Assignment_1

2024-02-04

dataset source: https://www.kaggle.com/datasets/benroshan/factors-affecting-campus-placement/data

Importing the dataset

This R code is using the `readr` package to read a CSV file named "Placement_Data_Full_Class.csv" and it stores the data in a variable called `dataset`.

```r
library(readr)
dataset<- read_csv("C:/Users/Abhinav Reddy/Desktop/FML/Placement_Data_Full_Class.csv")
```

```
## Rows: 215 Columns: 15
## -- Column specification -------------------------------------------------------
## Delimiter: ","
## chr (8): gender, ssc_b, hsc_b, hsc_s, degree_t, workex, specialisation, status
## dbl (7): sl_no, ssc_p, hsc_p, degree_p, etest_p, mba_p, salary
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Descriptive statistics:

Displaying descriptive statistics for quantitative variables

This R code summarizes and prints descriptive statistics for a specified quantitative variables in a dataset.

```r
quantitative_var<-c("sl_no","ssc_p","hsc_p","degree_p","etest_p","mba_p","salary")
quantitative_summary<- summary(dataset[quantitative_var])
print(quantitative_summary)
```

```
##      sl_no            ssc_p           hsc_p          degree_p        etest_p
## Min.   :  1.0    Min.   :40.89   Min.   :37.00   Min.   :50.00   Min.   :50.0
## 1st Qu.: 54.5    1st Qu.:60.60   1st Qu.:60.90   1st Qu.:61.00   1st Qu.:60.0
## Median :108.0    Median :67.00   Median :65.00   Median :66.00   Median :71.0
## Mean   :108.0    Mean   :67.30   Mean   :66.33   Mean   :66.37   Mean   :72.1
## 3rd Qu.:161.5    3rd Qu.:75.70   3rd Qu.:73.00   3rd Qu.:72.00   3rd Qu.:83.5
## Max.   :215.0    Max.   :89.40   Max.   :97.70   Max.   :91.00   Max.   :98.0
##
##      mba_p           salary
## Min.   :51.21   Min.   :200000
## 1st Qu.:57.95   1st Qu.:240000
## Median :62.00   Median :265000
## Mean   :62.28   Mean   :288655
## 3rd Qu.:66.25   3rd Qu.:300000
## Max.   :77.89   Max.   :940000
##                 NA's   :67
```

Displaying descriptive statistics for categorical variables

This R code shows the frequency distribution of each category within the variables listed in categorical_var.

```
categorical_var <- c("gender","ssc_b","hsc_b","degree_t","workex","specialisation")
 categorical_summary<-table(dataset[categorical_var])
print(head( categorical_summary))
```

```
## , , hsc_b = Central, degree_t = Comm&Mgmt, workex = No, specialisation = Mkt&Fin
##
##       ssc_b
## gender Central Others
##      F       8      2
##      M      17      1
##
## , , hsc_b = Others, degree_t = Comm&Mgmt, workex = No, specialisation = Mkt&Fin
##
##       ssc_b
## gender Central Others
##      F       4      5
##      M       3     14
##
## , , hsc_b = Central, degree_t = Others, workex = No, specialisation = Mkt&Fin
##
##       ssc_b
## gender Central Others
##      F       1      0
##      M       0      1
##
## , , hsc_b = Others, degree_t = Others, workex = No, specialisation = Mkt&Fin
##
##       ssc_b
## gender Central Others
##      F       0      0
##      M       0      0
##
## , , hsc_b = Central, degree_t = Sci&Tech, workex = No, specialisation = Mkt&Fin
##
##       ssc_b
## gender Central Others
##      F       0      1
##      M       0      0
##
## , , hsc_b = Others, degree_t = Sci&Tech, workex = No, specialisation = Mkt&Fin
##
##       ssc_b
## gender Central Others
##      F       0      1
##      M       2      9
##
## , , hsc_b = Central, degree_t = Comm&Mgmt, workex = Yes, specialisation = Mkt&Fin
##
##       ssc_b
## gender Central Others
```

```
##      F       3       1
##      M       8       0
##
## , , hsc_b = Others, degree_t = Comm&Mgmt, workex = Yes, specialisation = Mkt&Fin
##
##        ssc_b
## gender Central Others
##      F       2       5
##      M       7       6
##
## , , hsc_b = Central, degree_t = Others, workex = Yes, specialisation = Mkt&Fin
##
##        ssc_b
## gender Central Others
##      F       0       0
##      M       0       0
##
## , , hsc_b = Others, degree_t = Others, workex = Yes, specialisation = Mkt&Fin
##
##        ssc_b
## gender Central Others
##      F       0       1
##      M       0       1
##
## , , hsc_b = Central, degree_t = Sci&Tech, workex = Yes, specialisation = Mkt&Fin
##
##        ssc_b
## gender Central Others
##      F       0       0
##      M       4       0
##
## , , hsc_b = Others, degree_t = Sci&Tech, workex = Yes, specialisation = Mkt&Fin
##
##        ssc_b
## gender Central Others
##      F       0       3
##      M       3       7
##
## , , hsc_b = Central, degree_t = Comm&Mgmt, workex = No, specialisation = Mkt&HR
##
##        ssc_b
## gender Central Others
##      F       6       0
##      M       7       1
##
## , , hsc_b = Others, degree_t = Comm&Mgmt, workex = No, specialisation = Mkt&HR
##
##        ssc_b
## gender Central Others
##      F       6       7
##      M       6      13
##
## , , hsc_b = Central, degree_t = Others, workex = No, specialisation = Mkt&HR
##
```

```
##       ssc_b
## gender Central Others
##      F       2      0
##      M       1      0
##
## , , hsc_b = Others, degree_t = Others, workex = No, specialisation = Mkt&HR
##
##       ssc_b
## gender Central Others
##      F       0      1
##      M       0      1
##
## , , hsc_b = Central, degree_t = Sci&Tech, workex = No, specialisation = Mkt&HR
##
##       ssc_b
## gender Central Others
##      F       7      0
##      M       2      0
##
## , , hsc_b = Others, degree_t = Sci&Tech, workex = No, specialisation = Mkt&HR
##
##       ssc_b
## gender Central Others
##      F       0      3
##      M       2      7
##
## , , hsc_b = Central, degree_t = Comm&Mgmt, workex = Yes, specialisation = Mkt&HR
##
##       ssc_b
## gender Central Others
##      F       1      0
##      M       3      0
##
## , , hsc_b = Others, degree_t = Comm&Mgmt, workex = Yes, specialisation = Mkt&HR
##
##       ssc_b
## gender Central Others
##      F       0      3
##      M       2      4
##
## , , hsc_b = Central, degree_t = Others, workex = Yes, specialisation = Mkt&HR
##
##       ssc_b
## gender Central Others
##      F       1      0
##      M       1      0
##
## , , hsc_b = Others, degree_t = Others, workex = Yes, specialisation = Mkt&HR
##
##       ssc_b
## gender Central Others
##      F       0      0
##      M       0      0
##
```

```
## , , hsc_b = Central, degree_t = Sci&Tech, workex = Yes, specialisation = Mkt&HR
##
##       ssc_b
## gender Central Others
##      F       0      0
##      M       5      0
##
## , , hsc_b = Others, degree_t = Sci&Tech, workex = Yes, specialisation = Mkt&HR
##
##       ssc_b
## gender Central Others
##      F       1      1
##      M       1      0
```

Transformation of a variable

This R code is transforming the salary variable by taking the logarithm and creating a new variable log_transformed_salary.

```
dataset$log_transformed_salary <- log(dataset$salary)
head(dataset)
```

```
## # A tibble: 6 x 16
##   sl_no gender ssc_p ssc_b    hsc_p hsc_b  hsc_s degree_p degree_t workex etest_p
##   <dbl> <chr>  <dbl> <chr>    <dbl> <chr>  <chr>    <dbl> <chr>    <chr>    <dbl>
## 1     1 M       67   Others   91    Others Comm~      58  Sci&Tech No         55
## 2     2 M       79.3 Central  78.3  Others Scie~      77.5 Sci&Tech Yes       86.5
## 3     3 M       65   Central  68    Centr~ Arts       64  Comm&Mg~ No         75
## 4     4 M       56   Central  52    Centr~ Scie~      52  Sci&Tech No         66
## 5     5 M       85.8 Central  73.6  Centr~ Comm~      73.3 Comm&Mg~ No         96.8
## 6     6 M       55   Others   49.8  Others Scie~      67.2 Sci&Tech Yes       55
## # i 5 more variables: specialisation <chr>, mba_p <dbl>, status <chr>,
## #   salary <dbl>, log_transformed_salary <dbl>
```

This R code is transforming the mba_p variable by taking the square root and creating a new variable sqrt_transformed_mba_p.

```
dataset$sqrt_transformed_mba_p <- sqrt(dataset$mba_p)
head(dataset)
```

```
## # A tibble: 6 x 17
##   sl_no gender ssc_p ssc_b    hsc_p hsc_b  hsc_s degree_p degree_t workex etest_p
##   <dbl> <chr>  <dbl> <chr>    <dbl> <chr>  <chr>    <dbl> <chr>    <chr>    <dbl>
## 1     1 M       67   Others   91    Others Comm~      58  Sci&Tech No         55
## 2     2 M       79.3 Central  78.3  Others Scie~      77.5 Sci&Tech Yes       86.5
## 3     3 M       65   Central  68    Centr~ Arts       64  Comm&Mg~ No         75
## 4     4 M       56   Central  52    Centr~ Scie~      52  Sci&Tech No         66
## 5     5 M       85.8 Central  73.6  Centr~ Comm~      73.3 Comm&Mg~ No         96.8
## 6     6 M       55   Others   49.8  Others Scie~      67.2 Sci&Tech Yes       55
## # i 6 more variables: specialisation <chr>, mba_p <dbl>, status <chr>,
## #   salary <dbl>, log_transformed_salary <dbl>, sqrt_transformed_mba_p <dbl>
```

Plotting a quantitative variable

This R code generates a histogram for the mba_p variable in the dataset.

```
quantitative_var1 <- "mba_p"
hist(dataset[[quantitative_var1]],
     main = "Histogram of mba_p",
     xlab = quantitative_var1,
     col = "lightblue")
```

Histogram of mba_p
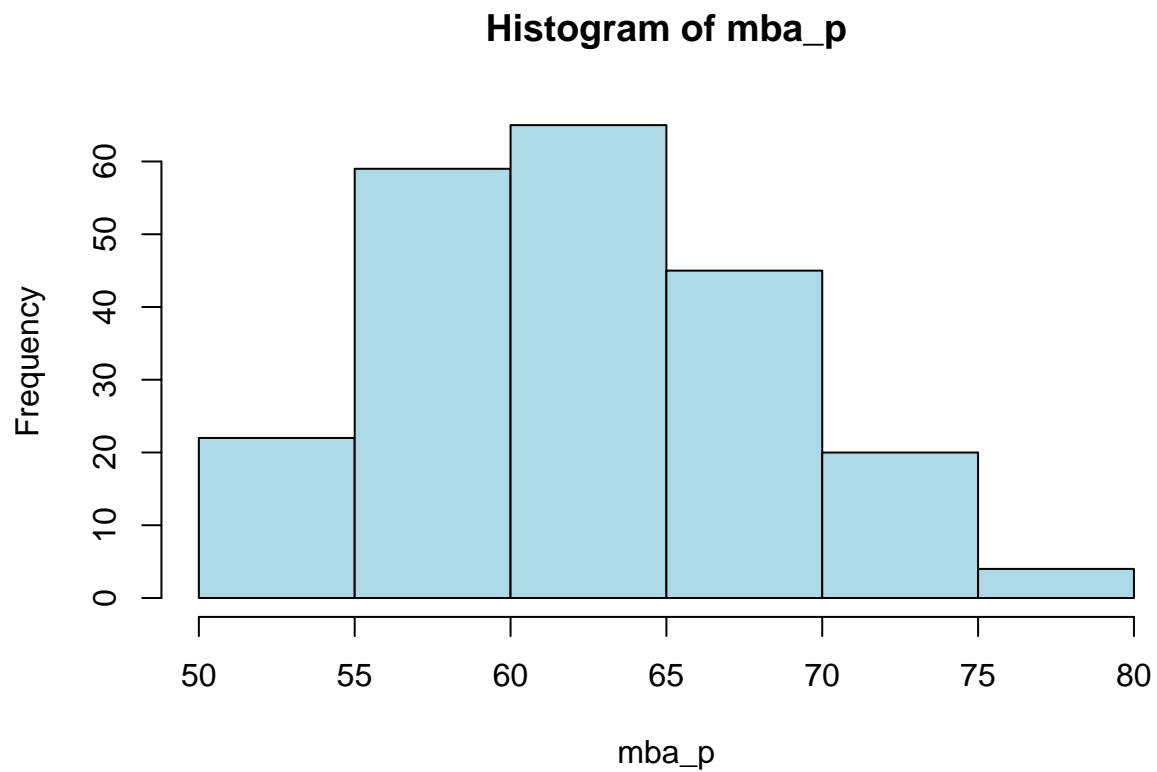
This R code generates a boxplot for the degree_p variable in the dataset.

```
boxplot(dataset$degree_p,
        main = "Box Plot of a degree_p",
        ylab = "Quantitative_Variable",
        col = "lightgreen", border = "black")
```
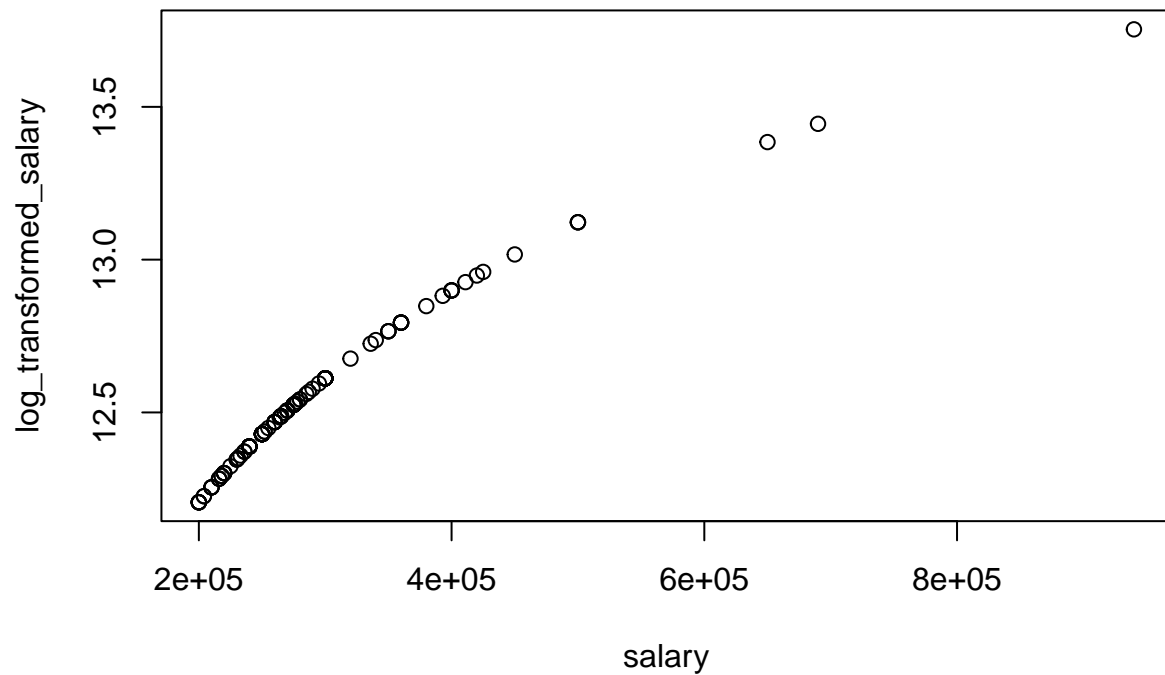
Box Plot of a degree_p
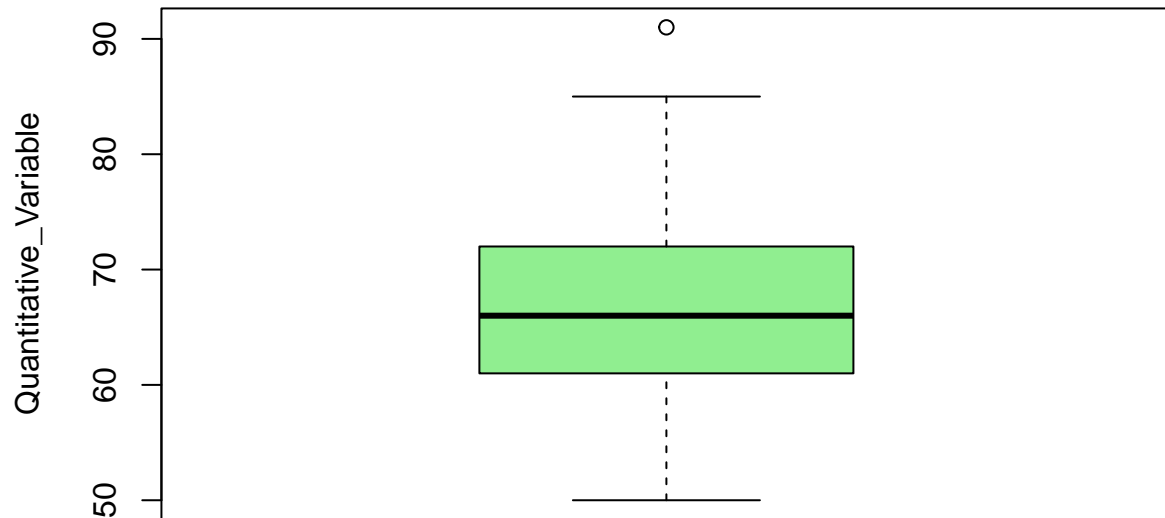
Plotting a Scatterplot

This scatter plot shows the relationship between salary variable and log_transformed_salary variable.

```
plot(x = dataset$'salary', y = dataset$log_transformed_salary,
     xlab = "salary", ylab = "log_transformed_salary",
     main = "Scatterplot of Salary vs Transformed Salary")
```

Scatterplot of Salary vs Transformed Salary

This scatter plot shows the relationship between of ssc_p variable and degree_p variable

```
plot(x = dataset$ssc_p, y = dataset$degree_p,
     xlab = "SSC_Percentage", ylab = "Degree_Percentage",
     main = "Scatter Plot of SSC Percentage vs Degree Percentage")
```

Scatter Plot of SSC Percentage vs Degree Percentage