

Analysis of A Wages' Sample for US Male Workers

1. Introduction

The Empirical Economics conducted a Current Population Survey, which reflected real weekly wages (deflated by personal consumption expenditures) for US male workers in 1988. In this study, 2000 male workers were sampled, in terms of their years of education, races, years of working experience, living areas and whether having part time jobs. All of these nine variables are worth taking to make a comparison of wages between different groups.

Paying attention to representative variables like different living areas and races, the main analysis looks at differences in wages (for US male workers) in such two aspects. Based on data of education background, an initial conclusion of relationship between wages and length of education years is also given.

2. Analysis

First, we will have an overview of statistical features of the whole dataset by numerical summary. Next is an analysis of some specific variables through graphical summary.

2.1 Numerical Summary

wage	educ	exper	race	smsa
Min. : 50.39	Min. : 0.00	Min. : -2.00	white:1844	no_smsa: 488
1st Qu.: 308.64	1st Qu.:12.00	1st Qu.: 8.00	black: 156	smsa :1512
Median : 522.32	Median :12.00	Median :15.00		
Mean : 608.12	Mean :13.11	Mean :18.41		
3rd Qu.: 783.48	3rd Qu.:16.00	3rd Qu.:27.00		
Max. :7716.05	Max. :18.00	Max. :59.00		
ne	mw	so	we	pt
no_ne:1542	no_mw:1503	no_so:1375	no_we:1580	no_pt:1815
ne : 458	mw : 497	so : 625	we : 420	pt : 185

Table 1

As is shown in Table 1, the lowest wage in the 2000 samples is \$50.39, whereas the highest one is \$7716.05, which shows a remarkable gap. However, 3/4 of people surveyed earned less than \$800, that is to say only a small portion of people had an extremely high salary. Years of working experiences vary a lot as well, a minimum of -2 suggested that there maybe exist some missing values. Over 3/4 of the respondents came from metropolitan statistical areas. We notice that people surveyed are almost evenly divided among such four areas: ne, mw, so, we, which stand for North East, Middle West, South and West respectively. Therefore we take "living area" as the first representative variable to make more

specific analysis below.

2.2 Graphical Summary

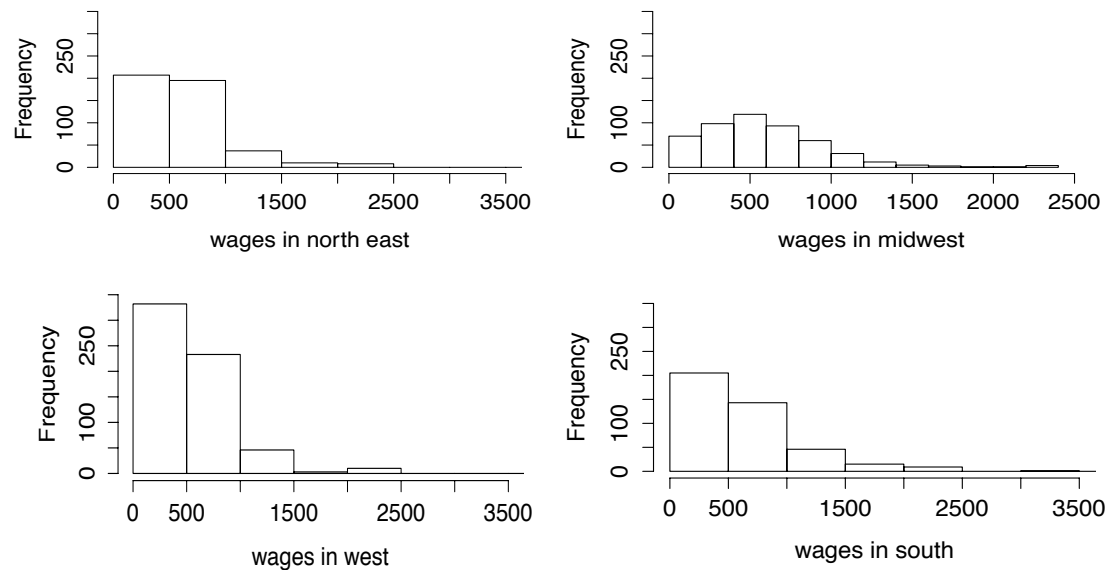


figure 1

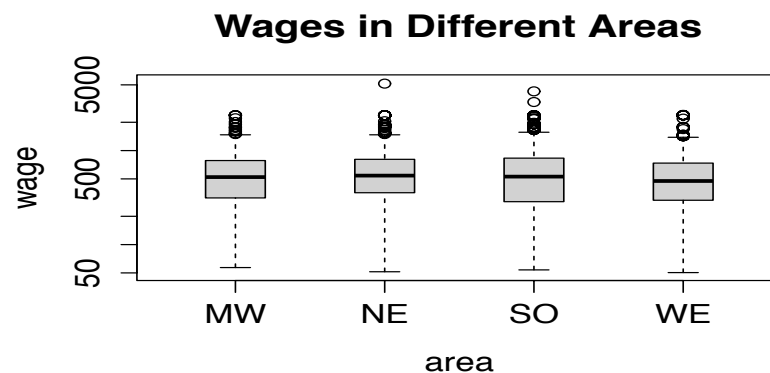


figure 2

It is clear to see frequency distribution of wages in a certain area and compare wages between areas in figure 1. People's numbers in all of the areas except in Midwest see a drop as the wages increase. A great proportion of Male workers in Midwest earned wages around \$500, which is higher than the other three's modes. People in the west has the largest proportion of wages under \$500 among the four areas. In figure 2, the average wages in the four areas are all around \$500, but wages in NE exhibit a more obvious central tendency.

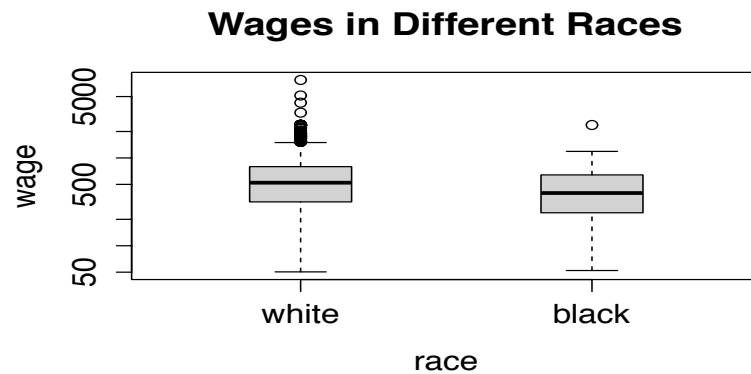


figure 3

Another boxplot shows the difference in wages between two races. It tells that the overall level of white people's wages is higher than black people's, though the lowest wages in the two groups have a slight difference.

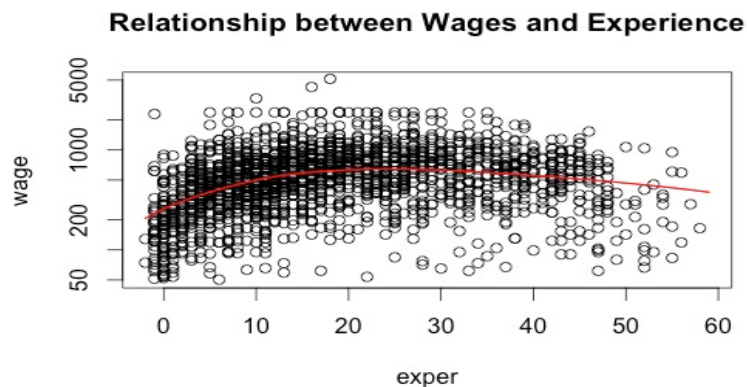


Figure 4

Based on the scatters, the last figure shows the relationship between wages and years of working experience. Below 20 years of working experience, wages will increase as the working years extend. But people's wages keep stable or even show a little downward trend above 30 years of working experience, probably due to problems with people's age growing.

3. Conclusion

Given the dataset contains a limited number of samples, and differences between groups in the same variable are still not dramatic, it is not so easy to come to any solid conclusion. But based on the analysis above, we get to know roughly about the features of the set of data. Such as 1) wages in NE exhibit a higher average and more obvious central tendency. 2) White people's wages are slightly higher than black people's in general. 3) Below 20 years, the length of working time shows a positive correlation with wages.

R-code:

```
##Load the library
library(faraway)

##read in the data
data(uswages)
attach(uswages)

uswages
summary(uswages)

##summarize areas
area<-vector(mode = "character",length=0)
for(i in 1:2000){
  if(ne[i]==1)
    area[i]<-"NE"
  else if(mw[i]==1)
    area[i]<-"MW"
  else if(so[i]==1)
    area[i]<-"WE"
  else
    area[i]<-"SO"
}

##Careforial Variable
uswages$race=factor(uswages$race)
uswages$smsa=factor(uswages$smsa)
uswages$ne=factor(uswages$ne)
uswages$mw=factor(uswages$mw)
uswages$we=factor(uswages$we)
uswages$so=factor(uswages$so)
uswages$pt=factor(uswages$pt)
summary.factor(race)
levels(uswages$race)=c("white","black")
levels(uswages$smsa)=c("no_smsa","smsa")
levels(uswages$ne)=c("no_ne","ne")
levels(uswages$mw)=c("no_mw","mw")
levels(uswages$we)=c("no_we","we")
levels(uswages$so)=c("no_so","so")
levels(uswages$pt)=c("no_pt","pt")
summary(uswages)

boxplot(wage~race,data=uswages,log='y',boxwex=0.35,col="lightgrey",main="Wages in Different
Races",ylab="wage",xlab="race")

boxplot(wage~area,data=uswages,log='y',boxwex=0.45,col="lightgrey",main="Wages in Different
Areas",ylab="wage",xlab="area",ylim=c(50,5000))

plot(exper,wage,log = 'y',main="Relationship between Wages and Experience",ylim=c(50,5000))
```

```
lines(lowess(exper,wage),col='red',lwd=1.5)
```

```
wage_ne<-vector(mode = "numeric",length=0)
wage_mw<-vector(mode = "numeric",length=0)
wage_we<-vector(mode = "numeric",length=0)
wage_so<-vector(mode = "numeric",length=0)
```

```
j<-0
k<-0
m<-0
n<-0
for(i in 1:2000){
  if(area[i]=="NE"){
    j<-j+1
    wage_ne[j]<-wage[i]}
  else if(mw[i]=="MW"){
    k<-k+1
    wage_mw[k]<-wage[i]}
  else if(we[i]=="WE"){
    m<-m+1
    wage_we[m]<-wage[i]}
  else{
    n<-n+1
    wage_so[n]<-wage[i]}
}
```

```
wage_ne<-vector(mode = "numeric",length=0)
wage_mw<-vector(mode = "numeric",length=0)
wage_we<-vector(mode = "numeric",length=0)
wage_so<-vector(mode = "numeric",length=0)
j<-0
k<-0
m<-0
n<-0
```

```
for(i in 1:2000){
  if(area[i]=="NE"){
    j<-j+1
    wage_ne[j]<-wage[i]}
  else if(area[i]=="MW"){
    k<-k+1
    wage_mw[k]<-wage[i]}
  else if(area[i]=="WE"){
    m<-m+1
    wage_we[m]<-wage[i]}
  else{
```

```
n<-n+1
wage_so[n]<-wage[i]}
}
hist(wage_mw,main = "Histogram of Wages in Midwest",xlab="wages in midwest",xlim=c(0,2500),ylim=c(0,350))
hist(wage_ne,main = "Histogram of Wages in North East",xlab="wages in north east",xlim=c(0,3500),ylim=c(0,350))
hist(wage_so,main = "Histogram of Wages in South",xlab="wages in south",xlim=c(0,3500),ylim=c(0,350))
hist(wage_we,main = "Histogram of Wages in West",xlab="wages in west",xlim=c(0,3500),ylim=c(0,350))
```