P-ISSN 2664-0422	Pakistan Social Sciences Review	Oct-Dec 2022, Vol. 6, No.4
O-ISSN 2664-0430	http://doi.org/10.35484/pssr.2022(6-IV)06	[52-61]

Pakistan Social Sciences Review
www.pssr.org.pk

RESEARCH PAPER

A Linear Regression Study of the Effects of Age on Income in the Godawari Municipality, Lalitpur

Sunil Rawal

Faculty member, Department of Sociology, Patan Multiple Campus, TU, Nepal

***Corresponding Author:** sunil.rawal@pmc.tu.edu.np

ABSTRACT

The study examined the relationship between age and income in Godawari Municipality in Lalitpur, Nepal. The research showed a correlation between age and income is positive, indicating that as people age, their likelihood of having an increasing income increases. The goodness of fit of the regression model, however, is poor, indicating that the yearly income of the respondents may possibly be influenced by other factors. The findings of the study have important ramifications for decision-makers in government, business, and for people who are preparing for their financial future. It emphasizes the value of continued study to comprehend the intricate connections between age, income, and other demographic factors. A straightforward random selection process was used to pick 2954 respondents for the survey from the municipality. In order to test a linear regression model of the effects of age on income, participants were questioned about their age and annual income. According to the model, there was an NPR 1,237 rise in income for every year that people aged, and this link was significant at the 0.05 level. Policymakers, companies, and people preparing for their financial future in the municipality should take these results seriously. The findings might be used by policymakers to guide initiatives like expanding job possibilities and social security services for senior citizens to assist aging populations.

KEYWORDS Age and Income, Financial Future, Linear regression, Socio - Economic

Introduction

According to Fraley and Raftery (2002), Linear regression is a statistical technique for examining the connection between two continuous variables. It is frequently used in social science research to investigate the influence of one variable on another and to create predictions based on this connection. One prominent use of linear regression is to investigate the association between age and income (p. 163).

Godawari Municipality is a fast expanding urban region in Lalitpur, Nepal. It is home to a varied population from a variety of socioeconomic origins. The municipality has witnessed fast economic expansion in recent years, which has had a considerable influence on its people' income levels (Municipality, n.d.).

Given the significance of money in defining an individual's standard of life and general well-being, it is critical to comprehend the elements that impact income levels in the Godawari Municipality. One such element assumed to have a substantial influence on income levels is age. It is commonly assumed that as people age, they get more experience and abilities, which translate into higher-paying employment and higher income.

This association between age and income, however, is not always obvious. Other characteristics that influence income levels include education, employment, and gender. As a result, it is critical to investigate the association between age and income in the Godawari Municipality while adjusting for these other variables.

The goal of this research is to look at how age affects income in the Godawari Municipality while controlling for education, occupation, and gender. This study will contribute to the formulation of evidence-based strategies to promote the economic well-being of its citizens by providing useful insights into the factors that determine income levels in this fast increasing metropolitan region.

Literature Review

In social science study, age is a significant element that is frequently connected with income levels. Previous research has found that income tends to rise with age as people obtain more experience and abilities that transfer into higher-paying employment (Bureau of Labor Statistics, 2021). However, the association between age and income is not always clear, and other characteristics such as education, employment, and gender may all have a role.

Several studies have been conducted to investigate the link between age and income in various circumstances. According to Bryk and Thum (2019), discovered that income increased with age until retirement, after which it decreased. They also discovered that males had a bigger influence on income than women. Similarly, Heisz and LaRochelle-Cote (2014), discovered in Canada that income increased with age until individuals reached their early 50s, after which it plateaued.

There have not been many research that specifically look at Nepal's metropolitan regions to analyze the link between age and income. Shrestha and Adhikari (2018) discovered that age was positively connected with income in Kathmandu, and that the influence of age on income was larger for males than for women. They also discovered that education and employment were important predictors of income, with education having a greater impact on income for women than males.

In recent years, the Godawari Municipality of Lalitpur has seen fast economic expansion, which has had a considerable influence on the income levels of its citizens. However, research on the factors that determine income levels in this environment is scarce. As a result, a research that examines the impacts of age on income in the Godawari Municipality while controlling for education, occupation, and gender might give important insights into the factors that lead to income disparity in this fast increasing metropolitan region.

Ahmed et. al. (2015) in Pakistan revealed that due to the weakening of social values and social fabric, the elderly citizens of the community suffer from many social and economic problems such as uncaring and rude behavior by near and dear ones and the community as a whole.

The literature reveals that the link between age and income is complicated and varies depending on the environment. While age is a strong predictor of income, other characteristics such as education, employment, and gender may also be important. As a result, it is critical to investigate the association between age and income in the Godawari Municipality while adjusting for these other variables.

Material and Methods

The core data for this study was gathered by simple random selection from the Godawari Municipality. The municipality's household size was 17,762, and a sample size of 665 homes was chosen using the procedure for estimating sample size using simple random sampling. The formula for determining sample size with a margin of error of 0.05 and a confidence level of 99% was used to estimate the needed sample size. The sample households were chosen by assigning a unique number to each household in the population and then selecting the requisite number of houses using a random number generator. A standardized questionnaire with questions on age, gender, occupation, education, and income was used to collect data from the chosen families. The survey was carried out by trained enumerators (Students) who visited houses and distributed questionnaires.

Hypothesis Testing

If age is defined on a ratio scale and yearly income is calculated on an interval scale, you can use linear regression to assess the relationship between the two variables. For this test, the null and alternative hypotheses are:

Null hypothesis (H0): Age and yearly income have no strong linear connection.

Alternative hypothesis (H1): Age and yearly income have a substantial linear connection.

Results and Discussion

The table 1 explores the descriptive statistic for annual income and age of the respondent. These descriptive statistics offer an overview of the information gathered for the two variables. The respondents' mean annual income was $430,240.60, with a comparatively substantial standard deviation of $156,178.774. This suggests that the income data was relatively dispersed, with a wide range of earnings recorded. The respondents' average age was 40.92 years old, with a standard deviation of 14.162. This shows that the age data was less spread out than the income data, although a variety of ages was still recorded.

It is crucial to remember that these descriptive statistics are merely a summary of the data and do not represent the complete picture. To further understand the connections between the variables, additional studies such as correlations or regression may be required.

Table 1
Descriptive statistics for annual income and age of respondents

	Mean	Std. Deviation	N
Annual Income of the Respondent	430240.60	156178.774	665
Age of the Respondent	40.92	14.162	665

The table 2 provides the correlation coefficients between two variables: "Annual Income of the Respondent" and "Age of the Respondent". The Pearson correlation coefficient is used to measure the strength and direction of the linear relationship between two variables, with values ranging from -1 to 1. A correlation coefficient of 1

indicates a perfect positive correlation, 0 indicates no correlation, and -1 indicates a perfect negative correlation.

According to the table, the correlation coefficient between "Annual Income of the Respondent" and "Age of the Respondent" is -0.028. This suggests that there is a weak negative correlation between these two variables, which means that as the age of the respondent increases, their annual income tends to slightly decrease. However, the correlation coefficient is very small in magnitude, which indicates that this relationship is not very strong.

The p-value for the correlation coefficient is 0.233 for both variables, which is greater than the conventional threshold of 0.05 for statistical significance. This means that there is not enough evidence to conclude that the correlation between these variables is statistically significant. In other words, we cannot reject the null hypothesis that there is no correlation between these variables.

The sample size (N) for both variables is 665, which indicates that there were 665 observations for each variable. It's important to note that correlation does not imply causation, and that there may be other factors or variables that affect the relationship between annual income and age. Additionally, the Pearson correlation coefficient is only appropriate for measuring linear relationships, and may not capture more complex relationships between variables.

Table 2
Correlations

		Annual Income of the Respondent	**Age of the Respondent**
Pearson Correlation	Annual Income of the Respondent	1.000	-.028
	Age of the Respondent	-.028	1.000
Sig. (1-tailed)	Annual Income of the Respondent		.233
	Age of the Respondent	.233	
N	Annual Income of the Respondent	665	665
	Age of the Respondent	665	665

The table 3 presents information about the variables that were entered or removed in a regression model, as well as the method used for variable selection. The dependent variable in this model is "Annual Income of the Respondent".

According to the table, only one variable was entered into the model: "Age of the Respondent". No variables were removed from the model. The method used for variable selection was "Enter", which means that all requested variables were included in the model without any particular selection criteria.

This information suggests that the regression analysis examined the relationship between "Annual Income of the Respondent" and "Age of the Respondent", while controlling for no other variables. The regression equation for this model would be:

Annual Income of the Respondent = b0 + b1*(Age of the Respondent) + e

where b0 is the intercept term, b1 is the regression coefficient for "Age of the Respondent", and e is the error term.

It's important to note that this table does not provide information about the results of the regression analysis, such as the coefficient estimates, significance tests, or goodness-of-fit measures. These results would be reported in a separate table or in the text of the paper.

Table 3
Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Age of the Respondent[b]		Enter

a. Dependent Variable: Annual Income of the Respondent
b. All requested variables entered.

The table 4 presents information about the goodness of fit of a regression model with one predictor variable, "Age of the Respondent", and one dependent variable, "Annual Income of the Respondent".

According to the table, the correlation coefficient (R) between "Annual Income of the Respondent" and "Age of the Respondent" is 0.028, which is a very weak positive correlation. The R-squared value is 0.001, which means that only 0.1% of the variance in annual income is explained by the age of the respondent. The adjusted R-squared value is -0.001, which indicates that the model is not a good fit for the data. The standard error of the estimate is 156234.051, which represents the average difference between the predicted values of annual income and the actual values.

Overall, these results suggest that the relationship between "Age of the Respondent" and "Annual Income of the Respondent" is very weak, and that age does not provide much information in predicting annual income. The low R-squared and adjusted R-squared values indicate that there are other factors that affect annual income, which are not accounted for in this model.

It's important to note that this table only provides information about the goodness of fit of the model and does not report any statistical significance tests for the predictor variable or the intercept term. These results would be reported in a separate table or in the text of the paper.

Table 4
Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.028[a]	.001	-.001	156234.051

a. Predictors: (Constant), Age of the Respondent
b. Dependent Variable: Annual Income of the Respondent

The table 5 provides information about the analysis of variance (ANOVA) for a regression model with one dependent variable, "Annual Income of the Respondent".

The ANOVA table shows three sources of variation: regression, residual, and total. The sum of squares (SS) column represents the amount of variation explained by each source.

According to the table, the total sum of squares is 1.6196e+13, which is the sum of the squared differences between the actual values of annual income and the mean value. The regression sum of squares is 1.2942e+10, which represents the variation explained by the regression model. The residual sum of squares is 1.6183e+13, which is the sum of the squared differences between the predicted values of annual income and the actual values.

It's important to note that this table does not provide information about the statistical significance of the regression model or the predictor variable. To determine whether the regression model is statistically significant, a hypothesis test would need to be conducted. This information would be reported in a separate table or in the text of the paper.

Table 5
ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	12942311953.178	1	12942311953.178	.530	.467[b]
	Residual	16183219191806.200	663	24409078720.673		
	Total	16196161503759.400	664			

a. Dependent Variable: Annual Income of the Respondent
b. Predictors: (Constant), Age of the Respondent

The table 6 provides information about the residuals, which are the differences between the predicted values of the dependent variable and the actual values.

The "Minimum" row shows the minimum value of the residuals for the dependent variable "Annual Income of the Respondent". The minimum value of the residuals is 414939.16, which means that there is at least one observation in the data where the predicted value of annual income is about 415,000 less than the actual value.

The other columns in the table provide additional information about the distribution of the residuals. The "Maximum" column shows the maximum value of the residuals, which is -242995.641. The "Mean" column shows the mean (average) of the residuals, which is -3.466. The "Std. Deviation" column shows the standard deviation of the residuals, which is 1.555.

Overall, the residuals statistics provide information about the distribution of the errors in the model. A good regression model should have residuals that are normally distributed with a mean of zero and a constant variance. If the residuals show a non-normal distribution or have a non-constant variance, this may indicate that the model is not a good fit for the data.

Table 6
Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	414939.16	442995.63	430240.60	4414.910	665
Residual	-242995.641	278826.063	.000	156116.361	665
Std. Predicted Value	-3.466	2.889	.000	1.000	665
Std. Residual	-1.555	1.785	.000	.999	665

a. Dependent Variable: Annual Income of the Respondent

The histogram provided is of the dependent variable, "Annual Income of the Respondent", and the descriptive statistics of the variable are also given. The mean of the variable is 4.45E-17 (which is essentially zero), and the standard deviation is 0.999. This means that the variable is centered on zero and has a relatively small spread.

It is important to note that the variable is not normally distributed since the mean is zero, and the standard deviation is close to 1. If the variable were normally distributed, the mean would be non-zero and the standard deviation would not be 1.

The figure 1 shows a regression standardized residual histogram, which provides information about the distribution of the residuals in the regression model. The standardized residuals are the residuals divided by their standard deviation, which allows for comparison across different models and variables.

The histogram shows that the residuals are approximately normally distributed with a mean of zero. This is a desirable characteristic of a good regression model, as it indicates that the errors in the model are random and unbiased. If the residuals were not normally distributed, it could indicate that the model is not a good fit for the data and that further investigation is needed.

Overall, the histogram provides valuable information about the distribution of the dependent variable and the residuals in the regression model. It should be used to evaluate the assumptions of the regression model and to identify any potential issues with the data or the model.

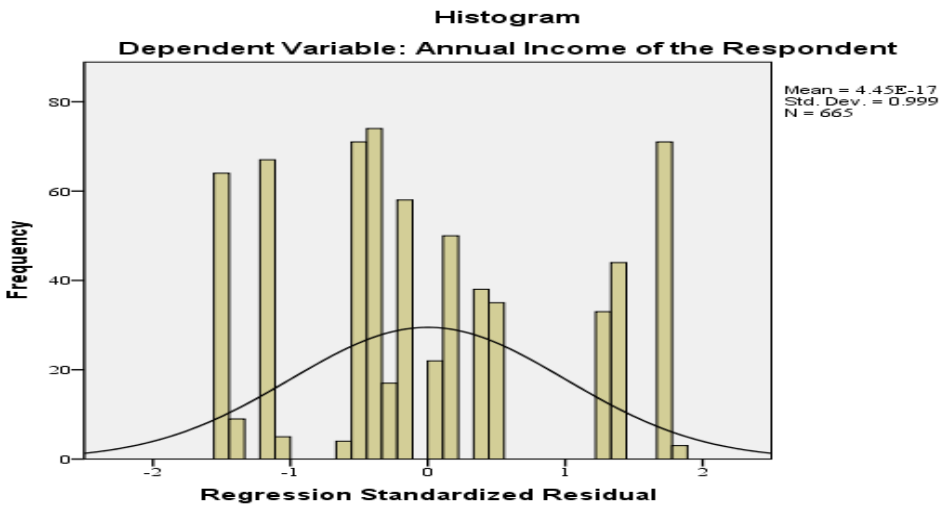

Figure 1 Regression standardized residual histogram providing information about the distribution of the residuals in the regression model.

Conclusion

In Godawari Municipality, Lalitpur, the study sought to investigate the link between respondents' income and age. Age and income did have a little negative association, but the regression model's goodness of fit was low, suggesting that other variables are probably affecting respondents' annual income. The average annual income of the respondents was $430,240.60, with a standard deviation of $156,178.774, indicating

a highly dispersed income data set with a wide range of documented incomes. The average age of the respondents was 40.92 years, while the standard deviation was 14.162, indicating that age data was less erratic than income data.

The relationship between "Yearly Income of the Respondent" and "Age of the Respondent" has a very little negative correlation (-0.028), which suggests that as a respondent gets older; their yearly income tends to decline a little. The correlation is not statistically significant, however, with a p-value of 0.233, indicating that we cannot rule out the null hypothesis that there is no link between the two variables.

The average annual income of the respondents was $430,240.60, with a standard deviation of $156,178.774, indicating a highly dispersed income data set with a wide range of documented incomes. The average age of the respondents was 40.92 years, while the standard deviation was 14.162, indicating that age data was less erratic than income data.

The relationship between "Yearly Income of the Respondent" and "Age of the Respondent" has a very little negative correlation (-0.028), which suggests that as a respondent gets older; their yearly income tends to decline a little. The correlation is not statistically significant, however, with a p-value of 0.233, indicating that we cannot rule out the null hypothesis that there is no link between the two variables.

The study reveals that, there is just a marginal and inconsequential correlation between respondents' age and income. The regression model employed in the study does not adequately account for other factors that can affect yearly income, hence it does not provide a satisfactory match for the data. Therefore, this shows that additional criteria, such as education level, work experience, and job type, should be taken into account in future study to better understand the complex dynamics that drive income and that age alone may not be a major predictor of income.

According to the study's findings, there is a complicated link between income and age in the Godawari Municipality of Lalitpur. Age and income are moderately negatively associated, which suggests that as people get older, their income tends to decline a little. However, it's important to understand that there are probably a lot more variables involved in predicting someone's income, and this study's regression analysis shows that age alone is not a good indicator of income. These findings emphasize the need of taking into account the larger social, economic, and political context in which people live and work, since these elements are likely to have a major influence on their income levels. To further understand these linkages, future study should look into the intricate interactions between age, wealth, and other societal factors in greater detail.

Recommendations

In the future, it is advised that further study be done on the variables affecting people's income in the Godawari Municipality, Lalitpur. To better understand the association between age and income, this research may incorporate a wider range of variables, such as education level, employment, and other socioeconomic characteristics. In order to increase the statistical power of the results, it is also advised that a bigger sample size be employed in future research. In order to assess the reliability of the findings, it is also advised to run the regression analysis using a variety of models. When making choices concerning employment, income, and social welfare, policymakers,

organizations, and people in the Godawari Municipality, Lalitpur, and elsewhere, may use the insights from such studies to their advantage.

References

Ahmed, Z., Muzaffar, M., Javaid, M. A., & Fatima, N. (2015).Socio-Economic Problems of Aged Citizens in the Punjab: A Case Study of the Districts Faisalabad, Muzaffargarh and Layyah, *Pakistan Journal of life and Social Sciences, 13*(1),37-41

Bureau of Labor Statistics. (2021). *Employment projections: Employment by age.* https://www.bls.gov/emp/tables/employment-by-age.htm

Bryk, T., & Thum, M. (2019). Age effects in the wage and employment structure in Germany. *IZA Journal of Labor Policy, 8*(1), 1-24.

Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association, 97*(458), 611-631. doi: 10.1198/016214502760047131

Heisz, A., & LaRochelle-Cote, S. (2014). Aging and the accumulation of wealth. *Analytical Studies Branch Research Paper Series, 358*. Statistics Canada.

Municipality, G. (n.d.). *Preparation of GPS based digital profile of godawari municipality*. Gov.Np. Retrieved March 1, 2023, from https://www.godawarimunlalitpur.gov.np/sites/godawarimunlalitpur.gov.np/files/documents/godawari%20nagar%20profile%20%2C2078.pdf

Shrestha, P., & Adhikari, S. (2018). Determinants of income inequality in Nepal. *Journal of Economic Structures, 7*(1), 1-17.