Pay Gap

Student Name:  
Student Registration Number:

Institution:

**Abstract**

Pay gap has been in existence from any given organizational setting or company and it is influenced by factors such as gender, age, ethnicity, disability. As far as this is concerned, this study was aimed at determining is age is still a significant factor in determining pay gap using ethnicity and disability status as the control variables. The data set consisted of 718 observations and it was a pre-collected data set. STATA 13 was used for analysis whereby descriptive summary analysis, correlation, simple and multiple linear regression analysis were performed on the data set. From descriptive statistics, Shapiro-Wilk test had p<0.05 indicating all variables of the data set were normally distributed. Correlation analysis was used to check for multicollinearity and it was found that there was no multicollinearity as all of the independent variables had weak correlations (r<0.5/r<-0.5). For the simple linear regression, age (β=931.91, p>0.05). For multiple linear regression, age (β=877.46, p>0.05), ethnicity (β=110.97, p>0.05) and disability (β=-44355.36, p>0.05) indicating age and both control variables are not significant at determining pay gap. Also, age and ethnicity have a positive relationship with pay gap while disability status have negative association with pay gap. In future studies, focus should be on determining factors that can significantly explain or predict pay gap.

**Introduction**

*Background and Literature*

In the modern day organizational setting, data analytics is a significant factor of consideration because it provides insights that help decision makers make well-informed and appreciate decisions for the prosperity of their respective companies or organization (Akter et al. 2016). In this case study, we will look at the issue of pay gap or pay disparities and which factor or factors are significantly behind it. At any company or employment place, issue of pay gap exist and we cannot run away from it (Blau & Kahn, 2020). There are those employees who earn more than the others and this might be attributed to employee experience, employee role, gender, religion, age, nationality, ethnicity, disability status, socio-economic status and even hate to say it, unethical issues such as favourism, nepotism, tribalism (Longhi & Platt, 2008). For instance, let look at how age affects job disparities as according to Klomega & Fleming (2014), an older person will earn much salary compared to the younger employee because this older employee has experience and even holds senior positions at the job and also, most of the younger people are still attending schools than older people hence they will definitely have less income. Another significant factor leading to pay gap is gender. According to Klomega & Fleming (2014), men receive higher wages and high paying jobs compared to women due to a number of factors. A few of these factors are; income for women is lowered because they spend much time tending to children at home, allocation of jobs that are low paying to women, unlike men, some jobs do not require women and additionally, we have still existence of stereotypes about which work and job titles women should do hence significantly reducing women status in employment field. According to Hills (2010), economic outcome can be measured by status of employed adult, educational outcome, personal income, wealth as well as weekly or hourly earnings.

Among those factors leading to pay gap, this study was interested in determining the factor of age on earnings of individuals. This is because things have changed in the modern day whereby we find young business millionaires and multi-millionaires (Johnstone et al., 2016). This can be attributed to the advancement in technology and the opportunities brought by it for instance, Forex trading, drop shipping, you tubing, affiliate marketing and many more (Otamiri & Temelaye, 2022). Hence it can be seen that there is a significant evolution in the association between income, pay gap and age and hence the reason as to why this paper was interested in determining if there still exist a significant relationship between age and pay disparity. Also, there exist a research gap whereby majority of the studies have exhausted on the relationship between gender and pay gap but not much studies have been performed on the relationship between age and pay gap. Hence, this is another reason of the study, to maximize of this existing research gap.

*Hypothesis*

Ho: Age does not significantly explain or predict pay gap.

H1: Age significantly explain or predict pay gap.

**Data and Methods**

*Research Design*

The three mostly used research designs are qualitative, quantitative and mixed-method research designs. Qualitative design is whereby non-numerical data set is used for analysis while quantitative design is whereby numeric data set is used for analysis (Bullock et al. 2017). In this case study, qualitative design was used because the study analyzed numerical data.

*Data*

Primary and secondary data are the two main data sources. Primary data is one collected for the first time for instance through interviews, surveys, observations by the research while secondary data is one that has already been collected, stored and ready to be used for research (Chawla, 2020). In this study, secondary data was used whereby a pre-collected data set called Understanding Society was used. The data set has four variables namely age, ethnicity, disability status and ethnicity. The sample size for the data set was 718 large enough to avoid issues brought by small sample sizes such as biased results due to poorly under-represented population. The data set was collected across nations hence it had good representation. The data set had to go preprocessing such as removing missing observations as well as transformation whereby the variables ethnicity and disability status were transformed through mean function to find an average value of the variables.

The variable of interest or the dependent variable was ‘earn’ which was collected by asking various individuals to record their monthly salaries via surveys. ‘Age’ was the independent variable because it is the one we are interested if it can predict ‘earn’. An individual was supposed to record their age in years during the surveys as well. The control variables were ethnicity and disability status that were as well collected during the survey. This is because according to Brynin & Guveli (2012), the salary of employee can significantly be influenced by their ethnicity regardless of their age, for instance, in cases of favors at work though nepotism. For disability, we have cases whereby the disabled are given priorities whenever there is a job application and pay in some jobs and in other jobs they are denied the priorities regardless of age (Kruse, 2018).

Still on data, ethics was an important factor of consideration especially given that the data set involved human subjects. Therefore, confidentiality of the data was guaranteed whereby it was only used for the purpose of research and no record of an individual was exposed. Again to ensure confidentiality, there was no primary key that identified the participants hence keeping them unidentifiable.

*Statistical Analysis Techniques*

The preprocessed data set was analyzed in STATA 13 software. The first analysis was descriptive analysis in order to get a general understanding of the data set. The summary analysis statistics performed were mean, standard deviation and Shapiro-Wilk test of normality (Mishra et al. 2019). Still to descriptive analysis, exploratory analysis of the selected variables was also done whereby the histograms of the variables was done.

After descriptive analysis of the data, inferential analysis was then performed. The inferential analysis was regression analysis to determine the effect of age on earning. Before performing this predictive analytics, correlation analysis was done to find out the relationship among the independent variables (main independent variable and the two control variables). In regression modeling, relationship between independent variables is not supposed to be strong because it brings about multicollinearity effect in the regression model distorting the results. So in such cases of two strongly related independent variables, only one variable is included in the model and the other is dropped because it implies that both variables explains the same thing (Daoud, 2017). However, a strong correlation between dependent and independent variable is good. The criterion for strong correlation is when the correlation coefficient between two variables is greater than 0.7.

After testing for multicollinearity, single and multiple regression model was then done. For single regression model, the dependent variable was ‘earn’ while the independent variable was ‘age’. The purpose here was to determine the direct effect of age on the earning or pay of individuals. For multiple regression model, dependent variable was ‘earn’ and independent variables were ‘age’, ‘disability status’ and ‘ethnicity’, simply put, the multiple model included control variables as the independent variables to help in determining how they affect pay gap in the society.

**Findings and Discussion**

*Descriptive Statistics*

From table 1, the variable earn had (M=32479.76, SD=27473.57) and we can clearly see a significantly large standard deviation value in relation to the mean value. This has an implication that there was a large degree of fluctuation of income among the participants from the mean income values, it simply means income varies a lot across different people and this might be due to various factors underlined such as age, gender, ethnicity, disability status, experience (Mishra et al. 2019). For the variable age, we had (M=16.58, SD=0.51) and it can be seen that the standard deviation is not significantly large in relation to the mean value and this has an implication that there was no significant fluctuation in age among the participants of this research. For the variable ethnicity, we had (M=4.14, SD=6.69) and we clearly see a significant large standard deviation even greater than the mean value. This implies that ethnicity was highly volatile or had high degree of fluctuation and this means that the participants in the study were from various different ethnic groups around the world hence representing nationality. Lastly disability status had (M=0.08, SD=03) and we see standard deviation was not significantly large from the mean value. This means disability status of the individual did not have high degree of variation because after all, one is either disable or not and so there is no much variation (Mishra et al. 2019).

To determine normality of the variables, Shapiro-Wilk test was used (Aslam, 2021). The null hypothesis for this test is that data is not normally distributed while the alternative hypothesis is that data is normally distributed. If p<0.05, null hypothesis is rejected and hence data is normally distributed. In our case study, p<0.05 for all variables implying that earn, ethnicity, disability status and age were all normally distributed (Aslam, 2021).

On exploration of distribution of variables, histogram plots were done. Figures 1, 2 and 4 indicated that the variables earn, ethnicity and disability were skewed to the right side implying that the variables had larger tails on the right. Figure 3 indicated that the distribution for age was not clear to determine, neither was it skewed to the left or right side.

*Inferential statistics*

*Correlation analysis*

Pearson’s correlation was performed to determine the degree of association among the variables (Schober et al. 2018). The correlation coefficient lies between -1 and 1 implying that two variables can be negatively or positively correlated and a correlation coefficient of 1 shows correlation between a variable and itself. A correlation coefficient greater than 0.5 or -0.5 and approaching 1 or -1 shows a strong positive or negative association respectively (Schober et al. 2018). From figure 5 in the appendix, correlation between earn and ethnicity was 0.027, correlation between earn and age was 0.018, correlation between earn and disability was -0.05, correlation between ethnicity and age was 0.061, correlation between ethnicity and disability was 0.018 and correlation between age and disability was 0.018. It be can be seen there was no strong correlations between the independent variables confirming no threat of multicollinearity in the regression model if all of the variables were to be included in the modeling.

*Model 1: Simple Linear Regression Modeling.*

The variable of interest was ‘earn’ while primary independent variable was age. From figure 6 in the appendix, R-Squared value for the model was 0.03% implying that the model only explained 0.03% variability of the dependent variable and this shows a poor model (Sarstedt & Mooi, 2019). The F (1, 716) = 0.22, p=0.64 and since p>0.05, the model was not generally good in fitting the data set (Sarstedt & Mooi, 2019).

The independent variable age had (β=931.91, p>0.05). The positive coefficient of 931.91 meant that there is a positive relationship between age and earn, an increase or decrease in age will lead to an increase or decrease in earning amount or income (Sarstedt & Mooi, 2019). In our case study, an increase in age by 1 years will lead to an increase in monthly income or pay by $931.91. This finding agrees with the related works done by Klomega & Fleming (2014) who argues that older people will earn more than younger people because of the experience difference and also because younger people are more focused on studies. However, the p-value was less than alpha 0.05 indicating that age is not significant at determining income or pay gap (Sarstedt & Mooi, 2019). This means that we cannot rely on age to predict income of an individual.

*Model 2: Multiple Linear Regression Modeling*

From figure 7, R-Squared value was 0.38% which was an improvement from the simple linear regression model and this indicates that the control variables were significant in improving the model (Sarstedt & Mooi, 2019). However, the model is still poor because the independent variables explained only 0.38% variability of the dependent variable. F (3, 713) = 0.90, p= 0.44 and since p>0.05, the model generally did not fit the data well (Sarstedt & Mooi, 2019).

The independent variable age had (β=877.46, p>0.05) hence again a positive relationship between age and income (Sarstedt & Mooi, 2019). An increase of an individual by 1 year will lead to an increase in monthly income by $ 877.46 which still agrees with the various related work. The p-value is still less than 0.05 implying even in the presence of control variables, age is not significant at explaining or predicting income or pay gap (Sarstedt & Mooi, 2019). Therefore, the null hypothesis was true; Age does not significantly explain or predict pay gap. This agrees with our doubt that age is no longer important in determining the earning of an individual. As earlier stated, this can be explained by technological advancements that has brought many opportunities to the young people making them millionaires and owners of lucrative properties at a younger age. For instance, forex trading, TikTok, YouTube, Instagram are among platforms that have raised young millionaire influencers who are even approached by big brands to work together (Plebani, 2015). Technological advancements have also made innovations and creativity much easier for instance, the likes of Mark Zuckerberg who found Facebook at a younger age making him a millionaire, Elon Musk of Tesla who started by inventing PayPal making him a multi-millionaire at a younger. This proves that in the modern day environment, anyone can be whatever they want and they can decide their own wealth with no limitation as age does not matter anymore.

For the control variables, ethnicity had (β=110.97, p>0.05) hence ethnicity has a positive relationship with pay gap (Sarstedt & Mooi, 2019). It means that an increase in ethnicity by 1, we will have an increase in pay gap by $110.97. This agrees with related work done by Brynin & Guveli (2012) whereby due ethnicity advantage between an employee and employer, employee may receive higher salaries that the rest which similar happens with tribalism, nepotism and any kind of favourism. However, p<0.05 implying ethnicity is not significant at explaining or predicting pay gap. The variable disability had (β=-44355.36, p>0.05) hence a negative relationship with pay gap. From the results, an increase in the rates of disability by a value of 1 will lead a decrease in pay gap by $ 44355.36 per month. However, the p-value is greater than 0.05 implying disability status is not significant or reliable in predicting pay gap or income of individuals.

The results of this study have great implication to any interested stakeholder or the general public. From this results, individual will get illumination that one can be rich and wealth at any given age and that you do not wait to be old for you to become rich and wealth. The dynamics in the modern world have changed whereby the technological advancements has brought many opportunities making it possible for one to become rich and wealth at any given age. Hence age does no longer explain, determine or predict pay gap or wealth of an individual. For future studies, I suggest more researches to be conducted to determine why age, ethnicity and disability status are significant predictors of pay gap and find out which factors are significant at explaining or predicting pay gap.

**Conclusion**

There is no doubt that data analytics indeed is vital in providing insights that help decision makers make well-informed and appropriate decisions. Pay gap exist and we cannot run away from it due to several factors that affect it, for instance, age, ethnicity, disability status, gender. Therefore, the purpose of this study was determine if age is still significant at determining pay gap or income in the modern day environment. Ethnicity and disability status were used as the control variables. The analysis was done in STATA 13. From the summary analysis results, the Shapiro-Wilk test confirmed that all of the variables were normally distributed hence ready for analysis. Pearson’s correlation was used to check for any possibility of multicollinearity effect in the regression model. The results indicated that there was no strong correlation between the independent variables hence no multicollinearity implying all variables could be used in regression modelling. For the regression modelling results, both simple and multiple linear regression modelling indicated that age is not a significant factor in determining in explaining ae gap. Also, the control variables were not significant at explaining or predicting pay gap. From the results, anyone can be rich at any given age and this is best explained by technological advancement that has brought with it pleasant opportunities making them to earn huge sums of money at a younger age. Therefore, in this modern creative and innovative era, one does not wait to be old for them to start earning big, but rather, young individuals are utilizing the opportunity from technology and earning big, the likes of Elon Musk, Mark Zuckerberg, Jack Ma.

**References**

Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R., & Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment? *International Journal of Production Economics*, *182*, 113-131. <https://doi.org/10.1016/j.ijpe.2016.08.018>.

Aslam, M. (2021). Analysing Gray Cast Iron Data using a New Shapiro-Wilks test for Normality under Indeterminacy. *International Journal of Cast Metals Research*, *34*(1), 1-5. <https://doi.org/10.1080/13640461.2020.1846959>.

Bullock, R., Little, M., & Millham, S. (2017). The relationships between quantitative and qualitative approaches in social policy research. In *Mixing methods: qualitative and quantitative research* (pp. 81-99). Routledge. <https://www.taylorfrancis.com/chapters/edit/10.4324/9781315248813-4/relationships-quantitative-qualitative-approaches-social-policy-research-roger-bullock-michael-little-spencer-millham>.

Blau, F. D., & Kahn, L. M. (2020). The gender pay gap: Have women gone as far as they can? In *Inequality in the United States: A Reader* (pp. 345-362). Routledge. <https://www.taylorfrancis.com/chapters/edit/10.4324/9781003071709-31/gender-pay-gap-women-gone-far-francine-blau-lawrence-kahn>.

Brynin, M. & Güveli, A., 2012. Understanding the ethnic pay gap in Britain. *Work, employment and society*, *26*(4), pp.574-587. <https://doi.org/10.1177%2F0950017012445095>.

Chawla, P. (2020). Research methods to understand the ‘youth capabilities and conversions’: the pros and cons of using secondary data analysis in a pandemic situation. In *Researching in the Age of COVID-19* (pp. 95-104). Policy Press. <https://doi.org/10.51952/9781447360391.ch009>.

Daoud, J.I., 2017, December. Multicollinearity and regression analysis. In *Journal of Physics: Conference Series* (Vol. 949, No. 1, p. 012009). IOP Publishing. <https://iopscience.iop.org/article/10.1088/1742-6596/949/1/012009/meta>.

Hills, J., 2010. An anatomy of economic inequality in the UK-report of the national equality panel. *LSE STICERD Research Paper No. CASEREPORT60*. <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1546894>.

Johnston, L., Liu, X., Yang, M., & Zhang, X. (2016). Getting rich after getting old: China’s demographic and economic transition in dynamic international context. *China’s New Sources of Economic Growth*, *1*, 215-246.

Klomegah, R. & Fleming, N., 2014. Pay Inequity: A Comparative Analysis of Pay Inequality in the United States by Selected Correlates. *The Journal of Public and Professional Sociology*, *6*(1), p.3. Available at: <https://digitalcommons.kennesaw.edu/jpps/vol6/iss1/3>.

Kruse, D., Schur, L., Rogers, S. & Ameri, M., 2018. Why do workers with disabilities earn less? Occupational job requirements and disability discrimination. *British Journal of Industrial Relations*, *56*(4), pp.798-834. <https://doi.org/10.1111/bjir.12257>.

Longhi S., & Platt, L. (2008) Pay Gaps Across Equalities Areas: An analysis of pay gaps and pay penalties by sex, ethnicity, religion, disability, sexual orientation and age using the Labour Force Survey. Institute for Social and Economic Research, Research Report 9. University of Essex.

Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. *Annals of cardiac anaesthesia*, *22*(1), 67. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6350423/>.

Otamiri, S. A., & Temelaye, D. T. (2022). INFLUENCE OF ONLINE MARKETING ON YOUTH EMPOWERMENT IN LOCAL GOVERNMENT AREAS IN RIVERS STATE. *BW Academic Journal*, 12-12.

Plebani, M. (2015). Evaluating and using innovative technologies: a lesson from Theranos? *Clinical Chemistry and Laboratory Medicine (CCLM)*, *53*(7), 961-962.

Sarstedt, M., & Mooi, E. (2019). Regression analysis. In *A Concise Guide to Market Research* (pp. 209-256). Springer, Berlin, Heidelberg. <https://link.springer.com/chapter/10.1007/978-3-662-56707-4_7>.

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia*, *126*(5), 1763-1768. <https://www.ingentaconnect.com/content/wk/ane/2018/00000126/00000005/art00051>.

**Appendix.**

1. Descriptive Summary

*Table 1: Summary Analysis Output*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | N | Mean | Std.Dev | Shapiro-Wilk (p-value) |
| earn | 718 | 32479.76 | 27473.57 | 0.00 |
| age | 718 | 16.58 | 0.51 | 0.00 |
| ethnicity | 718 | 4.14 | 6.69 | 0.00 |
| Disability status | 718 | 0.08 | 0.03 | 0.00 |

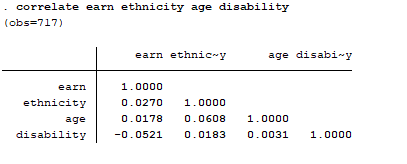
*Figure 1: Histogram Plot Earn Figure 2: Histogram Plot Ethnicity*

** **

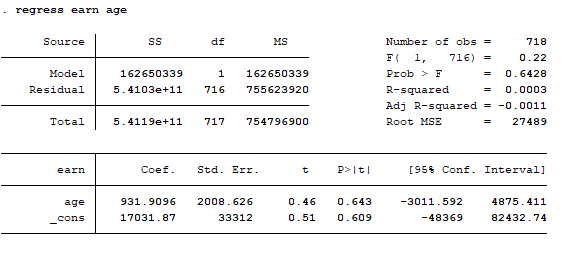
*Figure 3: Histogram plot age Figure 4: Histogram plot disability status*

*Figure 5: Correlation Matrix*



*Figure 6: Simple Linear Regression*



*Figure 7: Multiple Linear Regression*

