ECO3021S Project

Determinants of the Amount of Time Parents Spend on Childcare and how the COVID-19 Pandemic affected them.

Group Number:

60

Date of submission: 4th October 2022

Project members:

Surname	First name	Student number	EMPLID
Jameson	Liam	JMSLIA002	1697805
Jiao	Liren	JXXLIR001	1650850
Walbeck	Paul Joseph	WLBPAU001	1743131

1.   PLAGIARISM DECLARATION

1.   I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is my own.

2.   I have used the APA/HARVARD referencing guide for citation and referencing. Each contribution to, and quotations in this essay/report/project from the work(s) of other people has been acknowledged and has been cited and referenced.

3.   This essay/report/project is my own work.

4.   I have not allowed and will not allow anyone to copy my work.

Abstract

This research paper investigates the determinants of hours spent on childcare during wave five of the Covid-19 pandemic using the National Income Dynamics Study – Coronavirus Rapid Mobile Survey (NIDS-CRAM) - a cross-sectional data set. The main subject matter under scrutiny is that of the difference in gender, and whether this difference has a significant effect on the number of hours spent on childcare. This outcome is inspected in much the same way based on further factors such as education level, age, race, geographical location, number of children, access to early childhood development (ECD) and access to the Covid-19 Temporary Employee/Employer Relief Scheme (TERS). Policy recommendations including increased ECD, and specialized social grants are provided with the objective of decreasing the impact differences in race and gender have on hours spent on childcare. This is because the regression analysis showed these two variables to be the most statistically and practically significant.

Contents

1. Introduction

The childcare crisis has had detrimental effects across the world for decades. Apart from leaving numerous children without the adequate support they need to develop and the severely damaging their futures, the undue burden of childcare has negatively and adversely impacted multiple generations of women. Before the emergence of the Covid-19 pandemic, it had been estimated that over 35 million children below the age of five are sometimes left without any adult supervision (Gromada, Richardson & Rees, 2020). Coupled with this, it was found that women spend on average nearly five weeks more on childcare than their male counterparts each year in South Africa (Samman et al., 2021). Given the combination of the problematic pre-existing global state of childcare, paired with the onset of the Covid-19 pandemic – causing the closure of all schools in over 90% of the world's countries (Gates Foundation, 2021) - there is great value in analysing the consequences this has had on the amount of time parents, and predominantly mothers spend on childcare.

This research paper will primarily examine the significance that differences in gender bear on the number of hours spent on childcare in South Africa. Furthermore, this significance of this outcome will be investigated with respect to additional variables such as education level, age, race, geographical location, number of children, access to early childhood development (ECD) and access to the Covid-19 Temporary Employee/Employer Relief Scheme. The prior determinants of the number of hours spent on childcare (pre-Covid19) will be analysed and compared to research already conducted during the current Covid-19 impacted climate. There is currently a lack of research in this topic pertaining to the South African post-Covid context and as thus, by specifically analysing the effect of gender on the number of hours spent on childcare within the period of Wave 5 in South Africa, this paper will contribute to the gap relating to the effects of gender on childcare.

After outlining the methodology and empirical model specification of our analysis, a description of the data set used to make statistical inferences will be provided together with a preliminary data analysis. This is done to gauge an initial understanding of the relationships between key explanatory variables, most notably gender, and hours spent on childcare. A regression analysis will then be performed comparing the results of the simple and multiple regression models to the research outlined in the literature review. This will aid in determining consistency across the literature. Finally, conclusions will be drawn, and policy recommendations offered with the objective of improving the state of South Africa's childcare crisis post-Covid-19.

2. Literature Review

Globally, the childcare crisis is a significant problem that has persisted for several years. In fact, the research shows that across 53 developing countries, some 35.5 million children below the age of five were without adult supervision for at least an hour in a given week (Samman et al., 2021). In studies investigating the determinants of hours spent on childcare, it has been found that mothers spend roughly twice as much time in childcare than fathers do (Woolard, 2002) and even in well-educated households, care activities are still highly gendered as mothers supply a greater allocation of time to physical childcare (Gimenez-Nadal & Molina, 2012). Moreover, of parents raising two children, fathers only contribute around 22% on average to childcare, while mothers tend to contribute the remaining 78% (Zick & Bryant, 1996). A further study revealed that men with a bachelor's degree would spend 67 minutes alone performing child- care each day - much less than the three and a half daily hours an equivalent woman with a bachelor's degree who would spend (Craig, 2007). In Kenya, over 84% of women living in rural areas and over 80% of women living in urban areas report facing a trade-off between childcare and their livelihood ambitions (Samman et al., 2021). As reflected in the research, gender is the most prominent factor impacting hours spent on childcare.

Early childhood development encompasses a range of activities from school-based nursery education to community-based playgroups, to household childcare (Penn, 2004). It is thus no surprise that parents with access to ECD for their children positively affect their hours spent on childcare. (Spaull et al., 2021) have shown evidence that levels of attendance at early childhood development (ECD) programs had recovered remarkably by April/May 2021, edging towards pre-pandemic attendance levels. The survey revealed that of respondents living with children aged 0-6, 36% reported that at least one child aged 0-6 was attending an ECD program regularly. Thus, ECD is an important factor affecting hours spent on childcare.

According to a South African based study by (Budlender, Chobokoane & Mpetsheni, 2001), it has been found that white women spent 165 minutes per day, on average, on work in establishments, compared with the average of 138 minutes per day spent by Indian women, 123 minutes by coloured women and 63 minutes by Black/African women. This is indicative of the fact that Black/African women tend to have more available time to spend on caring for their children. This provides some insight into the impact that differences in race can have on hours spent on childcare. The ethnic diversity that South Africa offers, in combination with its history of racial prejudice with consequences playing out in real time, the explanatory variable - race - has the capacity to provide useful results in the context of hours spent on childcare in South Africa. Furthermore, the same study showed that both men and women tend to spend more time caring for others depending on their age, with participants over the age of 40

devoting far greater time than those aged 18-39. This identifies a further potential explanatory variable capable of positively impacting the hours parents spend on childcare.

Men with less than a high school education spend 48 minutes per day on childcare compared to the 1 hour 24 minutes men who have graduated college spend. On the other hand, women with less than a high school education spend 2 hours 24 minutes each day on childcare ascending to 3.7 for college graduates (England & Srivastava. 2013). Additionally, it has been found that highly educated mothers devote more time to childcare activities than their low-educated counterparts, in both the UK and Spain, (Gutièrrez-Domènech, 2010). These results indicate that both mothers and fathers with university qualifications contribute more time to childcare activities than non-qualified parents – evidence of the effect that education level has on hours of childcare, keeping differences in gender aside.

This research paper differs to that which has already been done in that it specifically focuses in on the factors that affect the time spent on childcare. This is unlike previous research conducted on the topic because it is being performed within a strictly South African based data framework, and it is being performed at a time in which the effects of the Covid-19 pandemic are still being felt nationally. Moreover, the determinants affecting childcare are specifically being targeted in this analysis, in a focused and intentional way. This paper will thus be contributing to the general research on the topic by providing an analysis of all relevant variables pertaining to childcare such as gender, education level, age, race, geographical location, number of children, access to early childhood development (ECD) and access to the Covid-19 Temporary Employee/Employer Relief Scheme (TERS).

3. Methodology and Empirical Model Specification

3.1. Methodology

This paper examines the determinants of the amount of time that is spent by parents, in South Africa, on childcare and how this was affected by the COVID-19 pandemic. To do this a multiple regression analysis is used, with the time spent on childcare being the dependent variable and the 10 independent variables being wage, age, gender, schooling, race, marital status, geographical status, early childhood development (ECD), Temporary Employee/Employer Relief Scheme (TERS) and the number of children in a family. The independent variables that will mainly be focused on are the levels of education, marital status and gender.

The data has been cleaned from the raw NIDS-CRAM survey data to remove any outliers that would skew our analysis. The regression analysis itself was conducted in Stata, with most of the variables being quantitative in nature but a few being transformed into binary variables.

This methodology is used based on other research papers which study similar variables and their effects on childcare. This use of multiple linear regression is useful as there are numerous variables that have previously been shown to affect the amount of time spent on childcare.

Based off literature that has been reviewed, there is evidence to suggest a significant positive relationship between education levels and childcare hours. Hence, it can be hypothesized that a similar trend would be observed within our dataset.
Furthermore, based off prior evidence, we might expect a somewhat significant difference in childcare between genders.

The relationships between the outcome variable and the explanatory variables are thought to be causal and not correlated, however the regressions that are used in this paper do not control for correlations. The data is nationally significant however it does all come from the same country and therefore could have some level of correlation from external sources from the observations that are available.

3.2.   Empirical model specification

$$\text{Hours Spent in Child Care} = \beta_0 + \beta_1(ParentsYearsSpentinSchool) + \beta_2(Age) + \beta_3(Gender) + \beta_4(Race)$$
$$+\beta_5(Location) + \beta_6(ECD) + \beta_7(TERS) + \beta_8(NumberofChildren) + u$$

In the model the variable TERS, represents the Temporary Employee/Employer Relief Scheme, while the ECD variable represents the Early Childhood Development Scheme. Both variables are set as binary variables as to whether a respondent to the survey received the grant or their children went to ECD classes.

The underlying assumptions to the multiple regression model are the 5 assumptions of multiple linear regression. Firstly, that the parameters of the model are linear, meaning that the coefficients of our independent variables are linear. Secondly, that the sample is taken randomly and is representative of the population. Thirdly, there must be no perfect collinearity, meaning that there must be no perfect linear relationships between the independent variables and none of the independent variables are constant. that the parameters of the model are linear, meaning that the coefficients of our independent variables are linear. Furthermore, the zero conditional mean assumption assumes that the factors in u, our unobserved term or residuals, are uncorrelated with our explanatory variables. Finally, homoscedasticity must be shown, which means that the explanatory variables must not be related to the variance of the error shown by our unobserved factor, u.

4. Data Description, Preliminary Data Analysis and Descriptive Statistics

4.1. Data Description

Our data comes from the National Income Dynamics Study – Coronavirus Rapid Mobile Survey (NIDS-CRAM) data set. This data set is a panel data set, that ran with over 28000 participants in 7300 families (NIDS-CRAM, 2022), that is nationally representative and was a survey gathered by the government every few months. The survey asked participants questions about their income and employment, their household welfare, receipt of grants, and about their knowledge and behaviour related to COVID-19. (NIDS-CRAM, 2022) For these reasons it is the perfect dataset to provide insight into the changes of childcare during COVID-19. From this dataset we will be extracting the variables of wage, age, gender, schooling, race, marital status, geographical status, early childhood development (ECD), Temporary Employee/Employer Relief Scheme (TERS) and the number of children in a family. Our dependent variable that we will be analysing is the amount of time parents spent on childcare, with education levels, gender and race being our main independent variables.

4.2. Preliminary Data Analysis

Figure 1: Graph showing Distribution of Hours Spent on Childcare in the Population.
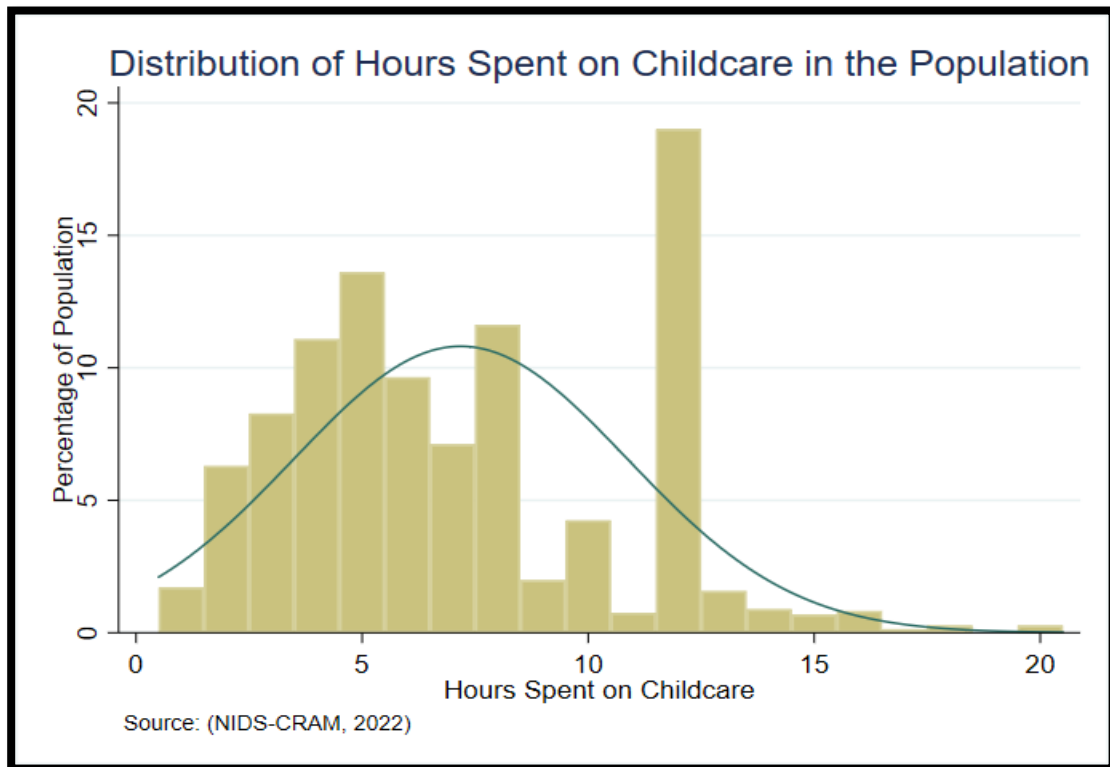
From this first graph there are initial insights into the amount of time spent on childcare for the whole sample space. There were two observations cleaned from the sample, one of 25 hours and one of 24. These were cleaned as they were deemed to be impossible, because they were either less hours than there are in a day, or a parent would have to engage in no other activity not even sleep. This graph provides good insight to the overall data, showing that there is a strong normal distribution to our data. We can clearly see here that the mode is 13 hours of childcare a day, and this is quite different from the grouping to the lower end of the rest of the data, but it does not adversely affect the overall normal distribution.

Figure 2: Graphs showing the Distribution of Hours Spent on Childcare by Gender in the Population.
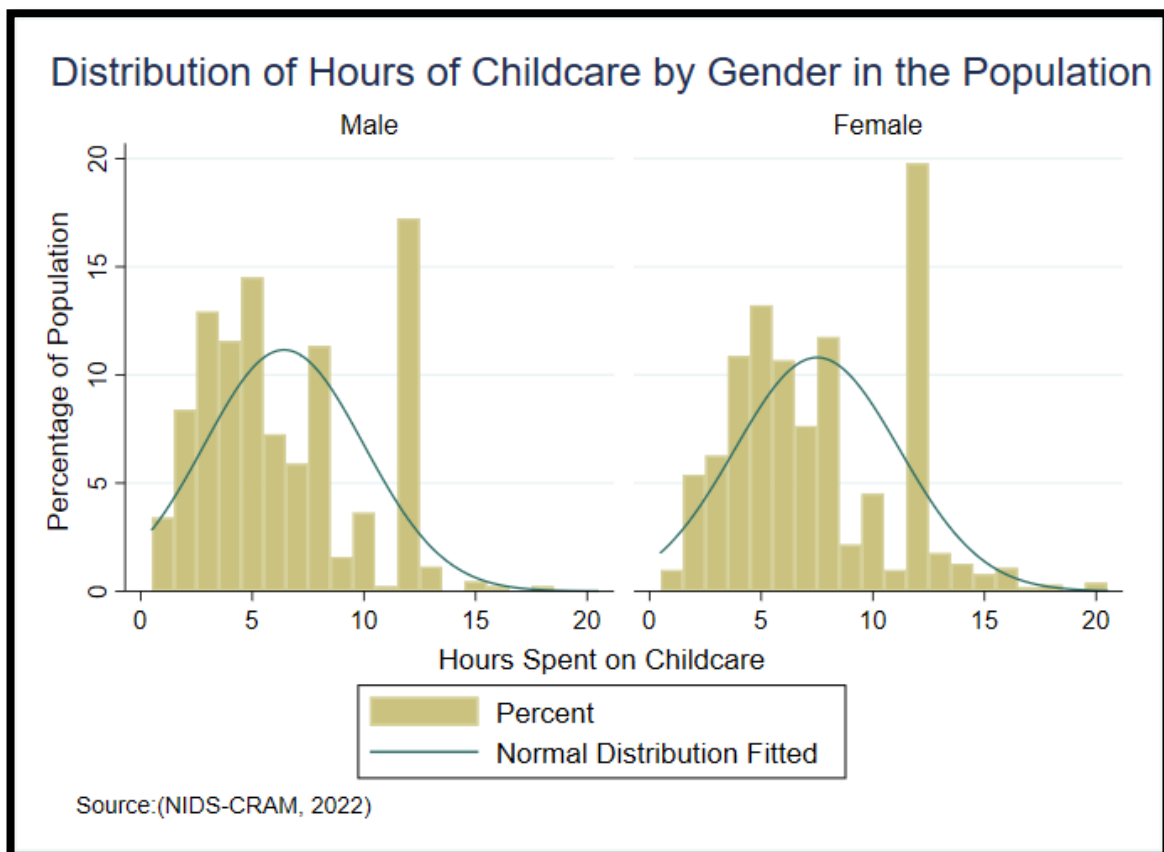
These graphs show that there is a difference in the amount time spent on childcare between genders, as can be seen by the normal distribution that have been fitted the hours of childcare done by females is higher than males. This corresponds with existing literature from other countries where it was found that mothers spent effectively double the amount of time caring for children than fathers. (Gimenez-Nadal & Molina, 2011) This is one the main variables in the regression model used and will be explored more deeply in the remaining parts of the paper.

Figure 3: Graphs showing Distribution of Hours Spent on Childcare by Geographical Location in the Population.
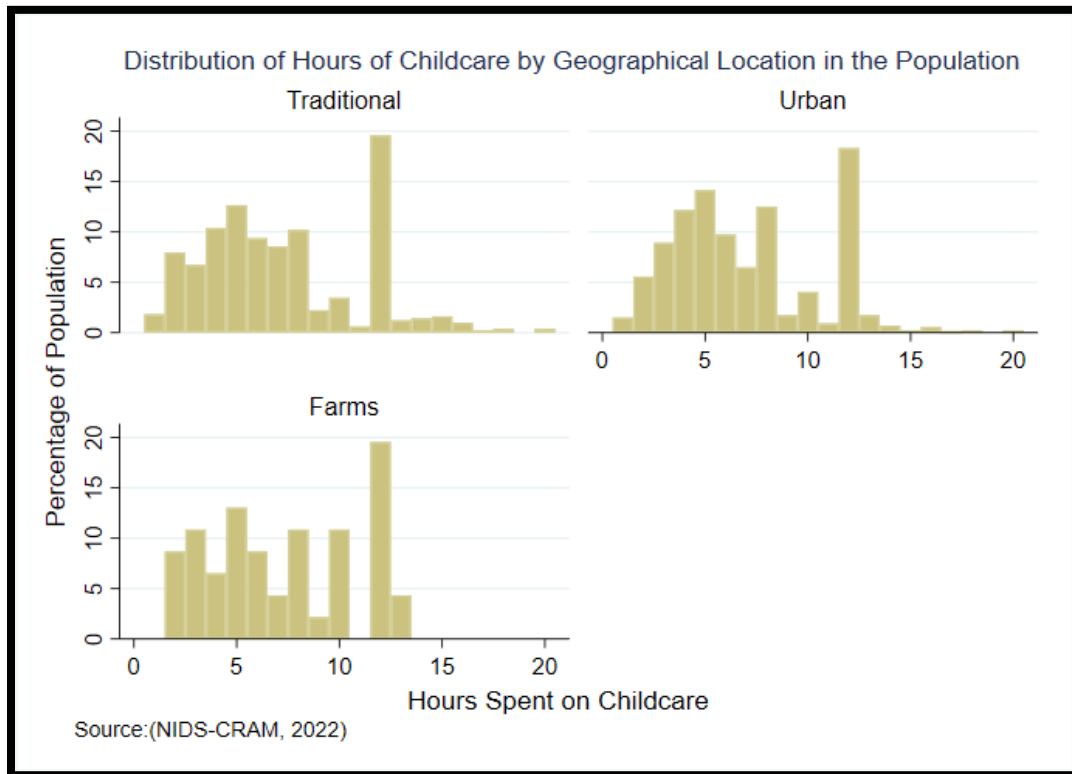
This graph represents the amount of time parents care for their children in different categorical geographic locations. Looking at geographical differences, as can be seen on the graphs, makes sense as people who live in different types of environments can spend vastly different amounts of time on different activities.

Figure 4: Graphs showing the Distribution of Hours Spent on Childcare by Race in the Population.
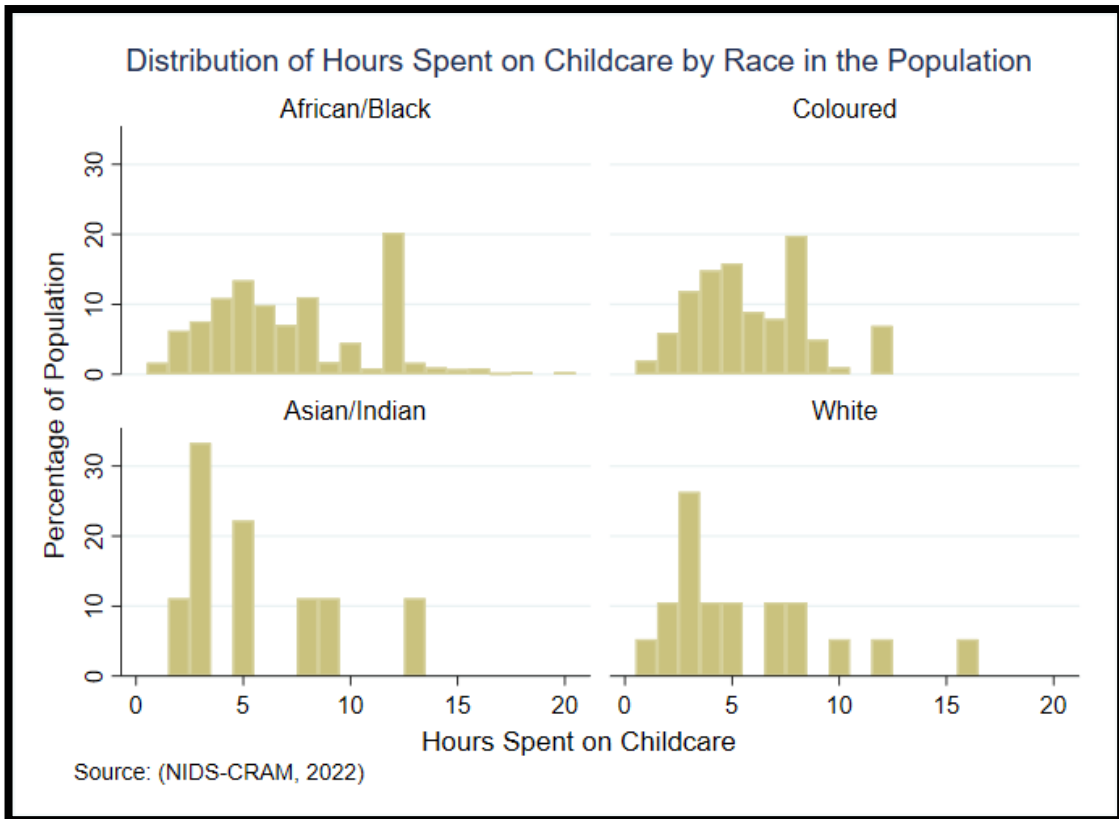
These graphs show how race impacts time spent on childcare. Within the South African economy there are large differences with how different races exist, economically, culturally and geographically, due to the apartheid regime. Seen here the amount of time spent in childcare is distributed closer to zero for white people comparatively to other races. This is not surprising as there is prolific inequality between White people and people of colour as can be seen by the levels of economic inequality with 52% of the African/Black population being poor and only 5% of the white (Woolard, 2002), there is an inequality present with white people spending less time in childcare. Because of this it is extremely important to see how childcare varies across races.

Figure 5: Distribution of Hours Spent on Childcare by Whether Children Attended ECD in the Population
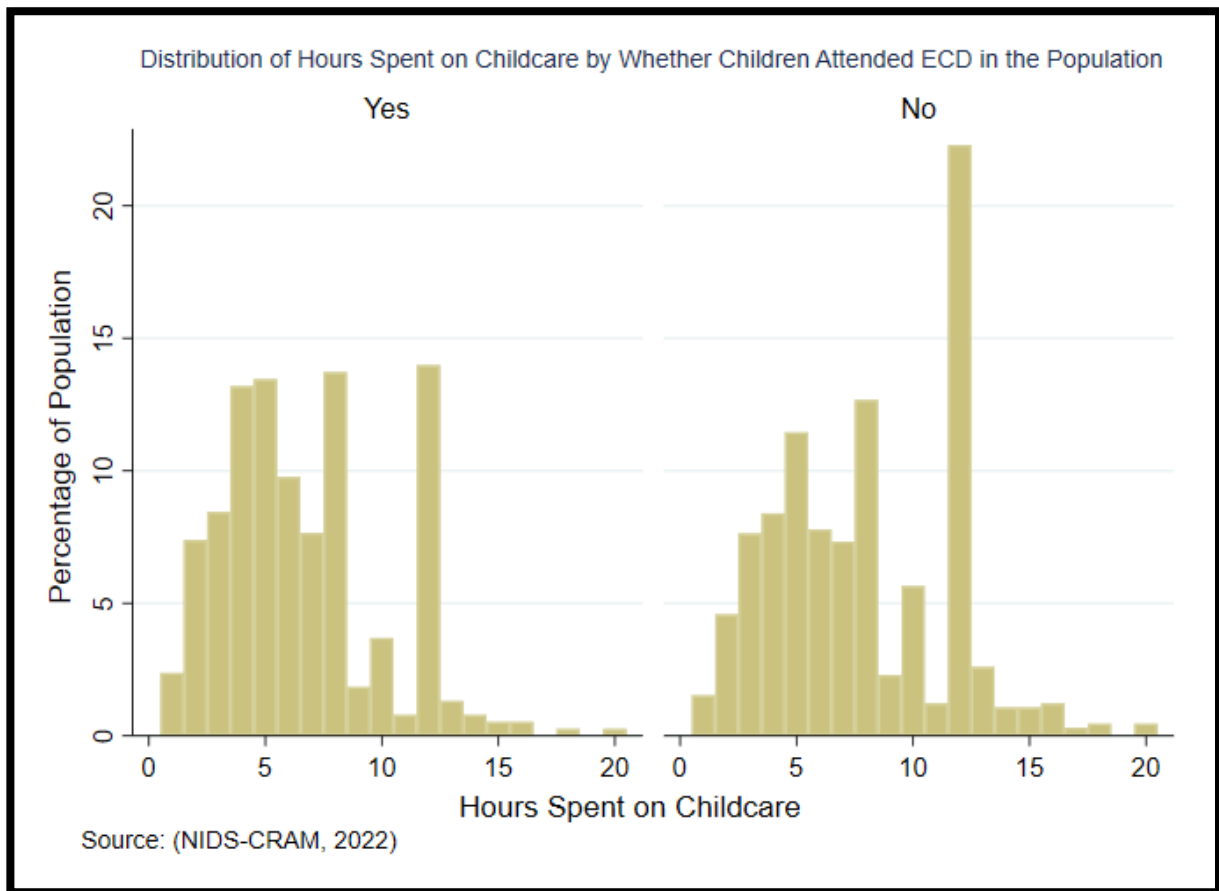
This graph shows the effects of a child attending the ECD program on how long a parent had to care for the child. This is an important variable as it begins to give insight into how policy around childcare can be created. This is because the ECD scheme is a policy that has been successfully established by the government and this means that it will provide useful insight into how we can write policy in the future to better aid young children. This will be further regressed and is part of important insights into a program that started in 2005 but could have been strongly affected by the COVID-19 pandemic.

4.3. Descriptive Statistics

Table 1: Descriptive Statistics Table by Race (Black/Non-Black)

	Obs.	Non-Black	Black	T-value	P-value	Significance
Total Number of Children	1448	2.24	2.97	-4.3	0	***
TERS	476	1.94	1.95	-.25	.8	
White	1448	0.15	0	15.1	0	***
Asian	1448	0.07	0	9.95	0	***
Coloured	1448	0.78	0	68.95	0	***
Parents Level of Education	1423	1.85	1.78	1.55	.13	
Age	1328	40.72	37.96	2.5	.01	**
Wage	501	12748.42	7723.43	2.7	.01	**
Urban	1383	0.95	.58	8.45	0	***
Farm	1383	0.03	.03	-.15	.9	
Traditional	1383	0.02	.38	-8.55	0	***
ECD	1028	1.58	1.64	-1	.33	
Female	1448	0.65	.7	-1.3	.2	
Male	1448	0.35	.29	1.3	.2	

Source: (NIDS-CRAM – Wave 5, 2022)

This table presents descriptive statistics for black and non-black individuals, testing for significance by race. The data sample is made up of four different categories for race. Significant differences are starred, *** p<0.01, ** p<0.05, * p <0.1.

5. Regression Results

Table 2: Regression Estimates for hours of childcare.

VARIABLES	(1) Simple Model	(2) Full Model	(3) Expected Difference	(4) Causal Relationship
Age		0.0521**	0.0602	-0.0365
		(0.0249)	(0.0447)	(0.0569)
Female	1.092***	0.964**		1.163
	(0.209)	(0.441)		(0.900)
ECD		0.878**	0.277	-0.796
		(0.443)	(0.738)	(1.002)
TERS		0.980	1.773	1.090
		(0.919)	(1.312)	(2.029)
Coloured		-1.872***	-2.333**	
		(0.644)	(0.991)	
Asian/Indian		-1.365	-3.551	
		(2.540)	(3.588)	
White		-4.164**	-4.341**	
		(2.110)	(2.147)	
Urban		0.0853	-0.0628	-2.405
		(0.486)	(0.794)	(3.003)
Farms		0.457	6.005**	-5.621
		(1.336)	(2.589)	(4.020)
Total Number of Children		-0.112	-0.360	0.136
		(0.127)	(0.236)	(0.414)
Parents Level of Education		0.427	0.0483	-1.398
		(0.648)	(1.217)	(1.618)
Constant	6.387***	0.594	1.211	9.493*
	(0.175)	(2.684)	(4.237)	(5.585)
Observations	1,448	286	112	39
R-squared	0.019	0.100	0.165	0.171

Source: (NIDS-CRAM – Wave 5, 2022)

Standard errors in parentheses. Significant differences are starred, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Bases category for race is Black, base case for location is traditional, base case for gender is male.

1) Is the simple regression model for gender with the base case of male.
2) Is our complete multiple regression model.
3) Is the multiple regression analysis on the sub-sample of female.
4) Is the multiple regression model on the sub-sample of Black/African.

6. Examination of the Regression Results

The first regression that has been run is the Simple Linear Regression (SLR) hours of childcare on gender (with male as the base case). The expectation based off prior literature is that gender should yield a statistically significant effect on hours spent on childcare, as per (England & Srivastava. 2013), (Gutièrrez-Domènech, 2010) and figure 2. The outcome of this regression proves this presumption to be true with a t-statistic of 5.19 producing an outputted p-value of 0.000***. This means that the null hypothesis that gender has no effect on the number of hours spent on childcare can be rejected at the 5% significance level. From the regression, it can be inferred that if the individual is female, the number of hours spent on childcare increases by 1.092 percentage points, ceteris paribus. Although there is clear evidence that gender is statistically significant, it is also practically significant. This is reflected by the size of the coefficient which estimates that females spend over one additional hour each day on childcare compared to males. The R-squared value describes the proportion of variation in hours spent on childcare that is explained by gender. This value is 0.019 (1.9%) implies that this model does not resemble a very good fit with the data used to modelling the SLR. Ultimately, these results are corroborated by (Gates Foundation, 2021) that find that women spend 45 minutes more than men per day on unpaid work, typically undertaking 75% of all childcare responsibilities. The difference equates to up to 5.7 weeks more work per year.

The second regression that was run was the Multiple Linear Regression (MLR) model. This utilizes all the independent/explanatory variables outlined in the empirical model specification to determine the hours spent on childcare. The expectation of this model is that statistical and practical significance will be derived from gender, ECD, race (being Black), education level and age. This is propagated by (Gutièrrez-Domènech, 2010), (Penn, 2004), (Budlender, Chobokoane & Mpetsheni, 2001) and (England & Srivastava. 2013) that claim these factors are crucial in improving hours spent on childcare.

According to the results of this regression and the coefficient estimates, it is predicted that being Coloured, Indian/Asian, White or having an additional child reduce the hours spent on childcare per day by 1.872, 1.365, 4.164 and 0.112 percentage points respectively, relative to being Black, ceteris paribus. Alternatively, it can also be predicted that an additional year of age, being a female (relative to a male), having access to ECD or TERS, living in urban areas or on farms (relative to traditional) or having an additional year of education increases the hours spent on childcare per day by 0.0521, 0.964, 0.878, 0.980, 0.085, 0.457, and 0.427 respectively, ceteris paribus. The null hypothesis that the explanatory variables have no effect on hours of childcare can only be rejected at the 5% level of significance for variables years of age, being female, accessing ECD, being Coloured or being White. The evidence for this is sufficiently shown in each of their p-values being less than 0.05 as per the

output (p<0.05). This conclusion suggests that these variables differ significantly from their hypothesized estimates and moreover, are practically significant given their size (which is large enough to be meaningful). Furthermore, the F-statistic of this regression is 2.89, yielding a p-value of 0.0013. The small p-value is sufficient evidence to reject the null hypothesis at the 5% level and conclude that the variables in this regression are jointly significant. Although some of the independent variables have small t-values and are not individually significant at the 5% level (p>0.05), it can be concluded that this model is a better fit for the data. Additionally, the R-squared statistic of 0.10 (10%) in this MLR is far greater than the 0.019 in the previous regression, meaning that this model explains a further 8.1% of the variation in hours of childcare compared to the SLR. The output of this regression confirms the prior literature outlined above.

The third regression model is a MLR model in which the regressor's effects are expected to output different estimates based on a particular sub-sample of the data. By regressing hours of childcare on all of the explanatory variables (except for gender), for the female sub-sample of the data, it is expected that significant results will be attained. This is because females are the most greatly affected participants of the sub-sample's regressors as explained in (Gutièrrez-Domènech, 2010) and figure 2. After running the regression, the output shows statistical and practical significance within the explanatory variables of; being Coloured or being White (relative to being Black) and living on a farm (relative to living in a traditional area). These are all significant at the 5% level with p-values less than 0.05 and affect hours spent on childcare by –2.333, -4.341 and +6.005 percentage points respectively, ceteris paribus. This is evidence supporting literature by ****. Moreover, the regression's F statistic of 1.97 yields a p-value of 0.0438, indicating significance at the 5% level. Despite the individual lack of significance in some of the explanatory variables, it can be concluded that the entire regression is significant and is a better goodness-of-fit for the dataset. This is further substantiated by the R-squared value of 0.165 indicating that 16.5% of the variation in hours of childcare can be attributed to the explanatory variables in this regression. This R-squared value is even greater than that of the previous regression and indicates that the variables hold greater explanatory power. By omitting gender, downward omitted variable bias is present seen through the far larger variance. These results indicate that living in more rural types of environments like farms; or being White or Coloured relative to being Black are of the greatest significance. Ultimately, this model reinforces the literature describing the significance of gender differences on hours of childcare (Zick & Bryant, 1996).

The final regression is run with objective of assessing causality between hours of childcare, and all of the explanatory variables outlined in the MLR specification except for the significant race variable – of which being Coloured or White (relative to being Black) have shown significance in the two prior MLR regressions. This regression is conducted on the sub-sample of being Black (and excludes all other

categories of race). The expectation based on the research suggests that hours spent on childcare is greater in Black participants, seen in (Budlender, Chobokoane & Mpetsheni, 2001). Upon analysis, it appears that years of age, being female (relative to male), having access to ECD and TERS, living in urban areas or on farms (relative to traditional), the number of children and level of education all remain unchanged/insignificant at the 5% level. These variables affect hours of childcare by -0.0365, +1.163, -0.796, +1.090, -2.405, -5.621, 0.136 and -1.398 percentage points, ceteris paribus. Based off of the low F statistic of 0.78 and high p-value of 0.63, this regression model reflects insignificance at the 5% level. The R-squared value of this sub-sample (excluding the Coloured, Asian/Indian and White race categories) increases even further to 0.171. This means the regression model now explains 17.1% of the variation in hours of childcare thereby gaining greater explanatory power. This output is indicative of the great impact that Race has on hours of childcare. This is because when the race categories were omitted from the regression, the remaining explanatory variables explained an even greater proportion of the variation in hours of childcare, than when race was included. Omitted variable bias is once again present in this model. This model gives credence to the literature by (Budlender, Chobokoane & Mpetsheni, 2001), describing the effect of race on the number of hours of childcare.

This data set was limited by the way in which in captured education level in the survey. This made it difficult to make any useful inferences about education level, which in prior analysis, seemed to have a positive effect on hours spent on childcare. For this reason, improvements can be made in the structure and categorization of certain questions, especially those relating to education level. The data was also limited in the number of total observations. The survey consisted of questions which some respondents failed to answer. By expanding the range of respondents and applying a greater degree of randomness when selecting a sample, more accurate and reliable results can be obtained, thereby improving the data set to be interrogated.

7. Policy Recommendations and Conclusion

To conclude, the results in this paper show evidence of statistically significant differences for the hours spent on childcare, according to gender, race and ECD. As previously stated, there are a few limitations to this dataset. These include the relatively small number of observations for some of the explanatory variables and the poor classification of some of the explanatory variables for the purpose of this paper, i.e. education levels. Increasing the number of observations may help in providing a more representative dataset.

Evidence shows that differences in hours of childcare due to gender exist in other countries outside of South Africa, as noted by Woolard (2002), who found that mothers in Spain do essentially twice the amount of childcare compared to fathers. The differences in hours of childcare due to race are arguably unique and more prevalent to South Africa environment when comparing with race are likely unique or at least more prevalent to the South African, most likely due to apartheid as has previously been discussed. Polatnick (2002) found that parents desire spending more time on childcare, and so the objective of any policy should attempt to create equality between males and females, while simultaneously increasing the amount of time parents as a whole spend on childcare.

For potential legislation we can firstly look at the existing programs shown in the regression, which are the ECD program and the TERS program. Both of these had positive impacts on the amount of time parents spend on childcare. This is likely due to the fact that both programs provide strong support systems to families, either in terms of direct monetary support or in easing the costs of raising children. Governments could increase the funding for these projects as they both seem to produce the desired effect on childcare.

Practically speaking governments would need to institute better social grants concerning children, for instance increasing the amount of ECD a child receives would mean that there would a further increase in the amount of time spent on childcare. When looking at the effects of racial differences on childcare, it is clear that based on South Africa's discriminatory past some races have suffered more than others. When considering that the problem is exaggerated in rural communities it makes sense to create policy that can be conducted in both urban and rural environments. As such, educational programs through community organizations, could attempt to alleviate the massive socio-economic differences between races. A specialized grant for example, could ensure that some level of racial inequality is alleviated, also there by improving childcare. Community based organization could find methods of educating fathers when it comes to the various educational, social and healthcare related problems associated with childcare.

Reference List

Budlender, D., Chobokoane, N. & Mpetshani, Y. (2001). A Survey of Time Use How South African women and men spend their time. *Retrieved From:* http://www.statssa.gov.za/publications/TimeUse/TimeUse2000.pdf

Craig, L. (2006). Parental education, time in paid work and time with children: an Australian time-diary analysis. *The British Journal of Sociology, 57, 4.* doi: 10.1111/j.1468-4446.2006. 00125.x

England, P. & Srivastava, A. (2013). Educational differences in US parents' time spent in child care: The role of culture and cross-spouse influence. *Social Science Research, 42, 971-988. Retrieved From:* https://doi.org/10.1016/j.ssresearch.2013.03.003

Gates Foundation. (2021). In the covid-19 era: an urgent need for change evidence review of the global childcare crisis and the road for post-covid-19 recovery and resilience executive summary. *Retrieved From:*
https://docs.gatesfoundation.org/Documents/Executive_Summary_Global_Childcare_Crisis_and_the_Road_Ahead_for_COVID-19_Recovery_and_Resilience_ENGLISH.pdf

Gimenez-Nadal, J.G. & Molina, J.A. (2013). Parents' education as a determinant of educational childcare time. *Journal of Population Economics, 26, 719-749.* doi: 10.1007/s00148-012-0443-7

Gromada, A., Richardson, D. & Rees, G. (2020). Childcare in a global crisis: the impact of COVID-19 on work and family life. *UNICEF Office of Research. Retrieved From:* https://www.unicef-irc.org/publications/pdf/IRB-2020-18-childcare-in-a-global-crisis-the-impact-of-covid-19-on-work-and-family-life.pdf

Gutièrrez-Domènech, M. (2010). Parental employment and time with children in Spain. *Review of Economics of the Household, 8, 371-391.* doi: 10.1007/s11150-010-9096-z

Ingle, K., Brophy, T., Daniels, R.C., (2021). National Income Dynamics Study – Coronavirus Rapid Mobile Survey (NIDS-CRAM) 2020 - 2021 Panel User Manual. Release July 2021. Version 1. Cape Town: Southern Africa Labour and Development Research Unit.

Penn, H. (2004). Childcare and Early Childhood Development Programmes and Policies: Their relationship to eradicating child poverty. *Retrieved From:* https://assets.publishing.service.gov.uk/media/57a08cbbed915d622c00152b/CHIPRep8.pdf

Polatnick, M.R. (2002). Quantity time: Do children want more time with their full-time employed parents? *Retrieved From:* https://dlib.bc.edu/islandora/object/bc-ir:101122

Samman, E., Presler-Marshall, E., Jones, N., Bhatkal, T., Melamed, C., Stavropoulou, M., Wallace, J. (2016). Women's work Mothers, children and the global childcare crisis. *Retrieved From:* https://cdn.odi.org/media/documents/10333.pdf

South African Government report. (2022). Two-year anniversary of the UIF's Covid-19 TERS. *Retrieved From:* https://www.gov.za/blog/two-year-anniversary-uif%E2%80%99s-covid-19-ters

Spaull, N., Daniels, R. C et al. (2021) NIDS-CRAM Wave 5 Synthesis Report. *Retrieved From:* https://cramsurvey.org/wp-content/uploads/2021/07/1.-Spaull-N.-Daniels-R.-C-et-al.-2021-NIDS-CRAM-Wave-5-Synthesis-Report.pdf

Woolard, I. (2002). An Overview of Poverty and Inequality in South Africa. Working Paper. Department for International Development, London. *Retrieved From:* https://repository.hsrc.ac.za/bitstream/handle/20.500.11910/8690/2166.pdf?sequence=1

Zick, C.D. & Bryant, W.K. (1996). A New Look at Parents' Time Spent in Child Care: Primary and Secondary Time Use. *Social Science Research, 25, 260-280. Retrieved From:* https://vula.uct.ac.za/access/content/group/4ec3754d-0aef-433a-9cff-30db08cefa3a/Project/Working%20Papers%20and%20Journal%20Articles/Topic%202/A%20New%20Look%20at%20Parents%E2%80%99%20Time%20Spent%20in%20Child%20Care.pdf

Appendix

Appendix Item A.1 Do File:

```
*ECO3021S Group 60 Project


*-------------------------------------------------------------------------------
                                            ***Starting Commands***
*-------------------------------------------------------------------------------


        cap log close

        clear all

        set more off

        numlabel, add


*-------------------------------------------------------------------------------
                                            ***Setting up Working Directory***
*-------------------------------------------------------------------------------


        cd "C:\Users\Public\Varsity\4th Year\DATA"


*-------------------------------------------------------------------------------
                                            ***Opening a log file***
*-------------------------------------------------------------------------------


        log using Project.log, replace text



*-------------------------------------------------------------------------------
                                            ***Loading Project Dataset***
*-------------------------------------------------------------------------------


        use dataset2022.dta, clear
```

```
*-----------------------------------------------------------------------------
                                    ***1. Cleaning and Recoding Variables***
*-----------------------------------------------------------------------------
/*


              TOPIC: Determinants of hours spent on child care in South Africa
              We are going to need the following variables for our Regression analysis:


*/


/*


              1. Wage
           2. Age
                      3. Gender
                      4. Schooling
                      5. Race
                      6. Marital Status
                      7. Geographical Location
                      8. ECD
                      9. TERS
                      10. no. CHildren
*/


*--------------------------
* Hours spend on child care
*--------------------------
* There is no specific variable named child care in our dataset so we must create it.
* There are one variable related to number of hours of child care in the dataset:
* w5_nc_numchild:  Number of hours a day looking after kids in March; w5_nc_hhinc: How much was
the total household income in March?


*------------------------------------------------------------------------------------------
              ***1. Preliminary Analysis******
*------------------------------------------------------------------------------------------


gen hourscare = w5_nc_chldcar_mar
```

```
drop if hourscare <= 0
drop if hourscare >= 24

histogram hourscare, percent normal

sum hourscare,d

ssc install pshare

pshare estimate hourscare, nq(4)

histogram hourscare, percent normal ytitle(Percent) xtitle(Hourscare) title(Distribution of hourscare in
the Population)



*------------------------
*        Schooling
*------------------------

lookfor school
codebook w5_nc_edschgrd
tab w5_nc_edschgrd

recode w5_nc_edschgrd        (-9 -3 20 13/18= .) (19 = 0) (17 = 12) (22=.)
ren w5_nc_edschgrd      education

recode education (1/7=1) (8/12=2), gen(schooling)
label var schooling "Schooling"
label val schooling schoolingv1
label define schoolingv1 0"0. No-schooling" 1"1. Primary" 2" 2.Secondary"

tab schooling
histogram hourscare, percent ytitle(Percent) xtitle(Hourscare) by(, title(Distribution of Hourscare by
schooling in the Population)) by(schooling)
reg hourscare schooling
```

```
*--------------------------
*        Age
*--------------------------

lookfor age

su  w5_nc_best_age_yrs
des w5_nc_best_age_yrs
tab w5_nc_best_age_yrs , m

recode w5_nc_best_age_yrs (65/100 = .), gen(age)
gen age_squared = age^2

sum age
histogram age, percent normal ytitle(Percent) xtitle(Age) title(Distribution of Age in the Population) //
Use the menu (Graphics)

reg hourscare age




*--------------------------
*        Gender
*--------------------------



recode w5_nc_best_gen ( 2 = 2) ( 1 = 1), gen(gender)
drop if 2< gender < 0
recode gender ( 2 = 0) (1 = 1), gen(Male)
tab Male

recode gender ( 1 = 0 ) ( 2 = 1 ), gen(Female)

tab gender



*--------------------------
*        Wage
```

```
*-------------------------

gen employed=w5_nc_hhincsrc1==1
gen wage=w5_nc_hhinc if employed==1
drop if wage<0

gen lwage = log(wage)

reg hourscare wage

gen w1=.
replace w1 = 1 if wage <= 10000

gen w2=.
replace w2 = 1 if wage <= 20000 & wage > 10000

gen w3=.
replace w3 = 1 if wage <= 30000 & wage > 20000

gen w4=.
replace w4 = 1 if wage <= 40000 & wage > 30000

gen w5=.
replace w5 = 1 if wage > 40000


*----------------------------
*          Marital Status
*----------------------------

lookfor married

tab w5_nc_mar, m
codebook w5_nc_mar
tab w5_nc_mar, m

recode w5_nc_mar (2 = 0 "not married") (1 = 1 "married"), gen(married)
```

```
tab w5_nc_mar married
label var married "Marital Status"
label val married married_lbl
label define married_lbl 0"0. not married" 1"1. married"
tab married


tab married


gen notmarried=.
replace notmarried = 1 if married==0


gen ismarried=.
replace ismarried = 1 if married==1


tab notmarried
tab ismarried


label var notmarried "Not Married"
label var ismarried "Married"




*------------------------
*        Race
*------------------------
recode w5_nc_best_race ( 1 = 1 "African/Black") ( 2 = 2 "Coloured") ( 3 = 3 "Asian/Indian") ( 4 = 4
"White"), gen (race)


by race, sort : sum wage
by race, sort : sum hourscare


tab race


recode race (2 3 4 = 0) (1 = 1), gen(Black)
recode race (1 3 4 = 0) (2 = 1), gen(Coloured)
recode race ( 1 2 4 = 0) (3 = 1), gen(Asian)
recode race (1 2 3 = 0) (4 = 1), gen(White)
```

label var Black "African/Black"

label var Asian "Asian"

label var Coloured "Asian/Indian"

label var White "White"

*----------------------------

*Location

*----------------------------

codebook w5_nc_geo2011

ren w5_nc_geo2011 location

tab location

recode location (-3 -9=.)

recode location (2 3 = 0) (1 = 1), gen(trad)

recode location (1 2 = 0) (3=1), gen(farm)

recode location (1 3 = 0) (2 = 1), gen(urban)

label var trad "Traditional"

label var urban "Urban"

label var farm "Farms"

histogram hourscare, percent ytitle(Percent) xtitle(Hourscare) by(, title(Distribution of Hourscare by Location in the Population)) by(location)

*----------------------------

*ECD

*----------------------------

lookfor ECD

tab w5_nc_ecdatt, m

tab w5_nc_ecdatt~20, m

ren w5_nc_ecdatt ecd

recode ecd (-3 -9=.)

fre ecd

```
ttest hourscare, by (ecd)

histogram hourscare, percent ytitle(Percent) xtitle(Hourscare) by(, title(Distribution of Hourscare by
ecd in the Population)) by(ecd)
reg hourscare ecd

tab ecd


gen yes=.
replace yes = 1 if ecd == 1

tab yes

gen no=.
replace no = 1 if ecd == 2

tab no

label var yes "Attended"
label var no "Did not Attend"

*----------------------------
*TERS
*----------------------------

lookfor TERS
tab w5_nc_eminc_t~s ,m
tab w5_nc_emsteer~r ,m

gen TERS = w5_nc_eminc_ters
tab TERS

drop if TERS<0

gen did=.
replace did = 1 if TERS == 1
```

```
tab did

gen didnt=.
replace didnt = 1 if TERS == 2

tab didnt

label var did "Recieved Money from TERS"
label var didnt "Did not Recieve Money from TERS"
*----------------------------
*Number of children in household
*----------------------------
lookfor residents
ren w5_nc_nou7res baby
ren w5_nc_no7to17~s teen
tab teen
tab baby
gen totalkids = teen + baby

tab totalkids

histogram hourscare, percent ytitle(Percent) xtitle(Hourscare) by(, title(Distribution of Hourscare by
number of kids in the Population)) by(totalkids)


*-----------------------
*Descriptive Statistics
*-----------------------


asdoc ttest totalkids, by(Black)replace
asdoc ttest TERS, by(Black) rowappend
asdoc ttest White, by(Black) rowappend
asdoc ttest Asian, by(Black) rowappend
asdoc ttest Coloured, by(Black) rowappend
asdoc ttest schooling, by(Black) rowappend
```

asdoc ttest age, by(Black) rowappend

asdoc ttest wage, by(Black) rowappend

asdoc ttest urban, by(Black) rowappend

asdoc ttest farm, by(Black) rowappend

asdoc ttest trad, by(Black) rowappend

asdoc ttest ecd, by(Black) rowappend

asdoc ttest Female, by(Black) rowappend

asdoc ttest Male, by(Black) rowappend

*--
         2. Regression Result ***
*--
Simple Linear Regression Model between Hours spent in Childcare and Gender

reg hourscare i.gender

outreg2 using Final.xls, excel append ctitle (Simple Model)

Full Multiple linear Regression Model

reg hourscare age i.gender ecd TERS i.race i.location totalkids schooling

outreg2 using Final.xls, excel append ctitle (Full Model)

* Multiple Linear Regression fro Sub-Sample with Expected Differences (GENDER)*

reg hourscare age ecd TERS totalkids i.race i.location schooling if gender==1

outreg2 using Final.xls, excel append ctitle (Expected Difference)

Multiple Linear Regression Model for a Sub-Sample with a Causal Relationship (RACE)

reg hourscare age i.gender ecd TERS totalkids i.location schooling if race==2

outreg2 using Final.xls, excel append ctitle (Causal Relationship)

*--
*Assumption Testing

*--

*MLR 3 Perfect Collinearity

pwcorr hourscare age gender ecd TERS race location totalkids schooling

*MLR 5 Heteroskedacity

reg hourscare age gender ecd TERS race location totalkids schooling

estat hettest

rvfplot, yline(0)