression and distress levels. Participants answer a four-point rating scale on the following questions; how often they felt depressed, felt that everything was an effort, slept restlessly, were happy, felt lonely, enjoyed life, felt sad, and could not get going.

1. Rarely or none of the time (less than 1 day)
2. Some or a little of the time (1-2 days)
3. Occasionally or a moderate amount of time (3-4 days)
4. Most or all of the time (5-7 days)

A sum of the responses is calculated with reference to the previous 7 days, with the scores being within the range of 0-24 or scores can be dichotomised with a score greater than or equal to 7 indicating a clinically significant level of psychological distress. A total of 194 cases were missing in the data. The continuous variable and dichotomised categorical variable (Not Depressed (CESD < 7) = 0, Depressed (CESD >= 7) = 1) were used in the analyses.

Marital Status

This categorical variable was coded as married and living with husband/wife = 1, married and separated with husband/wife = 2, Divorced/Widowed = 3, Never married = 5, Refusal = 8, Don’t know = 9.

The variable was recoded to the categories Married = 1, and Otherwise = 2. With variables married and living with husband/wife and married and separated with husband/wife being recoded to 1, and the rest otherwise.

Employment

This categorical variable was coded as employee = 1, self-employed = 2, farmer = 3, Student full time = 4, On State training scheme (FAS) = 5, Unemployed, actively looking for job = 6, Long term sickness or disability = 7 , Home duties = 8, Retired = 9, Refusal = 98, Don’t know 99.

This variable was recoded as to Employee = 1, House Duties/ Retired = 2, Self Employed = 3, Student = 4, Unemployed = 5

Gender of the Primary Carer

This categorical variable was coded as male = 1, female = 2.

Equivalent Annual Income

This variable is a numeric scale of the equivalent annual income, which is the total value of income earned during a fiscal year for the household.

Education level of the Primary Carer

This categorical variable was coded as No formal education = 1, primary education =2 , Lower secondary = 3, Upper secondary = 4, Technical or vocational qualification = 5, Both Upper secondary and Technical or vocational qualification = 6, Non Degree = 7 , Primary Degree = 8, Professional qualification ( of degree status at least ) = 9, Both a degree and a , Professional qualification = 10, Postgraduate Certification or Diploma = 11, Postgraduate Degree (Masters) = 12, Doctorate = 13 Refusal = 98, Don’t know 99.

For the purposes of the analyses this variable was recoded as Primary = 1, Secondary = 2, Third level = 3.

Age of Primary Carer

This is a numeric variable of the age of the primary carer, with the ages of 40 or over being coded as = 40.

After how many weeks of pregnancy was baby born

This numeric variable had a response of 11094, with births lasting 25 weeks or under being coded as = 25.

Primary Carer Smoking Status

This categorical was coded as Daily smokes=1, Occasionally smokes= 2, Never smokes= 3.

Gender of child

This categorical variable was coded with male = 1, female = 2.

**Strength & Difficulties Questionnaire**

The Strength & Difficulties Questionnaire (SDQ) is a 25-item behavioral screening questionnaire designed to gauge emotional health and problem behaviors in children. The SDQ consists of 5 subscales; emotional, conduct, hyperactivity, peer problems and prosocial, (Goodman, 1997). Primary carers were asked to give scores of their child on this questionnaire.

**Parental Stress Score**

Two subscales of the Parental Stress Scale, (Jones & Berry, 1995), were used to gauge both positive and negative aspects of parenthood. The parental stress score variable in this study used the six-item Parental Stressors sub-scale.

**Picture Similarities and Naming Vocabulary**

Children undertook two standardised cognitive tests which were conducted by the interviewer in the home, with the assistance of the Computer Assisted Personal Interviewing (CAPI). The two tests were Picture similarities and Naming Vocabulary scales from the British Abilities Scales, (McCulloch, et al., 1997) , which measured reasoning/problem solving and vocabulary skills respectively.

Child has Siblings

This categorical variable was coded as siblings = 1, no siblings = 2.

# Methodology

In this chapter there will be an explanation of the statistical methods used throughout the project, along with their formula and assumptions if applicable.



## Descriptive statistics

Descriptive statistics were used to summarise all variables of interest. Normally distributed data was summarised using the mean and standard deviation, while skewed data was summarised using the median and interquartile range. Categorical data was summarised using numbers and percentages.

## Between Group tests:

T-tests were used to compare the means between two unrelated groups on the same continuous, dependent variable. The Chi Square test was used to test for differences across categorical variables.

## Multiple Linear Regression

Multiple Linear Regression was founded by Adrien-Marie Legendre and Johann Carl Friedrich Gauss in the early 19th century. Legendre and Gauss applied multiple linear regression to make predictions of planetary movements. Multiple Linear Regression is a statistical technique that uses two or more explanatory variables (independent variables) to predict the outcome of a numeric response variable (dependent variable). The explanatory variables can be either continuous or categorical, but all categorical variables must be dummy coded to be included in the model. The fitted population model has the form:

* represents the independent variables, where i is from 1 to n
* is a constant, generated by the linear equation
* is an unknown that is to be estimated, where i is from 1 to n
* is the predicted outcome of the numeric response variable

Assumptions:

* There is a linear relationship between the outcome variable and the independent variables.
* The residuals are normally distributed with constant variance.
* The independent variables are not highly correlated with each other. Variance inflation factor must be less than 5 (VIF < 5).
* Observations are independent.

## Binary Logistic Regression

Pierre Francois Verhulst developed the logistic function as a means of modelling population growth back in the 1840’s. A paper by Cramer, (Cramer, 2002), gives a detailed history of the logistic regression. Binary logistic regression is used to explain the relationship between one dependent binary variable and independent variables, which can be categorical (included using dummy coded variables) or continuous. Odds ratios are estimated using the exponential of the unstandardised regression coefficients.

* represents the independent variables, where i is from 1 to n
* is a constant, generated by the linear equation
* is an unknown that is to be estimated, where i is from 1 to n

Assumptions:

* Binary logistic regression requires dependent variable to be binary
* Observations need to be independent of each other
* Little to no multicollinearity among the independent variables
* Linearity of independent variables and log odds, although this does not require the dependent and independent variables to be related linearly
* Generally large sample sizes are needed. General guideline is that a minimum of 10 cases in each outcome is needed

The odds ratio is calculated by taking the exponential of the from the binary logistic regression. The odds ratio is interpreted as the odds of times the odds of the comparison.

### Interpreting Odds Ratio

Odds are a way of representing probability. A paper by Szumilas, (Szumilas, 2010) gives an explanation of the Odds Ratio.

An odds ratio is a measure of association between a binary outcome variable, say A, and a second binary outcome variable, say B, in a population.

For example, if variable A has 2 outcomes; drug worked, or drug didn’t work. Let variable B be 2 groups; one a control group and the second a treatment group for some new drug. If the number in each cases equal to a,b,c,d, this can be illustrated as in the table below, Table 2.

|  |  |
| --- | --- |
| Drug worked | Drug didn’t work |
| Treatment | a | b |
| Control | c | d |

Table : Odds Ratio Table

If Odds Ratio = 1, Variable A does not affect the odds of Variable B

If Odds Ratio > 1, Variable A has higher odds of Variable B odds

If Odds Ratio < 1, Variable A has lower odds of Variable B odds

### Relative Risk

Relative risk is the probability of an event occurring in one group compared to the probability of an event occurring in the other group. It requires the examination of two binary variables, where one measures the event (drug worked vs drug didn’t work) and the other variable measures the group (treatment vs control). An example of this can be seen in Figure 1. Using Table 2, above, we can get a formula for relative risk.

If the relative risk = 1, no difference

If the relative risk > 1, have a higher risk of failing

If the relative risk < 1, have a lower risk of failing

**Odds ratios as an estimate of relative risk**

In the case of rare events odds ratios provide accurate approximations to relative risks. For rare event , since is very small compared to .

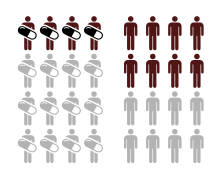


Figure :Source Wikipedia



## Poisson Regression

The Poisson distribution was founded by French mathematician, Simone Denis Poisson in 1838 as an approximation to the Binomial Distribution. A paper by Hu, (Hu, 2008), investigated and discussed the Poisson distribution and its application. The Poisson distribution is sometimes referred to as the law of small numbers as Poisson events tend to occur rarely even though there are many opportunities for these events to occur. For data to be Poisson distributed it’s mean must be equal to its variance. A study by Johnson and Kotz, (Johnson & Kotz, 1969), was one of the first uses and applications of Poisson regression. Poisson regression is a generalised linear model form of regression analysis. It is used to model count data and rare events, as rare events tend to follow a Poisson distribution. Poisson regression uses explanatory variables (independent variables) to predict the outcome of a count response variable (dependent variable). The model estimates the logarithm relationship between the explanatory variables and the response variable. To interpret the beta coefficients from the models you must take the exponential of the coefficients, which in turn can then be interpreted as follows: for a one unit change in the predictor variable it is estimated to change the mean of the dependent variable by the exponential of the respective regression coefficient.

* represents the independent variables, where i is from 1 to n
* is a constant, generated by the linear equation
* is an unknown that is to be estimated, where i is from 1 to n
* is the mean of Y, the count response variable

Assumptions:

* Outcome variable is count data
* Count data must be positive integers
* Count data must follow a Poisson distribution
* Explanatory variables must be categorical, dichotomous or continuous
* Observations must be independent

## Negative Binomial Regression

Negative binomial regression is a generalisation of Poisson regression. A study by Lawless, (Lawless, 1987), was conducted with the purpose of studying negative binomial regression properties and to fully understand the methodology. The Negative Binomial distribution was first studied by a Pierre Raymond de Montmort, a French mathematician in 1713. Montmort was investigating the distribution of the number of trials required in an experiment to obtain a given number of successes and founded the Negative Binomial distribution. Negative Binomial regression is used to model over-dispersed count data, which is when the conditional variance exceeds the conditional mean. Negative binomial regression has the same structure as Poisson regression, but it has an extra parameter to model the over-dispersion. To interpret the beta coefficients from the models you must take the exponential of the coefficients, which in turn can then be interpreted as follows: for a one unit change in the predictor variable the mean of the response variable is estimated to change by the exponential of the respective regression coefficient.

* represents the independent variables, where i is from 1 to n
* is a constant, generated by the linear equation
* is an unknown that is to be estimated, where i is from 1 to n
* is the mean of Y, the response variable

Assumptions:

* Outcome variable is count data
* Count data must be positive integers
* Count data must follow a negative binomial distribution
* Explanatory variables must be ordinal, dichotomous or continuous
* Observations must be independent

## Cluster Analysis

Cluster analysis was developed by Driver and Kroeber in 1932 when studying the field of anthropology, (Driver & Kroeber, 1932). Cluster analysis is a statistical methodology that classifies members of the sample into groups called clusters. The aim of cluster analysis is to explore multivariate data and try to detect whether the data can be divided into groups or clusters which are relatively distinct, as seen below in Figure 2. Cluster analysis involves formulating a problem, selecting a distance measure, selecting a clustering procedure, deciding the number of clusters, interpreting the profile clusters and finally, assessing the validity of clustering. Clustering procedures in cluster analysis may be two step, hierarchical or non-hierarchical, with the most popular non-hierarchical method being k-means clustering.

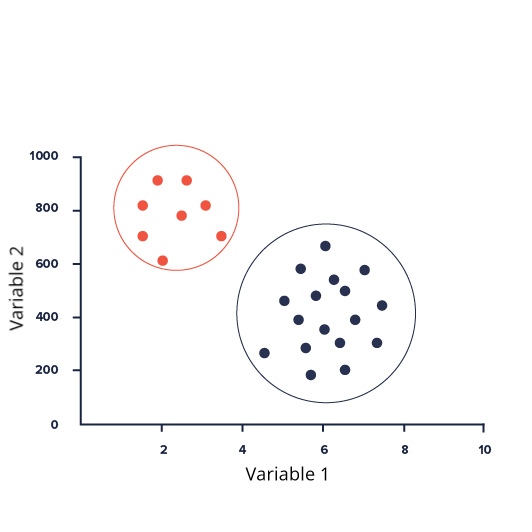


Figure : AN Example of Clusters in Data

### Methods of Distance Measures for Cluster Analysis

Figure : Source SlidePlayer Cluster Analysis by Dr. Chayada Bhadrakom

#### **Between groups Linkage**

In between groups linkage, the two clusters a and b are merged such that the average pairwise distance within the newly formed cluster is minimum. Consider a new cluster C, formed by merging a and b together. Then the distance between cluster a and b, D(a,b) is calculated as follows:

* Objectand are in cluster *C,* which is formed by merging clusters a and b
* is the distance between cluster a and cluster b

|  |  |
| --- | --- |
| **Pros** | **Cons** |
| * Between groups method is still efficient if there is noise between clusters | * Biased towards globular clusters |

#### **Nearest Neighbour**

The distance between groups is defined as the distance between the closest pair of the objects in each cluster, where only pairs consisting of one cluster from each group are considered

* Object is in cluster *a* and object is in cluster b
* is the distance between cluster a and cluster b

The distance between the two clusters is given by the value of the shortest link between the clusters

|  |  |
| --- | --- |
| **Pros** | **Cons** |
| * This method can separate non-elliptical shapes if the distances between two clusters is not small | * This method cannot separate accurately if there is noise between the clusters |

#### **Furthest Neighbour**

This clustering method is the converse of nearest neighbour. The distance between clusters is now defined as the distance between the most distant pair of objects, one from each group.

* Object is in cluster *a* and object is in cluster b
* is the distance between cluster a and cluster b

|  |  |
| --- | --- |
| **Pros** | **Cons** |
| * This method can efficiently separate clusters if there is noise between clusters | * Tends to break large clusters |

#### **Centroid Clustering**

The centroid method uses the centroid of each clusters to determine the average distance between clusters of cases.

* Where u is the centroid of cluster a and v is the centroid of cluster b
* is the distance between cluster a and cluster b

|  |  |
| --- | --- |
| **Pros** | **Cons** |
| * Method can handle very large data sets | * Outliers can’t be studied |

#### **Wards Method**

This Clustering method is the same as Between Groups Linkage except that Ward’s method calculates the sum of the squared distances .

* Objectand are in cluster *C,* which is formed by merging clusters a and b
* is the distance between cluster a and cluster b

Wards Method clustering procedure seeks to form partitions in a manner that minimises the loss associated with each grouping, and to quantify that loss in a form that is readily interpretable. At each step in the analysis, the union of every possible cluster pair is considered and the two clusters whose merging results in a minimum increase in information loss are combined. Information loss is defined by Ward in terms of an error sum-of-squares criterion (ESS).

* is the score of the th case

|  |  |
| --- | --- |
| **Pros** | **Cons** |
| * method does well in separating clusters if there is noise between clusters | * Biased towards globular clusters |

### Two Step Cluster Analysis

Two-step cluster analysis begins by identifying groups by running pre-clustering first and then by running hierarchical methods on the pre-clustering in the second step. It can be used on continuous or categorical data. Two step cluster analysis can handle large data sets that would take a long time to compute with hierarchical cluster methods because of its use of a quick cluster algorithm in the first step. This quick cluster algorithm in the first step is achieved by using a Cluster Feature Tree, as seen below, Figure 4. The first case goes at the root of the tree as a leaf node. The algorithm then proceeds to go through the rest of the data on a case by case basis and assigns the case to an existing node or creates a new node depending on how similar the case is to the already existing nodes. A Distance Measure, usually the log-likelihood or Euclidean measure, is used to determine how similar the case is to the existing nodes.

The second step is to group the tree nodes together using a hierarchical clustering algorithm.



Figure : Example of a Cluster Tree

### Hierarchical Cluster Analysis

Hierarchical clustering begins by treating each observation as a separate cluster. It then begins to repeatedly follow two steps;

1. Identify the two clusters that are closest together
2. Merge the two most similar clusters

This process is continued until all the clusters are merged together. There are multiple measures of distance(similarity) that can be used, such as Euclidean distance or Ward’s method. The criterion for choosing the pair of clusters to merge at each step is based on the optimal value of an objective function.

***Wards Method:***  the distance between two clusters A and B, is based on how much the sum of squares will increase when they are merged.

* is the centre of the cluster *j,*
* is the number of points in cluster *j*
* is called the merging cost of combining the clusters A and B
* is the data point

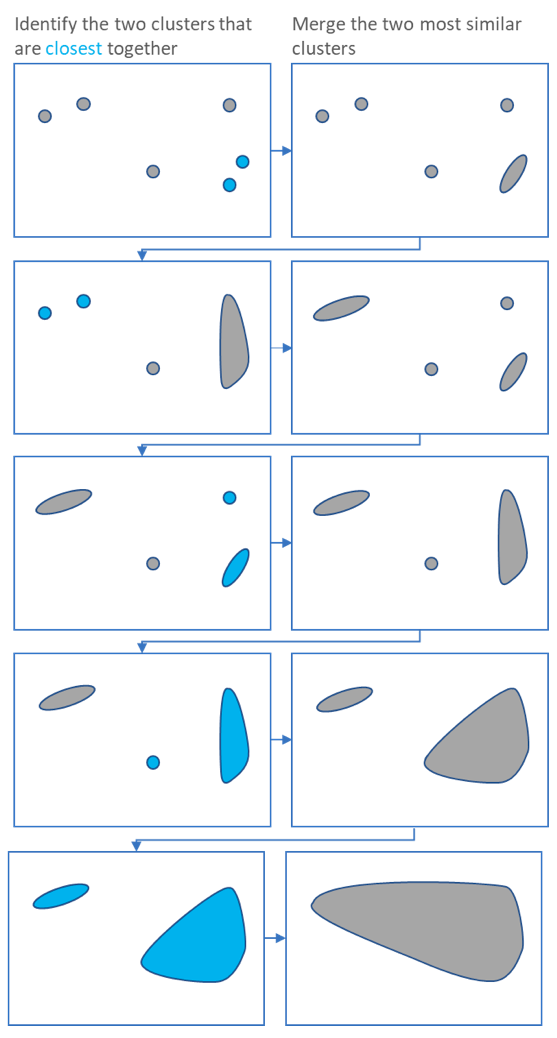


Figure : Source Towards Data Science

There are two main types of Hierarchical Clustering algorithms:

1. *Agglomerative* – starts by considering every case as its own cluster. At each step of the algorithm the two clusters that are the most similar are combined into a new bigger cluster. This is repeated until all cases are a member of one single big cluster (root node). Which is shown in Figure 5 and Figure 6.
2. Divisive – this is the inverse order of agglomerative, all cases are included in a single cluster. At each step of the algorithm, the most diverse cluster is divided into two. This is repeated until all cases are in their own individual cluster. This would be the reverse of Figure 5 starting from the bottom right square and working its way up to the top left.

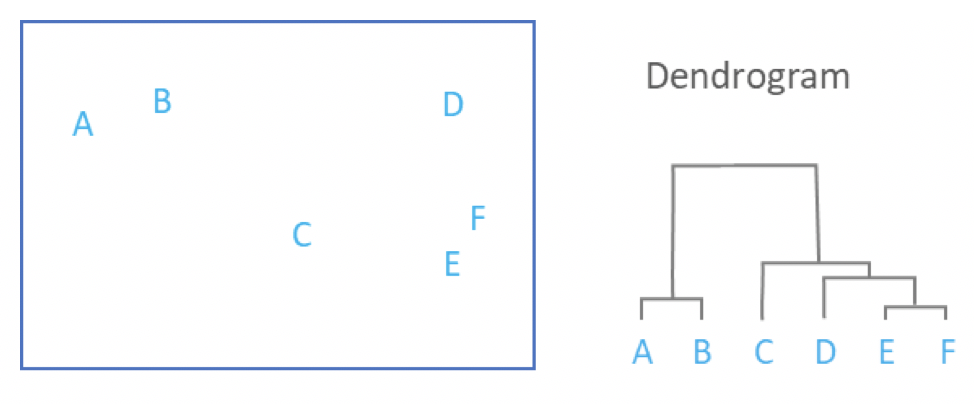


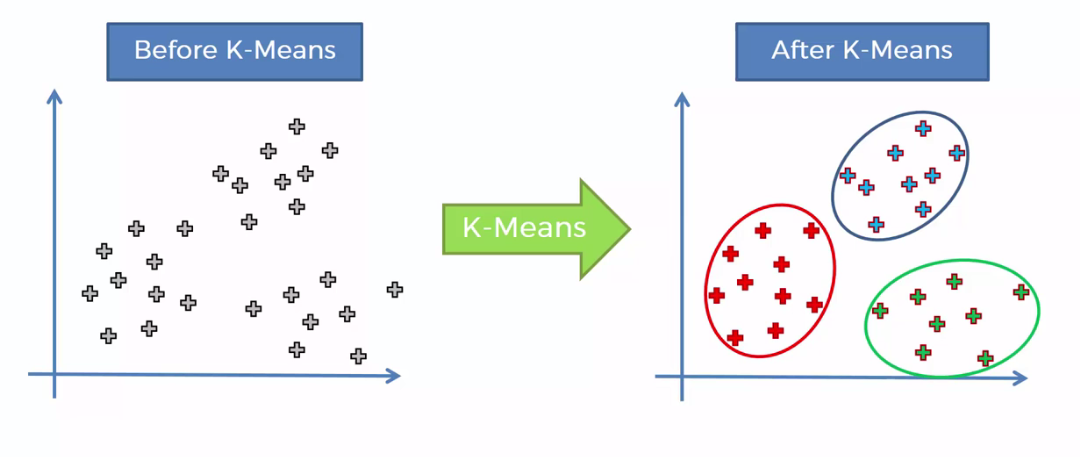
Figure : Example of a Dendrogram formed by Hierarchical Clustering

### K-Means Clustering

K-means clustering aims to assign cases to a set number of clusters defined by the user in such a way that maximises the separation of those clusters while minimizing intra-cluster distances relative to the cluster’s mean or centroid. Euclidean distances are usually the measure of distance for the clusters. Its algorithm is presented below.

1. An initialisation step creates *k* centroids. Some algorithms use an existing case as a starting centroid while others randomly assign cases into *k* clusters in order to calculate the first centroids. It is often possible to define your own custom initial placements for centroids.
2. An assignment step places each case into a cluster whose centroid (mean) is closest to it.
3. An update or re-assignment step then calculates new centroids based on the membership of each cluster.
4. Step 2 and 3 are repeated until the solution converges, i.e. when the centroid positions no longer change.

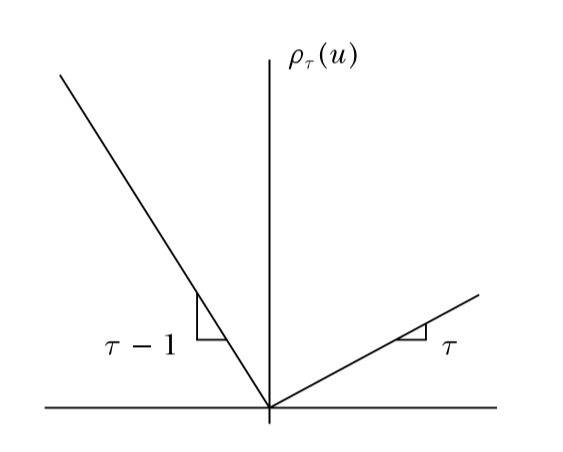
* J is the objective function
* k is the number of clusters
* n is the number of cases
* is the data point i
* is the centroid for cluster j



## Quantile Regression

Ruder Josip Boskovic, a mathematician from Dubrovnik, was the first to propose the idea of estimating a median regression slope back in 1760, which is noted in an article by Stigler who found and wrote about the manuscript note on fitting a linear regression by Boscovich (Stigler, 1984). Koenker and Hallock were the first to introduce and use quantile regression in their study, (Koenker & Hallock, 2001) on the analysis of birth weights of infant babies in the US. Quantile regression is a statistical method that uses explanatory variables (independent variables) to predict the outcome of a response variable (dependent variable). The methodology is an extension of linear regression, where quantile regression is generally used when the data is skewed or has extreme outliers as it estimates the conditional quantile required. This contrasts with multiple linear regression which estimates the conditional mean. Quantile regression estimates are more accurate and robust against outliers in the response measurements than linear regression.

Quantile regression uses the fact that we can define the median as the solution to the problem of minimising a sum of absolute residuals. The symmetry property of the piecewise linear absolute value function implies that the minimization of the sum of the absolute residuals must equal the number of positive and negative residuals, which means that there are an equal number of observations above and below the median. For any 0 < τ < 1, define the piecewise linear function as seen below;



To obtain estimates of the other conditional quantile functions, the following minimization problem is proposed:

The minimization problem, where is the data point and is formulated as a linear function of parameters, is solved by linear programming methods.

Assumptions:

* There is a linear relationship between the outcome variable and the independent variables.
* The independent variables are not highly correlated with each other. Variance inflation factor must be less than 5 (VIF < 5).

# Exploratory Data Analysis and Between Group Tests

In this section, a descriptive analysis of the variables of interest in the project are presented.

## Descriptive Statistics for Wave 1 Cohort Data

In , the descriptive statistics summarise the variables of interest separately for singletons and non-singletons for the Wave 1 cohort data. The accompanying figures, and box plots of the numeric variables of interest, illustrate the differences between singletons and non-singletons. T-tests were used to compare the means between two unrelated groups on the same continuous dependent variable. T-tests were performed on the following variables; age of primary carer, how many weeks until the baby was born and equivalised annual income. The Chi Square test was used to test for differences across categorical variables. Chi-square tests were performed on the following variables; employment status of primary carer, partner in the home, highest level education of primary carer, ethnicity, marital status, depressed or not and finally if mother smokes.

Some interesting findings from the below table, , Singleton primary carers have a higher percentage in employment compared to non-singleton primary carers (53.4% vs 39.9%), while more non-singleton primary carers are retired/home duties than their singleton counterparts (47.9% vs 36%). When looking at marital status of the primary carer, more non-singleton primary carers tended to be married than singleton primary carers (79.2% vs 69.7%), some of this difference might be explained by the fact that older mothers tend to be more likely to have non-singleton pregnancies. When looking at the dichotomised depression variable, non-singleton primary carers tended to be depressed more than singleton primary carers (17.8% vs 10.5%).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Characteristic | Singleton | | | Non-Singleton | | | p-value |
|  | n | % | Mean (SD) or median [IQR] | n | % | Mean (SD) or median [IQR] |  |
|  |  |  |  |  |  |  |  |
| Age of Primary Carer |  |  | 32 [8.0] |  |  | 34 [7.0] | <.001 |
| Gender of Primary Carer  *Female* | 10699 | 99.7 |  | 397 | 99.7 |  |  |
| *Male* | 37 | 0.3 |  | 1 | 0.3 |  |  |
| Employment status  *Employee* | 5737 | 53.4 |  | 159 | 39.9 |  | <.001 |
| *Home duties/retired* | 3869 | 36.0 |  | 191 | 47.9 |  |  |
| *Self Employed* | 486 | 4.5 |  | 23 | 5.8 |  |  |
| *Student* | 171 | 1.6 |  | 5 | 1.3 |  |  |
| *Unemployed* | 473 | 4.4 |  | 20 | 5.0 |  |  |
| Partner in the home  *Yes* | 9414 | 87.7 |  | 361 | 90.7 |  | .071 |
| *No* | 1322 | 12.3 |  | 37 | 9.3 |  |  |
| Highest level of education  *Third* | 3893 | 36.3 |  | 138 | 34.7 |  | .605 |
| *Secondary* | 6560 | 61.2 |  | 252 | 63.3 |  |  |
| *Primary* | 273 | 2.5 |  | 8 | 2.0 |  |  |
| Ethnicity  *Irish* | 6538 | 78.9 |  | 270 | 83.3 |  | .058 |
| *Non-Irish* | 1741 | 21.1 |  | 54 | 16.6 |  |  |
| Marital Status  *Married* | 7387 | 69.7 |  | 312 | 79.2 |  | <.001 |
| *Not Married* | 3217 | 30.3 |  | 82 | 20.8 |  |  |
| Depressed or Not depressed  *Not Depressed* | 9437 | 89.5 |  | 323 | 82.2 |  | <.001 |
| *Depressed* | 1110 | 10.5 |  | 70 | 17.8 |  |  |
| How many weeks until baby was born |  |  | 40 [2.0] |  |  | 37 [3.0] | <.001 |
| Gender of child  *Male* | 5480 | 51.0 |  | 199 | 50.0 |  |  |
| *Female* | 5256 | 49.0 |  | 199 | 50.0 |  |  |
| Mother Smokes  *Daily* | 1696 | 20.9 |  | 57 | 18.0 |  | .107 |
| *Occasionally* | 1275 | 15.7 |  | 62 | 19.6 |  |  |
| *Never* | 5127 | 63.3 |  | 197 | 62.3 |  |  |
| First Child    *First Child* | 4175 | 38.9 |  | 229 | 57.5 |  |  |
| *Has older sibling* | 6561 | 61.1 |  | 169 | 42.5 |  |  |
| Equivalised annual income |  |  | 18868.11  [15402.73] |  |  | 16981.13 [12721.39] | <.001 |

Table : Descriptive Statistic

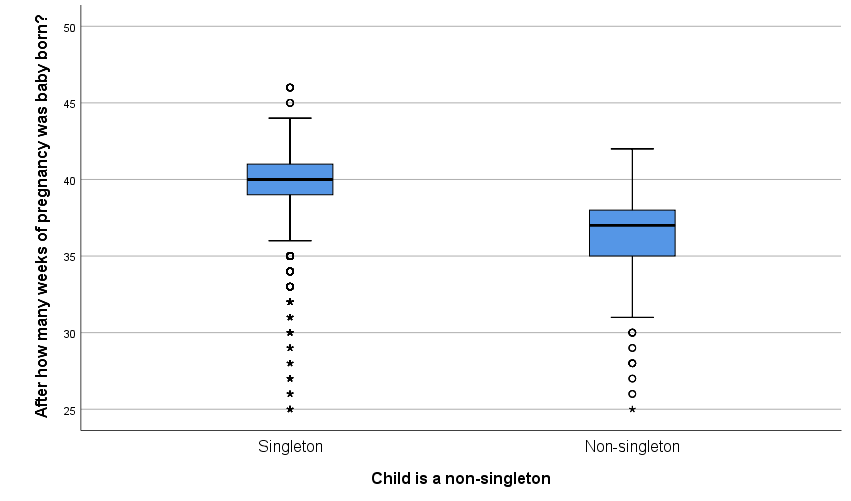


Figure : Box Plot of length of pregnancy controlling for Non-singleton

From , it can be seen that mothers who are expecting a singleton tend to have on average a longer number of weeks of pregnancy than a non-singleton pregnancy, with the median being estimated to be 40 weeks for singletons and 37 weeks for non-singletons. The histograms in and , show the length of pregnancy for singleton and non-singleton births to be slightly negatively skewed with the presence of pre-term births.

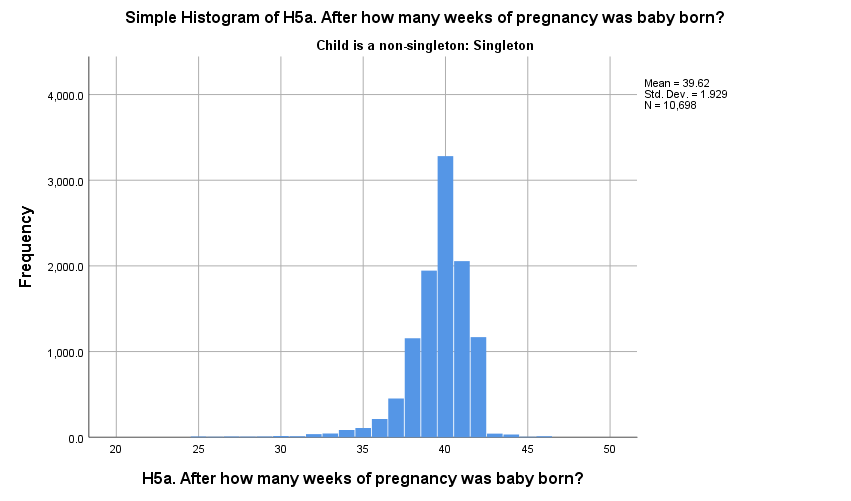
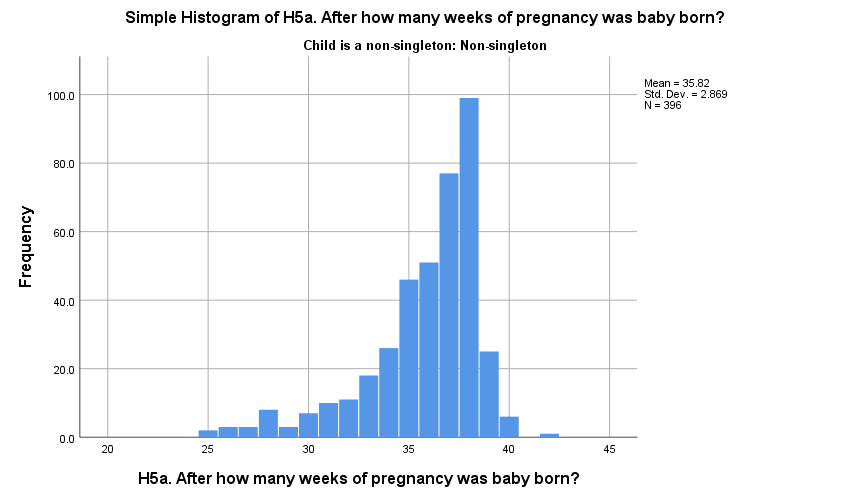


Figure : Length of Pregnancy For Non-singletoN

Figure : Length of Pregnancy For Singleton

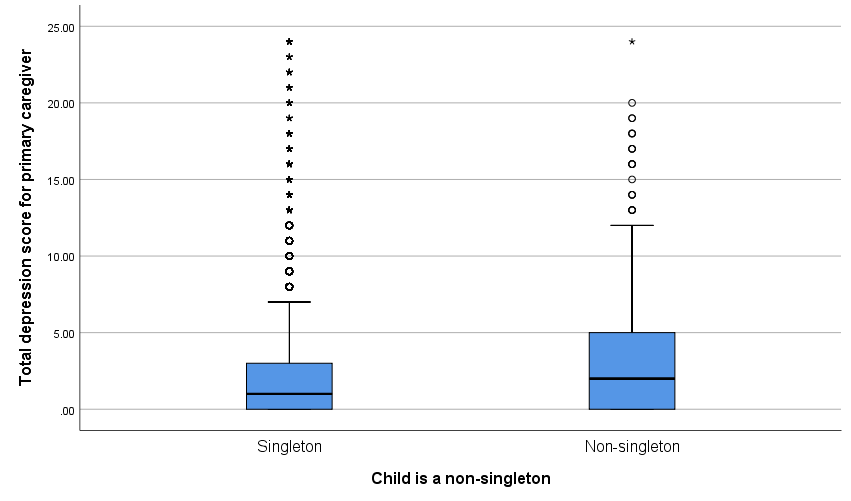


Figure : box plot of Depression score of primary carer controlling for non-singleton

From Figure 10, it can be seen that non-singleton primary carers tend to show higher scores on the CESD-8 scale than singleton carers, with the percentage of Primary Carers reporting depressive symptoms for non-singleton carers being 17.8% compared to the 10.5% of singleton carers. From the histograms in Figure 11 and Figure 12, we can see the data is skewed to the right with a high frequency at 0 for both singletons and non-singletons, indicating that the majority of primary carers show little to no depressive symptoms.

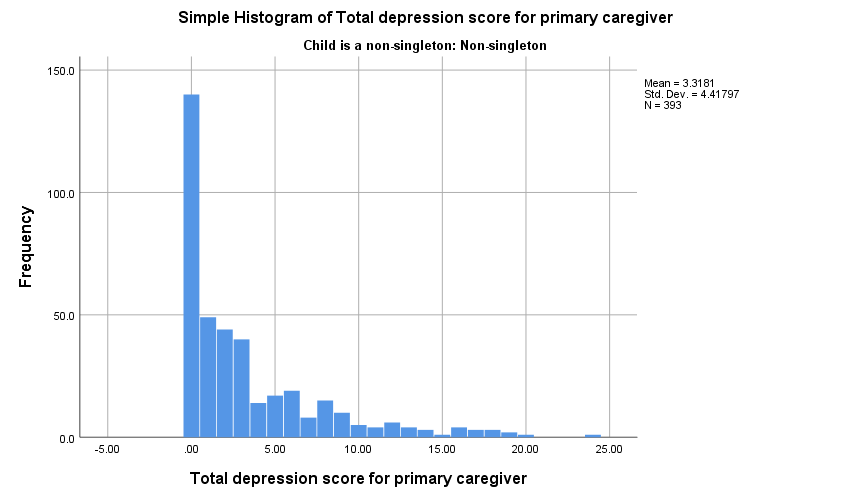
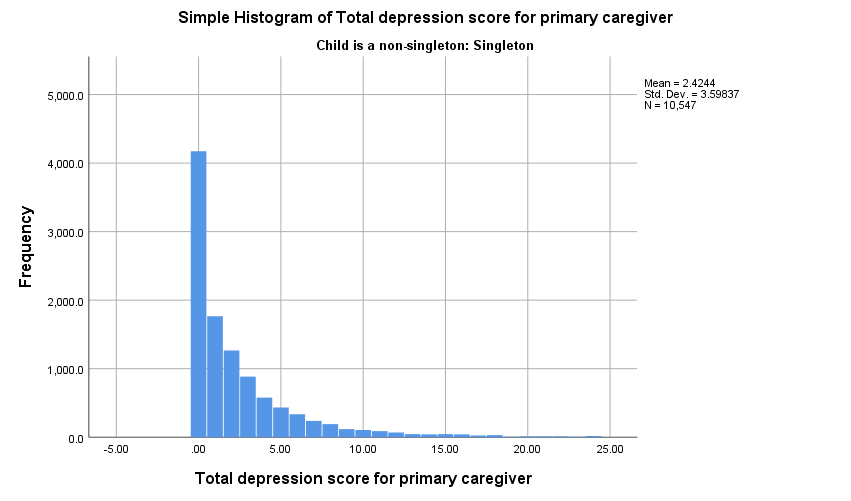


Figure : Depression Scores of Primary Carers For singleton

Figure : Depression Scores of Primary Carers For Non-singleton

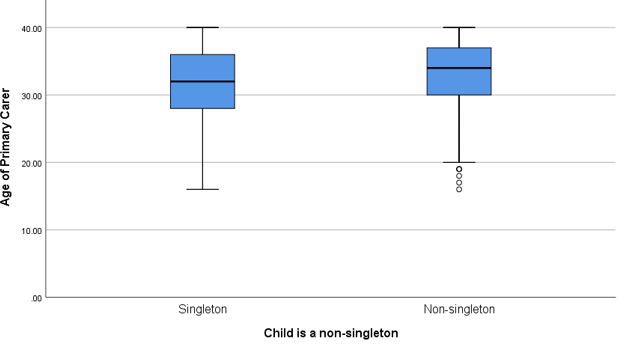


Figure : Box Plot of Age of Primary Carer Controlling for Non-singleton

From Figure 13,**Error! Reference source not found.Error! Reference source not found.Error! Reference source not found.** it can be seen that the age of non-singleton primary carers tends to be higher than those of singleton primary carers, with the median age for non-singletons being 34 years old while the median age for singletons being 32 years old. The tails for both boxplots would imply that the data is skewed to the left. The histograms below, Figure 14 and Figure 15 , show the data is in fact skewed to the left.

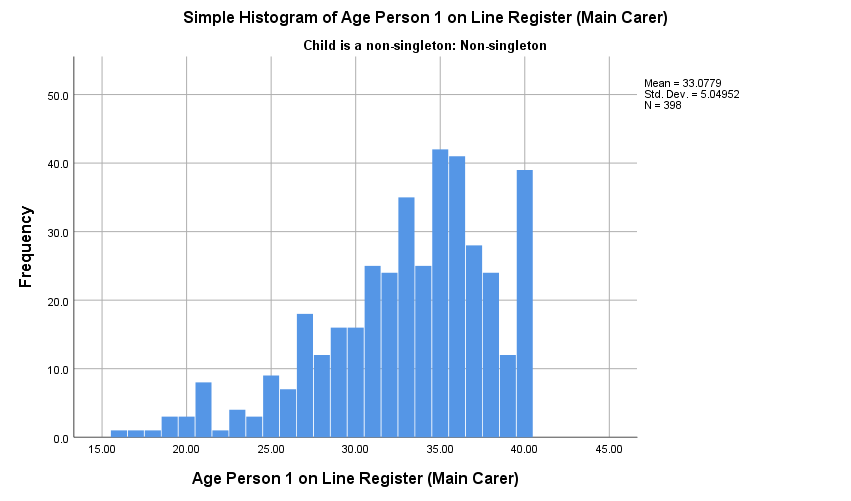
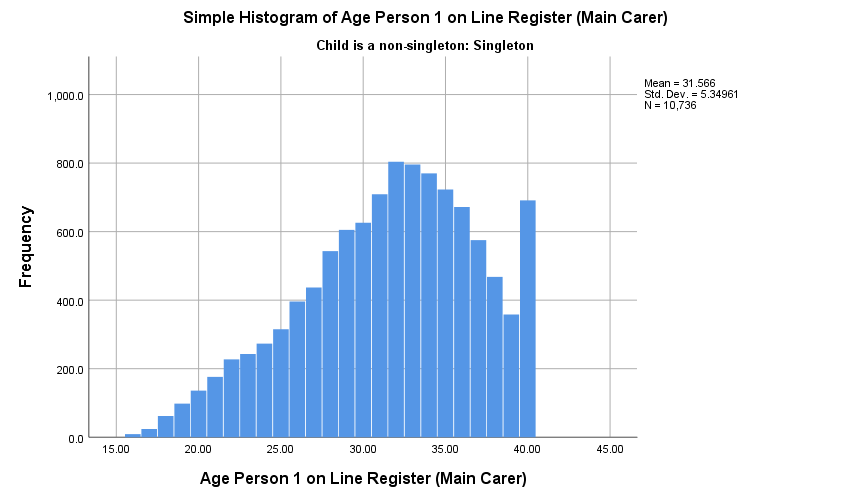


Figure : Age of Primary Carer for Non-singleton

Figure : Age of Primary Carer for Singleton

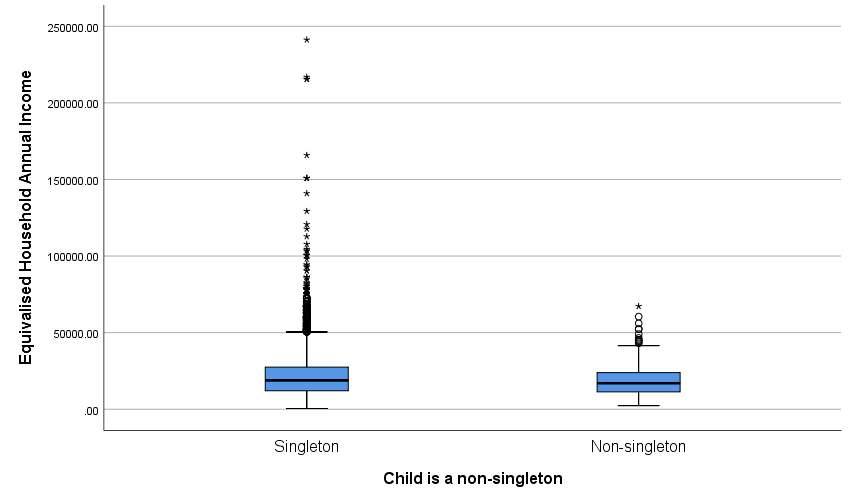


Figure : Box Plot of Annual Income Controlling for Non-singleton

From Figure 16, the equivalised household income seems to be relatively close for both singleton and non-singleton families. The median income for singleton households is €18,868.11 with an interquartile range of €15,402.73 and the median for non-singleton households being €16,981.13 with an interquartile range of €12,721.39. However, singletons have a far greater number of outliers than non-singletons. The histograms in Figure 17 and Figure 18, show the data for both singletons and non-singletons to be skewed to the right.

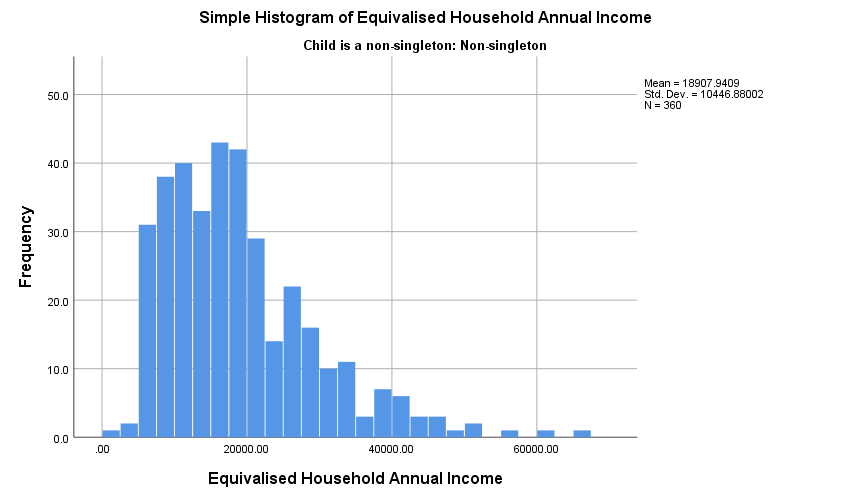
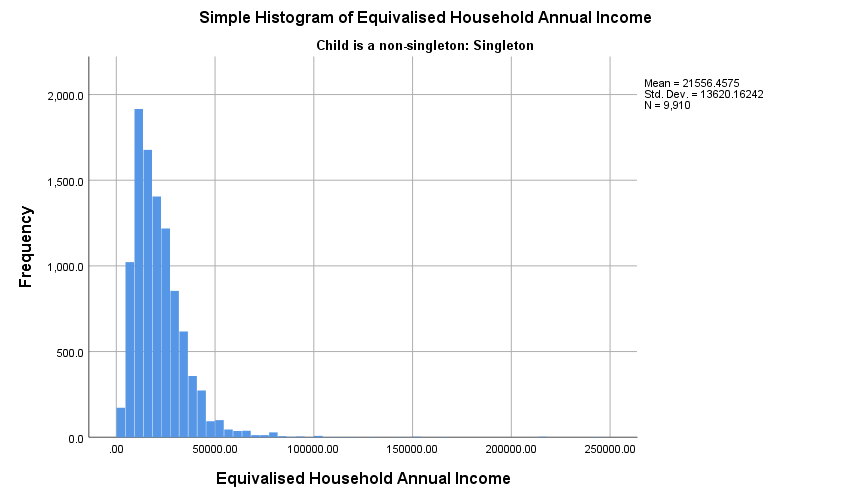


Figure : Income For Singleton Households

Figure : Income for Non-Singleton Households

## Descriptive Statistics for Wave 3 Cohort Data

In Table 4, the descriptive statistics summarise the variables of interest separately for singletons and non-singletons for the Wave 3 cohort data. The accompanying figures and box plots of the numeric variables of interest, visualise differences between singletons and non-singletons. T-tests were used to compare the means between two groups on the continuous outcomes; stress score of primary carer, SDQ peer problems, SDQ emotional, SDQ conduct, SDQ hyperactivity, SDQ prosocial, attitude scores from teacher reports, language scores from teacher reports, linking scores from teacher reports and reading scores from teacher reports. The Chi Square test was used to test for differences across categorical variables. Chi-square tests were performed on the following variables; ethnicity, highest level of education and marital status.

Some interesting findings from the descriptive statistics in the below table, Table 4, are that Singletons tend to perform better in the academic variables from the teacher reports than their non-singleton counterparts. Scores appear quite similar for singletons and non-singletons on SDQ scores, picture similarities and naming vocabulary scores. More graphs and boxplots of the descriptive statistics can be seen in the appendix.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Characteristic | Singleton | | | Non-Singleton | | | p-value |
|  | n | % | Mean (SD) or median [IQR] | n | % | Mean (SD) or median [IQR] |  |
|  |  |  |  |  |  |  |  |
| Stress Score of Primary Carer |  |  | 11 [5.0] |  |  | 11 [6.0] | .459 |
|  |  |  |  |  |  |  |  |
| SDQ Peer Problems Subscale |  |  | 1.0 [2.0] |  |  | 0.0 [2.0] | .987 |
|  |  |  |  |  |  |  |  |
| SDQ Emotional Subscale |  |  | 1.0 [2.0] |  |  | 1.0 [3.0] | .231 |
|  |  |  |  |  |  |  |  |
| SDQ Conduct Subscale |  |  | 1.0 [2.0] |  |  | 1.0 [2.0] | .760 |
|  |  |  |  |  |  |  |  |
| SDQ Hyperactivity Subscale |  |  | 3.0 [4.0] |  |  | 3.0 [4.0] | .664 |
|  |  |  |  |  |  |  |  |
| SDQ Prosocial Subscale |  |  | 9.0 [3.0] |  |  | 9.0 [3.0] | .006 |
|  |  |  |  |  |  |  |  |
| Gender of Primary Carer  *Female* | 10699 | 99.7 |  | 397 | 99.7 |  |  |
| *Male* | 37 | 0.3 |  | 1 | 0.3 |  |  |
|  |  |  |  |  |  |  |  |
| *Attitude Score Teacher Report* |  |  | 9.0 [2.0] |  |  | 8.0 [2.0] | .188 |
|  |  |  |  |  |  |  |  |
| *Language Score Teacher Report* |  |  | 9.0 [2.0] |  |  | 8.0 [3.0] | .027 |
|  |  |  |  |  |  |  |  |
| *Linking Score Teacher Report* |  |  | 9.0 [2.0] |  |  | 8.0 [3.0] | <.001 |
|  |  |  |  |  |  |  |  |
| *Reading Score Teacher Report* |  |  | 8.0 [2.0] |  |  | 8.0 [3.0] | .013 |
|  |  |  |  |  |  |  |  |
| Highest level of education  *Third* | 5370 | 62.0 |  | 189 | 58.5 |  | .605 |
| *Secondary* | 3194 | 36.8 |  | 133 | 41.2 |  |  |
| *Primary* | 102 | 1.2 |  | 8 | 0.31 |  |  |
|  |  |  |  |  |  |  |  |
| *Picture Similarities Scores* |  |  | 85.0 [13.0] |  |  | 85 [13.0] | .152 |
|  |  |  |  |  |  |  |  |
| *Naming Vocabulary Scores* |  |  | 112.0 [20.0] |  |  | 112.0 [16.0] | .337 |
|  |  |  |  |  |  |  |  |
| Ethnicity  *Irish* | 6538 | 78.9 |  | 270 | 83.3 |  | .058 |
| *Non-Irish* | 1741 | 21.1 |  | 54 | 16.6 |  |  |
| Marital Status  *Married* | 5648 | 88.1 |  | 211 | 91.7 |  | .150 |
| *Not Married* | 768 | 12.9 |  | 19 | 8.3 |  |  |

Table : Descriptive Statistics for Wave 3

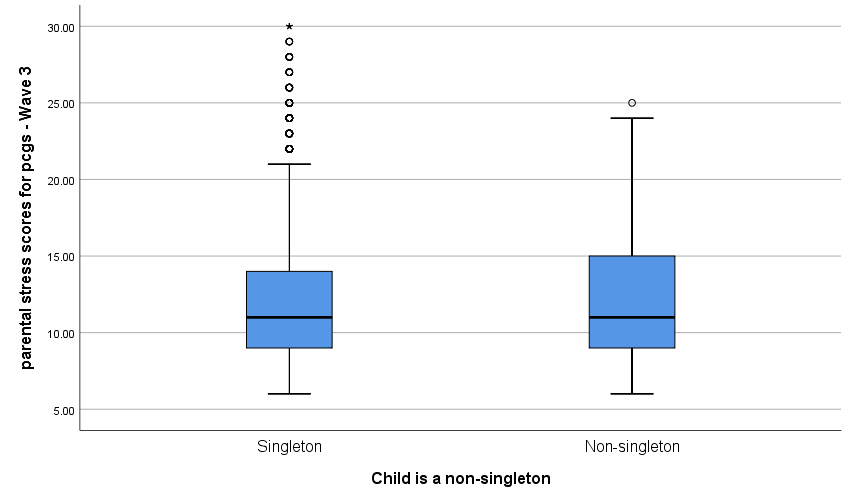


Figure : Box Plot of Parental Stress Scores Controlling for Non-singleton

From Figure 19Figure 15: Age of Primary Carer for Non-singleton

, the parental stress scores have the same median, but non-singletons have a slightly higher interquartile range than singletons primary carer’s parental stress scores, which can be seen by the longer blue box for non-singleton. The median parental stress score for singleton primary carers is 11 with an interquartile range of 5 while and the median for non-singleton primary carer’s parental stress scores is 11 with an interquartile range of 6. A histogram of the parental stress scores is seen below in Figure 20 and Figure 21.

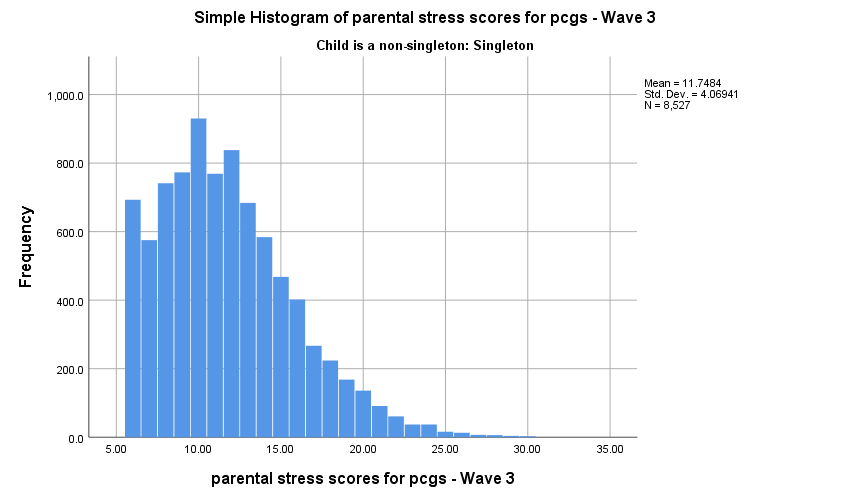
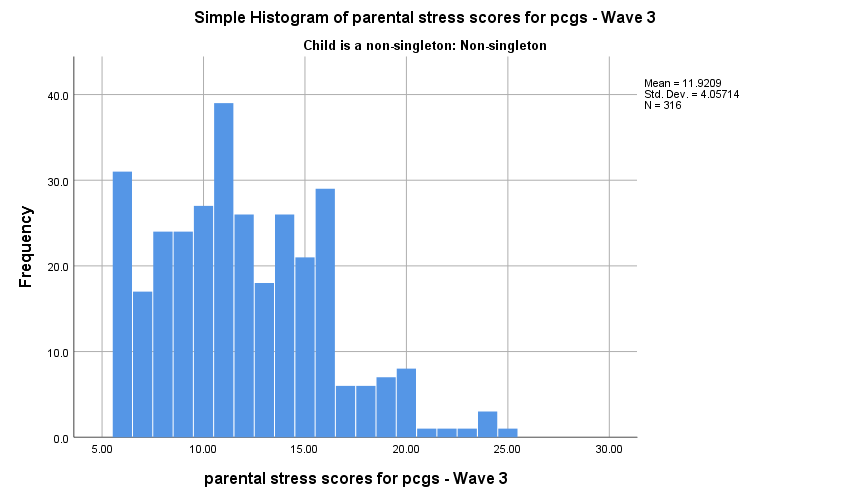


Figure : Parental stress Scores of Primary Carer for Singleton

Figure : Parental Stress Scores of Primary Carer for Non-singleton

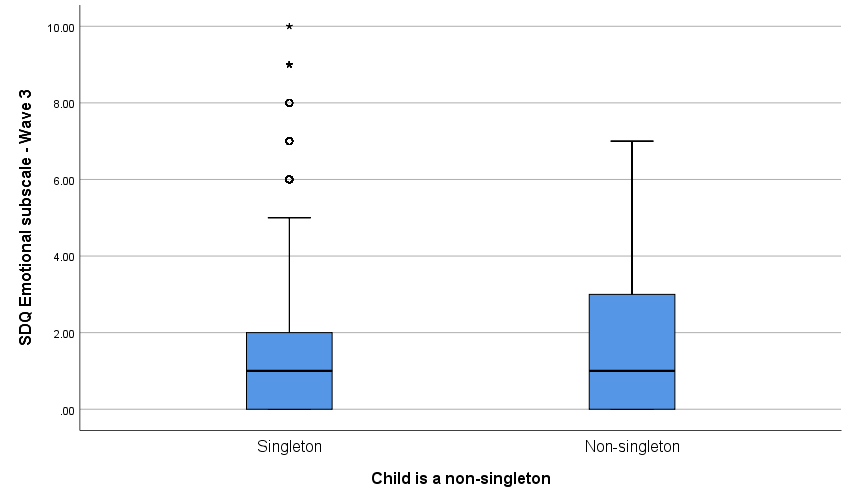


Figure : Box Plots of SDQ Emotional Scores Controlling for Non-singleton

From Figure 22Figure 15: Age of Primary Carer for Non-singleton

, the SDQ emotional subscale scores have the same median, but non-singletons have a slightly higher interquartile range than singletons, which can be seen by the longer blue box for non-singleton. The median SDQ emotional subscale scores for singletons is 1 with an interquartile range of 2 while the median for non-singleton SDQ emotional subscale scores is 1 with an interquartile range of 3. Histograms of the SDQ emotional scores for singletons and non-singletons can be seen in Figure 23 and Figure 24.

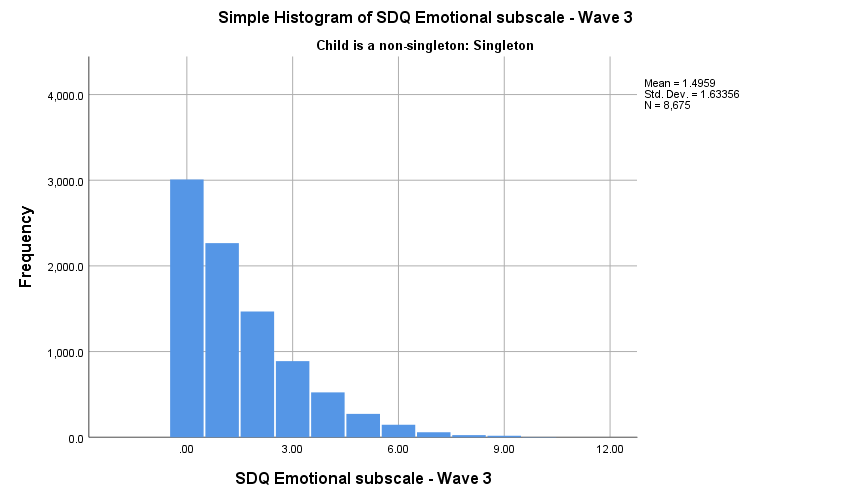
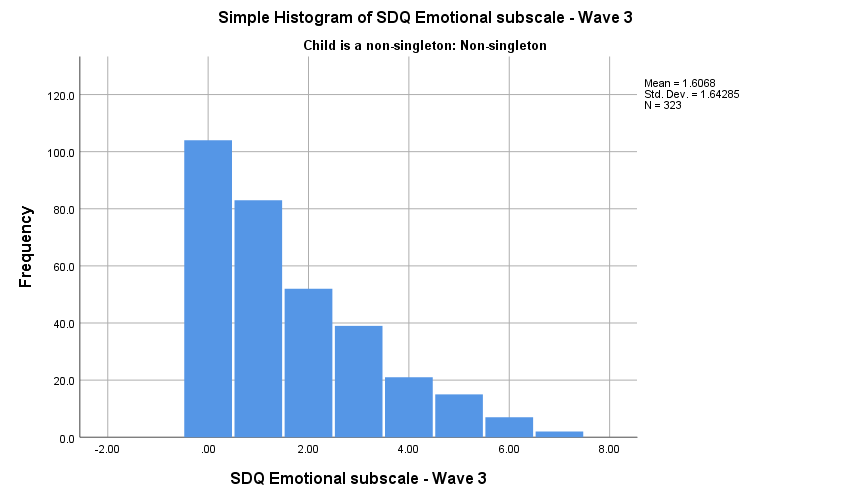


Figure : SDQ Emotional Score for Non-singleton

Figure : SDQ Emotional Score for Singleton

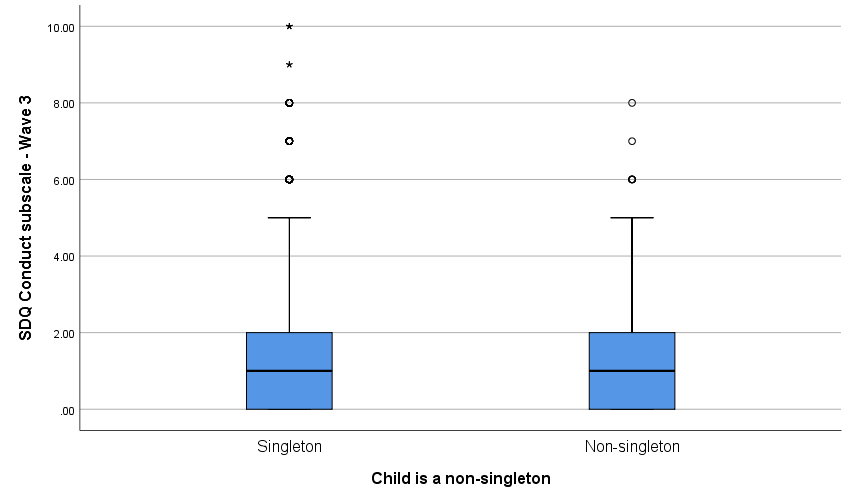


Figure : Box Plots of SDQ conduct Scores Controlling for Non-Singleton

From Figure 25Figure 15: Age of Primary Carer for Non-singleton

, the SDQ conduct subscale scores have the same median and interquartile range as singletons. However, singletons have a few more outliers in their data than non-singletons. The median SDQ conduct subscale scores for singletons and non-singletons is 1 with an interquartile range of 2. Histograms of the SDQ conduct scores for singletons and non-singletons can be seen in Figure 26 and Figure 27.

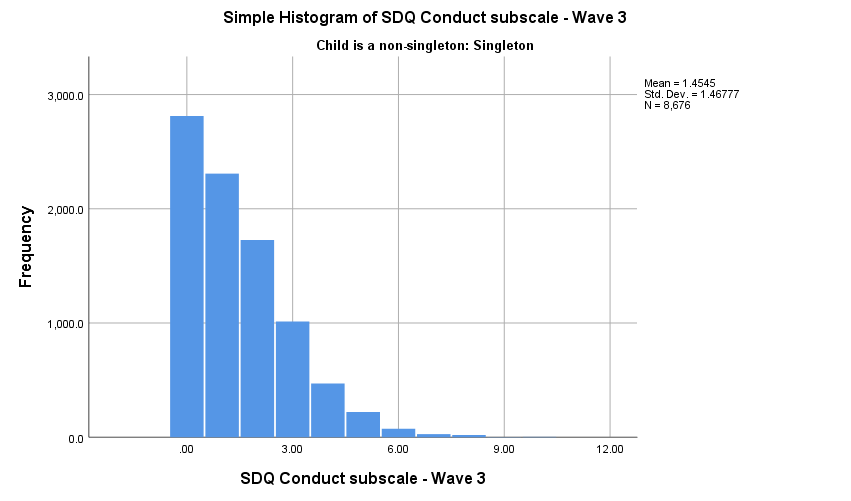
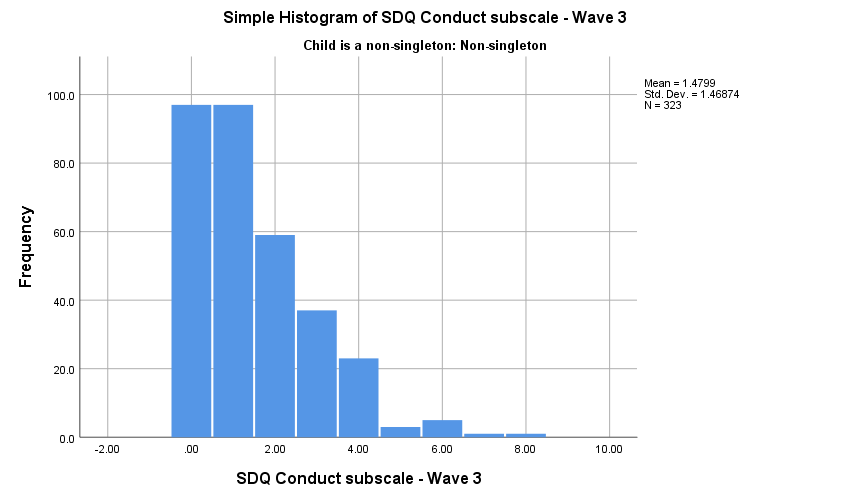


Figure : SDQ Conduct Scores for Singleton

Figure : SDQ Conduct Scores for Nonsingleton

# Analysis of Wave 1 data

## Length of Pregnancy with key predictors non-singleton and singleton pregnancy

There is evidence to suggest twins have a death rate four times higher than singletons with this figure worsening for triplets as they have a death rate six times higher than those of singletons (Wimalasundera , et al., 2003). An explanation for the higher death rates in multiple births than singletons is the preterm birth of multiples which in turn results in lower birth weights in the children (Beatrice, et al., 2002). A multiple regression of Length of Pregnancy was undertaken to see if non-singletons tended to have shorter pregnancies than singletons in our Growing Up in Ireland data, with key predictor nonsingleton with reference category singleton, while controlling for variables; gender of child, employment, education, age of primary carer, equivalent annual income, ethnicity, parent smokes and marital status as informed by the literature review. For the model, the shortened factor levels for employment, marital status, ethnicity and education were used to avoid breaching any assumptions of Multiple Linear Regression along with keeping consistent controls across all statistical tests.

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | β | Standard Error | P-value |
| Constant | 40.28 | 0.204 | <0.001 |
| Age | -0.026 | 0.006 | <0.001 |
| Gender of Child  *Female* | 1.00 | 1.00 | Reference |
| *Male* | 0.064 | 0.048 | .179 |
| Employment Status  *Employee* | 1.00 | 1.00 | Reference |
| *Home duties/retired* | -0.042 | 0.056 | 0.447 |
| *Self Employed* | 0.201 | 0.106 | 0.056 |
| *Student* | 0.151 | 0.276 | 0.584 |
| *Unemployed* | -0.173 | 0.142 | 0.224 |
| Highest Level of Education  *Third* | 1.00 | 1.00 | Reference |
| *Secondary* | -0.109 | 0.052 | 0.053 |
| *Primary* | -0.213 | 0.222 | 0.336 |
| Ethnicity  *Non-Irish* | 1.00 | 1.00 | Reference |
| *Irish* | 0.201 | 0.062 | 0.001 |
| Marital Status    *Married* | 1.00 | 1.00 | Reference |
| *Not married* | 0.123 | 0.069 | 0.766 |
| Singleton or Non-singleton  *Singleton* | 1.00 | 1.00 | Reference |
| *Non-Singleton* | -3.763 | 0.123 | <0.001 |
| Mother Smokes  *Never* | 1 | 1 | Reference |
| *Daily* | 0.045 | 0.060 | 0.456 |
| *Occasionally* | 0.098 | 0.067 | 0.139 |
| Equivalised annual income | 6.03e-7 | 1.923e-6 | 0.754 |

Table : Multiple Regression of length of pregnancy

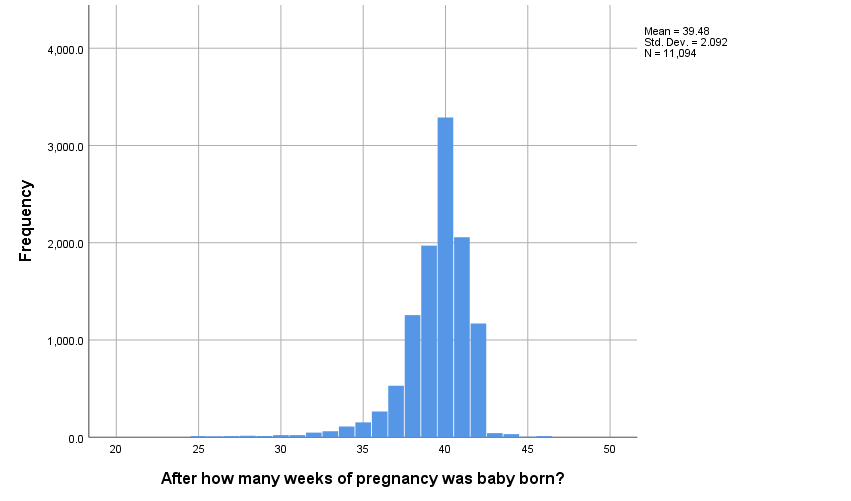
### Results/Conclusion of Length of Pregnancy

From Table 5

|  |  |  |
| --- | --- | --- |
| **Variable** | **Odds ratio** | **95% confidence interval** |
| **Age** | .997 | .960, .994 |
| **Gender of Child**  *Female* | 1.00 | Reference |
| *Male* | 1.001 | .855, 1.171 |
| **Employment Status**  *Employee* | 1.00 | Reference |
| *Home duties/retired* | 1.340 | 1.118, 1.605 |
| *Self Employed* | 1.563 | 1.110, 2.200 |
| *Student* | 1.233 | .593, 2.562 |
| *Unemployed* | 1.550 | 1.061, 2.263 |
| **Highest Level of Education**  *Third* | 1.00 | Reference |
| *Secondary* | 1.243 | 1.034, 1.493 |
| *Primary* | .900 | .481, 1.686 |
| **Ethnicity**  *Non-Irish* | 1.00 | Reference |
| *Irish* | 1.290 | 1.044, 1.594 |
| **Marital Status**    *Married* | 1.00 | Reference |
| *Not married* | 1.604 | 1.326, 1.941 |
| **Singleton or Non-singleton**  *Singleton* | 1.00 | Reference |
| *Non-Singleton* | 1.872 | 1.308, 2.679 |
| **How many weeks until the baby was born** | .985 | .949, 1.023 |
| **Equivalized annual income** | 1.00 | 1.00 |
|  |  |  |

, after running the multiple regression for length of pregnancy, the β (or coefficient) for a non-singleton was calculated to be -3.763 after controlling for the variables listed above. From this model, the length of the pregnancy for a non-singleton primary carer is 3.763 weeks shorter than a singleton pregnancy.

A histogram plot of the length of pregnancy is shown below. The data appears to be centered about 40 weeks with a slight skewness to the left.



For the multiple regression model, the R-Squared is 0.1384, so 13.84% of the variation of the length of pregnancy is explained by the model. As the model only explained 13.84% of variation, it is not a great predictor of the length of pregnancies. However, it does show the effects of having a non-singleton pregnancy on the length of pregnancies, which on average is 3.763 weeks shorter for non-singletons than singleton pregnancies.

## Depression of Primary Carer with key predictor non-singleton and singleton pregnancy

In the literature review there are several international studies that identified higher depression rates in non-singleton mothers compared to singleton mothers. A study by Hay (Hay, et al., 1990) found 29.7% of the mothers of twins aged 3 months reported depression (5 times higher rate than singleton mothers).

Depression levels of primary carers in the GUI study were measured using the eight-item Centre of Epidemiological Studies Depression Scale (CESD-8). The CESD-8 is a self-report screening survey used to gauge depression and distress levels. Participants answer a four-point rating scale;

1. Rarely or none of the time (less than 1 day)
2. Some or a little of the time (1-2 days)
3. Occasionally or a moderate amount of time (3-4 days)
4. Most or all of the time (5-7 days)

A sum of the responses is calculated with reference to the previous 7 days, with the scores being within the range of 0-24, or scores can be dichotomised with a score greater than or equal to 7 indicating a clinically significant level of psychological distress. A total of 194 cases were missing in the data. I decided to use the dichotomised category to run a binary logistic regression with Not Depressed (CESD < 7) = 0, Depressed (CESD >= 7) =1. I also ran a poisson regression and a negative binomial regression on this variable.

The key predictor variable for these regression models is non singleton with reference category singleton, controlling for the factors; gender of child, employment, education, marital status, age of primary carer, equivalent annual income, length of pregnancy and ethnicity as informed by the literature review.

For the binary logistic, poisson and negative binomial models, the shortened factor levels for employment, marital status, ethnicity and education are used so that adequate counts are within each of the subcategories defined by the independent variables. Firstly, a binary logistic regression is run to attain an odds ratio. Odds ratios are easily interpreted and give useful insights into the variables of interest. However, there is a loss of information when dichotomizing a variable. To combat this loss of information a poisson regression was conducted. As the variable is over dispersed, meaning the variance is greater than the mean, it is no longer appropriate to use a poisson regression. A negative binomial regression must be fitted instead. This ensures that a full insight into the original variable, and no information is lost by dichotomizing the CESD-8 scores.

## Binary Logistic Regression of Primary Carer Depression

Using the dichotomised variable for primary carer depression allows us to attain an odds ratio and get some insight into the effect of having a non-singleton has on the mental health of primary carers. A binary logistic regression was fitted with key predicator non-singleton with reference category singleton, while controlling for variables; gender of child, employment, first child, education, age of primary carer, length of pregnancy, equivalent annual income, ethnicity, first child and marital status as informed by the literature review.

For the model the shortened factor levels for employment, marital status, ethnicity and education were used to avoid breaching any assumptions of Binary Logistic Regression along with keeping consistent controls across all statistical tests.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Odds ratio** | **95% confidence interval** |
| **Age** | .997 | .960, .994 |
| **Gender of Child**  *Female* | 1.00 | Reference |
| *Male* | 1.001 | .855, 1.171 |
| **Employment Status**  *Employee* | 1.00 | Reference |
| *Home duties/retired* | 1.340 | 1.118, 1.605 |
| *Self Employed* | 1.563 | 1.110, 2.200 |
| *Student* | 1.233 | .593, 2.562 |
| *Unemployed* | 1.550 | 1.061, 2.263 |
| **Highest Level of Education**  *Third* | 1.00 | Reference |
| *Secondary* | 1.243 | 1.034, 1.493 |
| *Primary* | .900 | .481, 1.686 |
| **Ethnicity**  *Non-Irish* | 1.00 | Reference |
| *Irish* | 1.290 | 1.044, 1.594 |
| **Marital Status**    *Married* | 1.00 | Reference |
| *Not married* | 1.604 | 1.326, 1.941 |
| **Singleton or Non-singleton**  *Singleton* | 1.00 | Reference |
| *Non-Singleton* | 1.872 | 1.308, 2.679 |
| **How many weeks until the baby was born** | .985 | .949, 1.023 |
| **Equivalized annual income** | 1.00 | 1.00 |
|  |  |  |

Table : Logistic Regression of Depression of Primary Carer

### Results/Conclusion of Binary Logistic Regression on Depression of Primary Carers

In Table 6, the results from the logistic regression are presented. The odds ratio for primary carer depression for a non-singleton compared to a singleton birth was calculated to be 1.872, with 95% confidence interval [1.308, 2.679], after taking account of potential confounding variables; gender of child, employment, education, marital status, age of primary carer, equivalent annual income, length of pregnancy and ethnicity. From this model, the odds of being depressed for a non-singleton primary carer is estimated to be 1.872, with 95% confidence interval [1.308, 2.679], times the odds of being depressed for a singleton primary carer.



## Poisson regression of Primary Carer Depression

There is an associated loss of information when dichotomizing a variable. So as to retain the full insight and information of the CESD-8 scores, a Poisson regression of Primary Carer Depression with the original variable with scores ranging from 0-24 was undertaken, with key predicator nonsingleton with reference category singleton, while controlling for variables; gender of child, employment, first child, education, age of primary carer, length of pregnancy, equivalent annual income, ethnicity, first child and marital status as informed by the literature review. For the model, the shortened factor levels for employment, marital status, ethnicity and education were used to avoid breaching any assumptions of Poisson Regression along with keeping consistent controls across all statistical tests.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** |  | **Std. Error** | **P-value** |
| **Age** | -.124 | 1.74e-03 | <.001 |
| **Gender of Child**  *Female* | 1.00 | Reference | Reference |
| *Male* | -0.419 | 1.515e-02 | .005 |
| **Employment Status**  *Employee* | 1.00 | Reference | Reference |
| *Home duties/retired* | 1.823 | 1.744e-02 | <.001 |
| *Self Employed* | 1.718 | 3.425e-02 | <.001 |
| *Student* | 1.162 | 7.552e-02 | 0.124 |
| *Unemployed* | 3.289 | 3.704e-02 | <.001 |
| **Highest Level of Education**  *Third* | 1.00 | Reference | Reference |
| *Secondary* | -2.28 | 6.399e-02 | <.001 |
| *Primary* | .384 | 1.721e-02 | .025 |
| **Ethnicity**  *Non-Irish* | 1.00 | Reference | Reference |
| *Irish* | 1.44 | 2.022e-02 | <.001 |
| **Marital Status**    *Married* | 1.00 | Reference | Reference |
| *Not married* | 2.577 | 1.901e-02 | <.001 |
| **Singleton or Non-singleton**  *Singleton* | 1.00 | Reference | Reference |
| *Non-Singleton* | 2.431 | 3.778e-02 | <.001 |
| **How many weeks until the baby was born** | -.0879 | 3.737e-03 | .0186 |
| **Equivalized annual income** | -5.460e-5 | 7.255e-07 | <.001 |
| **First Child**    *Has older sibling* | 1.00 | Reference | Reference |
| First Child | -.478 | 1.564e-02 | .002 |
| **AIC: 43778** |  |  |  |

Table : Poisson regression of Primary Carer Depression

## Negative Binomial Regression of Primary Carer Depression

A Negative binomial regression of Primary Carer Depression, the original variable with scores ranging from 0-24, was undertaken to handle the over-dispersion in the variable Primary Carer Depression, with key predicator nonsingleton with reference category singleton, while controlling for variables; gender of child, employment, first child, education, age of primary carer, length of pregnancy, equivalent annual income, ethnicity, first child and marital status as informed by the literature review. For the model, the shortened factor levels for employment, marital status, ethnicity and education were used to avoid breaching any assumptions of Negative Binomial Regression along with keeping consistent controls across all statistical tests.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** |  | **Std. Error** | **P-value** |
| **Age** | -.118 | 3.927e-03 | .002 |
| **Gender of Child**  *Female* | 1.00 | Reference | Reference |
| *Male* | -0.486 | 3.344e-02 | .147 |
| **Employment Status**  *Employee* | 1.00 | Reference | Reference |
| *Home duties/retired* | 1.957 | 3.889e-02 | <.001 |
| *Self Employed* | 1.972 | 7.558e-02 | .009 |
| *Student* | .85 | 1.831e-01 | 0.642 |
| *Unemployed* | 3.356 | 9.126e-02 | <.001 |
| **Highest Level of Education**  *Third* | 1.00 | Reference | Reference |
| *Secondary* | -1.794 | 1.358e-01 | .186 |
| *Primary* | .466 | 3.726e-02 | .211 |
| **Ethnicity**  *Non-Irish* | 1.00 | Reference | Reference |
| *Irish* | 1.135 | 4.388e-02 | .009 |
| **Marital Status**    *Married* | 1.00 | Reference | Reference |
| *Not married* | 2.631 | 4.431e-02 | <.001 |
| **Singleton or Non-singleton**  *Singleton* | 1.00 | Reference | Reference |
| *Non-Singleton* | 2.613 | 9.198e-02 | .004 |
| **How many weeks until the baby was born** | .057 | 8.509e-03 | .498 |
| **Equivalized annual income** | -4.557e-5 | 1.456e-06 | .002 |
| **First Child**    *Has older sibling* | 1.00 | Reference | Reference |
| First Child | -.554 | 3.459e-02 | .109 |
| **AIC: 30807** |  |  |  |

Table : Negative Binomial Regression of Primary Carer Depression

### Results/Conclusion of Primary Carer Depression Models

In Table 8, the results from the Negative Binomial regression are presented. The mean difference for primary carer depression scores for a non-singleton compared to a singleton birth was calculated to be 2.613, after taking account of potential confounding variables; gender of child, employment, education, marital status, age of primary carer, equivalent annual income, length of pregnancy, first child and ethnicity. The negative binomial model is a more accurate model than the Poisson model, where the Poisson Regression results can be seen in Table 7, as the Negative Binomial Regression was able to handle the over-dispersion in the variable Primary Carer Depression. The Akaike’s Information Criterion (AIC) which compares the quality of a set of statistical models to each other, with a lower score indicating a more accurate model, had a lower score of 30807 for the negative binomial model compared to the AIC score of the Poisson model which was 43778.

# Analysis of Wave 3 data

The aim is to statistically analyze the differences between non-singleton and singleton children developments such as, cognitive and social involvements using the wave 3 cohort data, when the children are 5 years old. One of the areas of interest would be to look at academic results to get an insight into the cognitive developments of both the non-singleton and singleton children. Following on from academic scores, an analysis on Strength and Difficulties scores using cluster analysis methods such as hierarchical and k-mean, would allow for a better perception of social development of the children. Using multiple regression and quantile regression on naming vocabulary scores and picture scores would allow for better insights of the academic and cognitive ability of the children. It was also of interest to investigate the relationship between the mental wellbeing of the primary carer and having a non-singleton pregnancy in wave 3 to get an insight into how their mental states vary from wave 1 to wave 3 using negative binomial regression. Finally, to see if potentially having non-singletons has any effect on the parent’s marital status.

## Negative Binomial Regression of Primary Carer Parental Stress Wave 3

A Negative binomial regression of Primary Carer Parental Stress scores in wave 3 was undertaken to see the effect of having a non-singleton pregnancy vs a singleton pregnancy has on the stress scores of their respective primary carers and in turn mental health. The negative binomial regression would allow us to further investigate if the trend of non-singletons mothers tending to be more depressed from wave 1, when the child is 9 months old, is still evident in wave 3 when the child is 5 years old. A negative binomial regression was used to handle the over-dispersion in the variable Primary Carer stress in wave 3, with key predicator non-singleton with reference category singleton, while controlling for variables; gender of child, employment, first child, education, age of primary carer, length of pregnancy, equivalent annual income, ethnicity, first child and marital status as informed by the literature review.

For the model, the shortened factor levels for employment, marital status, ethnicity and education were used to avoid breaching any assumptions of Negative Binomial Regression along with keeping consistent controls across all statistical tests.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** |  | **Std. Error** | **P-value** |
| **Age** | -.015 | 9.744e-03 | .119 |
| **Gender of Child**  *Female* | 1.00 | Reference | Reference |
| *Male* | -0.12 | 8.142e-03 | .134 |
| **Employment Status**  *Employee* | 1.00 | Reference | Reference |
| *Home duties/retired* | -.041 | 9.627e-03 | .667 |
| *Self Employed* | -.070 | 1.829e-02 | .702 |
| *Student* | .214 | 4.447e-02 | .631 |
| *Unemployed* | .348 | 2.309e-02 | .132 |
| **Highest Level of Education**  *Third* | 1.00 | Reference | Reference |
| *Secondary* | -.461 | 3.709e-02 | .214 |
| *Primary* | -.453 | 8.989e-03 | <.001 |
| **Ethnicity**  *Non-Irish* | 1.00 | Reference | Reference |
| *Irish* | -1.06 | 1.099e-02 | <.001 |
| **Marital Status**    *Married* | 1.00 | Reference | Reference |
| *Not married* | 2.631 | 4.431e-02 | <.001 |
| **Singleton or Non-singleton**  *Singleton* | 1.00 | Reference | Reference |
| *Non-Singleton* | .0476 | 2.285e-02 | .037 |
| **How many weeks until the baby was born** | .020 | 2.075e-03 | .333 |
| **Equivalized annual income** | -4.558e-6 | 3.470e-07 | .189 |
| **First Child**    *Has older sibling* | 1.00 | Reference | Reference |
| First Child | .642 | 8.431e-03 | <.001 |
| **AIC: 35996** |  |  |  |

Table : Negative Binomial Regression of Primary Carer Parental Stress

### Results/Conclusion of Negative Binomial Regression of Primary Carer Parental Stress

From Table 9, the results from the Negative Binomial regression for Primary Carer Parental Stress in wave 3 are presented. The mean difference for primary carer parental stress for a non-singleton compared to a singleton birth was calculated to be 0.0476, after taking account of potential confounding variables; gender of child, employment, education, marital status, age of primary carer, equivalent annual income, length of pregnancy, first child and ethnicity. This would suggest that the higher depression rates of non-singleton primary carers as seen in the wave 1 analysis, may no longer be as prevalent in the wave 3 data.

## Hierarchical Cluster Analysis of Wave 3 Academic Variables

A study by Day (1932), found that twins were less advanced in language development at the ages of 2, 3, 4 and 5 years of age in comparison to singletons of the same age and that twins also scored on average 10 points lower on an IQ test than a singleton child. A Hierarchical cluster analysis was performed on the wave 3 academic variables using Ward’s method, as this was the most suitable for our data to investigate and see if these trends found by Day (1932) are evident in the Growing Up In Ireland data. The following variables were used; attitude, language, reading and linking scores from their teachers’ reports. After graphing the dendrogram a three-cluster solution was identiﬁed based on maximising the similarity within the clusters and variability between clusters. The clusters were identified as following; Cluster 1 are the best performing in academic scores, Cluster 2 have lower academic scores than Cluster 1 and finally Cluster 3 was the lowest performing cluster in academic scores across all three clusters.

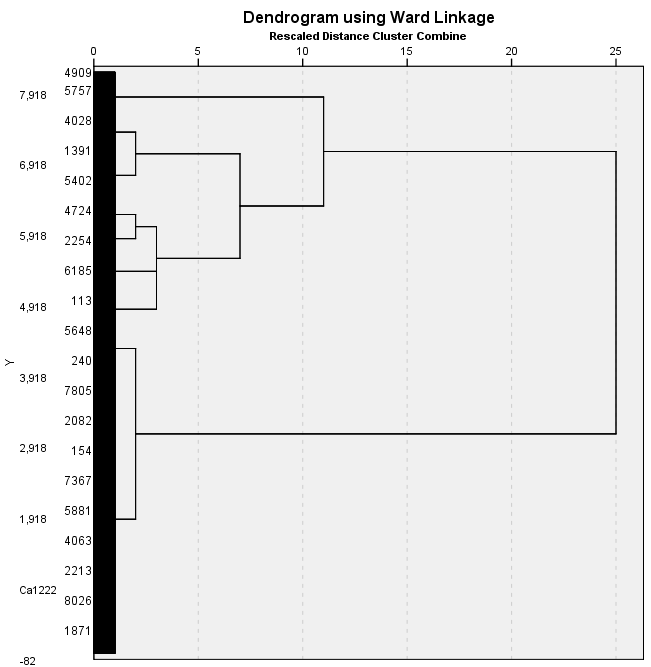


Figure : Dendrogram of Academic Variables

Using a Ward Linkage distance of 8.5, on the above, Figure 28: Dendrogram of Academic Variables, a three-cluster solution was identiﬁed based on maximising the similarity within the clusters and variability between clusters.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  | **Cluster1**  (n = 4901)  (60.0%) | **Cluster 2**  (n=2355)  (28.8%) | **Cluster 3**  (n=909)  (11.1%) |
|  |  |  |  |
| Attitudes Total Teacher Report  *Median* | 9.0 | 8.0 | 5.0 |
| *Interquartile Range* | 1.0 | 3.0 | 4.0 |
|  |  |  |  |
| Language Total Teacher Report  *Median* | 9.0 | 7.0 | 3.0 |
| *Interquartile Range* | .00 | 4.0 | 2.0 |
| Linking Total Teacher Report  *Median* | 9.0 | 7.0 | 4.0 |
| *Interquartile Range* | .00 | 3.0 | 2.0 |
| Reading Total Teacher Report  *Median* | 9.0 | 7.0 | 4.0 |
| *Interquartile Range* | 1.0 | 2.0 | 3.0 |

Table : Cluster Analysis of Academic Variables

From Table 10, Cluster 1 had a median score of 9 across each of the variables; attitude scores, language scores, linking scores and finally reading scores from their teachers’ report, with 9 being the highest score possible in each of the variables, and the higher the score the better the ability. Cluster 1 also has the smallest interquartile range across all the variables in each cluster. This indicates that cluster 1 is the best performing cluster of the 3 in the academic variables seen above. In cluster 2, the median scores across the variables are 7 and 8, with interquartile ranges between 2 - 4, indicating a lower performance in the academic variables than cluster 1. Finally, cluster 3 were the lowest performing cluster with median scores spanning from 3 - 5 across the variables and interquartile ranges between 2 - 4.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  | **Cluster1**  (n = 4901)  n (%) | **Cluster 2**  (n=2355)  n (%) | **Cluster 3**  (n=909)  n (%) | Cramer V | P-Value |
|  |  |  |  |  |  |
| Male  *Singleton* | 2169 (54.5) | 1264(31.8) | 544(13.7) | .024 | .307 |
| *Non-Singleton* | 66(49.6) | 43(32.3) | 24(18) |  |  |
|  |  |  |  |  |  |
| Female  *Singleton* | 2593(66.2) | 1002(25.6) | 323(8.2) | .051 | .006 |
| *Non-Singleton* | 73(53.3) | 46(33.6) | 18(13.1) |  |  |

Table : CLuster Analysis of Academic Variables Controlling for Gender and Nonsingleton

|  |  |  |  |
| --- | --- | --- | --- |
|  | Cluster1  (n = 4901)  n (%) | Cluster 2  (n=2355)  n (%) | Cluster 3  (n=909)  n (%) |
|  |  |  |  |
| *Ethnicity*  *Irish* | 3437 (63.8) | 1484 (27.5) | 469 (8.7) |
| *Non-Irish* | 607 (53.1) | 358 (31.3) | 178 (15.6) |
|  |  |  |  |
| Employment Status  *Employee* | 2897 (63.8) | 1250 (27.5) | 395 (8.7) |
| *Home duties/retired* | 1510 (54.2) | 849 (30.5) | 425 (15.3) |
| *Self Employed* | 237 (61.1) | 119 (30.7) | 32 (8.2) |
| *Student* | 73 (61.9) | 35 (29.7) | 10 (8.5) |
| *Unemployed* | 184 (55.3) | 102 (30.6) | 47 (14.1) |
| Highest Level of Education    *Third* | 2010 (65.5) | 833 (27.2) | 225 (7.3) |
| *Secondary* | 2826 (57.3) | 1477 (30.0) | 625 (12.7) |
| *Primary* | 67 (39.9) | 43 (25.6) | 58 (34.5) |
|  |  |  |  |

Table : Cluster Analysis of Academic Variables Controlling for Social Factors

### Results/Conclusion of the Cluster Analysis on Academic Variables

From Table 11

|  |  |  |
| --- | --- | --- |
| **Variable** | **Odds ratio** | **95% confidence interval** |
| **Age** | .997 | .960, .994 |
| **Gender of Child**  *Female* | 1.00 | Reference |
| *Male* | 1.001 | .855, 1.171 |
| **Employment Status**  *Employee* | 1.00 | Reference |
| *Home duties/retired* | 1.340 | 1.118, 1.605 |
| *Self Employed* | 1.563 | 1.110, 2.200 |
| *Student* | 1.233 | .593, 2.562 |
| *Unemployed* | 1.550 | 1.061, 2.263 |
| **Highest Level of Education**  *Third* | 1.00 | Reference |
| *Secondary* | 1.243 | 1.034, 1.493 |
| *Primary* | .900 | .481, 1.686 |
| **Ethnicity**  *Non-Irish* | 1.00 | Reference |
| *Irish* | 1.290 | 1.044, 1.594 |
| **Marital Status**    *Married* | 1.00 | Reference |
| *Not married* | 1.604 | 1.326, 1.941 |
| **Singleton or Non-singleton**  *Singleton* | 1.00 | Reference |
| *Non-Singleton* | 1.872 | 1.308, 2.679 |
| **How many weeks until the baby was born** | .985 | .949, 1.023 |
| **Equivalized annual income** | 1.00 | 1.00 |
|  |  |  |

, after running a cross tabs on the cluster analysis, it seems that more singleton children, both male and female tend to be in the above average cluster with this evidence being stronger in the female cases; Male Singletons vs Non-singletons, 54.5% vs 49.6%, and Female Singletons vs Non-singletons, 66.2% vs 53.3%. Non-singletons were also more frequent in the below average class with Male Singletons vs Non-Singletons 13.7% vs 18% and Female Singletons vs Non-Singletons 8.2% vs 13.1%.

Looking at the control variables of social factors, Table 12, it would suggest that children whose primary carer has a third level education tended to be in cluster 1, with 65.5% of the children whose primary carer had a third level education being in cluster 1, the best performing in academic scores cluster. Closely followed by second level education making up for 57.3% in cluster 1. Primary Carers whose education level was a primary school education tended to have the highest percentage in cluster 3, the lowest performing cluster in academic scores across all three clusters, with 34.5% of the children whose primary carer had a primary school education being in cluster 3.

The evidence would suggest that for primary carers ethnicity, Irish primary carers children tended to be in cluster 1, with 63.8% of the children whose primary carer was of Irish ethnicity being in cluster 1, the best performing in academic scores cluster. While non-Irish primary carers children tended to be in cluster 1, with 53.1% of the children whose primary carer was of non-Irish ethnicity being in cluster 1. However, for non-Irish primary carers children they made up 15.6% in cluster 3 vs 8.7% for Irish primary carers children being in cluster 3, the lowest performing cluster in academic scores across all three clusters.

For employment status of the primary carer two categories stand out, home duties/retired and unemployed. Both these categories have the least percentage of children in cluster 1, 54.2% and 55.3% respectively, compared to employed, student and self-employed (63.8%, 61.9% and 61.1%). On the other hand, both home duties/retired and unemployed have a higher percentage of children in cluster 3, the lowest performing cluster with 15.3% and 14.1% compared to employed, student and self-employed (8.7%, 8.5% and 8.2%).

## K-means Cluster Analysis of Strength and Difficulties on Wave 3 Data

A study by Wood (Wood, et al., 1996), found that twins were more likely to develop Attention Deficit Hyperactivity Disorder than singletons, while a study by DiLalla, (DiLalla, 2006), found that twins aged 5 years old were less social when they were placed in a peer play situation with an unfamiliar, same age, same sex peer. Using a Hierarchical cluster and then K-means cluster analysis, the intentions were to see if the above trends were evident in the Growing Up in Ireland data. A hierarchical cluster analysis was performed on the wave 3 Strength and Difficulties Questionnaire (SDQ) variables; SDQ emotional subscale, SDQ conduct subscale, SDQ peer problems as judged by their primary carer, SDQ prosocial subscale as judged by the primary carer and finally SDQ hyperactivity subscale as judged by the primary carers. After graphing the dendrogram, a four-cluster solution was identiﬁed based on maximising the similarity within the clusters and variability between clusters. After identifying the number of clusters, a k-means cluster analysis was run with a specific 4 cluster solution using Ward’s method, as this was the most suitable for our data. The clusters were identified as following; Cluster 1 was emotional children, Cluster 2 was the worst SDQ scoring children, Cluster 3 was hyperactive children and finally Cluster 4 was the best SDQ scoring children.

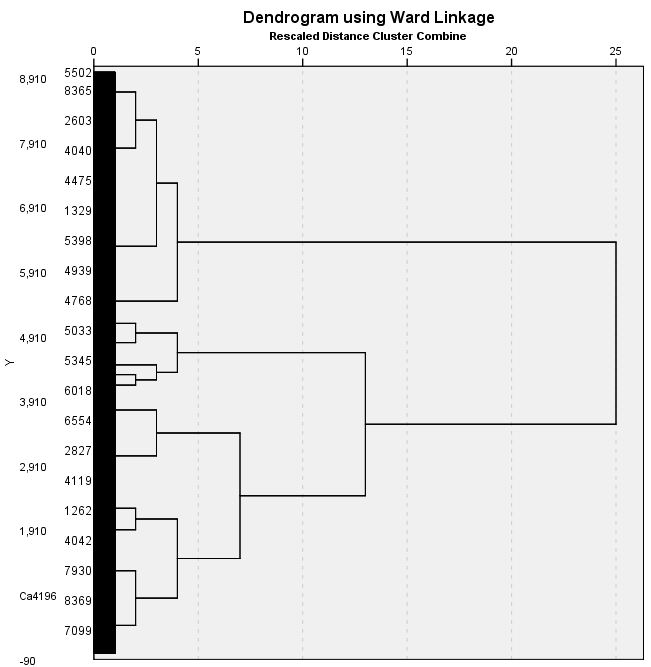


Figure : Dendrogram of SDQ Variables

Using a Ward Linkage distance of 5, on the above, Figure 29, a four-cluster solution was identiﬁed based on maximising the similarity within the clusters and variability between clusters.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Cluster1  (n = 2094)  (25%) | Cluster 2  (n=712)  (8.5%) | Cluster 3  (n=1508)  (18%) | Cluster 4  (n=4048)  (48.5%) |
|  |  |  |  |  |
| SDQ emotional subscale    *Median* | 1.0 | 5.0 | 0.0 | 0.0 |
| *Interquartile Range* | 3.0 | 2.0 | 1.0 | 1.0 |
|  |  |  |  |  |
| SDQ conduct subscale    *Median* | 0.0 | 2.0 | 1.0 | 0.0 |
| *Interquartile Range* | 1.0 | 3.0 | 2.0 | 0.0 |
| SDQ hyperactivity subscale    *Median* | 3.0 | 6.0 | 7.0 | 1.0 |
| *Interquartile Range* | 3.0 | 4.0 | 2.0 | 2.0 |
| SDQ peer problems    *Median* | 1.0 | 3.0 | 1.0 | 0.0 |
| *Interquartile Range* | 2.0 | 3.0 | 2.0 | 0.0 |
| SDQ Prosocial subscale    *Median* | 7.0 | 5.0 | 6.0 | 10.0 |
| Interquartile Range | 2.0 | 3.0 | 3.0 | 1.0 |

Table : Cluster Analysis of SDQ Variables

From Table 13, for the variables; SDQ emotional subscale, SDQ conduct subscale, SDQ peer problems, SDQ hyperactivity subscale as judged by their primary carers, a lower score is more desirable, with high scores indicating problems. For SDQ prosocial subscale as judged by the primary carer, higher scores are desirable and lower scores indicate anti-social traits.

Cluster 4 had the lowest median scores across 4 of the variables; SDQ emotional, SDQ conduct subscale, SDQ peer problems, SDQ hyperactivity subscale while also having the highest median in the prosocial subscale. The interquartile ranges across the variables span from 0 - 2, indicating that Cluster 4 is the best SDQ scoring cluster, and the most desirable cluster to be in.

Cluster 3 seems to have a higher median for hyperactivity than the rest of the clusters, while showing no other higher than average scores across all other variables. This would suggest that the children in cluster 3 are hyperactive.

Cluster 2 seems to have the worst SDQ median scores across all the variables, tending to have higher SDQ scores for the first 4 variables along with the lowest prosocial median. This cluster is the least desirable to be in.

Cluster 1 seems to have normal SDQ median scores for, SDQ conduct, SDQ peer problems, SDQ prosocial but for variables SDQ emotional and SDQ hyperactivity the median and interquartile range are higher in these 2 variables compared to the other 3, suggesting that Cluster 1 has emotional problems as well as more hyperactive children than in clusters 2 and 4.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  | **Cluster 1**  (n = 2094)  n (%) | **Cluster 2**  (n=712)  n (%) | **Cluster 3**  (n=1508)  n (%) | **Cluster 4**  (n=4048)  n (%) | Cramer V | P-Value |
|  |  |  |  |  |  |  |
| Male  *Singleton* | 1008 (24.7) | 408 (10.0) | 1006 (24.6) | 1660 (40.7) | .017 | .737 |
| *Non-Singleton* | 34 (25.4) | 17 (12.7) | 30 (22.4) | 53 (39.6) |  |  |
|  |  |  |  |  |  |  |
| Female  *Singleton* | 1017 (25.4) | 276 (6.9) | 447 (11.2) | 2261 (56.5) | .036 | .141 |
| *Non-Singleton* | 35 (24.1) | 11 (7.6) | 25 (17.2) | 74 (51.0) |  |  |

Table : CLuster Analysis of SDQ Variables Controlling for Gender and Non-Singleton

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Cluster1  (n = 2094)  n (%) | Cluster 2  (n=712)  n (%) | Cluster 3  (n=1508)  n (%) | Cluster 4  (n=4048)  n (%) |
|  |  |  |  |  |
| *Ethnicity*  *Irish* | 1402 (25.4) | 413 (7.5) | 890 (16.1) | 2819 (51.0) |
| *Non-Irish* | 306 (26.2) | 96 (8.2) | 250 (21.4) | 517 (44.2) |
|  |  |  |  |  |
| Employment Status  *Employee* | 1146 (24.7) | 323 (7.0) | 8.8 (17.4) | 2363 (50.9) |
| *Home duties/retired* | 727 (25.5) | 316 (11.1) | 532 (18.6) | 1281 (44.9) |
| *Self Employed* | 99 (24.6) | 33 (8.2) | 63 (15.6) | 208 (51.6) |
| *Student* | 34 (27.4) | 10 (8.1) | 28 (22.6) | 52 (41.9) |
| *Unemployed* | 88 (26.0) | 30 (8.8) | 77 (22.7) | 144 (42.5) |
| Highest Level of Education    *Third* | *816 (25.9)* | 207 (6.6) | 500 (15.9) | 1624 (51.6) |
| *Secondary* | *1232 (24.4)* | 479 (9.5) | 958 (19.0) | 2370 (47.0) |
| *Primary* | *46 (26.7)* | 25 (14.5) | 48 (27.9) | 53 (30.8) |

Table : Cluster Analysis of SDQ Variables Controlling for Social Factors

### Findings from the K-Means Cluster Analysis

The analysis presented in Table 14

|  |  |  |
| --- | --- | --- |
| **Variable** | **Odds ratio** | **95% confidence interval** |
| **Age** | .997 | .960, .994 |
| **Gender of Child**  *Female* | 1.00 | Reference |
| *Male* | 1.001 | .855, 1.171 |
| **Employment Status**  *Employee* | 1.00 | Reference |
| *Home duties/retired* | 1.340 | 1.118, 1.605 |
| *Self Employed* | 1.563 | 1.110, 2.200 |
| *Student* | 1.233 | .593, 2.562 |
| *Unemployed* | 1.550 | 1.061, 2.263 |
| **Highest Level of Education**  *Third* | 1.00 | Reference |
| *Secondary* | 1.243 | 1.034, 1.493 |
| *Primary* | .900 | .481, 1.686 |
| **Ethnicity**  *Non-Irish* | 1.00 | Reference |
| *Irish* | 1.290 | 1.044, 1.594 |
| **Marital Status**    *Married* | 1.00 | Reference |
| *Not married* | 1.604 | 1.326, 1.941 |
| **Singleton or Non-singleton**  *Singleton* | 1.00 | Reference |
| *Non-Singleton* | 1.872 | 1.308, 2.679 |
| **How many weeks until the baby was born** | .985 | .949, 1.023 |
| **Equivalized annual income** | 1.00 | 1.00 |
|  |  |  |

, found a slightly higher proportion of non-singleton females were assigned to cluster 3, the hyperactive cluster in comparison to their singleton counterparts, 17.2% vs 11.2%. For males there were no significant difference for non-singletons vs singletons across each cluster.

Differences across the demographic variables, Table 15, suggests that children whose primary carer has a third level education tended to be in cluster 4 with 51.6% of the children whose primary carer had a third level education being in cluster 4, the best scoring in SDQ scores cluster. Closely followed by second level education making up for 47.0% in cluster 1. Primary Carers whose education level was a primary school education tended to have the highest percentage in cluster 2 and 3, which are the worst SDQ scoring children and hyperactive children clusters with 14.5% and 27.9% respectively.

The evidence would suggest that for primary carer’s ethnicity, Irish primary carers children tended to be in cluster 4 with 51.0% of the children whose primary carer was of Irish ethnicity being in cluster 4, the best SDQ scores cluster. For non-Irish primary carers children tended to be in cluster 4 with 44.2% of the children whose primary carer was of non-Irish ethnicity being in cluster 4. However, for non-Irish primary carers children they made up 21.4% in cluster 3 vs 16.1% for Irish primary carers children being in cluster 3, the hyperactive children cluster.

For employment status clusters 1, 2 and 3 only vary by a small number of percentages across each variable. Home duties/retired has the highest percentage of children in cluster 2 with 11.1%, with cluster 2 being the lowest SDQ scores cluster. Self-employed primary carers have the smallest percentage in cluster 3 with 15.6%, with cluster 3 being hyperactive children. Finally, for primary carers who are employees or self-employed, have the highest percentage in cluster 4, the best SDQ scoring cluster with 50.9% and 51.6%, followed by home duties/retired, unemployed and student which have percentages respectively 44.9%, 42.5% and 41.9%.

## Multiple Regression of Wave 3 Naming Vocabulary Scores with key predictors non-singleton and singleton

A study by Webbink (Webbink, et al., 2008) which used a cohort study undertaken in the Netherlands that investigated and compared the longitudinal IQ scores for approximately 188,000 singletons and 6000 twins who attended primary school in the Netherlands from 1994 to 2003, found that after controlling for variables of interest such as; gender, ethnicity and parent’s education, language scores for twins aged 6 years old scored 16% of a standard deviation lower than non-singletons in the language test. A multiple regression of naming vocabulary scores from wave 3 data was undertaken to see if these trends were evident in the Growing Up in Ireland data, with key predicator variable nonsingleton with reference category singleton, while controlling for variables; gender of child, employment status of primary carer, education, age of primary carer, equivalent annual income of household, ethnicity, if study child had siblings and marital status of the primary carer, as informed by the literature review.

For the model, the shortened factor levels for employment, marital status, ethnicity and education were used to avoid breaching any assumptions of Multiple Linear Regression along with keeping consistent controls across all statistical tests.

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Β | Standard Error | P-value |
| Constant | 89.24 | 2.037 | <0.001 |
| Age of Primary Carer | 0.302 | 0.056 | <0.001 |
| Gender of Child  *Female* | 1.00 | 1.00 | Reference |
| *Male* | 1.93 | 0.46 | <0.001 |
| Employment Status of Primary Carer  *Employee* | 1.00 | 1.00 | Reference |
| *Home duties/retired* | -1.69 | 0.539 | 0.001 |
| *Self Employed* | -0.926 | 0.994 | 0.352 |
| *Student* | 2.346 | 2.638 | 0.374 |
| *Unemployed* | -1.129 | 1.387 | 0.416 |
| Highest Level of Education    *Third* | 1.00 | 1.00 | Reference |
| *Secondary* | -1.574 | 0.494 | 0.001 |
| *Primary* | -3.687 | 2.336 | 0.115 |
| Ethnicity  *Non-Irish* | 1.00 | 1.00 | Reference |
| *Irish* | 13.07 | 0.631 | <0.001 |
| Marital Status    *Married* | 1.00 | 1.00 | Reference |
| *Not married* | 0.820 | 0.676 | 0.225 |
| Child has Sibling  *Yes* | 1.00 | 1.00 | Reference |
| *No* | .126 | .469 | 0.788 |
| Singleton or Non-singleton  *Singleton* | 1.00 | 1.00 | Reference |
| *Non-Singleton* | -.761 | 1.192 | 0.523 |
| Equivalized annual income | 9.626e-5 | 1.893e-6 | <0.001 |

Table : Multiple Regression of Wave 3 Naming Vocabulary Scores with key predictors non-singleton and singleton

### Results/Conclusion of Multiple Regression of Naming Vocabulary Scores

From Table 16

|  |  |  |
| --- | --- | --- |
| **Variable** | **Odds ratio** | **95% confidence interval** |
| **Age** | .997 | .960, .994 |
| **Gender of Child**  *Female* | 1.00 | Reference |
| *Male* | 1.001 | .855, 1.171 |
| **Employment Status**  *Employee* | 1.00 | Reference |
| *Home duties/retired* | 1.340 | 1.118, 1.605 |
| *Self Employed* | 1.563 | 1.110, 2.200 |
| *Student* | 1.233 | .593, 2.562 |
| *Unemployed* | 1.550 | 1.061, 2.263 |
| **Highest Level of Education**  *Third* | 1.00 | Reference |
| *Secondary* | 1.243 | 1.034, 1.493 |
| *Primary* | .900 | .481, 1.686 |
| **Ethnicity**  *Non-Irish* | 1.00 | Reference |
| *Irish* | 1.290 | 1.044, 1.594 |
| **Marital Status**    *Married* | 1.00 | Reference |
| *Not married* | 1.604 | 1.326, 1.941 |
| **Singleton or Non-singleton**  *Singleton* | 1.00 | Reference |
| *Non-Singleton* | 1.872 | 1.308, 2.679 |
| **How many weeks until the baby was born** | .985 | .949, 1.023 |
| **Equivalized annual income** | 1.00 | 1.00 |
|  |  |  |

, after running the multiple regression, the β (or coefficient) for a non-singleton was calculated to be -0.761 after controlling for the variables listed above. From this model, the naming vocabulary scores for non-singleton children is on average 0.761 points lower than naming vocabulary scores of singleton children with 95% CI [-3.09, 1.57]. However, since 0 is contained in the confidence interval, this result is not statistically significant.

## Quantile Regression on Naming Vocabulary Wave 3

A quantile regression of naming vocabulary scores from wave 3 data was undertaken as it shows more insight into the naming vocabulary scores, with key predicator variable non-singleton with reference category singleton, while controlling for variables; gender of child, employment status of primary carer, education, age of primary carer, equivalent annual income of household, ethnicity, if study child had siblings and marital status of the primary carer, as informed by the literature review. With the figure below, Figure 30, we can see how the effect of naming vocabulary scores changes by percentile for non-singletons.

For the model, the shortened factor levels for employment, marital status, ethnicity and education were used to avoid breaching any assumptions of Quantile Regression along with keeping consistent controls across all statistical tests. In Figure 30, a summary of the quantile regression results is shown. Each of the dependent variables in the quantile regression has its own graph, with the x axis being the percentile τ and the predictive values of the variables being on the y axis. For each of the dependent variables the quantile regression estimates are plotted starting with τ = 0.05 and finishing at 0.85, as seen below as the shaded grey region with dotted points. For every dependent variable, these point estimates can be interpreted as the impact of a one-unit change on the dependent variable has on naming vocabulary scores, while holding other dependent variables fixed. The x-axis is the ranges of τ, the quantiles, while the y-axis is the effect of the dependent variables has on naming vocabulary scores. The red solid lines in every plot represents the ordinary least square estimates of the mean effect, while the red dashed lines show the 90% confidence interval for the least square estimates. The shaded grey region portrays a 90% pointwise confidence band for the quartile regression estimates. The first plot labelled Intercept is the naming vocabulary scores when the reference dependent variables are used, that is; the naming vocabulary scores of a female singleton child, who has siblings, when the child’s primary carer is married, as well as the primary carer’s ethnicity being non-Irish who is employed and has a third level education.

Looking at non-singleton variable plot in Figure 30, at the lower quantiles of τ, say for example when τ=10, there is a notable difference in the 90% confidence interval for the ordinary least squares estimate and the 90% pointwise confidence band for the quartile regression estimates with the least squares estimating a difference of roughly ±1 point in naming vocabulary scores while the quantile regression estimate is ±3.5. This is also true at the higher quantiles as for when τ ≥ 65, there is a notable difference in the 90% confidence interval for the ordinary least squares estimate and the 90% pointwise confidence band for the quartile regression estimates with the least squares estimating a difference of roughly ±1 point in naming vocabulary scores while the quantile regression estimate gives only negative effects.

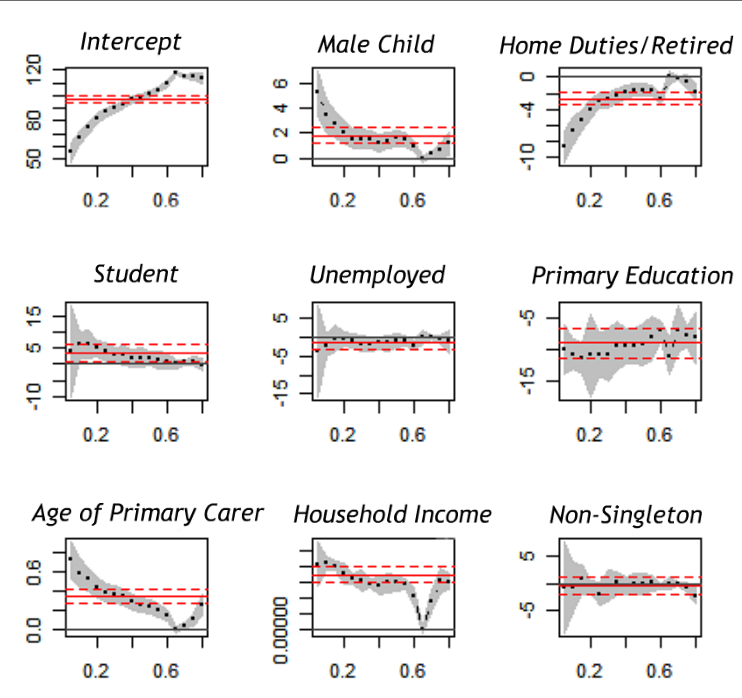


Figure : Output of Quantile Regression

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Quantile |  | | |  | | |  | | |
|  | **β** | **Std. Error** | **P-value** | **β** | **Std. Error** | **P-value** | **β** | **Std. Error** | **P-value** |
| Constant | 87.04 | 2.31 | <.001 | 101.22 | 1.56 | <.001 | 114.9 | 1.53 | <.001 |
|  |  |  |  |  |  |  |  |  |  |
| Age of Primary Carer | 0.37 | .05 | <.001 | 0.24 | .041 | <.001 | .104 | .038 | .006 |
| Gender of Child  *Female* | Reference |  |  |  |  |  |  |  |  |
| *Male* | 1.54 | 0.49 | .002 | 1.60 | .376 | <.001 | .601 | .373 | .106 |
| Employment Status of Primary Carer    *Employee* | Reference |  |  |  |  |  |  |  |  |
| *Home duties/retired* | -2.95 | 0.61 | <.001 | -1.58 | .422 | <.001 | -.611 | .485 | .208 |
| *Self Employed* | -1.34 | 0.95 | .158 | -1.09 | 1.19 | .358 | .102 | 1.46 | .944 |
| *Student* | 3.83 | 1.74 | .028 | 1.86 | 1.28 | .149 | .986 | .906 | .276 |
| *Unemployed* | -0.88 | 1.07 | .413 | -.845 | 1.16 | .468 | -.60 | .606 | .321 |
| Highest Level of Education    *Third* | Reference |  |  |  |  |  |  |  |  |
| *Secondary* | -0.77 | 0.54 | .153 | -.073 | .435 | .091 | -1.45 | .588 | .014 |
| *Primary* | -10.78 | 2.08 | <.001 | -9.14 | 1.98 | <.001 | -7.85 | .950 | <.001 |
|  |  |  |  |  |  |  |  |  |  |
| Marital Status  *Married* | Reference |  |  |  |  |  |  |  |  |
| *Not married* | 1.27 | 0.65 | .049 | .769 | .442 | .082 | .446 | .443 | .314 |
| Child has Sibling  *Yes* | Reference |  |  |  |  |  |  |  |  |
| *No* | -.464 | .491 | .344 | .508 | .388 | .191 | -.008 | .409 | .983 |
| Singleton or Non-singleton  *Singleton* | Reference |  |  |  |  |  |  |  |  |
| *Non-Singleton* | -1.98 | 1.28 | .120 | -0.231 | 1.177 | .844 | -.072 | .653 | .269 |
|  |  |  |  |  |  |  |  |  |  |
| Equivalized annual income | 1.6e-4 | 1.0e-5 | <.001 | 1.5e-4 | 2.0e-5 | <.001 | 1.6e-4 | 2e-5 | <.001 |

Table : Quantile Regression of Naming Vocabulary Scores

## Results/Conclusion of Quantile Regression on Naming Vocabulary Scores

From Table 17

|  |  |  |
| --- | --- | --- |
| **Variable** | **Odds ratio** | **95% confidence interval** |
| **Age** | .997 | .960, .994 |
| **Gender of Child**  *Female* | 1.00 | Reference |
| *Male* | 1.001 | .855, 1.171 |
| **Employment Status**  *Employee* | 1.00 | Reference |
| *Home duties/retired* | 1.340 | 1.118, 1.605 |
| *Self Employed* | 1.563 | 1.110, 2.200 |
| *Student* | 1.233 | .593, 2.562 |
| *Unemployed* | 1.550 | 1.061, 2.263 |
| **Highest Level of Education**  *Third* | 1.00 | Reference |
| *Secondary* | 1.243 | 1.034, 1.493 |
| *Primary* | .900 | .481, 1.686 |
| **Ethnicity**  *Non-Irish* | 1.00 | Reference |
| *Irish* | 1.290 | 1.044, 1.594 |
| **Marital Status**    *Married* | 1.00 | Reference |
| *Not married* | 1.604 | 1.326, 1.941 |
| **Singleton or Non-singleton**  *Singleton* | 1.00 | Reference |
| *Non-Singleton* | 1.872 | 1.308, 2.679 |
| **How many weeks until the baby was born** | .985 | .949, 1.023 |
| **Equivalized annual income** | 1.00 | 1.00 |
|  |  |  |

, after running the quantile regression, when τ = 25, i.e the 25th percentile, the β (or coefficient) for a non-singleton was calculated to be -1.98 after controlling for the variables listed above. At this level for τ = 25, the naming vocabulary scores for non-singleton children is on average -1.98 lower than naming vocabulary scores of singleton children. when τ = 50, i.e the 50th percentile, the β (or coefficient) for a non-singleton was calculated to be -0.231 after controlling for the variables listed above. At this level for τ = 50, the naming vocabulary scores for non-singleton children is on average -0.231 lower than naming vocabulary scores of singleton children. Finally, when τ = 75, i.e the 75th percentile, the β (or coefficient) for a non-singleton was calculated to be -0.72 after controlling for the variables listed above. At this level for τ = 75, the naming vocabulary scores for non-singleton children is on average -0.72 lower than naming vocabulary scores of singleton children.

## Marriage Status of Primary Carer from Wave 1 to Wave 3 Crosstabs

A study by Stephen McKay (McKay, 2010) identified in a study of English babies the financial, martial and mental stress associated with having multiple births vs singletons. One of the findings in this study was that there was evidence to suggest that families with a multiple birth are more likely to separate or divorce than singleton birth families. To investigate if this trend is in the Growing Up in Ireland data a cross tabs was run for wave 1 and wave 3 marital status’ obtaining the below table, Table 18.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  | Wave 3 Status | |  |  |
|  | ***Married and living with husband/wife***  Count n(%) | **Married and separated from husband/wife**  Count n(%) | ***Divorced/ Widowed***  Count n(%) | ***Never Married***  Count n(%) | ***Cramer V***  Count n(%) | ***P-value***  Count n(%) |
| Wave 1 Status |  |  |  |  |  |  |
| Singleton  *Married and living with husband/wife* | 5733 (96.7) | 147 (2.5) | 25 (0.4) | 23 (0.4) | .588 | <.001 |
| *Married and separated from husband/wife* | 31 (22.3) | 65 (46.8) | 29(20.9) | 14(10.1) |  |  |
| *Divorced/*  *Widowed* | 23 (23.5) | 3 (3.1) | 53(54.1) | 19(19.4) |  |  |
| *Never Married* | 628 (27.4) | 14(0.6) | 9(0.4) | 1645(71.6) |  |  |
|  |  |  |  |  |  |  |
| Non-Singleton  *Married and living with husband/wife* | 241(95.3) | 11 (4.3) | 1 (0.4) | 0 | .661 | <.001 |
| *Married and separated from husband/wife* | 0 | 3 (75) | 1 (25) | 0 |  |  |
| *Divorced/*  *Widowed* | 0 | 0 | 4 | 0 |  |  |
| *Never Married* | 12(22.6) | 3 (5.7) | 2 (3.8) | 36 (67.9) |  |  |

Table : Cross Tab of Marital Status from Wave 1 to Wave 3

### Results/Conclusion of Marital Status from Wave 1 to Wave 3 Crosstabs

From Table 18, the percentage of primary carers who were married and living with husband/wife in wave 1 but are now divorced is 2.5% for singleton primary carers vs 4.3% for non-singleton primary carers. This would not suggest any strong evidence that there are higher divorce rates for non-singleton families than singleton families as the percentage increases are so small. The sample size for non-singleton is quite small, so more data would be needed to make statistically significant conclusions of the effects having a non-singleton has on marital status of the primary carer.

## Multiple Regression of Wave 3 Picture Similarities Scores Squared with key predictors non-singleton and singleton

A multiple regression of picture similarities scores from wave 3 data was undertaken to quantify and investigate if non-singletons performed worse in picture similarities scores. A study by Day (1932), found that twins were less advanced in language development at the ages of 2, 3, 4 and 5 years of age in comparison to singletons of the same age and that twins also scored on average 10 points lower on an IQ test than a singleton child. The multiple regression of picture similarities scores with key predicator variable non-singleton with reference category singleton, while controlling for variables; gender of child, employment status of primary carer, education, age of primary carer, equivalent annual income of household, ethnicity, if study child had siblings and marital status of the primary carer, as informed by the literature review. For the model, the shortened factor levels for employment, marital status, ethnicity and education were used. Looking at the histogram of picture similarities scores, Figure 31 and Figure 32, the data was skewed to the right. Following on from this and running a multiple regression on picture scores and looking at the histogram of the residuals, Figure 33, it is clear the residuals are not normally distributed. So, because of the non-normally distributed residuals, the variable will need to be transformed. Taking the square of the picture scores transformed the data to be more normally distributed to then conduct a multiple regression without violating any assumptions of multiple regression.

There are down sides to taking the transformation of variables, as you will lose interpretability of the coefficients, except when taking natural log transform. When a natural log transform is used the coefficients in the regression output can be interpreted as the predicted percentage increase in y (outcome variable) for a 1 unit increase in x (predictor variable), controlling for all other variables in the model. While the interpretability is lost when taking the squared transformation, it is still possible to identify and classify how predictor variables affect the outcome variables. This topic of transformations is very much being tested with machine learning techniques and models. The machine learning methods are trying to fit complex models and data by any means possible to try and classify binary outcomes, with the expense of losing the interpretability of the coefficients of the predictor variables.

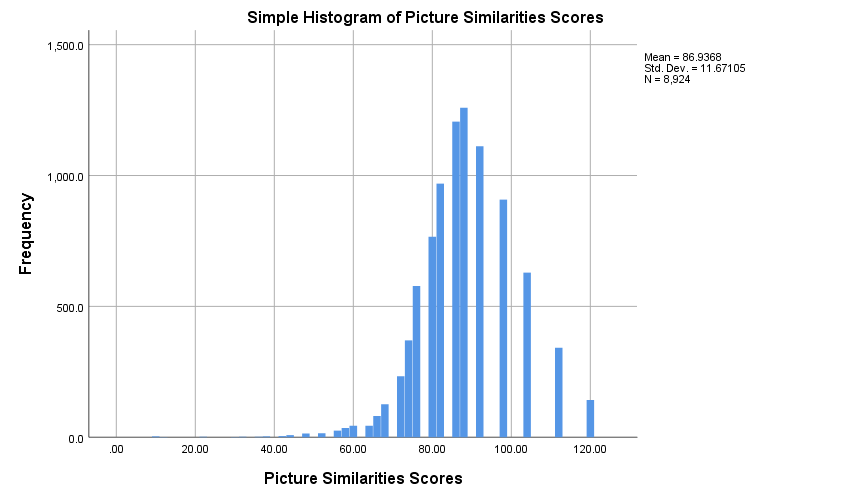
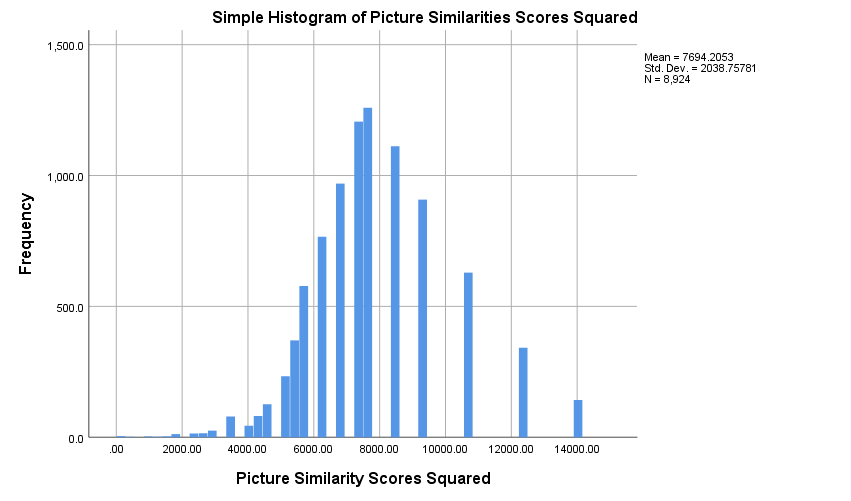
 Figure : Picture Similarities Scores Squared

Figure : Picture Similarities Scores

From the above figures, Figure 31 and Figure 32, we can see that picture similarity scores squared tend to be more normally distributed than just the picture similarity scores. After running a multiple regression on both picture scores and picture scores squared, we get the below; an for picture similarity scores, but the residuals of the picture similarity scores is not normally distributed, Figure 33, thus violating an assumption of multiple linear regression.. For picture similarity scores squared the and its residuals are normally distributed, Figure 34.

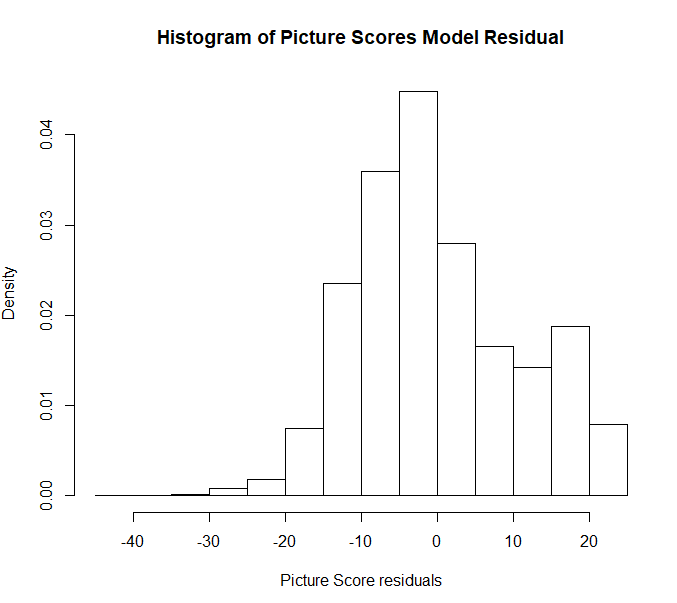
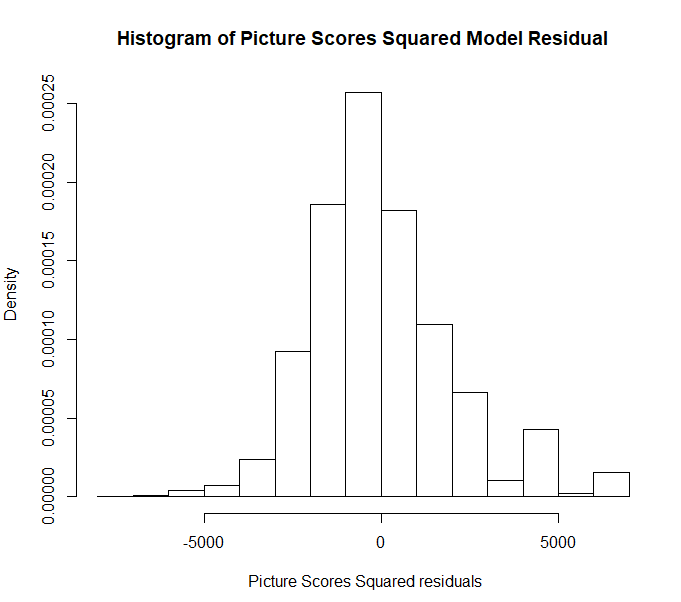


Figure : Residuals of Picture Similarities Scores

Figure : Residuals of Picture Similarities Scores Squared

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | β | Standard Error | P-value |
| Constant | 8027 | 248.1 | <0.001 |
| Age of Primary Carer | -10.0 | 6.891 | .147 |
| Gender of Child  *Female* | 1.00 | 1.00 | Reference |
| *Male* | 264.9 | 55.64 | <0.001 |
| Employment Status of Primary Carer  *Employee* | 1.00 | 1.00 | Reference |
| *Home duties/retired* | -108.2 | 65.91 | 0.101 |
| *Self Employed* | -26.64 | 122.2 | 0.827 |
| *Student* | 260.0 | 323.3 | 0.421 |
| *Unemployed* | -180.8 | 169.4 | 0.286 |
| Highest Level of Education    *Third* | 1.00 | 1.00 | Reference |
| *Secondary* | -181.0 | 60.43 | 0.003 |
| *Primary* | -312.8 | 279.0 | 0.262 |
| Ethnicity  *Non-Irish* | 1.00 | 1.00 | Reference |
| *Irish* | 62.39 | 76.69 | .416 |
| Marital Status    *Married* | 1.00 | 1.00 | Reference |
| *Not married* | -152.4 | 82.61 | 0.06 |
| Child has Sibling  *Yes* | 1.00 | 1.00 | Reference |
| *No* | -64.19 | 57.4 | 0.264 |
| Singleton or Non-singleton  *Singleton* | 1.00 | 1.00 | Reference |
| *Non-Singleton* | -195.1 | 145.8 | 0.181 |
| Equivalized annual income | 5.014e-3 | 2.319e-3 | .031 |

Table : Multiple Regression of Picture Similarities Scores Squared

### Results/Conclusion of Picture Similarity Scores

From Table 19

|  |  |  |
| --- | --- | --- |
| **Variable** | **Odds ratio** | **95% confidence interval** |
| **Age** | .997 | .960, .994 |
| **Gender of Child**  *Female* | 1.00 | Reference |
| *Male* | 1.001 | .855, 1.171 |
| **Employment Status**  *Employee* | 1.00 | Reference |
| *Home duties/retired* | 1.340 | 1.118, 1.605 |
| *Self Employed* | 1.563 | 1.110, 2.200 |
| *Student* | 1.233 | .593, 2.562 |
| *Unemployed* | 1.550 | 1.061, 2.263 |
| **Highest Level of Education**  *Third* | 1.00 | Reference |
| *Secondary* | 1.243 | 1.034, 1.493 |
| *Primary* | .900 | .481, 1.686 |
| **Ethnicity**  *Non-Irish* | 1.00 | Reference |
| *Irish* | 1.290 | 1.044, 1.594 |
| **Marital Status**    *Married* | 1.00 | Reference |
| *Not married* | 1.604 | 1.326, 1.941 |
| **Singleton or Non-singleton**  *Singleton* | 1.00 | Reference |
| *Non-Singleton* | 1.872 | 1.308, 2.679 |
| **How many weeks until the baby was born** | .985 | .949, 1.023 |
| **Equivalized annual income** | 1.00 | 1.00 |
|  |  |  |

, after running the multiple regression, the β (or coefficient) can be interpreted as a one unit increase has an effect size of β on picture scores squared. The β for a non-singleton was calculated to be -195.1 after controlling for the variables listed above. From this model, the picture similarity scores squared for non-singleton children is on average -195.1 lower than picture similarity scores of singleton children. As discussed above, the interpretability is lost when taking a transformation, but it is clear, that on average non-singletons score lower than their singleton counterparts.

# Summary

The main findings of this study are set out below;

|  |
| --- |
| * Primary Carer Depression is higher for non-singletons compared to singletons, when children are 9 months of age.   Odds ratio = 1.87 with 95% CI [1.308, 2.679] |
| * Mean CESD-8 depression scores of primary carers of non-singletons is estimated to be 2.61 points higher than that of singletons, when children are 9 months of age with 95% CI [0.802, 4.472] |
| * Length of pregnancy for a non-singleton is estimated to be on average 3.763 weeks shorter than a singleton pregnancy, with 95% CI [3.522, 4.004] |
| * Primary carer parental stress scores of non-singletons is estimated to be 0.0476 points higher than that of singletons, when the children were aged 5 years old with 95% CI [0.003, 0.092] |
| * Cluster analysis of Strengths and Difficulties (SDQ), as judged by the Primary Carers of the child, suggest non-singleton females may be more likely to be seen as hyperactive than singleton females. The evidence would suggest for the males that there was no significant difference for non-singletons vs singletons across each cluster, when the children were aged 5 |
| * There was no statistically significant evidence to suggest that families with a multiple birth are more likely to separate or divorce than any other families, but more data would be needed to further investigate this |
| * Non-singletons tended to perform worse in academic tests in school as judged by their teachers, with more non-singleton males being in the lowest performing cluster than their singleton counterparts (18% vs 13.7% respectively). The evidence is stronger in the female case, with more non-singleton females being in the lowest performing cluster than their singleton counterparts (13.1% vs 8.2% respectively), when the children were aged 5 years old. |
| * The naming vocabulary scores for non-singleton children is on average 0.761 points lower than naming vocabulary scores of singleton children, when aged 5 years with 95% CI [-3.09, 1.57]   However, since 0 is contained in the confidence interval, this result is not statistically significant |
| * On average non-singletons score lower than their singleton counterparts on picture similarities scores, at the age of 5 |

In the literature review there are several international studies that identified higher depression rates in non-singleton mothers compared to singleton mothers. A study by Hay (Hay, et al., 1990) found 29.7% of the mothers of twins aged 3 months reported depression (5 times higher rate than singleton mothers). The trend of non-singletons primary carers showing higher depression rates is evident in our wave 1 analysis with the mean CESD-8 depression scores of primary carers of non-singletons is estimated to be 2.61 points higher than that of singletons, although not as severe as the study by Hay above. Looking at wave 3 parental stress scores of the primary carer, non-singletons primary carers have a higher stress score than singleton primary carers. Although the severity between non-singleton primary carers and singleton primary carers is no longer as prevalent as wave 1 results. The mean difference for primary carer parental stress for a non-singleton compared to a singleton birth was calculated to be 0.0476, when the children were aged 5.

The shorter length of pregnancies for non-singletons in our wave 1 analysis follows the trend of the international studies discussed in the literature review, which in turn is potentially harmful to the health of the children and mothers. From the literature review twins were estimated to have a stillbirth and neonatal death rate four times higher than singletons with this figure worsening for triplets as they have a death rate six times higher than those of singletons (Wimalasundera , et al., 2003). An explanation for the higher death rates in multiple births than singletons is the preterm birth of multiples which in turn results in lower birth weights in the children (Beatrice, et al., 2002).

A study by Wood (Wood, et al., 1996), found that twins were more likely to develop Attention Deficit Hyperactivity Disorder than singletons. After running a cluster analysis of Strengths and Difficult (SDQ), as judged by the primary carers of the child, of wave 3 data, we found that this trend was only significant in the female case as the results showed that non-singleton females may be more likely to be seen as hyperactive than singleton females. The evidence would suggest for the males there was no significant difference for non-singletons vs singleton, when the children were aged 5.

There was no statistically significant evidence to suggest that families with a multiple birth are more likely to separate or divorce than any other families in the Growing Up in Ireland study, but more data would be needed to further investigate this. This result contrasts the study by Stephen McKay (McKay, 2010) carried out in England which yielded evidence to suggest that families with a multiple birth are more likely to separate or divorce than singleton birth families.

As IQ scores were not directly surveyed in the Growing Up in Ireland study, to get a feel of IQ and cognitive abilities, picture similarities scores, naming vocabulary scores and other academic variables were used to gauge the intellectual and cognitive ability of the children. It was found that non-singletons tended to perform worse in academic tests in school as judged by their teachers, with more non-singleton males being in the lowest performing cluster than their singleton counterparts (18% vs 13.7% respectively). The evidence is stronger in the female case, with more non-singleton females being in the lowest performing cluster than their singleton counterparts (13.1% vs 8.2% respectively). For picture similarities scores non-singletons scored on average lower than their singleton counterparts on picture similarities scores, at the age of 5. These results all follow the trends shown in international studies discussed in the literature review, such as singletons tending to perform better in IQ tests when the children are aged 5 years old. A study by Day (1932), found that twins were less advanced in language development at the ages of 2, 3, 4 and 5 years of age in comparison to singletons of the same age, and that twins also scored on average 10 points lower on an IQ test than a singleton child.

In conclusion, in Ireland non-singletons tend to be disadvantaged in comparison to their singleton counterparts in academic and health developments up until the age of 5 years old. Non-singleton mothers have shorter pregnancies and therefore lower birth weights which puts the non-singleton babies potentially at adverse health risks. These adverse health risks are not just specifically for the children, as non-singleton pregnancies can potentially have significant effects on the primary carers mental health. As outlined in the introduction, Ireland has some government benefits such as increased child benefits payments, but more is needed to combat these problems. Longer parental leave and the setting up of better support networks for primary carers of non-singletons may be potential solutions here in Ireland. Additional classes for non-singletons could be offered in schools in an aid to help non-singletons develop and nurture their academic and cognitive development.

# Bibliography

Albert, P., 2015. Why is depression more prevalent in women?. *Journal of Psychiatry Neuroscience,* pp. 219-221.

Almond, D., K. Y., C. & D. S., L., 2005. The Costs of Low Birth Weights. *The Quartely Journal of Econmics,* pp. 1031-1083.

Beatrice, B. et al., 2002. Impact of the Increasing Number of Multiple Births on the Rates of Preterm Birth and Low Birthweight: An International Study. *American Journal of Public Health,* pp. 1323-1330.

Campbell, D., Teijlingen, E. & Yip, L., 2004. Economic and Social Implications Of Multiple Births. *Best Practice & Research Clinical OBstetrics & Gynaecology,* pp. 657-668.

Cramer, J., 2002. The Origins of Logistic Regression. *Tinbergen Institute Discussion Paper.*

CSO, 2019. [Online]   
Available at: https://statbank.cso.ie/px/pxeirestat/Statire/SelectVarVal/saveselections.asp

CSO, 2019. *Central Statistics Office.* [Online]   
Available at: https://statbank.cso.ie/multiquicktables/quickTables.aspx?id=vsa17

Daniel, Y., Ochshorn, Y. & Fail, G., 2000. Analysis of 104 twin pregnancies conceived with assisted reproductive technologies and 193 spontaneously concieved twin pregnancies. *Fertility and Ster,* pp. 683-689.

Day, E., 1932. The Development Of Langugage in Twins. A Comparison of Twins and Single Children. *Child Development,* pp. 179-199.

Department of Health and Children, 2005. *Request for Tenders (RTF) for Proposals to Undertake a National Longitudinal Study of Children in the Republic Of Ireland,* s.l.: National Children's Office of the Department of Health and Children and the Department of Social and Family Affairs.

DiLalla, L., 2006. Social Development of Twins. *Twin Research and Human Genetics,* p. 102.

Driver, H. & Kroeber, A., 1932. Quantitative Expression of Clutural Relationships. *University of California Publications in AmericanArchaeology and Ethnology,* pp. 211-256.

Epperson, N., 1999. Postpartum major depression: Detection and treatment. *American Family Physician,* pp. 2247-2254.

Fisher , J. & Stocky, A., 2003. Maternal Perinatal Mental Health and Multiple Births: Implications for Practice. *Twin Research,* pp. 506-513.

Goodman, R., 1997. The Strength and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry,* Issue 38, pp. 581-586.

Haworth, C., Davis, O. & Plomin, R., 2013. Twins Early Development Study (TEDS): AGenetically Sensitive Investagation of Cognitive and Behavioral Development From Childhood to Young Adulthood. *Twins registries worldwide: An important resource for scientific research,* pp. 117-125.

Hay, D. et al., 1990. What information should the multiple birth family recieve before, during and after the birth?. *Acta Geneticae Medicae et Gemellologiae,* pp. 259-269.

Hu, H., 2008. Poisson distribtution and application. *Department of Physics and Astronomy, University of Tennesse.*

Johnson, N. & Kotz, S., 1969. *Developments in Discrete Distributions,* s.l.: Department of Statistics, University of North Carolina.

Jones, W. & Berry, J., 1995. The Parental Stress Scale: Initial Psychometric evidence. *Journal of Social and Personal Relationships,* pp. 463-472.

Koenker, R. & Hallock, K. F., 2001. Quantile Regression. *Journal of Economic Perspectives,* 15(4), pp. 143-156.

Kulkarni, A. et al., 2013. Fertility Treatments and Multiple Births in the United States. *The New England Journal of Medicine,* , 369(23), pp. 2218-2225.

Lawless, J. F., 1987. Negative binomial and mixed Poisson regression. *The Canadian Journal Of Statistics,* Volume 15, pp. 209-225.

MacDorman, M. & Atkinson, J., 1999. Infant Mortality Statistics From the 1996 Period. *National Center For Health Statistics.*

McCulloch, K., Smith, P. & Elliott, C., 1997. *British Ability Scales Second Edition (BAS II).* London: TEchnical Manual.

McKay, S., 2010. The Effects of Twins and Multiple Births on Families and their Living Standards. *University of Birmingham.*

Meltzer, H. et al., 2012. Relationship Between Personal Debt and Specific Common Mental Disorders. *European Journal Of Public Health,* pp. 108-113.

Pison, G., Monden, C. & Smits, J., 2019. [Online]   
Available at: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1728-4457.2015.00088.x

Robin, M., Bydlowski, M., Cahen, F. & Josse, D., 1991. Maternal reactions to the birth of triplets. *Acta Geneticae Medicae et Gemellologiae,* pp. 41-51.

Scottish Intercollegiate Guidlines Network, 2002. *Postnatal depression and puerperal psychosis,* Edinburgh: Royal College of Physicians.

Senat, M. & Ancel, P., 1998. How Does Multiple Pregnancy Affect Maternal Mortality and Morbility. *Clinical Obstretrics and Gynecology,* pp. 79-83.

Stigler, S., 1984. Boscovich, Simpson and a 1760 manuscript note on fitting a linear relation. *Biometrika,* Volume 71, pp. 615-620.

Sutcliffe, A. & Derom, C., 2006. Follow-up of twins: Health, behaviour, speech, language outcomes and implicaitions for parents. *Early Human Development,* pp. 379-386.

Szumilas, M., 2010. Explaining Odds Ratio. *Journal of the Canadian Academy of Child and Adolescent Pyschiatry,* pp. 227-229.

Tandberg, A., Bjorge, T., Bordahl, P. & Skjaerven, R., 2007. Increasing Twinning Rates in Norway, 1967-2004: the Influence of Maternal Age and Assisted Reproductive Technology. *Acta Obstetricia et Gynecologica Scandinavica ,* pp. 833-839.

Thorpe, K., Golding, J., MacGillivray, I. & Greenwood, R., 1991. Comparison of prevalence of depression in mothers of twins and mothers of singletons. *British Medical Journal,* pp. 875-878.

Webbink, D. et al., 2008. Do twins have lower cognitive ability than singletons?. *Intelligence,* 36(6), pp. 539-547.

Wimalasundera , R., Trew, G. & Fish, N., 2003. Reducing the incidence of twins and triplets. *Best Practice & Research Clinical Obstetrics & Gynaecol,* pp. 309-329.

Wood, C., Hay, D. & McLaughlin, M., 1996. Twin-Sibling Differences in Parental Reports of ADHD, Speech, Reading and Behaviour Problems. *Child Psychology and Psychiatry,* pp. 569-578.

Yadav, D., Chaudhary, U. & Shrestha, N., 2011. Risk Factors Associated with Low Birth Weight. *Journal of Nepal Health Research Council,* pp. 159-164.

# Appendix

## Data Preparation file

# Data Prep

#load quantreg package #load MASS package

# bringing in all varaibles

attach(GUI\_Data\_9MonthCohort)

# MMa5ap2.f - gender of child

# MMH5a - length of pregnancy

# MMagep1 - age of primary

# NoSiblings - 0 = has siblings

# 1 = first child

NoSiblings.f <- factor(NoSiblings)

#setting factors for multiple regression for length of pregnancy

EmploymentShort.f <- factor(EmploymentShort)

EducationShort.f <- factor(EducationShort)

MMa5ap2.f <- factor(MMa5ap2) #gender child

EthnicityShort.f <- factor(EthnicityShort)

Nonsingleton.f <- factor(Nonsingleton)

MaritalShort.f <- factor(MaritalShort)

MMJ10.f <- factor(MMJ10) #smoke

EducationShort.f <- relevel(EducationShort.f, ref="3") # making third level education ref

EthnicityShort.f <- relevel(EthnicityShort.f, ref="2") # making non irish reference cat

Smoking.f <- relevel(MMJ10.f, ref="3") # making non smoker reference cat

# Converting categorical data with factor

MMa5ap1\_factor <- factor(MMa5ap1) # gender

MML10\_factor <- factor(MML10) # employment

WorkingShort\_factor <- factor(WorkingShort) # employment in shorter factors

Partner\_factor <- factor(Partner) # partner in household

MML34\_factor <- factor(MML34) #education

EducationShort\_factor <- factor(EducationShort) #education in short factors

MS12\_factor <- factor(MS12) # martial status

DepressedPCG\_factor <- factor(DepressedPCG) #Depressed or not

Nonsingleton\_factor <- factor(Nonsingleton) # singleton

FF26\_factor <- factor(FF26) # ethnicity

# converting numeric

MMagep1\_numeric <- as.numeric(MMagep1) #age

Equivinc\_numeric <- as.numeric(Equivinc) # income

## Multiple Regression for Length of Pregnancy code in Wave 1 in R

Regression <- lm ( MMH5a ~ MMa5ap2.f + EmploymentShort.f + Smoking.f + EducationShort.f + MMagep1 + Equivinc + EthnicityShort.f + Nonsingleton.f + MaritalShort.f )

summary(Regression)

confint(Regression) # getting 95% CI

plot(Regression)

ModelResiduals <- Regression$residuals

hist(Modelresiduals, prob=TRUE)

#add normal curve

curve(dnorm, add=TRUE)

## Poisson Regression for Depression Scores in Wave 1 code in R

Pos\_Regression <- glm ( CES\_TOT\_PCG ~ MMa5ap2.f + EmploymentShort.f + NoSiblings.f + EducationShort.f + MMagep1 + MMH5a + Equivinc + EthnicityShort.f + Nonsingleton.f + MaritalShort.f , family= poisson )

summary(Pos\_Regression)

## Negative Binomial Regression for Depression Scores in Wave 1 code in R

negBin\_Regression <- glm.nb ( CES\_TOT\_PCG ~ MMa5ap2.f + EmploymentShort.f + NoSiblings.f + EducationShort.f + MMagep1 + MMH5a + Equivinc + EthnicityShort.f + Nonsingleton.f + MaritalShort.f )

summary(negBin\_Regression)

confint(negBin\_Regression) # Getting 95% CI

## Quantile Regression for Naming Vocabulary Scores in Wave 3 code in R

# load package quantreg

Y = b3\_nvabscore

qr25 <- rq( Y ~ MMa5ap2.f + EmploymentShort.f + EducationShort.f + MMagep1 + Equivinc + Nonsingleton.f + MaritalShort.f + NoSiblings.f , data=fyp\_Merged1and3\_GUI\_Data\_9MonthCohort, method = "br" , tau = 0.25, model=TRUE)

sumqr25 = summary(qr25)

sumqr25

qr50<- rq( Y ~ MMa5ap2.f + EmploymentShort.f + EducationShort.f + MMagep1 + Equivinc + Nonsingleton.f + MaritalShort.f + NoSiblings.f , data=fyp\_Merged1and3\_GUI\_Data\_9MonthCohort, method = "br", tau = 0.5, model=TRUE)

sumqr50 = summary(qr50)

sumqr50

qr75<- rq( Y ~ MMa5ap2.f + EmploymentShort.f + EducationShort.f + MMagep1 + Equivinc + Nonsingleton.f + MaritalShort.f + NoSiblings.f , data=fyp\_Merged1and3\_GUI\_Data\_9MonthCohort, method = "br", tau = 0.75, model=TRUE)

sumqr75 = summary(qr75)

sumqr75

qrseq <- rq( Y ~ MMa5ap2.f + EmploymentShort.f + EducationShort.f + MMagep1 + Equivinc + Nonsingleton.f + MaritalShort.f + NoSiblings.f, data=fyp\_Merged1and3\_GUI\_Data\_9MonthCohort, seq(0.05, 0.85, by =0.05))

sumqr\_seq = summary(qrseq)

plot(sumqr\_seq)

## Negative Binomial Regression of Parental Stress Scores in Wave 3 code in R

#neg binomial of stress score in wave 3 data

Wave3\_negBin\_Regression <- glm.nb ( bpc3\_stress ~ MMa5ap2.f + EmploymentShort.f + NoSiblings.f + EducationShort.f + MMagep1 + MMH5a + Equivinc + EthnicityShort.f + Nonsingleton.f + MaritalShort.f )

summary(Wave3\_negBin\_Regression)

confint(Wave3\_negBin\_Regression)

## Multiple Regression of Naming Vocabulary Scores in Wave 3 code in R

NameVoc\_Score\_Regression <- lm ( b3\_nvabscore ~ MMa5ap2.f + EmploymentShort.f + Smoking.f + EducationShort.f + MMagep1 + Equivinc + EthnicityShort.f + Nonsingleton.f + MaritalShort.f + NoSiblings.f )

summary(NameVoc\_Score\_Regression)

plot(NameVoc\_Score\_Regression)

NameVoc\_Score\_ModelResiduals <- NameVoc\_Score\_Regression$residuals

hist(NameVoc\_Score\_ModelResiduals, prob=TRUE)

## Multiple Regression for Picture Scores Squared in Wave 3 code in R

Pic\_Score\_Regression <- lm ( Square\_PictureScores ~ MMa5ap2.f + EmploymentShort.f + Smoking.f + EducationShort.f + MMagep1 + Equivinc + EthnicityShort.f + Nonsingleton.f + MaritalShort.f + NoSiblings.f )

summary(Pic\_Score\_Regression)

confint(Pic\_Score\_Regression) # getting 95% CI

plot(Pic\_Score\_Regression)

Pic\_Score\_ModelResiduals <- Pic\_Score\_Regression$residuals

hist(Pic\_Score\_ModelResiduals, prob=TRUE, xlab = " Picture Scores Squared residuals", main="Histogram of Picture Scores Squared Model Residual")

## Further graphs and boxplots for Wave 3 descriptive statistics

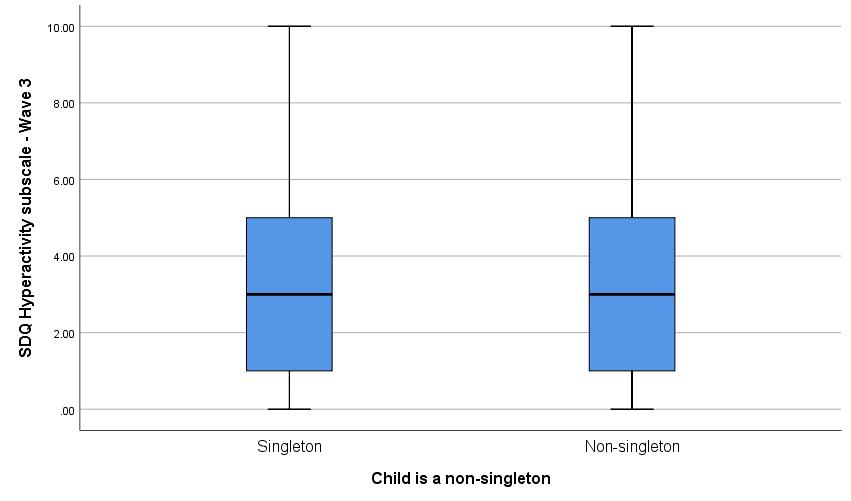


Figure : Boxplot of SDQ Hyperactivity Controlling for Non-singleton

From Figure 35Figure 15: Age of Primary Carer for Non-singleton

, the SDQ hyperactivity subscale scores have the same median and interquartile range as singletons. The median SDQ hyperactivity subscale scores for singletons and non-singletons is 1 with an interquartile range of 2. Histograms of the SDQ hyperactivity for non-singletons and singletons can be seen below in Figure 36 and Figure 37.

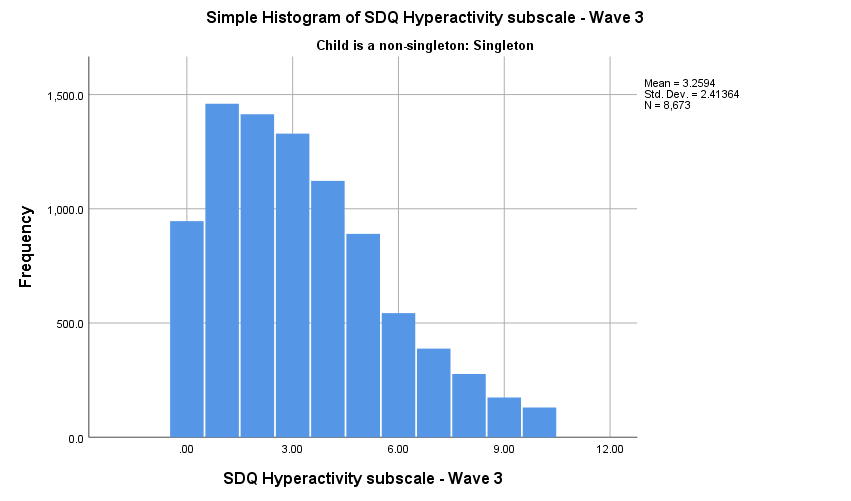
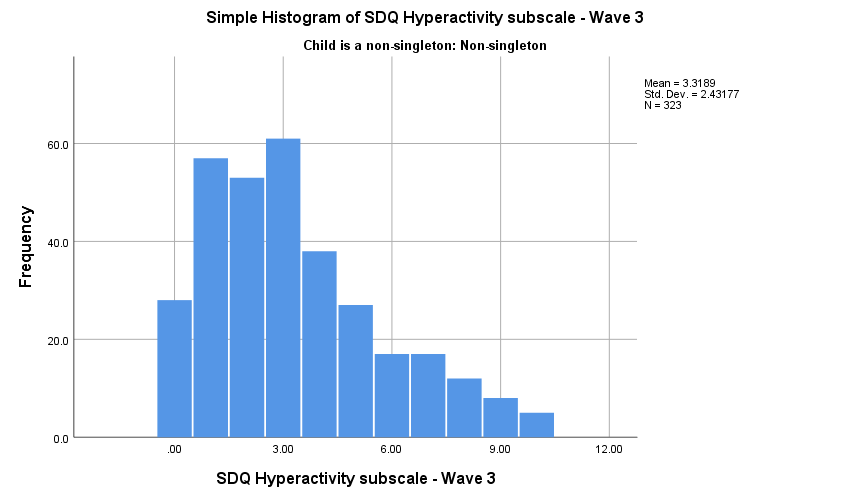


Figure : SDQ Hyperactivity Scores for Singleton

Figure : SDQ Hyperactivity Scores for Non-singleton

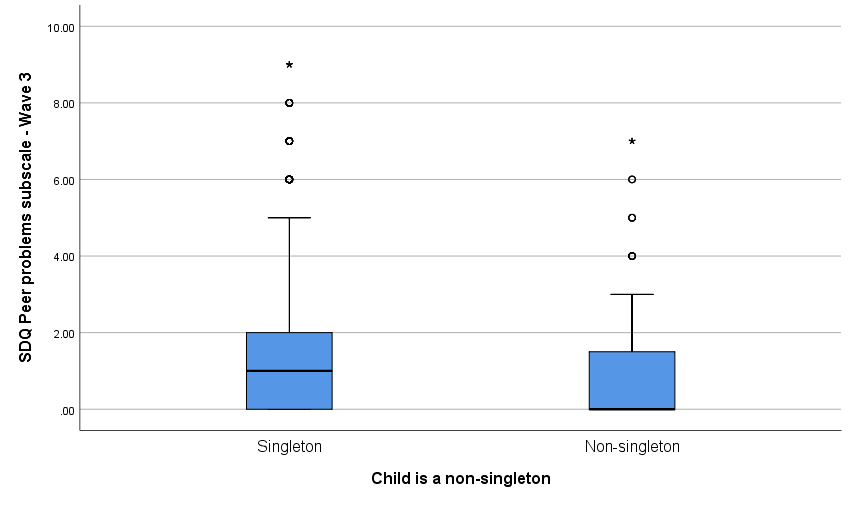


Figure : SDQ Peer Problems Score Controlling for Non-Singletons

From Figure 38Figure 15: Age of Primary Carer for Non-singleton

, the SDQ peer problems subscale scores, singletons have a higher median and interquartile range than non-singletons, the higher interquartile range for singletons can be seen by the longer blue box for singletons compared to non-singletons. The median SDQ peer problems subscale scores for singletons is 1 with an interquartile range of 2 while the median for non-singleton SDQ peer problems subscale scores is 0 with an interquartile range of 2. Indicating singletons have more peer problems than their non-singletons counterparts. Histograms of the SDQ hyperactivity for non-singletons and singletons can be seen below in Figure 39 and Figure 40.

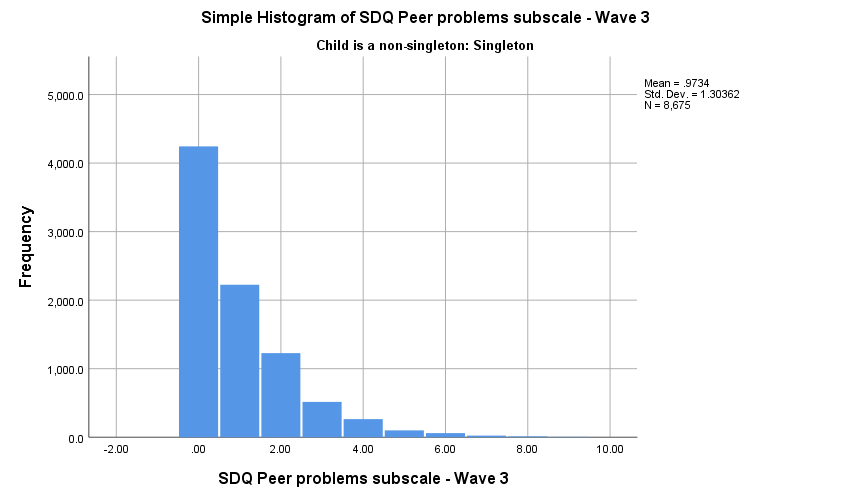
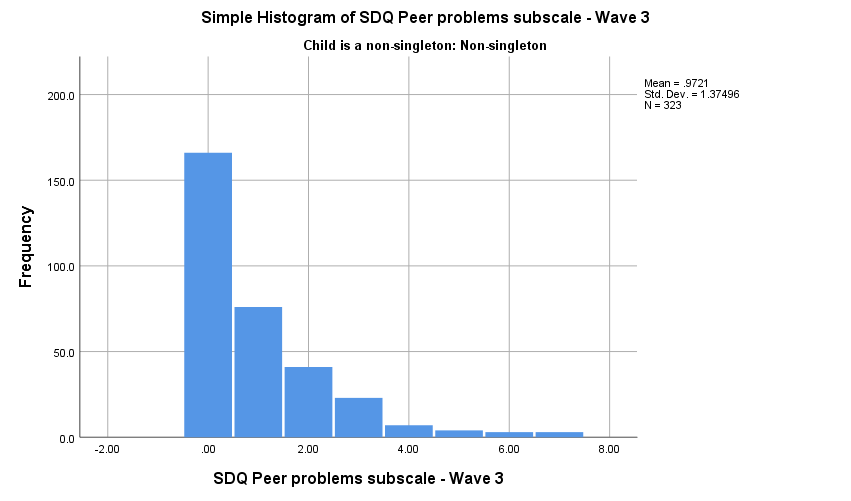


Figure : SDQ Peer Problem scores for Singleton

Figure : SDQ Peer PRoblem Scores for Non-singleton

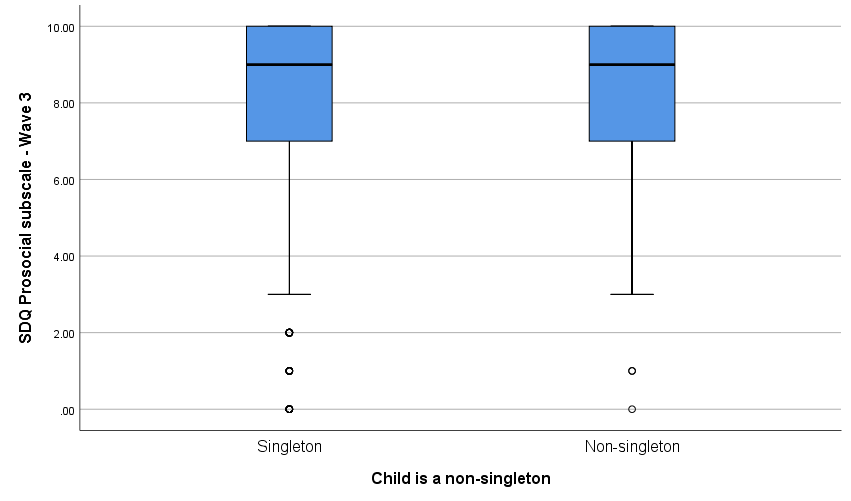


Figure : SDQ Prosocial Scores Controlling for Non-singleton

From Figure 41Figure 15: Age of Primary Carer for Non-singleton

, the SDQ prosocial subscale scores have the same median and interquartile range as singletons. The median SDQ prosocial subscale scores for singletons and non-singletons is 9 with an interquartile range of 3. Histograms of the SDQ prosocial for non-singletons and singletons can be seen below in Figure 42 and Figure 43.

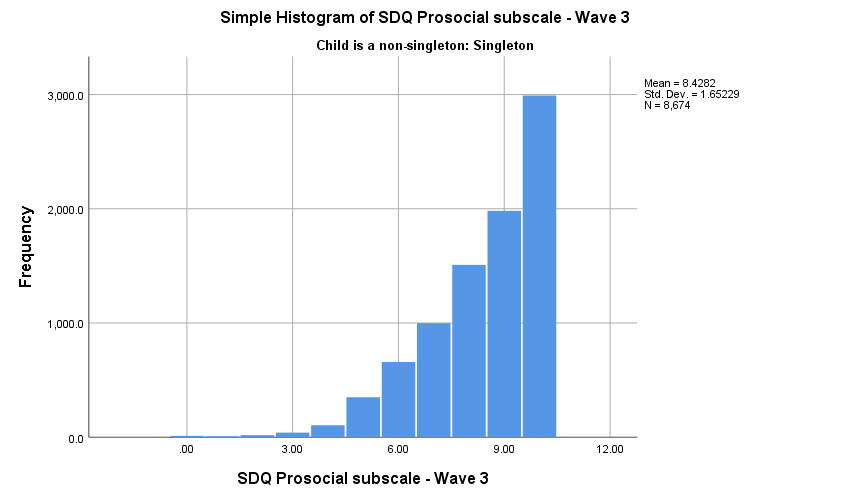
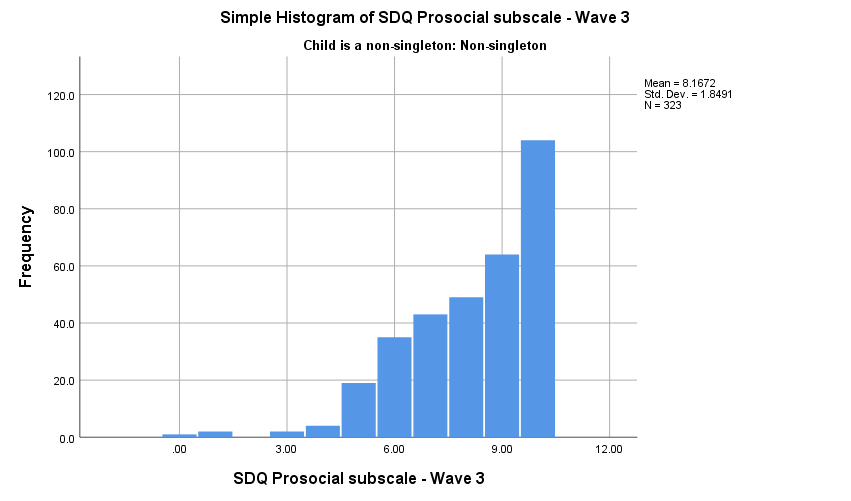


Figure : SDQ Prosocial Scores For Singleton

Figure : SDQ Prosocial Scores For Non-Singleton

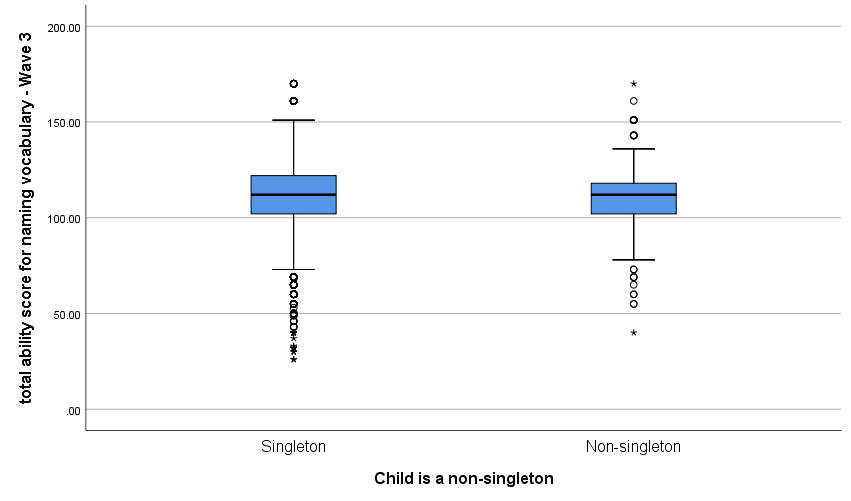


Figure : Naming Vocabulary Scores Controlling for Non-Singleton

From Figure 44Figure 15: Age of Primary Carer for Non-singleton

, the naming vocabulary scores have the same median, but singletons have a slightly higher interquartile range than non-singletons, which can be seen by the longer length of the blue box for singleton. The median naming vocabulary scores for singletons is 120 with an interquartile range of 20 while the median for non-singleton naming vocabulary scores is 120 with an interquartile range of 16. Histograms of the naming vocabulary scores for non-singletons and singletons can be seen below in Figure 45 and Figure 46.

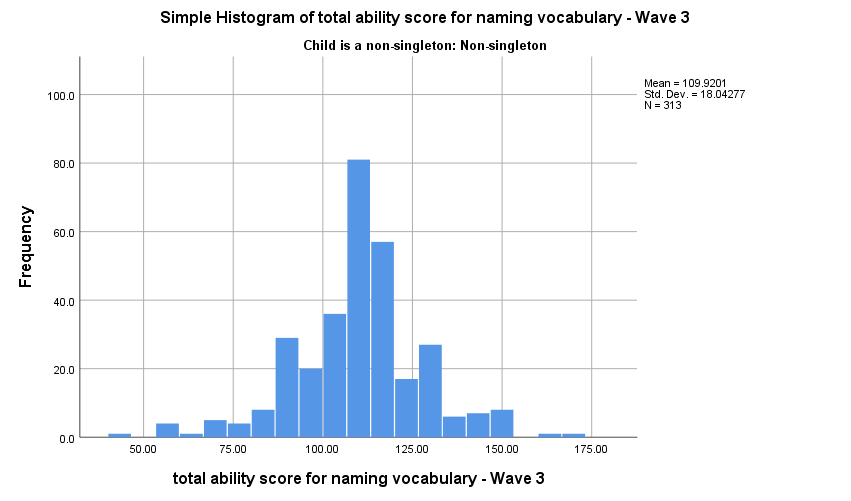
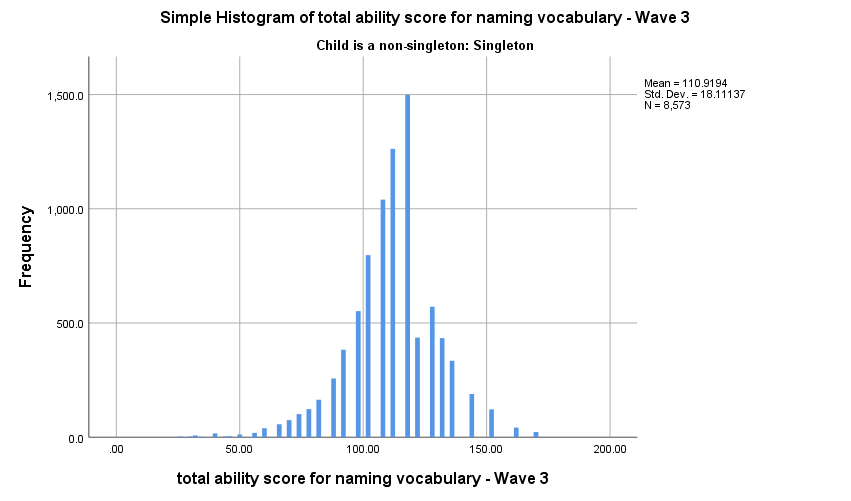


Figure : Naming Vocabulary Scores for Singleton

Figure : Naming Vocabulary Scores for Non-Singleton

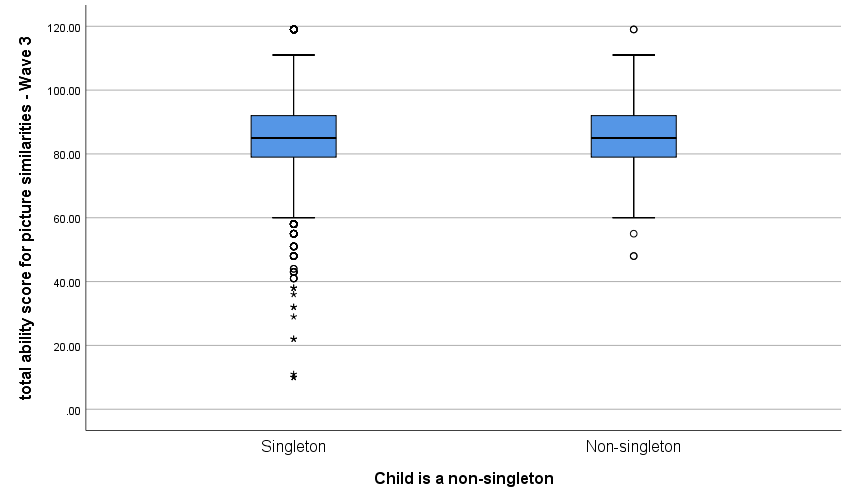


Figure : Picture Similarities Scores Controlling For Non-singleton

From Figure 47Figure 15: Age of Primary Carer for Non-singleton

, the picture similarities scores have the same median and interquartile range as singletons. However, singletons have a few more outliers in their data than non-singletons. The median picture similarities scores for singletons and non-singletons is 85 with an interquartile range of 13. Histograms of the picture similarities scores for non-singletons and singletons can be seen below in Figure 48 and Figure 49.

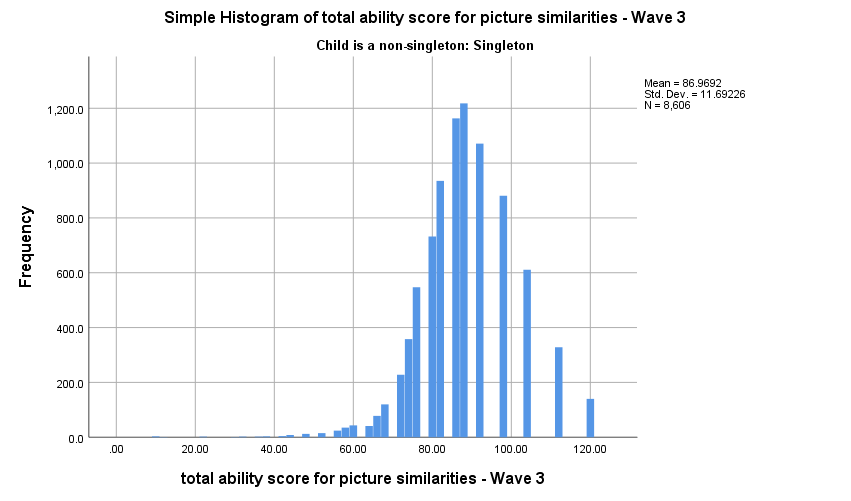
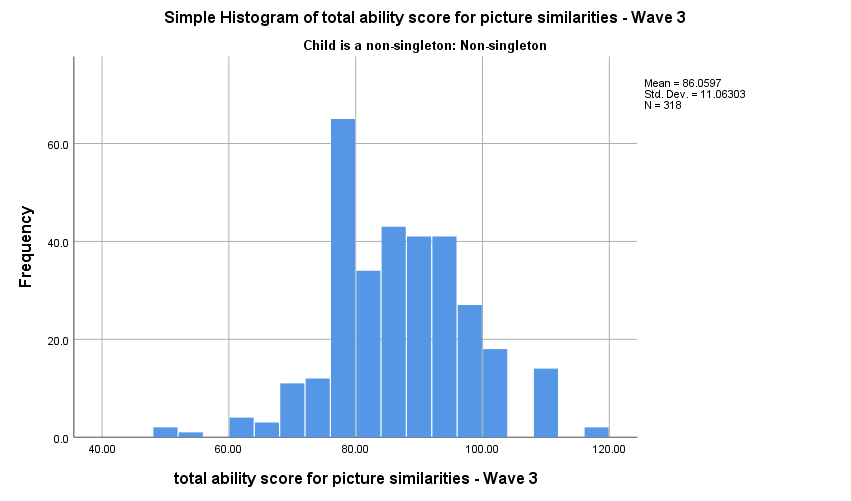


Figure : Picture Similarities Scores for Non-singleton

Figure : Picture Similarities Scores For Singleton

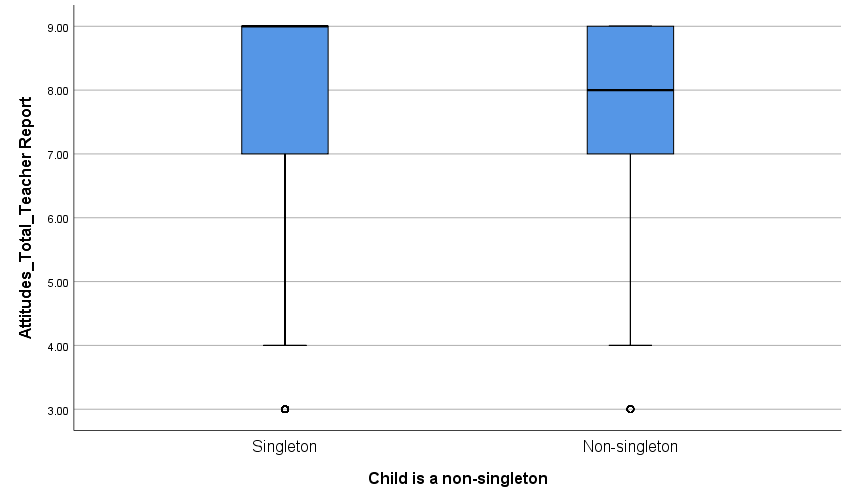


Figure : Attitudes Scores Controlling for Non-singleton

From Figure 50Figure 15: Age of Primary Carer for Non-singleton

, the attitude scores from teacher reports have the same interquartile range, but singletons have a slightly higher median than non-singletons. The median attitude scores for singletons is 9 with an interquartile range of 2 while the median for non-singleton attitudes scores is 8 with an interquartile range of 2. Histograms of the attitude scores from teacher reports for non-singletons and singletons can be seen below in Figure 51 and Figure 52.

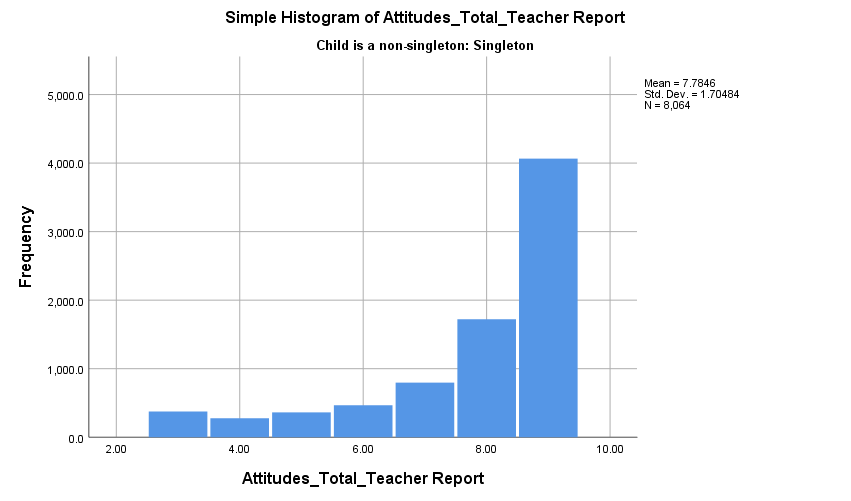


Figure : Attitude Scores for Singleton

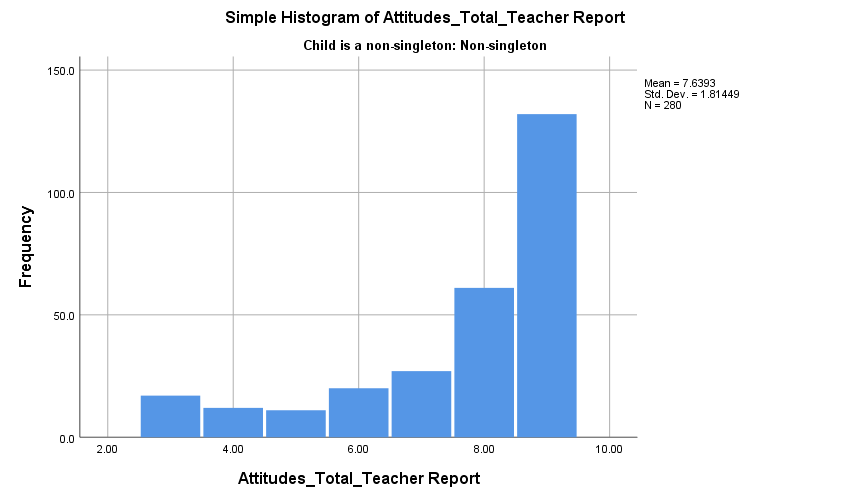


Figure : Attitude Scores for Non-singleton

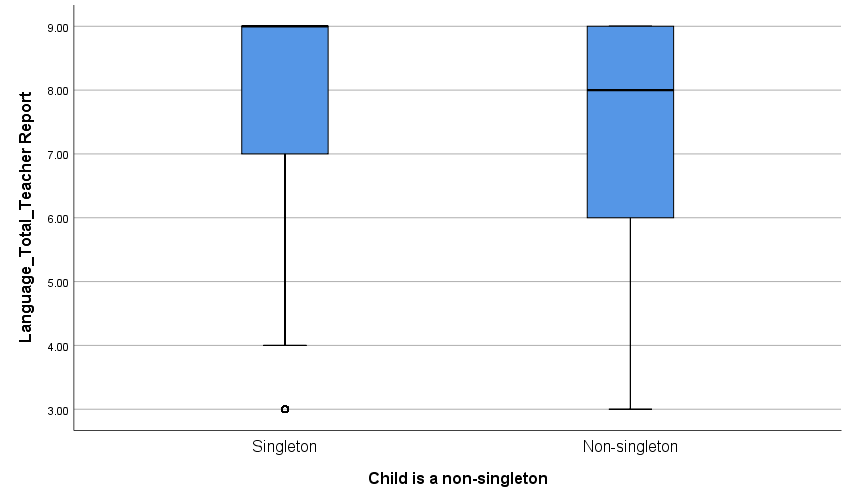


Figure : Language Scores Controlling for Non-singletons

From Figure 53Figure 15: Age of Primary Carer for Non-singleton

, the language scores from teacher reports, singletons have a higher median and a smaller interquartile range than non-singletons, the higher interquartile range for non-singletons can be seen by the longer blue box for non-singletons compared to singletons. The median language scores for singletons is 9 with an interquartile range of 2 while the median for non-singleton language scores is 8 with an interquartile range of 3. Indicating singletons perform on average better in the language scores than their non-singletons counterparts. Histograms of the language scores from teacher reports for non-singletons and singletons can be seen below in Figure 54 and Figure 55.

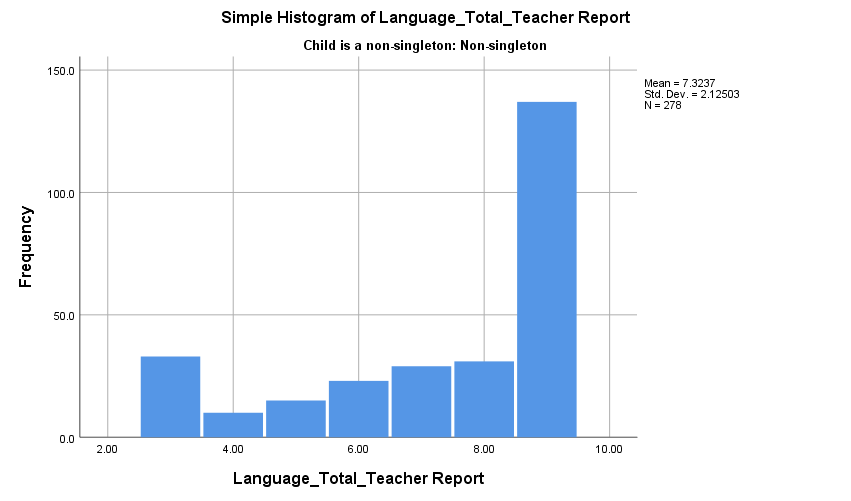
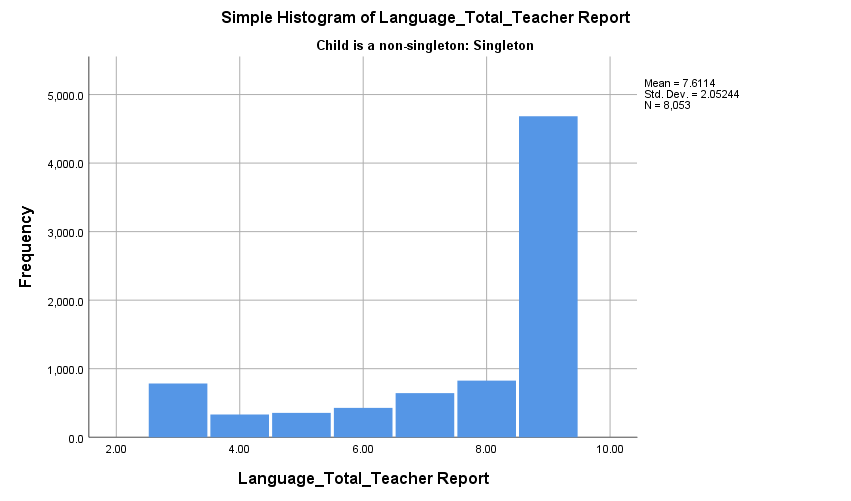


Figure : Language Scores for Singleton

Figure : Language Scores for Non-singleton

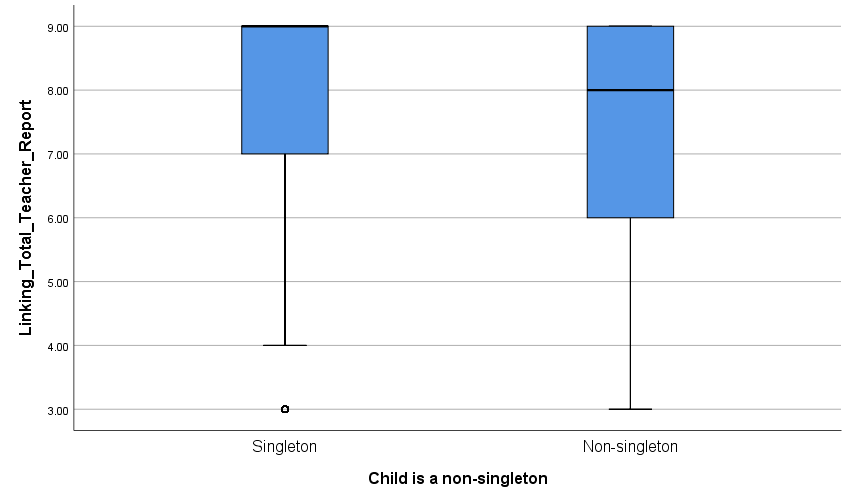


Figure : Linking Scores Controlled for Non-singleton

From Figure 56Figure 15: Age of Primary Carer for Non-singleton

, the linking scores from teacher reports, singletons have a higher median and a smaller interquartile range than non-singletons. The higher interquartile range for non-singletons can be seen by the longer blue box for non-singletons compared to singletons. The median linking scores for singletons is 9 with an interquartile range of 2 while the median for non-singleton linking scores is 8 with an interquartile range of 3, indicating singletons perform on average better in the linking scores than their non-singletons counterparts. Histograms of the linking scores from teacher reports for non-singletons and singletons can be seen below in Figure 57 and Figure 58.

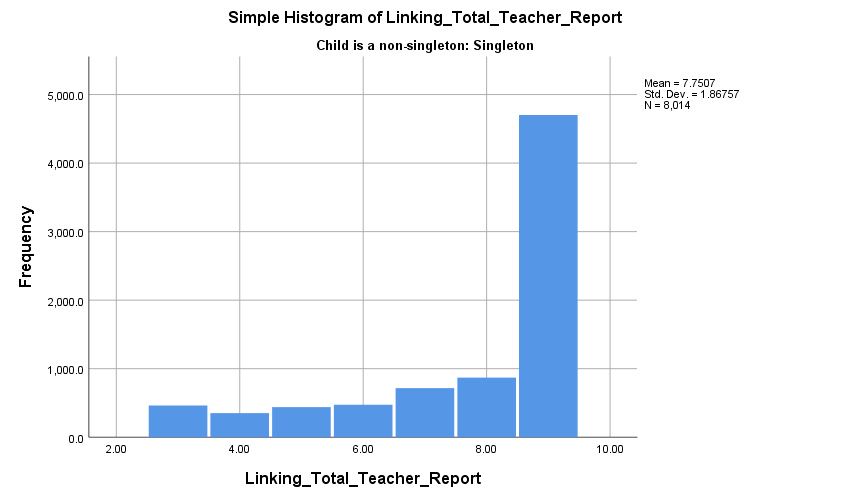
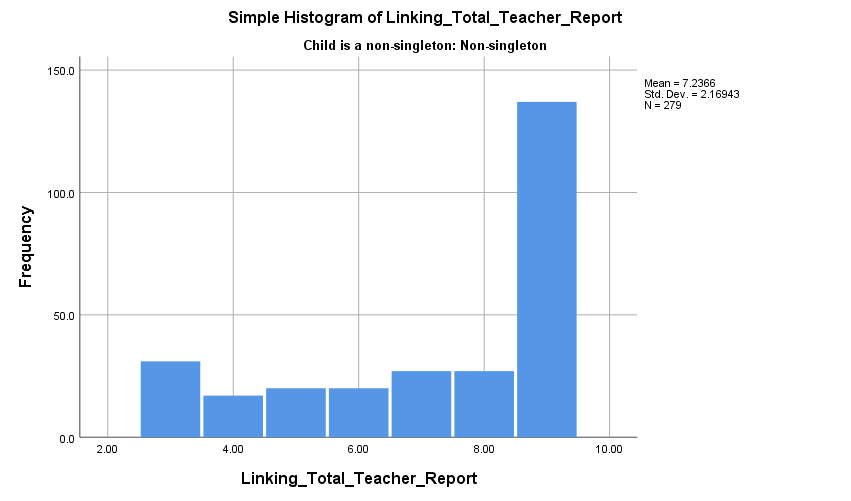


Figure : Linking Scores for Singleton

Figure : Linking Scores for Non-singleton

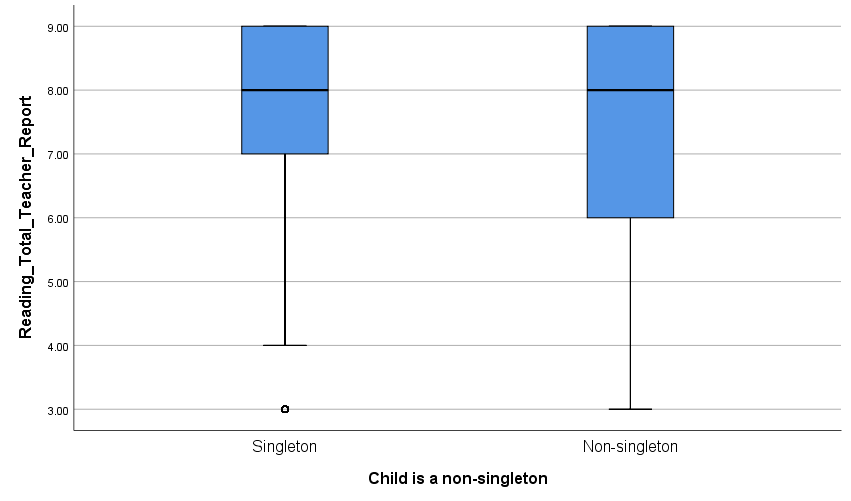


Figure : Reading Scores Controlling for Non-singleton

From Figure 59Figure 15: Age of Primary Carer for Non-singleton

, the reading scores from teachers report, both singletons and non-singletons have the same median, but non-singletons have a slightly higher interquartile range than singletons, which can be seen by the longer blue box for non-singleton. The median reading scores for singletons is 8 with an interquartile range of 2 while the median for non-singleton reading scores is 8 with an interquartile range of 3. Histograms of the reading scores from teacher reports for non-singletons and singletons can be seen below in Figure 60 and Figure 61.

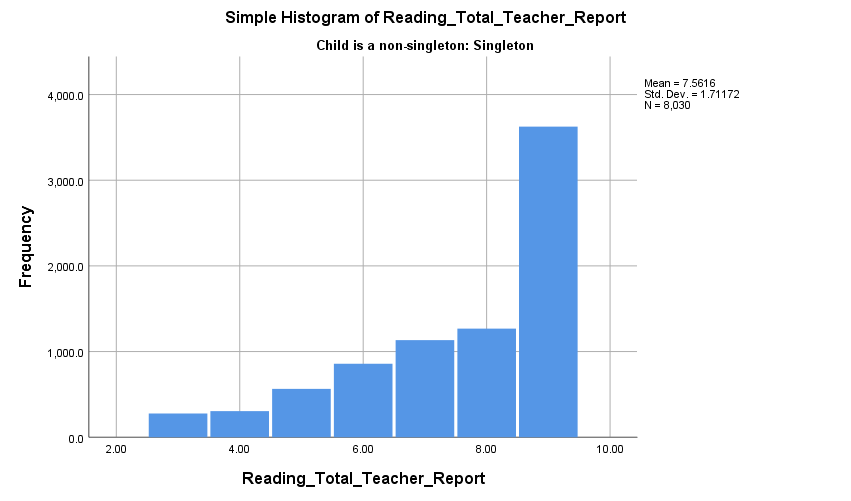
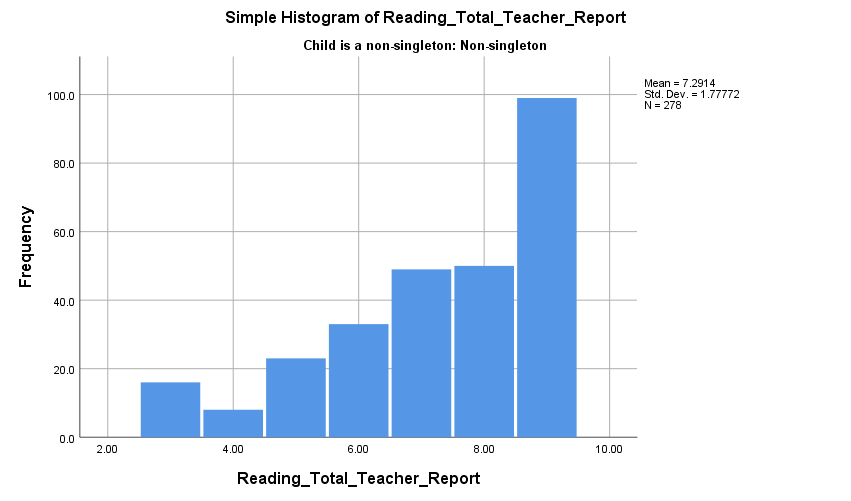


Figure : Reading Scores for Singleton

Figure : Reading Scores for Non-Singleton