**Analysis of the average hourly wage data**

1. Bullet point executive summary

For the given data, by analyzing, we find some interesting facts, here we list three main characters as below:

(1) The distribution of average hourly wages in the sample almost presents an inverted pyramid shape.

(2) Among the other five variables, years of education, nonwhite and female have obvious correlations with the average hourly wage.

(3) Among the three variables with correlations with the average hourly wage, by calculation, the correlation relationship between years of education and average hourly wage is positive which corresponds to our observation.

2. Analysis

(1) Point 1

The table of frequency distribution of average hourly wages is as shown in Table.1.

Table.1 Frequency distribution of average hourly wages

| Average hourly wages | Frequency |
|---|---|
| 0.53 but less than 2.13 | 12 |
| 2.13 but less than 3.73 | 164 |
| 3.73 but less than 5.33 | 128 |
| 5.33 but less than 6.93 | 91 |
| 6.93 but less than 8.53 | 37 |
| 8.53 but less than 10.13 | 42 |
| 10.13 but less than 11.73 | 12 |
| 11.73 but less than 13.33 | 17 |
| 13.33 but less than 14.93 | 4 |
| 14.93 but less than 16.53 | 4 |
| 16.53 but less than 18.13 | 4 |
| 18.13 but less than 19.73 | 4 |

| | |
|---|---|
| 19.73 but less than 21.33 | 2 |
| 21.33 but less than 22.93 | 4 |
| 22.93 but less than 24.53 | 0 |
| 24.53 but less than 26.13 | 1 |

The histogram of average hourly wages from the sample data is plotted in Figure.1.

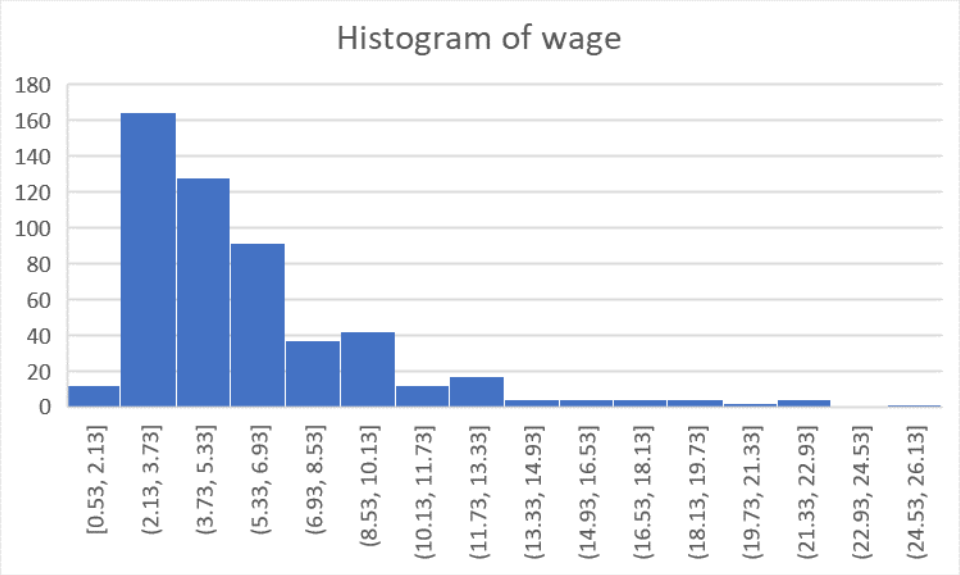


Figure.1 Histogram of average hourly wages

It can be observed that among all the wage intervals, the frequency of minimum wage interval [0.53, 2.13] is much smaller than the frequency of its adjacent interval [2.13, 3.73], while the frequencies of intervals at its right side have a common overall downward trend.

Usually, the inverted pyramid shape of people's income structure is easy to accept. Due to economic development and limited resources, the proportion of people with higher income levels in the total population will be lower. But how to explain the much lower frequency of the minimum wage interval, a possible explanation is due to existence of the minimum wage standard set by the government which has played a bottom-up role in the income level structure, when the income level structure above the minimum income level is still consistent with our consistent understanding.

(2) Point 2

We plot scatter graphs of wage versus other five variables respectively in Figure.2.

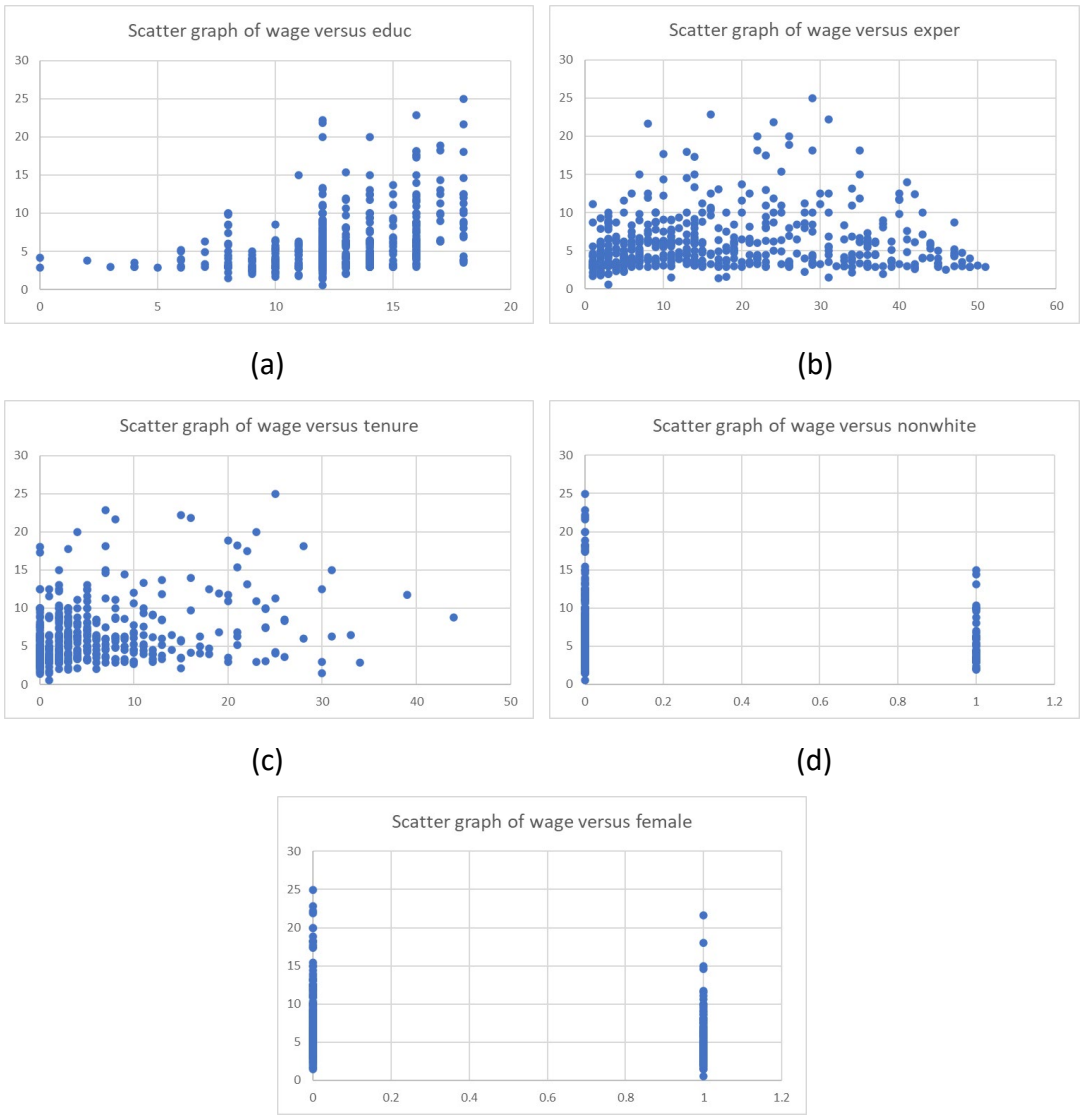


Figure.2 Scatter graphs of wage versus other five variables

From graph (a), an interesting fact can be drawn which is that education of quite long time does not prevent from low income levels, while it does greatly raise the upper limit of income levels, which can be proved by observation that those samples with higher average hourly wages basically have education experiences of many years.

From graphs (b) and (c), there are no apparent trend can be found, indicating that number of years of potential experience and years with current employer can not guarantee high levels of incomes. This is a profound lesson, despite of the loyalty spirit advocated by employers and enterprises nowadays, they seem not to really award their loyal employees. Therefore, from the perspective of the employees, learning to be flexible and find true value points may be really vital.

From graphs (d) and (e), the wage differences caused by race and gender can be observed, though the average hourly wages of nonwhite people are not necessarily lower than white people, as well that the average hourly wages of female people are not necessarily lower than male people, it can be observed that size of nonwhite samples with high income levels is much smaller than size of white samples with high income levels, and size of female samples with high income levels is much smaller than size of male samples with high income levels. This fact reveals that race and gender factors still play a significant role in the workplace, which needs urgent improvement.

(3) Point 3

Among the three variables with correlations with the average hourly wages, we choose years of education to analyze the covariance and the coefficient of correlation between the average hourly wage and it.

Let $X$ denotes years of education and $Y$ denotes average hourly wages, then the covariance is calculated by

$$\operatorname{cov}(X,Y)=\frac{\sum_{i=1}^{526}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)}{526-1}\approx 4.143$$

Since the covariance can have any value, it is improper to use it to determine the relative strength of the relationship, then the coefficient of correlation is calculated by

$$r=\frac{\operatorname{cov}(X,Y)}{S_X S_Y}\approx 0.406$$

The coefficient of correlation is positive, indicating that small values of $X$ tend to be paired with small values of $Y$, and large values of $X$ tend to be associated with large values of $Y$. Yet, the value is only 0.406, which only represents weak positive correlation. With connection to our analysis in part (2), years of education can bring with higher upper limit but no higher lower limit, it corresponds to a weak positive correlation.

3. Summary

In this report, we analyze the data given and mainly study the structure of average

hourly wages and possible relationships between average hourly wages and the other five variables with charts and tables, it turns out that the distribution of average hourly wages presents an inverted pyramid shape, and years of education, nonwhite as well as female have obvious correlations with the average hourly wages.

From our analysis, interesting facts are found, some are consistent with people's common understanding while some are not, we not only analyze possible reasons behind those phenomena but also draw some suggestions for employees from them. Through this report, the power of statistics on data is well perceived.