International Journal of Advanced Statistics and Probability, 4 (1) (2016) 27-30

International Journal of Advanced Statistics and Probability

Website: www.sciencepubco.com/index.php/IJASP
doi: 10.14419/ijasp.v4i1.5655
Research paper

The relationship between household income and educational level. (south Darfur rural areas-Sudan) statistical study

Sofian A. A. Saad [1]*, Amin I. Adam [2], Afra H. Abdelateef [3]

[1] University of Nyala, Faculty of Economics, Department of Economics, Nyala-Sudan
[2] Omdurman Islamic University, Faculty of Economics, Department of Statistics, Omdurman-Sudan
[3] Sudan University of Science & Technology, Faculty of Science, Department of Statistics
**Corresponding author E-mail:sofianabuelbacher@yahoo.com*

Abstract

The main aim of this paper is to find out the relationship between the household level of income and the level of education for the household being lived in southern Darfur state (Sudan) since they were seriously affected by the war conflict and lost most of their income sources. One-way Analyses of Variance (ANOVA) have been used to asses this relation where the independent variable and dependent variables are categorical and continuous respectively. Data gathered from Household survey (HHS) is analyzed. The results indicate the existence of relationship between certain education groups and the level of income, mainly high level of education comparing with the low level.

Keywords: *Analysis of Variance (ANOVA); Household Income; Household Survey (HHS); Internally Displaced People (IDPs); Least Significance Difference (LSD).*

1. Introduction

It is often said that education is the key to success. Education broadens one's mind, builds confidence to makedecisions, face challenges and accept failures, and open the door to new and better job opportunities.

During the first decade of this century, all Darfur states especially Southern Darfur state has faced horrible militant crisis by some military movement targeting the civilians in the region which leads to serious social and economic collapse, especially in terms of the livelihood for the affected people, these situations together with the scarcity of the natural resources at that time force a majority of the people for losses their main sources of income and then hindering their ability to have a better style of life. However, due to this situation thousands of people become homelessness leaving their areas of origin and lived around the main localities near the big cities as an internally displaced peoples (IDPs) in camps seeking better situation of life. As a result of that, this situation makes hard pressure to the limited resources in cities and inflates the phenomena of the unemployment.

This study would be conducted to see the main factors that affected determination of income level for the affected people in Southern Darfur state targeting two localities (Bolbol and Beliel) in concentration on the direction of the relationship between the education levels as an independent categorical variable with four levels and the level of income as a continuous dependent variable.

Problem and objectives of the study

No one can deny the role played by income into human life, as it is known, the money is the lifeblood. However, it will make it difficult for everyone to have a better life style in the absence of basic services such as food, clothes, health, education and security, which were impossible to obtain if an appropriate source of income are available. The main objectives of this paper are:

- To find out the relationship between household head level of income and their education level.
- To reject one of two opposing mutually exclusive hypotheses based on observed data. The null hypothesis states that the population means are equal; the alternative hypothesis states that not all the populations' means are the same.

These objectives would be achieved using the ANOVA test through the calculation of the probability which is called the p -value. A large p -value (i.e., $p > 0.05$) implies that the observed data represent a likely event under the null hypothesis؛ hence there is no evidence to reject the null hypothesis as false.

On the other hand, a small p -value (i.e., $p \leq 0.05$) suggests that the observed data represent a highly improbable event if the null hypothesis is true.Therefore, the observed data contradict the initial assumption and the null hypothesis should be rejected.

Assumptions

One assumption is tested (i.e., There is a direct relation between the level of income and the level of education)

2. Literature review

2.1. Income inequality and education

The high and positive correlation between education and income is a well-established fact. In the theory of the human capital, [1] showed that acquiring education increases the skills and competencies of individuals and their productivity. Since in a competitive labor market wages equal workers' productivity, higher productivity leads to higher wage.This means that a more educated society holds greater welfare. Since its conceptualization, this theory was the focus of increasing scientific research. Supporting as well as opposing views have encouraged the production of countless empirical and theoretical studies. Nowadays, the ac-

Copyright © 2016 Sofian A. A. Saad et. al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

knowledgment of a causal relationship between education and earning is a well-established result, and it is one of the most important achievements in economics. Conversely things are less clear cut when analyzing the link between income inequality and educational attainments. On the one hand, rising wage inequality should encourage investments in education mainly because it raises the return to education. [2] Observes a faster skill accumulation as a result of rising returns. This increase in the supply of skills should eventually mitigate the increase in inequality. On the other hand, increasing income inequality affects also the resources that households have been available to finance education. The intergenerational theory claims that there exists a perfect correlation between income and education distributions. This entails that barriers, e.g. liquidity constraints, family background, might prevent the investment in education for the fraction of the population belonging to the bottom of the income distribution. If the intergenerational mechanism is persistent, then the same parts of population are trapped at low levels of education and income for more than one generation.

2.2. Measures of education and income

The main measure of educational attainment used in the literature is the highest degree of education an individual has completed. Education can be recorded as years of completed education or as an ordinal variable with the obvious ordering going from the lowest to the highest level of education. An alternative measure used by several empirical studies is also the enrollment rate for the three educational levels (primary, secondary, and tertiary education). Regarding the macro studies, income inequality is measured by the Gini index as well as the income quartiles, while the micro studies have used the family income as a determinant of educational choices.

2.3. Empirical evidence

Papers analyzing the effect of income (wealth) inequality on educational attainments can be divided into two broad groups: the first one related to the macroeconomic literature analyzes the more general relationship between inequality and growth, and considers education as a key factor to increase growth. The second group of studies focuses on the effect of family income on children's outcomes and applies a microeconomic approach. However, both groups attempt to provide evidence and/or theoretically support for the idea that unequal society might harm investments in education. Amongst others, the papers [3, 4, and 5] pertain to the macroeconomic approach. In particular, and ("as stated by Galor & Zeira [3]") in the presence of imperfect credit markets, the wealth distribution affects investments in human capital. By developing an overlapping generation model with intergenerational transmissions and suggest that the initial distribution of wealth is crucial to determine individuals' educational choices and the aggregate output both in the short and in the long run. Along the same line of reasoning, [4] end up with similar conclusions. Their theoretical model suggests that the pattern of occupational (educational) choice is shaped by the initial distribution of wealth. [5] Investigates the relationship between income distribution, democratic institution and growth. The paper mainly aimed at addressing data and estimation issues. One of the main conclusions of the paper is that there is strong empirical support for the link going from income distribution to education decisions, i.e. more equal societies have higher rates of investment in education. In addition, [6] perform an empirical analysis using household surveys for 35 countries. They demonstrate that the poverty index, their proxy for economic status of the household, is correlated with reduced school attainmentinthepoorest40percentofthepopulation. [7] Investigates the issue using an unbalanced panel of 108 countries for the period 1960-95. His main finding is a robust negative correlation between income inequality and secondary education enrolment. The effect is stronger when considering female's access to any level of education. These results support the view that poor families are prevented from accessing school by their low incomes. Thus, greater income inequality reduces access to school. Except for the theoretical papers [3, 4 and 5], the empirical macro-studies lack in properly addressing the endogeneity of the inequality variable; that is, when other omitted factors are correlated with both the education and inequality measure, or when the causation goes to the other way around (education causes inequality). Thus, caution is needed when interpreting these results.

3. Methodology

Before starting the analysis one must consider the method when designing a study. The method of analysis depends on the nature of the data and the purpose of the study. However, since our data is both categorical (independent variable with four levels) and continuous (dependent variable) the suitable method to conduct the study is the statistical one-way analysis of variance (ANOVA). One-way analysis of variance (ANOVA) is used to compare several means. This method is often used in scientific or medical experiments when treatments, processes, materials or products are being compared.

ANOVA postulated mathematical equation that relates the measured response of the elements to the sources of variation using a number of assumptions. In ANOVA the sum of squares of a source of variation is a measure of the variability due to that source. Sum of squares is usually denoted as ''SS'' with subscript identifying the corresponding source.

3.1. Degrees of freedom

The degrees of freedom refer to the number of independent observations that are calculated in the sum of squares (Keppel 1973). It is often denoted as ''df. '' Li (1964: 35–36) postulated that the degree of freedom for the difference between (or sum of) two quantities is equal to the difference between (or sum of) the two corresponding number of degrees of freedom. The degrees of freedom of a source of variation can also be determined from its structure, regardless of whether it contains a single factor or several factors combined in a nested or crossed manner.

The following are rules for determining degrees of freedom of a source in balanced designs.

1) A source containing a single factor has degrees of freedom one less than its number of levels
2) A source containing nested factors has degrees of freedom equal to the product of the number of levels of each factor inside the parentheses، and the number of levels minus one of each factor outside the parentheses.
3) A source containing crossed factors has degrees of freedom equal to the product of the number of levels minus one of each factor in the source.
4) The total degrees of freedom in a model is one less than the total number of observations.

Mean square is an ''average'' of the sum of squares. It is an estimate of the variance of that source. For instance, dividing the sum of squares by its degrees of freedom yields the mean sum of squares (MS).

3.2. Anova assumptions

The objective of a study using the method of ANOVA cannot be achieved unless some assumptions are satisfied [13] However, Analysis of variance assumes the following:

1) Equality of variance (homoscedasticity). Errors should have a common variance (σ^2)
2) Normality. Error should be normally distributed with zero mean and σ^2 variance.
3) Independence. Errors are independent for all. Therefore, one should check the availability of the assumptions before

starting the analysis through the suggestive tests (Bartlett and Cochran 1947), and (Hahn and Meeker 1993).

4. Data, outcomes and inference

This paper will examine the following question: "What is the relationship between one's highest education level attained and current income?" Do all levels of education lead to higher income? Or do certain education qualifications lead to greater increases in income?

To examine the study questions, data from household survey (HHS) in certain locality in South Darfur state (Beleil & Bolbol) is used. The HHS is a survey used to collect data on socio-economic characteristics of residents of the South Darfur State (Sudan).

Data collection for the HHS was conducted through face-to-face interviews.For this paper, the two variables studied are the level of education attained ("household head education level") and total family income in Sudanese currency SDG ("income"). Education level is a categorical variable with 4 levels (i.e., "Illiteracy", "Primary School", "High Secondary School" and "University"). Income is a continuous variable.

Full probability sampling, where every household had a chance of being selected is conducted. The sampling method was stratified sampling; the state was stratified first by locality followed by admin unit and then by villages. With regard to experimental design, there was no random assignment of households to different conditions or treatments. The population of interest is the Southern Darfur population. As full probability sampling was conducted, the findings can be generalised to the entire Southern Darfur population. A majority of the Sothern Darfur state population (household head) has an education level of primary and less. Summary and Density Distribution of HHS Current Income is as follow:

Table 1: Descriptive Statistics

Source	N	Range	Min	Max	Mean	S.D
illiteracy	60	1350	650	2000	1226.7	422.1
Primary	60	1900	600	2500	1230.0	406.1
h.secondary	60	1850	650	2500	1313.3	450.3
university	32	800	1200	2000	1468.8	272.9

For the illiteracy the min is 650, max is 2000, mean is 1227, and the range income is 1350. For the primary level the min is 600, max is 2500, mean is 1230, and the range income is 1900. For the high secondary level the min is 650, max is 2500, mean is 1313, and the range income is 1850. For the university level the min is 1200, max is 2000, mean is 11469, and the range income is 800.

The hypotheses for this study are as stated below:

Null Hypothesis: The mean income is the same across all education levels.

Alternative Hypothesis: At least one mean income is different from each other.

As the means between more than two groups (i.e., four) will be compared, the study will use the ANOVA. The ANOVA analysis will compare the means across the four groups and determine if the observed differences are due to between-group variability (i.e., education) or within-group variability (other factors).

As stated in the above methodology there are three conditions for analysis of variance ("ANOVA") to be met, namely (i) independency, (ii) approximate normality, and (iii) equal variance. For (i), the data was randomly sampled with full probability sampling, and the sample size of each education group is unequal independent of each other. For (ii), the sample size is greater than 30, even between each group, according to the statistical central theorem this will represent the condition of normality even if the normality condition is violated. For (iii), the variances between each group is equal to each other, checked by levene's test with (p-value = 0.241).

Table 2: Anova of Current Income and Education

Source	SS	df	MS	F	Sig
Between Groups	1511413.5	3	503804.5	3.04	0.03
Within Groups	34516416.7	208	165944.3		
Total	36027830.2	211			

The p-value from the ANOVA, table (2) is equal to 0.03 (i.e., < 0.05). Thus, we reject the null hypothesis, at the alpha = 5% significance level, and conclude that the data provides convincing evidence that at least one income means for all groups are different from each other. To determine which education levels differ in mean incomes, we did a multiple comparisons and examine the pairwise tests with a modified significance level of 5% which based on a post-hoctest (i.e., Least Significance Difference) (LSD), table (3). At the 5% significance level, p-values from two pairwise tests are significant, this is mainly between the group illiteracy and university, primary and university, which are indicated that there is a significance difference between the amounts of income earning by each group. For the rest of combinations between the other groups, p-value showed no significance differences in terms of the mean income. Thus, we conclude that the data provides convincing evidence that mean income is different across certain education levels (i.e., illiteracy & university, primary & university) while it provides no significance relations between the rest of education level (i.e., illiteracy & primary, illiteracy & high secondary, primary& high secondary, high secondary & university).

Table 3: Multiple Comparisons

(I) code	(J) code	Mean difference (I-J)	Std.Error	Sig.	95.0% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-3.333	74.374	0.9643	-149.95	143.290
	3	-86.667	74.374	0.2452	-233.29	59.957
	4	-242.083	89.171	0.0072	-417.87	-66.288
2	1	3.333	74.374	0.9643	-143.29	149.957
	3	-83.333	74.374	0.2638	-229.95	63.290
	4	-238.750	89.171	0.0080	-414.54	-62.955
3	1	86.667	74.374	0.2452	-59.957	233.290
	2	83.333	74.374	0.2638	-63.290	229.957
	4	-155.417	89.171	0.0828	-331.212	20.378
4	1	242.083	89.171	0.0072	66.288	417.878
	2	238.750	89.171	0.0080	62.955	414.545
	3	155.417	89.171	0.0828	-20.378	331.212

LSD output

5. Conclusion

To summarize the findings of the HHS data, there is a significant and positive relationship between higher-educationlevels, mainly the university level and income. In regard with low education level, higher-educationqualifications "University level" leads to higher income. Notwithstanding, it should be noted that there is no significant difference in income between (i.e., illiteracy & primary, illiteracy & high secondary, primary & high secondary, high secondary & university). However, this analysis does not imply that income is dependent solely on education level.

Acknowledgement

Special thanks to my colleague Elsiddig Alsadig for his greatest efforts that helps me more in the publication of this paper.

References

[1] Gary S. Becker, Human Capital, A Theoretical and Empirical Analysis, Volume Publisher: NBER, Volume ISBN: 0-226-04109-3, (p. 13 - 44), http://www.nber.org/books/beck75-1

[2] Topel, Robert (1994a), "Regional Labor markets and the Determinants of Wage Inequality". American Economic Review; V.84-#2, pp. 17-22.

[3] Galor, O. and J. Zeira (1993), "Income Distribution and Macroeconomics," Review of Economic Studies, 60: 35 −52. http://dx.doi.org/10.2307/2297811.

[4] Chinhui Juhn, Kevin M. Murphy and Brooks Pierce, "Wage Inequality and the Rise in Returns to Skill", Journal of Political Economy, 1993, vol. 101, issue 3, pages 410-442 http://dx.doi.org/10.1086/261881.
[5] Perotti (1993), "Income distribution, political instability and investment". NBER working paper # 4486
[6] Filmer and Pritchett (1999), "The Effect of Household Wealth on Educational Attainment", Evidence from 35 countries. Population and development review: 85-120 http://dx.doi.org/10.1111/j.1728-4457.1999.00085.x.
[7] Checchi, D (2006), the Economics of Education. Human Capital, Family Background and Inequality. http://dx.doi.org/10.1037/h0041412.
[8] Boneau, C. A. (1960). "The effects of violations of assumptions underlying the t test", Psychological Bulletin, 57, 49-64.
[9] Rosie Cornish, 2006. Mathematics learning, statistics, one way analysis of variance.
[10] Beatrice d'Hombres, Anke Weber and Leandro Elia (2012), "Literature review on income inequality and the effects on social outcomes", JRC scientific and policy reports. http://www.jrc.ec.europa.eu/
[11] Checchi D. (2003). Inequality in Incomes and Access to Education. A Cross-country Analysis (1960-95). Labour, 17(2), 153-201.http://dx.doi.org/10.1111/1467-9914.00226.
[12] Chevalier, A., Harmon, C., O'Sullivan, V., & Walker, I. (2005). The Impact of Parental Income and Education on the Schooling of Their Children. WP05, Institute for Fiscal Studies.
[13] Cochran (1947), "Some Consequences When Assumptions for the Analysis of Variances are not satisfied", Article in Biometrics 3(1):22-38.http://dx.doi.org/10.2307/3001535.