Wage-Productivity-Analysis

Paula Ramirez

22/09/2024

Wage-Productivity-Analysis

Basic Manipulation.

Basic configurations

```
#Clearing all the plots, the console and the workspace.
#Setting the overall format for numbers.
if(!is.null(dev.list())) dev.off()
```

```
## null device
##           1
```

```
cat("\014")
```

```
rm(list=ls())
options(scipen=9)
```

1. Read in the text file and change to a data frame

Reading the file and converting it to dataframe.

```
PR_data_employment <- read.table(here("Wage-Productivity-Analysis", "Analysis_Emp_Rate_CA.txt"),
                                 header = TRUE, sep = ",")
PR_data_employment <- as.data.frame(PR_data_employment)
#Showing first results
head(PR_data_employment)
```

```
##                   Province Year UnEmp Part  Emp
## 1 Newfoundland and Labrador 1976  13.4 49.4 42.8
## 2 Newfoundland and Labrador 1977  15.4 50.6 42.8
## 3 Newfoundland and Labrador 1978  15.9 51.7 43.5
## 4 Newfoundland and Labrador 1979  14.8 53.4 45.5
## 5 Newfoundland and Labrador 1980  13.3 53.2 46.2
## 6 Newfoundland and Labrador 1981  13.5 53.5 46.3
```

2. Append my initial in the *dataframe* `PR_data_employment`.

```
#Append PR initials to all variables in the dataframe
colnames(PR_data_employment) <- paste(colnames(PR_data_employment), "PR", sep = "_")
head(PR_data_employment)
```

```
##                 Province_PR Year_PR UnEmp_PR Part_PR Emp_PR
## 1 Newfoundland and Labrador    1976     13.4    49.4   42.8
## 2 Newfoundland and Labrador    1977     15.4    50.6   42.8
## 3 Newfoundland and Labrador    1978     15.9    51.7   43.5
## 4 Newfoundland and Labrador    1979     14.8    53.4   45.5
## 5 Newfoundland and Labrador    1980     13.3    53.2   46.2
## 6 Newfoundland and Labrador    1981     13.5    53.5   46.3
```

3. Change each character variable to a factor variable

Changing to factor

```
#To Factor
PR_data_employment$Province_PR <- as.factor(PR_data_employment$Province_PR)
str(PR_data_employment)
```

```
## 'data.frame':    480 obs. of  5 variables:
##  $ Province_PR: Factor w/ 10 levels "Alberta","British Columbia",..: 5 5 5 5 5 5 5 5 5 5 ...
##  $ Year_PR    : int  1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 ...
##  $ UnEmp_PR   : num  13.4 15.4 15.9 14.8 13.3 13.5 16.2 18.1 20.1 20.2 ...
##  $ Part_PR    : num  49.4 50.6 51.7 53.4 53.2 53.5 53.1 53.3 53.5 54.2 ...
##  $ Emp_PR     : num  42.8 42.8 43.5 45.5 46.2 46.3 44.5 43.7 42.8 43.3 ...
```

4. Create a new variable showing the level of unemployment

This new variable should be discrete with three levels

```
#Creating new Level Variable with ifelse
PR_data_employment$Lev_UnEmp_PR <- ifelse(PR_data_employment$UnEmp_PR < 11.6, "M" , "H")
PR_data_employment$Lev_UnEmp_PR <- ifelse(PR_data_employment$UnEmp_PR < 6.5, "L" , PR_data_employment$L
PR_data_employment$Lev_UnEmp_PR <- as.factor(PR_data_employment$Lev_UnEmp_PR)
#Showing Results
head(PR_data_employment,10)
```

```
##                  Province_PR Year_PR UnEmp_PR Part_PR Emp_PR Lev_UnEmp_PR
## 1  Newfoundland and Labrador    1976     13.4    49.4   42.8            H
## 2  Newfoundland and Labrador    1977     15.4    50.6   42.8            H
## 3  Newfoundland and Labrador    1978     15.9    51.7   43.5            H
## 4  Newfoundland and Labrador    1979     14.8    53.4   45.5            H
## 5  Newfoundland and Labrador    1980     13.3    53.2   46.2            H
## 6  Newfoundland and Labrador    1981     13.5    53.5   46.3            H
## 7  Newfoundland and Labrador    1982     16.2    53.1   44.5            H
## 8  Newfoundland and Labrador    1983     18.1    53.3   43.7            H
## 9  Newfoundland and Labrador    1984     20.1    53.5   42.8            H
## 10 Newfoundland and Labrador    1985     20.2    54.2   43.3            H
```

```
str(PR_data_employment$Lev_UnEmp_PR)
```

```
##  Factor w/ 3 levels "H","L","M": 1 1 1 1 1 1 1 1 1 1 ...
```

The columns has created with the three levels. I taked in the first condicion 11.6 to get from 11.5 values

5. What are the dimensions of the dataset (rows and columns)?

dimension function dim(x)

```
#Structure
dim(PR_data_employment)
```

```
## [1] 480   6
```

Data has 480 rows and 6 columns

2. Summarizing Data.

1. Means and Standard Deviations

a. mean and standard deviation for Unemployment

```
#Calculating mean
meanUnE_PR <- mean(PR_data_employment$UnEmp_PR)
meanUnE_PR
```

```
## [1] 9.220417
```

```r
#Calculating standard deviation
sdUnE_PR <- sd(PR_data_employment$UnEmp_PR)
sdUnE_PR
```

```
## [1] 3.553075
```

The mean is higher than the standard deviation, which means the data is relatively consistent and not strongly spread out.

b. Use the results above to calculate the coefficient of variation

```r
#Coefficient of Variation
cvUnE_PR <- sdUnE_PR/meanUnE_PR
#round the result
round(cvUnE_PR,3)
```

```
## [1] 0.385
```

The standard deviation is appr. 38.5% of the mean.The variability is moderate.

c. Calculate the mean and standard deviation for Participation Rate

```r
#Calculating mean for Part_PR
meanPart_PR <- mean(PR_data_employment$Part_PR)
meanPart_PR
```

```
## [1] 64.42875
```

```r
#Calculating standard deviation
sdPart_PR <- sd(PR_data_employment$Part_PR)
sdPart_PR
```

```
## [1] 4.790083
```

d. Also calculate the coefficient of variation (rounded to 3 decimal places).

```r
##Coefficient of Variation Part_PR
cvPart_PR <- sdPart_PR/meanPart_PR
#round the result
round(cvPart_PR,3)
```

```
## [1] 0.074
```

The standard deviation is approximately 7.4%of the mean, indicating that variability is almost null.

e. Does the Unemployment or Participation have more variation? Comparing the coefficients of variation of both the unemployment and participation, the first one has a higher variation with 38.5% compared to the participation rate with 7.4%.

2. Calculate the 74 th percentile of the number of Employment Rate.

This calculation should be rounded to the nearest whole number (no decimal places)

```r
# Calculating the 75% of the data and round to 0 decimals
round(quantile(PR_data_employment$Emp_PR, c(.74)),0)
```

```
## 74%
##  62
```

That result means that the only 26% of the employment rates are above 62.

3. Organizing Data.

1. Summary Table

a. Create a table showing the average unemployment rate by province. This should be rounded to two decimal places

```r
#  Crating a Summary Table, aggregating by average (mean)
SummaryTable_PR <- aggregate(PR_data_employment[,3]
                             ,by=list(PR_data_employment$Province_PR)
                             ,FUN=mean
                             # Removing nulls
                             ,na.rm=TRUE)
#Round the x value
SummaryTable_PR$x <- round(SummaryTable_PR$x,2)
# Showing the table created previously
SummaryTable_PR
```

```
##                      Group.1     x
## 1                    Alberta  6.55
## 2           British Columbia  8.12
## 3                   Manitoba  6.31
## 4               New Brunswick 10.93
## 5  Newfoundland and Labrador 15.63
## 6                 Nova Scotia 10.10
## 7                    Ontario  7.39
## 8       Prince Edward Island 11.96
## 9                     Quebec  9.33
## 10              Saskatchewan  5.89
```

b. Which province has, on average, the highest unemployment rate?

The summary table show me that the Newfoundland and Labrador is the province that has the highest unemployed rate with 15.63%. That means that this province is 6.41 points above the general average.

15.63 - 9.22 (calculated before) = 6.41

2. Cross Tabulation

a. Create a table counting all levels of unemployment (the variable you created in Part 2: Q1.4) by province.

```r
# Getting the number of unemployedlevels for each province
SummaryTable2_PR <- table(PR_data_employment$Province_PR,PR_data_employment$Lev_UnEmp_PR)
SummaryTable2_PR
```

```
##
##                              H  L  M
##   Alberta                    0 25 23
##   British Columbia           6 12 30
##   Manitoba                   0 29 19
##   New Brunswick             19  0 29
##   Newfoundland and Labrador 46  0  2
##   Nova Scotia               13  1 34
##   Ontario                    0 15 33
##   Prince Edward Island      24  0 24
##   Quebec                     9  6 33
##   Saskatchewan               0 32 16
```

This result obtained shows the number of rows by each level by province.

b. Change the table to show the percentage of each Unemployment level in each Province. This should be rounded to three decimal places.

```
# ******Showing the percentage of each level in each province
round(prop.table(SummaryTable2_PR),3)
```

```
##
##                                 H     L     M
##   Alberta                   0.000 0.052 0.048
##   British Columbia          0.013 0.025 0.062
##   Manitoba                  0.000 0.060 0.040
##   New Brunswick             0.040 0.000 0.060
##   Newfoundland and Labrador 0.096 0.000 0.004
##   Nova Scotia               0.027 0.002 0.071
##   Ontario                   0.000 0.031 0.069
##   Prince Edward Island      0.050 0.000 0.050
##   Quebec                    0.019 0.013 0.069
##   Saskatchewan              0.000 0.067 0.033
```

Manitoba y Saskatchewan have a significant percentage low unemployment, meanwhile Newfoundland and Labrador has a rate considerably high.

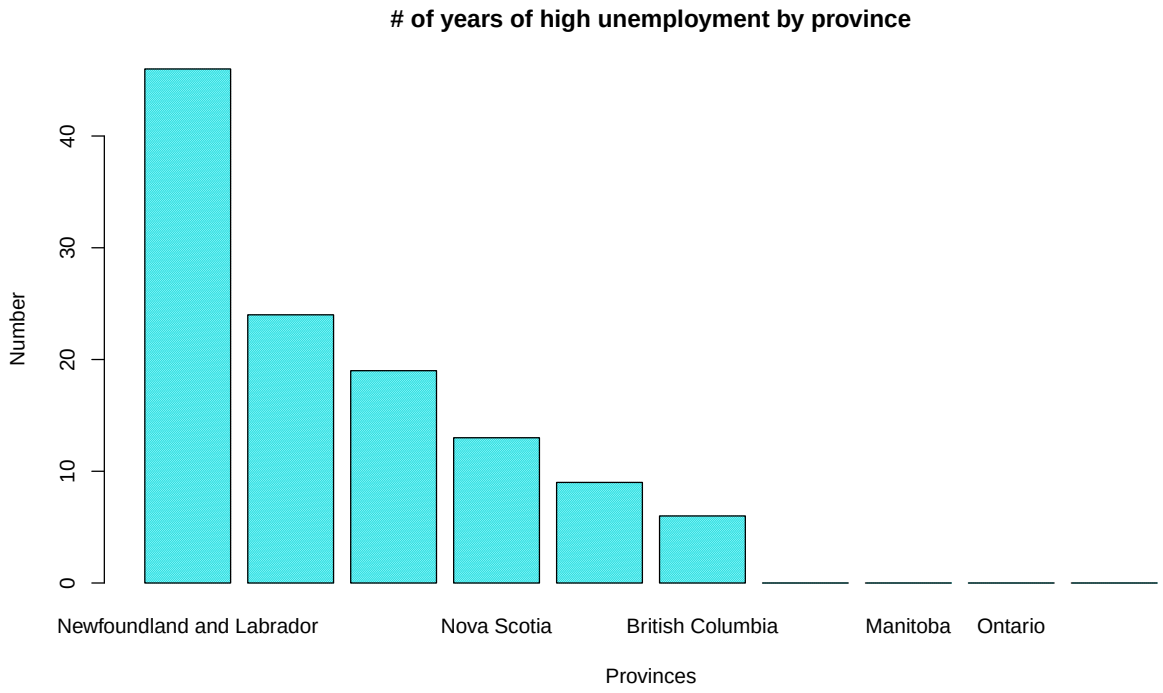
c. What percentage of high unemployment levels were in Prince Edward Island? This province has 54% of high unemployed. Means that in a little more than half of the of the observations(years), Prince Edward Island had a rate greater than 11.6.

3. Bar Plot

a. Create a column plot of years of high unemployment in each province b. The plot should be: i. Rank ordered by highest count of high unemploymentl. ii. Properly labeled (title, x-axis, etc) iii. The bars should have a different colour than the one shown in class.

```
# Filtering only H rates with the new column *********
high_unemp_PR <- PR_data_employment[PR_data_employment$Lev_UnEmp_PR =="H", ]
high_unemp_PR <- table(high_unemp_PR$Province_PR)
#Bar Plot order DESCENDING parameter in true
barplot(high_unemp_PR[order(high_unemp_PR,decreasing=TRUE)],
        col=5,   #colour
```

```
        density = 80, angle = 50,
        #Tittle
        main="# of years of high unemployment by province",
        #Properly labeled x and y axis
        xlab="Provinces",
        ylab="Number")
```



c. Based on the bar plot, (approximately) how many of years did Nova Scotia experience high unemployment?

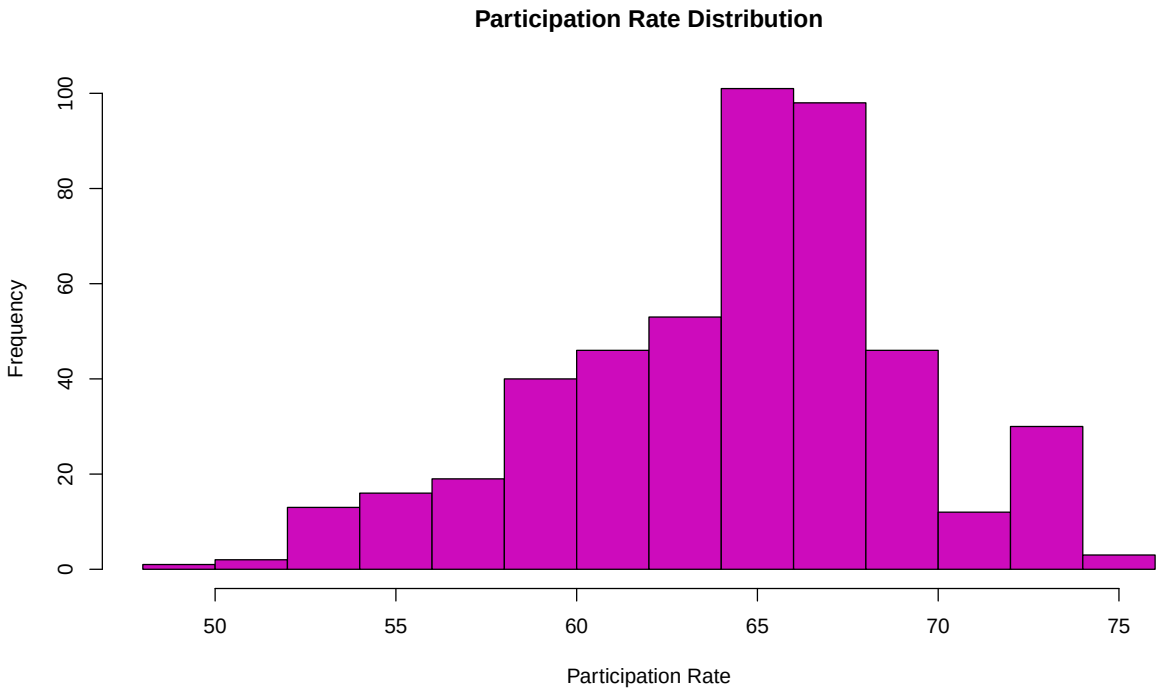
The top of New Scotia bar is located between 12 and 14 number of years.

4. Histogram

a. Create a histogram of Participation Rate b. The plot should be properly labeled and a unique colour and have 10 breaks.

```
#Histogram participation rate*******
hist(PR_data_employment$Part_PR,
     col=6,
     breaks = 10,
     #Tittle
     main="Participation Rate Distribution",
     #Properly labeled x and y axis
     xlab="Participation Rate",
     ylab="Frequency"

    )
```



c. Which range of Participation Rate is the most common?

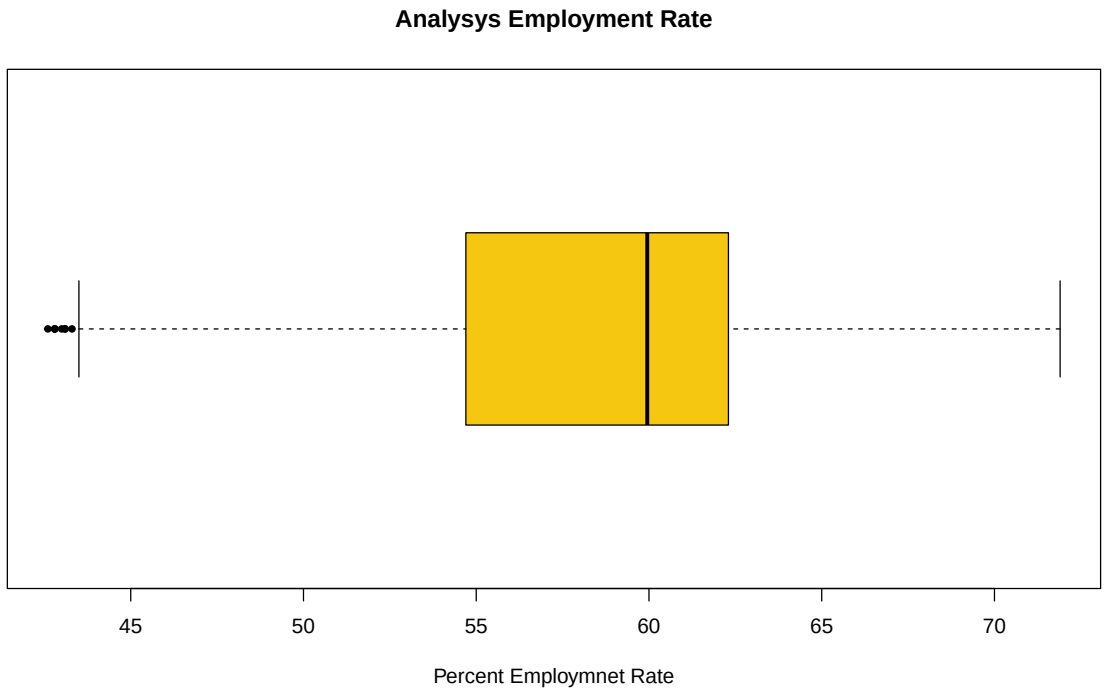
The histogram shows that the common range of participation rate is between 64 to 67, which means that the majority of the observations are concentrated in this range.

The percentage of labour force participants is predominantly in the high ranges.

5. Box plot

a. Create a horizontal box plot of number of Employment Rate b. The plot should be properly labeled and a unique colour

```
# Horizontal Box Plot Employment rate *****
boxplot(PR_data_employment$Emp_PR,
        #Tittle
        main="Analysys Employment Rate",
        xlab="Percent Employmnet Rate",
        col=7,
        horizontal=TRUE,
        pch=20)
```




```
quantile(PR_data_employment$Emp_PR,.60)
```

```
##  60%
## 60.7
```

c. Based on the box plot, approximately how many years had an Employment Rate less than ~ 60%?

Approximately 288 to 292.8 years had an Employment Rate of less than 60%.

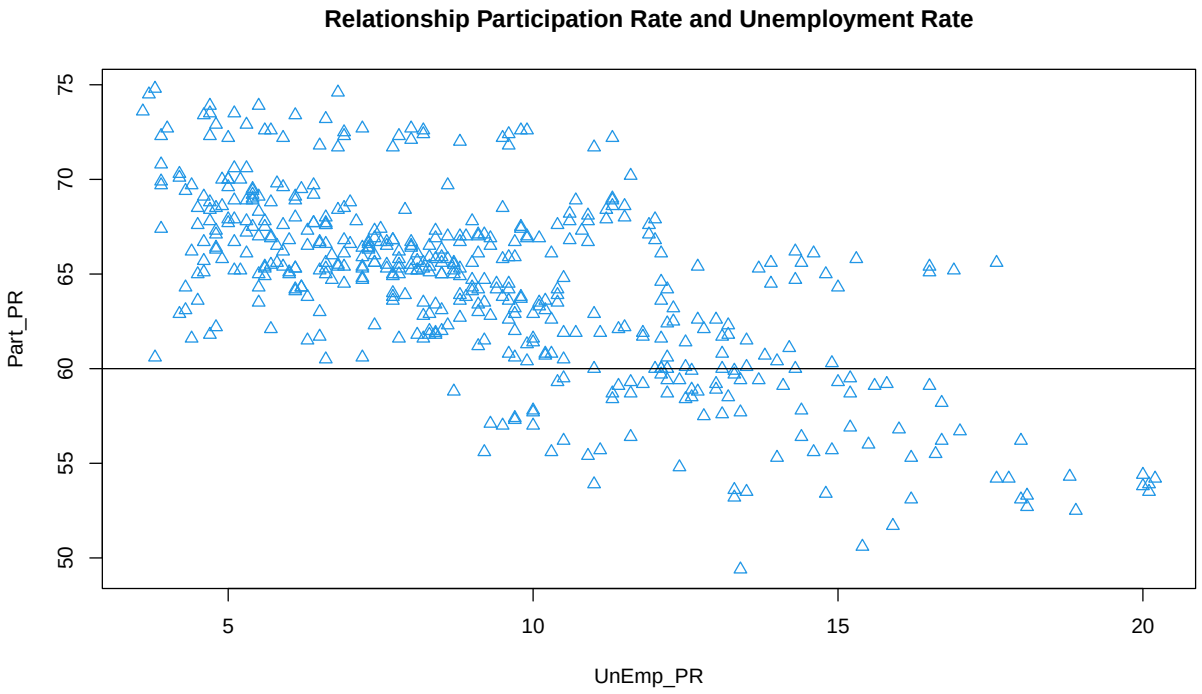
To get this result, I first identified the position 60% on the x axis of tjhe box plot, this corresponds to 60% to 61% of the total data approximately. Finally, my calculations was: 480 total observations * ~ 0.60 = 288 480 total observations * ~ 0.61 = 292.8

6. Scatter Plot

a. Create a scatter plot comparing Participation Rate and Unemployment Rate. b. The plot should be properly labeled with a marker type different than the one demonstrated in class

```
# Plot metrics participation rate and unemployed rate
plot(Part_PR ~ UnEmp_PR,
     col=4,
     data=PR_data_employment,
     #Defining marker type
     pch=2,
     main="Relationship Participation Rate and Unemployment Rate")

abline(coef = c(60,0)) #overlays a line, intercept=6, slope=0
```



c. Does there appear to be an association between Participation Rate and Unemployment Rate? It can ve observed that the markets are almost dispersed, which means that thre not a clear lineal relation between participation rate and unemployed rate. However, it can be observed a slight trend (negatively correlated); as the participation rate tends to increase, the unemployed rate decreases

References

R Core Team (2024). hist {graphics}. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.rdocumentation.org/packages/graphics/versions/3.6.2/topics/hist