Let's think about your assignment step by step.

1. **Introduction** – what is the issue you are looking at and why should the reader be interested in this topic? Then explain what you are doing to address this topic in your assignment.

Read through these example introductions. What do you think? What was good and what can be improved? What grade would you give it? Do you understand the comments the tutor gave to the author? Make sure you don't make the same mistakes as some of these authors!

A. Examples of Introduction

Example 1:

1. Introduction

Children's success in the early years is related to subsequent adaptive development (Jimerson, Egeland, & Teo, 1999). Therefore, it is necessary to know the determinants that support learning during the first years of school. Modifiable environmental factors are primary contributors to academic achievement, and the effects of genetic factors are negligible (Lemelin et al., 2007). Parental education has been shown to be associated with children success at school (Haveman and Wolfe, 1984; Davis-Kean, 2000; Carneiro, Costas and Matthias, 2007; Sabates and Duckworth, 2009). Moreover, the education level of the mother has a more significant influence on the child's educational achievement than that of the father (Haveman and Wolfe, 1995). Also, it was suggested that mathematical assessments were the strongest predictors of academic success throughout school education, followed by reading performance (Duncan et al., 2007).

In this study, I explore the impact of the mother's educational level on the child's math performance at age 5. In this way, the children's differences in mathematics achievement in the early years are reduced, and inequality is reduced. I focus on the mother's highest level of education, controlling the lone parent status and demographics (gender and month of birth). I will use an OLS model to look at the factors. I examine the relationship between maternal education and math scores at age 5. It will show determinants that are, on average, related to math performance at age 5.

There are two motivations for it. One motivation is that as a student, my classmates and I are very concerned about the factors associated with higher mathematics achievement in the early years. These results may contribute to making more children have better math performance in the first years. Another motivation is that mathematical success is related to increased academic performance and labour market development in the future. To maximize individual opportunities in later life, promoting early mathematics achievement is a good way.

What do you think of this introduction? See if you picked up on some of the issues raised by the tutor who marked this assignment.

Comments from the tutor:

Your introduction doesn't really convince the reader that what you are looking at is important. You say early success is related to adaptive development.....so what? Why is this important. What has it got to do with the main issue you are looking at maths achievement? You don't discuss these issues, instead start referencing your literature. You only mention maths development at the end of the second paragraph. You need a shorter, more focused introduction.

Example 2:

Introduction

“Children socialization is the lifelong process of inheriting and accepting norms, customs, and ideologies, providing an individual with the skills and habits necessary for participating within his or her own society,” which also involves self-ego and identity, with effects beginning prior to birth and continuing into adulthood. (Inkeles, 1968) Simultaneously, the family is not only the earliest social environment for children but also one of the most important social units to cultivate their further development. Children from different social class would follow a different pattern of socialisation Children social-emotional ability is the mirror of their behaviours. Coleman's (Coleman, 1988)typology of capital is the one most widely used to explain socioeconomic status. The index of it is generally classified into three aspects: XI income; XII education background; XIII occupation status. (Mueller and Parcel, 1981)

What do you think of this introduction? See if you picked up on some of the issues raised by the tutor who marked this assignment.

Comments from the tutor: The introduction covers some general points relating to the research topic, but you fail to demonstrate the relevance of your research question. Importantly, it is not clear from the introduction, what the exact outcome is in which you are interested.

Example 3:

Introduction

Research has shown that school readiness predicts later academic achievement (Duncan et al, 2007, Mclelland et al, 2013, Bradbury et al, 2015, Li-Grining et al., 2010) which emphasizes the importance of school readiness scores in generating potential lifelong inequalities and thus the need for policy to address any existing achievement gaps in early years. The government responded well to this research providing free early education to young children with the aim of reducing the gap on school readiness scores between disadvantaged children and their advantaged peers (Abrue and Roberts, 2016). In 1998, England introduced universal part-time preschool education for 4-year-olds, which was extended to include 3-year-olds in

2004, and in 2009 to 15% of the most disadvantaged two-year-olds (West and Noden, 2016). Recent reforms to early year education, however, have sparked controversy as the conservative government introduced a policy in which parents who are in work are eligible for 30 hours free childcare, in comparison to the 15 hours free childcare which is universal. This has been criticized for only being on offer to working parents rather than prioritizing the most deprived families who may not qualify, this means that achievement gaps may widen (The Sutton Trust, 2017) This goes against existing research which has found that there are larger benefits to early years education for children from families with lower incomes, children whose parents have less education, (The Sutton trust, 2017) and children from particular ethnic minorities (Social Mobility Commission, 2016).
This study seeks to confirm whether these factors are associated with school readiness scores. It will do this by conducting statistical analysis on data from the Millennium Cohort Study to examine the difference between school readiness scores at age 3 between different ethnic groups in the UK. It will then identify whether income and mothers level of education are associated with the achievement gap for these ethnic groups.

What do you think of this introduction? See if you picked up on some of the issues raised by the tutor who marked this assignment.

Comments from the tutor:

You have chosen an interesting topic and you do a good job showing its relevance.

Think about these things when you are writing your introduction. Don't make the same mistakes as some of these authors.

2. Literature Review – discuss the existing evidence in the area – theoretical, and empirical.

B. Examples of Literature review

Example 1:

Literature review

Some scholars realize that students' academic performance is closely related to their family background. (Carneiro, Meghir, and Parey 2007; Davis-Keane 2005; Glick and Sahn 2000; Jimerson, Egeland and Teo 1999; Klebanov, Brooks-Gunn and Duncan 1994; Haveman and Wolfe 1995; Smith, Brooks-Gunn and Klebanov 1997;). For example, children with parents from higher education backgrounds will reach higher qualifications (Black, Devereux, and Salvanes, 2005). There are two main reasons: One is that parents with higher qualification are more likely to make more money for their children to have more experiences. Another reason is that their children have a greater chance of getting a higher education like them.

Also, research evidence suggests that mother education has a more important effect on children's academic achievement than father's qualification (Haveman and Wolfe 1995). Especially in mathematics and reading, children's accomplishments in these two areas are related to the expectations of highly educated mothers for the children

(Halle, Kurtz-Coates and Mahoney 1997). For example, a mother's education can improve a child's performance in mathematics at 7-8 years old (Carneiro, Meghir and Parey, 2007). As for the reasons, when children are small, the mother is more likely to spend more time on them, but the father works in the office (Bell and Starkey, 1974). Notably, the mother can show the children how to calculate different mathematical algorithms including addition, subtraction, and so on. Besides, when children watch educational TV shows, the mother can answer children's questions about the course. These are effective ways to increase children's cognitive development (Coneus and Sprietsma, 2009).
This study focus on the effect of mother's education and gender on children's mathematical development. Considering above literature mentioned are the reviews of other researchers, most of them only said the effect of maternal education on children's performance, but did not think about whether the relationship between children's math performance and mother's education will vary across gender. In this study, I will examine not only how the mathematical development scores of children at age 5 change with the maternal education, but also the influence of gender on the association between children's math performance and mother's education. My primary assumption is that the children's math performance at age 5 is positively related to their mother's qualification, and gender affect the association between mother's qualification and children's mathematics achievement.

What do you think of this literature review? See if you picked up on some of the issues raised by the tutor who marked this assignment.

Comments from the tutor:

Your literature review is really limited. There is a huge amount of literature that looks at maths achievement and differences in maths achievement. You mention parents education, and particularly mothers education but nothing else. You also don't really expand on how education may relate to differences is children's maths success. You say higher ed = more money and therefore more experiences, and that children will have higher education if their parents do without offering an explanation at all. You could have mentioned genetic transmission, more money for tutors/private school, more educational resources such as books; being able to help more with homework; cultural capital etc etc. Also what about gender, ethnic, age or social class differences in maths achievement - there is a huge amount of literature on all of these areas.

Example 2:

Literature review

Extensive research since the 1990s has shown that SES is highly associated with a wide array of cognitive performance, social adjustment, and educational attainment in children and the key finding are emerging in psychology and social science. (Clouston et al., 2012)There is a general consensus that a combination of income education and occupation provides a better understanding of SES than any those indicators alone(White, 1982). From the perspective of household income, plenty of longitudinal studies proved that the path linking family income and children later

development (Boyd et al., 2012). It is believed that parental resources would transform from an economic area to their children educational experience which indirectly contributes to children's outcome. (Brooks-Gunn & Duncan, 1997)”The most commonly mentioned linkages between SES and well-being is lack of access to resources” (Klerman, 1991). A lower ses family is not usually able to afford to provide cognitively stimulating materials and experiences for their children. Consequently, those children are less likely to expose to various learning and recreational resources, which limits their cognitive growth. Willms (2003)takes a longitudinal to examine the negative impact on the early year (3-5) cognitive development when they live in poverty and observed that lower SES children are more likely to suffer "double jeopardy."(Willms, 2003)from their psychological development to academic achievement.

Davis (2005) examines from a national sample of 868 children from 8-12 living in America, suggesting that parental income and their education background indirectly related to their children academic achievements through their exception. This is because parental belief would guide their behaviours of interacting with their children. Compared to the average working woman, upper-class mothers spend more time patiently teaching their children and “high-SES parents engage children in more conversations and practice them deliberately” (Phillips & Shonkoff, 2000) trying to expand their children intellectual horizons.

Parents with higher education background commonly form a harmonic and social atmosphere at home, tender family environment, which is beneficial for socialisation in children (Steinberg).McLoyd(1998) holds the view that distress and poor living condition among low SES parents may trigger overuse of negative control strategies. .On the other hand, the relationship between parental occupation and children's social and emotional development is not so consistent far from now(McLoyd, 1998). Full-time employment mothers play a more visible and vital role in childhood, which has also been discussed by more scholars (Bradley et al., 1989; Korupp, Ganzeboom, & Van Der Lippe, 2002). Other researchers mention that father from the different working class is the major contributor to their offspring further development Thinking along this line, studies focus on the effects of SES on children educational and cognitive outcomes to explain the intergenerational reproduction of inequality(McLanahan & Percheski, 2008). By testing a cross-sectional model of how parent education influences child development during middle childhood,(El Nokali, Bachman, & Votruba-Drzal, 2010)indicates that the effect of SES will become more pronounced on children well-being with age, which suggests that interventions for enhancing children's social capability in the early stage may show better outcomes than if left to the later stage.

Apart from mental aspects, physical inequality has also been exploring. For instance, Hawley(1992) points out children health are strongly related to SES. Children from low SES family are more likely to experience a birth defect or be at low birth weight. (Crooks, 1995; Hawley & Disney, 1992) Particular attention is paid on the effects of single parenthood, which could be connected to a lower family SES (Sarsour et al., 2011). The disappointing academic attachment can be partly explained by the fact that low- SES parents cannot provide extra time and financial resource raising and monitoring their children. At the risk of inadequately supervising from parents, children tend to be undersocialized in the early years. Research have examined the impact on children personal development in terms of their community environment and family conflict in different socio-economic status (Habib et al., 2013). A central issue in research on this topic is whether the associations found between family

structure and child development outcomes are causal or driven by other characteristics. Therefore, the impact of family socioeconomic status on early childhood competence has not reached a consistent conclusion. Bradley states that existing evidence cannot reflect the precise relation between SES and socioemotional well-being in children. He argues that the plight of providing more accurate conclusion partly because low SES and other cofactors to influencing socialization are not easy to detect separately.

What do you think of this introduction and literature review? See if you picked up on some of the issues raised by the tutor who marked this assignment.

Comments from the tutor: The literature is too vague, poorly written, and does not seem to follow a consistent narrative. Some of the literature you review does not relate to educational outcome, but rather the effects of SES more generally. You should try to be more focussed and make sure the literature is relevant to your own work.

Example 3:

Literature review

Assessment scores of children's mastery of language and ability to read and write at age of school entry are positively correlated with later education outcomes (Bennett, Weigel, & Martin, 2002; Lonigan & Shanahan, 2009; Romano, Babchishin, Pagani, & Kohen, 2010). Literature on child development indicate critical windows for learning particularly before age five, as this is the most rapid period of brain development and language learning (Britto & Pérez-Escamilla, 2013; Fox & Rutter, 2010). However, a number of factors have been identified that negatively impact early child development, setting some children at a disadvantage with potential long-term effects.

Household income and having a lone parent are two factors shown to significantly influence child language development. Children born into low-income households tend to perform worse on measures of school readiness and language ability compared to children born into higher-income households (Duncan et al., 2007). A similar achievement gap is found when comparing children of lone parents to their peers with two parents: Children of lone parents also tend to have worse outcomes on reading and communication tests, controlling for income (Conley, Pfeiffer, & Velez, 2007; Janus & Duku, 2007). This achievement gap may exist for a number of reasons. One reason may be that children with lone parents might receive less verbal interaction early in life by having the time and attention of one parent instead of two (Castro et al., 2015; Muller, 1993). Another reason may be that lone parents are more likely to experience mental health challenges such as stress, anxiety, and depression, as they are stretched to maintain a household without a partner, and these challenges may negatively affect the way these parents interact with their children (Deater-Deckard, 2008; Weinraub & Wolf, 1983).

In the UK, children of lone parents are twice as likely to fall below the poverty line compared to the national average (Social Metrics Commission, 2018). National policy in the UK has focused on improving low-income children's educational

outcomes through targeted interventions to improve school readiness (Science and Technology Committee, 2018; Department for Education, 2013). Of these low-income children, nearly 40% live in lone parent households (Social Metrics Commission, 2018). Given that children of lone parents tend to have worse outcomes on communication and literacy skills controlling for household income, low-income children in lone parent households may face unique challenges relative to their two-parent low-income peers (Conley et al., 2007; Janus & Duku, 2007). The following analysis further explores the relationship between lone parent status and measures of children's communication and literacy skills.

What do you think of this introduction and literature review? I think you will agree this is the best one yet. It is well written and clearly structured. It introduces the reader to the issue and discusses the relevant literature thoroughly.

3. Data and results

In this section the main points to look out for are:

1. Have you described the data and variables you are using? This should link to you discussion in your literature review – this justifies how you construct your model.
2. Do you have a descriptive table of results that test your hypothesis? This will depend on what your hypothesis is but don't just show univariate analysis on your outcome variable.
3. Have you discussed the results clearly?
4. Have you explained your regression model? And have you displayed the results clearly – not using STATA output!
5. Have you described the results? Have you interpreted your coefficients?

C. Examples of data and results section

Example 1:

Data Description

The database we used is from Millennium Cohort Study (MCS) conducted by the Centre for Longitudinal Studies (CLS) at the Institution of Education at the University College London. We chose mathematical development from Foundation Stage Profile scores (age 5, MCS3) as the dependent variable. For analysing the influences from family, we used a set of dummy variables including mothers' smoking status (moth1eversmoke) and lone parent status (lp) and a set of categorical variables including mothers' occupations (moth1nssec), fathers' occupations (dad1nssec) and parents' highest level of education (moth1nvq, dad1nvq), as the independent variables. Gender effect of cohort members on mathematical development was represented by sex dummy variable dummy (sex=o Boys; sex=1 Girls).
The outcome variable mathematical development (md) included 8475 values with one missing value. Table1 shows the stats of md. The score measuring the mathematical development level is range from 0 to 27 with a mean value of 20 and

standard deviation of 4.85449. As we observed some missing values in our dataset, we tell the Stata just display the available and applicable values when building the regression models.

Table1. Summary stats of mathematical development

Age 5 outcome	Obs	Mean	Standard deviation	Range	25th percentile	50th percentile	75th percentile
md	8474	20.0431	4.85449	0-27	18	21	24

Methodology

Our evaluation was based on an Ordinary Least Squared Model. Firstly, we conducted Ordinary Least Squared model to test the relevance level of all the variables we chose. The basic static regression model that we hypothesized can be expressed as follows:

$md = \alpha+\beta_1*sex+ \beta_2*Moth1nvq+ \beta_3*Dad1nvq+ \beta_4*lp+ \ldots+\mu$

where md measures the mathematical development level of children at 5 in the UK; α is the interception coefficient; β is the elasticity of each explanatory variable with respect to mathematical development level; μ is the error term.

Before our regression process, we did some operations to our raw data. Firstly, for dealing with the categorical variable mothlnssec, we ran the regression on md and mothlnssec firstly to test its significant level as Table2 shows:

Table2. OLS estimates of the mother's occupation on mathematical development

```
. reg md moth1nssec if (md>=0&md<.)
```

Source	SS	df	MS	Number of obs	=	8,474
				F(1, 8472)	=	45.62
Model	1069.36626	1	1069.36626	Prob > F	=	0.0000
Residual	198605.912	8,472	23.4426242	R-squared	=	0.0054
				Adj R-squared	=	0.0052
Total	199675.278	8,473	23.5660661	Root MSE	=	4.8418

md	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
moth1nssec	.1636903	.0242361	6.75	0.000	.1161816	.211199
_cons	19.78649	.0648819	304.96	0.000	19.6593	19.91367

The result shows that the relationship between mother's occupation level and child's mathematical development is statistically significant at 99% confidence level, as the p-value is 0.000. This implies for every one unit level increased in mother's education, the mathematical development score will increase by 0.1637 unit.

Thus it is necessary for further evaluation of the different impacts of each strata of mothers' occupations, we are going to create dummy variables for each category of our moth1nssec variable.

This variable is originally coded as: 0= never worked or long term unemployment; 1= higher managerial & prof; 2= lower managerial & prof; 3= intermediate; 4= small employers; 5= lower supervisory & technical; 6= semi-routine; 7= routine. We

generated a set of new dummy variables (mt1-mt8) to represent each strata respectively. For example, we generate the strata of never worked or long term unemployed mothers with frequency of 4720 as a new variable named mt1, if mothlssec=0, this dummy equals to 1, otherwise equals to 0. We used the same way to deal with the variable of father's occupation. The OLS estimates show that the relationship of father's occupation and children's mathematical development is statistically significant and positive. Thus we also created dummies (dd1-dd8) for each occupation level of fathers for further analysis.

Furthermore, It is always found that influences of parents on children's behaviour may vary in gender. Thus, we want to examine the effect of mother's education level on children's mathematical development is the same for both males and females. To check this we will include an interaction term between our continuous variable (mthlnvq) and the sex dummy variable.

Secondly, in order to evaluate the marginal effect, we creating a binary dependent variable, which is low mathematical development level (lowmath). We set Lowmath=1, when children were scored at the 25th percentile or below and lowmath=1 when the scores are higher. The value at the 25th percentile is 18, so we used 18 as a cut off for our variable that measures low mathematical performance.

We used Linear Probability Model to interpret the change in the probability of achieving low mathematical development when the associated factors change. The relationship can be formulated as follows:

Probability (lowmath=1) = $\alpha+\beta_1$*sex+ β_2*Moth1nvq+ β_3*Dad1nvq+ β_4*lp+ β_5*mt2+ … +μ

Where the coefficient α is the baseline probability for a child with low mathematical development with all independent variables equal to zero; β can be interpreted as the change in the probability of getting low mathematical development score when the associated independent variable changes holding others constant.

Results and Analysis

The results of the ordinary least square model were shown in Table3. The OLS estimates of our original unrestricted model indicated several variables were statistically insignificant, including mother's smoke status, mothers' occupations who are smaller employers, lower supervisory& technical, semi-routine and routine. It may indicate that these independent variables cannot explain the change in children's mathematical performances.

Looking at the goodness-of-fit in this model, adjusted R-squared indicates that 13.09% of the response variable variation can be explained by the multiple linear regression model. The value is not high while we still drew some statistically significant predictors, the reason probably is it is hard to interpret human's behaviour.

The constant of 16.2453 indicates that when the children are boys living with both parents, mothers who never smoked, parents are never educated and parents have never worked or are long term unemployed, the boys' mathematical development score is 16.2453.

The coefficient of sex variable is 1.328613 with a probability of 0, which means the influence of gender on children's mathematical development is statistically significant. And girls are estimated to performance 1.3286 units better than boys holding other variables constant.

According to both the probabilities of coefficients of parents' education levels are zero, the relationship between mathematical development and parents' education levels are statistically significant respectively. The coefficient of mother's education is 0.5885 which implies a positive relationship between mother's education level and mathematical development. For example, the child who has a mother with a degree will achieve 0.5885 unit higher score than the child who has a mother with a foundation level education. The coefficient of father's education is 0.2424 which is also positive but lower than 0.5885. That indicates the influence of education level from father side is positive but not as important as mother side. The possible reason is mothers always spend more time on taking care of children.

As we have discussed before parents' occupations show strongly positive relationship with children's mathematical performance. Thus we look at the influence of each occupation level. The OLS estimates show that mt5-mt8 are not statistically significant at 90% significance level. In other words, the model predicted that mothers who worked as small employers, lower supervisory& technical, semi-routine and routine have no relationship with children's mathematical performance. While, mothers work as higher managerial& professor, lower managerial& professor and intermediate are predicted to have positive impact on mathematical development of children at age 5. All of fathers' occupation levels show a statistically significant relationship with mathematical development at 90% significance level. Comparing the coefficients of all the occupation categories, we found mothers who have lower managerial jobs or are professors (mt3) have the greatest positive impact on children's mathematical development, the coefficient of which is 1.0369. That indicates children with mothers worked as lower managers and professors will have a 1.0369 higher score than children with mothers who have never worked or are long term unemployment. And fathers who work as higher manager and professors (dd2) have the greatest positive influence on children's mathematical development. In other words, when fathers are higher manager and professors, children will perform a score of 1.9438 higher than children with fathers who have never worked or are long term unemployment when holding other variables constant.

In terms of the influence of marital status, the model estimated it was statistically significant at 95% level with probability of 0.043. The coefficient is -0.3413 which implies the children who live in a single parent family will have a score of 0.3413 lower than the children who live with both parents. In reality, it implies the education from both parents' sides is crucial for promoting children's mathematical development.

To address the influence of gender difference on mother side factors, we created an interaction between girls dummy and mother's education level. The coefficient is -0.3834 with a probability of 0.043 which means the interaction term is statistical significant and the association between mothers' education level and children's mathematical development is different between girls and boys. The negative sign indicates that the relationship between mothers' education levels and children's mathematical development is slightly less strong for girls than for boys. In other words, for girls, each one unit in mothers' education level is associated with a 1.5337(0.5885-0.3834+1.3286) point increase in mathematical development score.

Table3. OLS Model

Independent variables	Full OLS Model
Sex	1.3286(0.2328)***

Moth1eversmoke	0.0824(0.1268)
Moth1nvq	0.5885(0.0582)***
Dad1nvq	0.2424(0.0435)***
lp	-0.3834(0.1898)**
Mother's occupation	
Higher managerial&prof (mt2)	0.8860(0.2889)***
Lower managerial&prof (mt3)	1.0369(0.1840)***
Intermediate (mt4)	0.7010(0.1987)***
Small employers (mt5)	0.3879(0.3766)
Lower supervisory&technical (mt6)	0.1764(0.3981)
Semi-routine (mt7)	0.0214(0.2199)
Routine (mt8)	-0.0001(1)
Father's occupation	
Higher managerial&prof (dd2)	1.9438(0.2596)***
Lower managerial&prof (dd3)	1.6600(0.2384)***
Intermediate (dd4)	1.8433(0.3466)***
Small employers (dd5)	1.2026(0.2520)***
Lower supervisory&technical (dd6)	1.3308(0.2446)***
Semi-routine (dd7)	0.4764(0.2603)*
Routine (dd8)	0.8376(0.2547)***
Interaction term	
Sex#moth1nvq	-0.2588(0.0719)***
Constant	16.2453(0.2424)***
R-squared	0.1341

Notes. Marginal effects. Standard errors in parentheses. * p<.10, ** p<.05 , *** p<.01

What do you think of this section? See if you picked up on some of the issues raised by the tutor who marked this assignment.

Comments from the tutor:

You develop a model that controls for some of the things you don't discuss in your literature review (occupation) and leaves out 1 of the few things mentioned in the text (age). I think you need to justify the variables in your model better given this mismatch. Why have you included smoking for example? Why lone parenthood? Why not age of child, income. Why not a previous measure of ability to proxy for innate ability? You also use both mothers and fathers education and occupation. These variables are likely to be highly correlated. remember the aim is to build a flowing narrative that tells a consistent story from start to finish. Your text lacks flow due to the lack of consistency across the various sub sections.

Your descriptive table (1) shows univariate statistics on your outcome. This could have gone in the appendix and been replaces by a table that examines variation in maths scores across your independent variables. This would be a better test of your hypothesis and again would link better with the text and flow nicely on to your regression analysis.

In explaining your regression you say you interpret your coefficients in terms of elasticities. This is only true if you are running a log log model, which you don't say you are. Otherwise OLS coefficients are interpreted in unit differences in y given a 1 unit increase in x. Please read your notes from lecture 1 on this.

I am also unclear why your run a bivariate regression with one of your independent variables in Table 2? Also I am not sure why you enter a categorical variable as a continuous variable in this model.

You were asked to produce your results in tables that were user friendly, had title and labels and that were NOT Stata ouput. Yet your table in this section is Stata output using variable names as they appear in the dataset. There is a video in Moodle to help you produce appropriate tables, that you should have watched. Your OLS Table which is not in Stata output lacks a proper title. These things are important, the reader should be able to see what id going on by just looking at the table.

This whole methodology section is rather unfocused and most of it could be deleted. For your OLS model you have entered mothers education as though it were a continuous variable when it is categorical and should have been entered as a series of dummies. When you discuss the insignificance of some of the occupation dummies you should have mentioned possible correlations between some of your variables I mentioned earlier. You could have looked at this in your data too.
In your text you talk about marital status, but you do not include this in your model. Instead you have a dummy that indicates whether a child has ever lived in a lone parent family. It does not mean other children were living with a married mum and dad. Be accurate when you are talking about your variables.
The insignificance of smoking is not surprising, you needed to justify this variable in your model as otherwise the reader is thinking why should smoking be related to maths scores? It may be you were proxying social class or income?? But I think the model would be better without this variable.

NB - This section was part of an assignment that was unfortunately below the standard expected at MSc level and the student failed.

Example 2:

Data

The data used in this analysis are taken from the Millennium Cohort Study (MCS), a longitudinal study undertaken by the Centre for Longitudinal Studies at University College London. The MCS follows the lives of children born in the UK between 2000–2001, tracking the children's mental, behavioral, and physical development over time as well as gathering information about their family context (Centre for Longitudinal Studies, n.d.). Using longitudinal data is advantageous in allowing us to track change over time from initial baseline conditions. By controlling for relevant factors, we can identify correlations between various child and family characteristics and long-term child development outcomes, which may strengthen inferences of possible causal effects (Raudenbush, 2001).
The child and family characteristics (the independent variables) included in the analysis were gathered during Sweep 1, which was conducted between 2001 and 2002 when children were aged 9 months. Communication, language, and literacy

(CLL) outcomes (the dependent variable) were measured during Sweep 3, which was conducted between 2005 and 2006 when children were aged 5 years. These CLL scores are taken from the Foundation Stage Profile (FSP) survey conducted by the Department for Children, Schools, and Families and linked with the Millennium Cohort Study. The FSP is completed by teachers across England, reporting the children's performance in six domains, including CLL (hereafter, 'FSP–CLL'). This sample includes 7,917 children followed at both sweeps.[1]

Descriptive Statistics

Table 1 provides a cross-tabulation of children's FSP–CLL outcomes by lone parent status, showing mean score, the standard deviation, the minimum and maximum scores attained, and the scores for each intermediate quartile. (See Appendix 1 for additional descriptive statistics of the sample included in this analysis.)

Table 1. Age 5 outcomes: Foundation Stage Profile – communication, language, & literacy scores

	# of obs.	Mean	Standard deviation	Min–Max	25th percentile	50th percentile	75th percentile
ALL CHILDREN	7,917	24.8	7.1	0–36	20	26	31
Children with a lone parent at age 9 months	2,054	22.7	7.1	0–36	18	23	28
Children without a lone parent at age 9 months	5,863	25.5	7.0	0–36	21	27	31

Note. "# of obs." is the number of observations.

Analysis and Modeling

We will now explore the potential relationship between a child's language ability and being born into a lone parent household using regression analysis. Regression analysis allows us to identify correlations between score outcomes and factors that we know may influence those outcomes, based on a wealth of theoretical and empirical literature.

The model on which this analysis is based focuses on the relationship between a number of child and family characteristics (the independent variables) and children's FSP–CLL scores at age five (the dependent variable). Although I am particularly interested in the possible effect of having a lone parent on language skills development, it is important to include other potentially relevant factors to increase the explanatory power of this model and to control for possible confounds. The independent variables included in the model are the child's sex and ethnicity,

whether English is the primary language in the household, and household income. Each of these variables have been shown to correlate with children's cognitive development and education outcomes (Darling-Hammond, 1998; Halle, Hair, Wandner, McNamara, & Chien, 2012; Kiernan & Mensah, 2009); as such, these variables may also influence FSP–CLL outcomes and are relevant to this analysis. The dependent variable, FSP–CLL scores, is continuous. I run an ordinary least squares (OLS) regression to evaluate score outcomes across all independent variables. OLS regression generates beta coefficients that indicate changes in FSP–CLL scores in relation to these child and family characteristics. Examining the direction and magnitude of change in FSP–CLL outcomes associated with each characteristic may contribute to a growing body of literature and theory, from which causal inferences might be drawn.

An interaction term is also included in this OLS regression analysis to explore whether the possible effect of having a lone parent varies by household income. Even if children in lone parent households tend to have lower FSP–CLL scores, it may be that lone parents who are wealthy may be able to compensate for the absence of an additional parent in ways that lone parents who are poor cannot (e.g., through better childcare, nutrition, and greater autonomy to take time away from work). If this were the case, we would expect to see a statistically significant interaction coefficient for lone parent status and household income.

Understanding whether coming from a lone parent household is associated with a significant achievement gap in language and literacy ability, and having a sense of the magnitude of that possible impact, can inform the development of social policy solutions and educational support for the most vulnerable and disadvantaged children.

Limitations

There are limitations to this model and analysis that are important to keep in mind before turning to the results. First, this analysis cannot establish a direct causal relationship between child and family characteristics and children's language and literacy outcomes. However, regression analysis can reveal correlations that are statistically significant (i.e., unlikely to be due to chance). Theoretically informed models and careful reasoning can also help us ascertain the direction of the relationship (e.g., it does not make sense to say that children's FSP–CLL scores at age five caused lone parenthood at age 9 months, but an association between lone parenthood at age 9 months and FSP–CLL scores is plausible and backed by theory and a breadth of literature). Second, FSP–CLL scores are reported by teachers and not direct tests of children's ability. Poorer outcomes might reflect subconscious biases of teachers, particularly along gendered and racial lines (Gillborn, 2010). Third, this model does not (and cannot) account for all variables that may influence children's FSP–CLL scores, such as a child's genetics, the presence of other siblings and other family members, the gender of the lone parent, and environmental factors. Importantly, because all independent variables were measured in Sweep 1 when the children were 9 months old, this model does not account for changes in parent relationship status or household income after the time of initial data collection. Two-parent households may have divorced or separated, and lone parents may have found new partners. Household income could also have shifted between Sweeps 1 and 3 with a change in parents' employment status or wages. However, a large sample size can help reduce the margin of error, and I have also included several

relevant independent variables to help reduce the error term. By situating these findings within existing literature, this analysis may expand our understanding of factors influencing children's learning and development outcomes.

Results

I run a full, unrestricted model to look at the possible relationship between having a lone parent at age 9 months and children's FSP–CLL scores at age 5 years, controlling for household income, the child's sex and ethnicity, and whether the household speaks English as a second language.

Table 2 shows the results of the OLS regression, giving us a sense of the direction of change and magnitude of impact each of these characteristics may have. The relationship between having a lone parent (LP) and FSP–CLL is statistically significant at the 0.1% ($p < 0.001$) level, meaning the correlation between LP and FSP–CLL is less than 0.1% likely to be the result of chance (although I cannot be entirely certain of ruling out all possible confounds, I have notably controlled for household income, which [as detailed above] the literature has shown to be the most likely confound). On average, children who have a lone parent are given an FSP–CLL score 1.8 points lower than their peers with two parents (recall from Table 1 that the range of CLL scores is from 0–36). The relationship between household income and FSP–CLL scores is also statistically significant at the 0.1% level. On average, children are given an FSP–CLL score that is 0.009 points higher for every incremental increase of household income. Given that income is measured as a continuous variable, the difference between the poorest and richest children is considerable. Children with a household income at the 25th percentile score on average 2.18 points lower than children with a household income at the 75th percentile and 6.8 points lower than children in households at the 99th percentile.

Results do not suggest a statistically significant interaction between lone parent and household income on FSP–CLL scores. This indicates that the possible effect being in a lone parent household has on FSP–CLL scores does not significantly vary by household income.

Of the control variables included, a child's sex and ethnicity are associated with a statistically significant difference in FSP–CLL outcomes. On average, teachers assigned females FSP–CLL scores 2.3 points higher than their male peers. On average, teachers gave Pakistani and Bangladeshi children FSP–CLL scores lower than their white peers, with Pakistani and Bangladeshi children on average receiving scores 2.1 points lower than white children. However, being of Indian, Black, mixed, or 'other' ethnicity was not associated with a statistically significant difference in FSP–CLL score. Interestingly, coming from a household with English as a second language is not associated with a statistically significant difference in FSP–CLL scores. The goodness-of-fit measure, R-squared, indicates this model accounts for 13.1% of the variance in observed FSP–CLL scores.

Table 2. OLS regression of FSP–CLL scores on child and family factors

Independent variables	(1) Full OLS model

Lone parent	–1.815***
	(0.311)
Household income	0.00886***
	(0.000458)
Lone parent × Household income	0.00161
	(0.00117)
Sex (female)	2.263***
	(0.149)
Ethnicity	
Indian	–0.274
	(0.492)
Pakistani & Bangladeshi	–2.120***
	(0.414)
Black	–0.248
	(0.391)
Mixed/Other	–0.163
	(0.340)
Household speaks English as an additional language	–0.417
	(0.333)
Constant	21.69***
	(0.205)
Observations	7,917
R^2	0.131

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

What do you think of this data, methods and results section? I hope you can see why it is much better than the previous example. In fact this is part of an assignment that got an A overall. See if you agree with the points raised by the tutor who marked this assignment.

Comments from the tutor:
I love the way you write a consistent story and weave in your analysis in a seemless way, thoroughly justifying your research question, the data you use, your analysis and results so that a layman could understand. Very well done.

A couple of minor points:

Try not to start a sentence with Table 1.....

You show the descriptive but don't talk about the results.

Your model might have included age at test and mothers education.

Finally you need to add a conclusion – which summarises you findings and discusses the implications they have.

Then add a reference section for all of your citations.