

Descriptive Statistics Report

Anna Buchnowska

7 grudnia 2016

Contents

Introduction	1
Descriptive (summary) analysis	1
Modality distributions	1
Histogram	2
Boxplot	6

Introduction

Following output is generated by RStudio. The data set for this case study was intended to answer the question:

What is the inter-relation between these three variables? How much total information do they provide?

Descriptive (summary) analysis

The purpose of the descriptive statistics is to provide the reader with an idea about the basic elements of the groups being studied.

group	grades	IQ	earnings
:80	Min. :2.000	Min. : 67.00	Min. : 600
A: 8	1st Qu.:3.000	1st Qu.: 86.25	1st Qu.: 1500
B:12	Median :3.500	Median : 97.00	Median : 1800
	Mean :3.525	Mean : 97.01	Mean : 3718
	3rd Qu.:4.000	3rd Qu.:107.25	3rd Qu.: 2200
	Max. :5.000	Max. :134.00	Max. :40000
	NA's :80		NA's :75

Modality distributions

Shapiro-Wilk test of normality. The significance level of the test - 0.05 (α)

1. grades

H_0 : The sample comes from a normal distribution.

H_1 : The sample doesn't come from a normal distribution.

Shapiro-Wilk normality test

```
data: data$grades
```

```
W = 0.93587, p-value = 0.2001
```

P-value is greater than $\alpha = 0.05$. Observed data is consistent with the null hypothesis.

2. IQ

H_0 : The sample comes from a normal distribution.

H_1 : The sample doesn't come from a normal distribution.

Shapiro-Wilk normality test

```
data: data$IQ
```

```
W = 0.98692, p-value = 0.4322
```

P-value is greater than $\alpha = 0.05$. Observed data is consistent with the null hypothesis.

2. earnings

H_0 : The sample comes from a normal distribution.

H_1 : The sample doesn't come from a normal distribution.

Shapiro-Wilk normality test

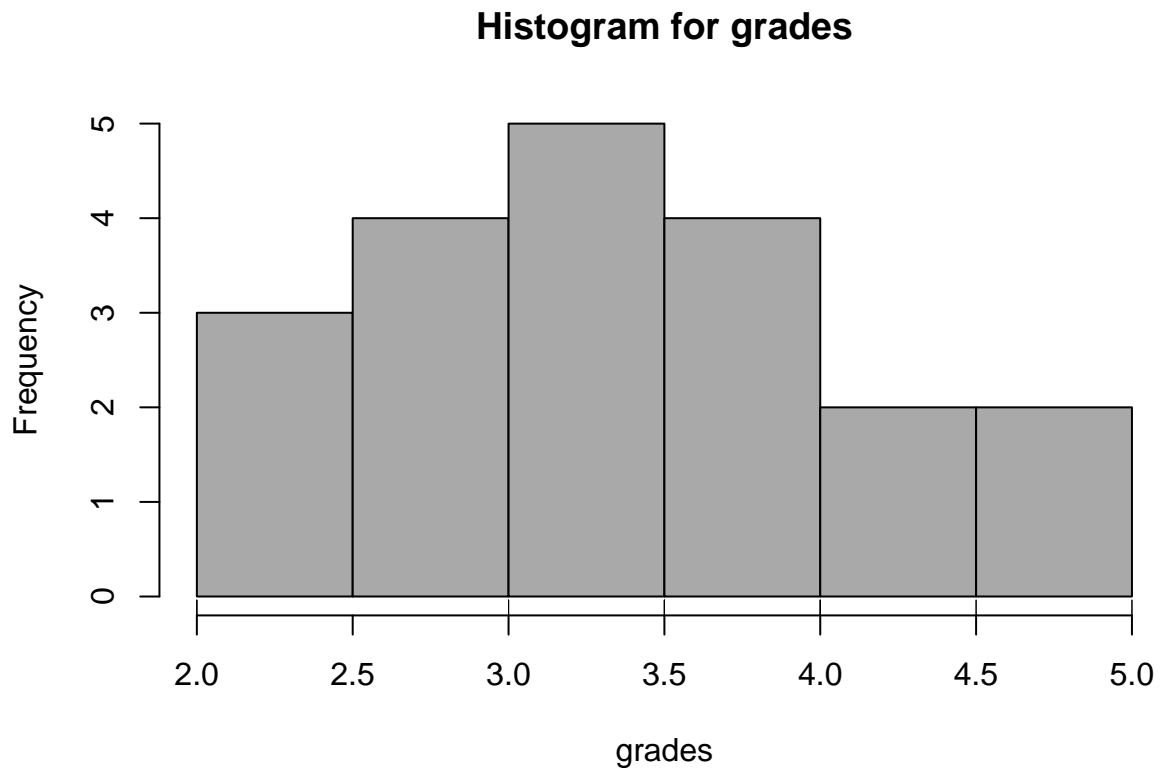
```
data: data$earnings
```

```
W = 0.33525, p-value = 1.25e-09
```

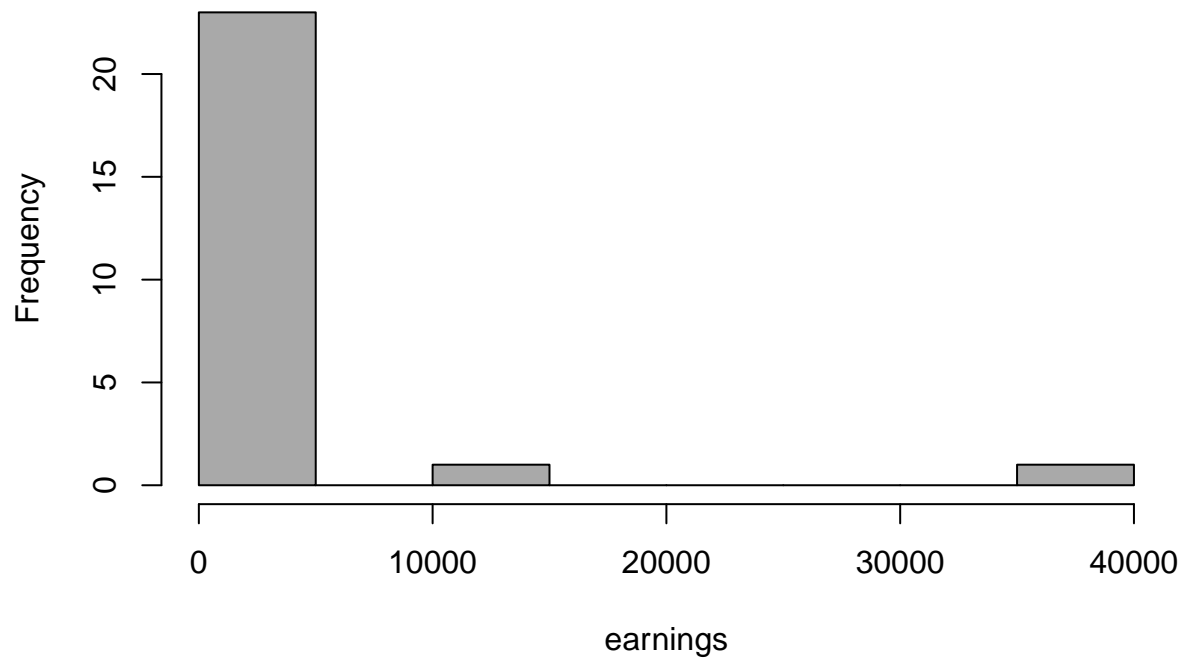
P-value is effectively close to zero. observed data is inconsistent with the null hypothesis.

Reject the null hypothesis H_0 in favor of the alternative hypothesis

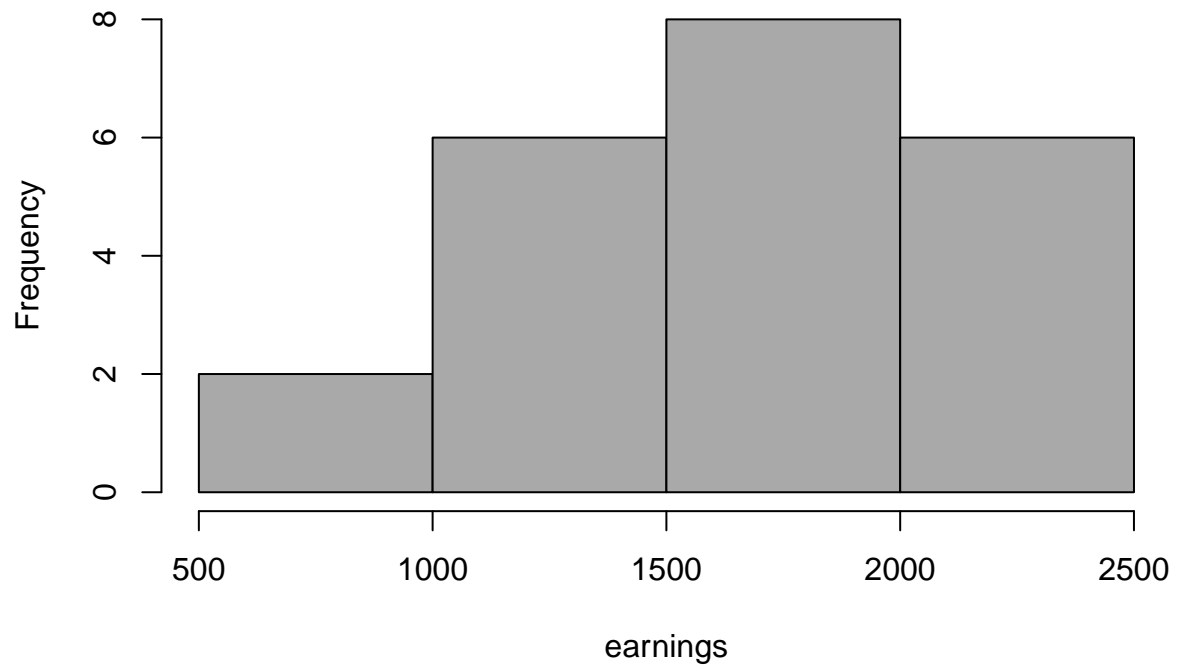
Histogram



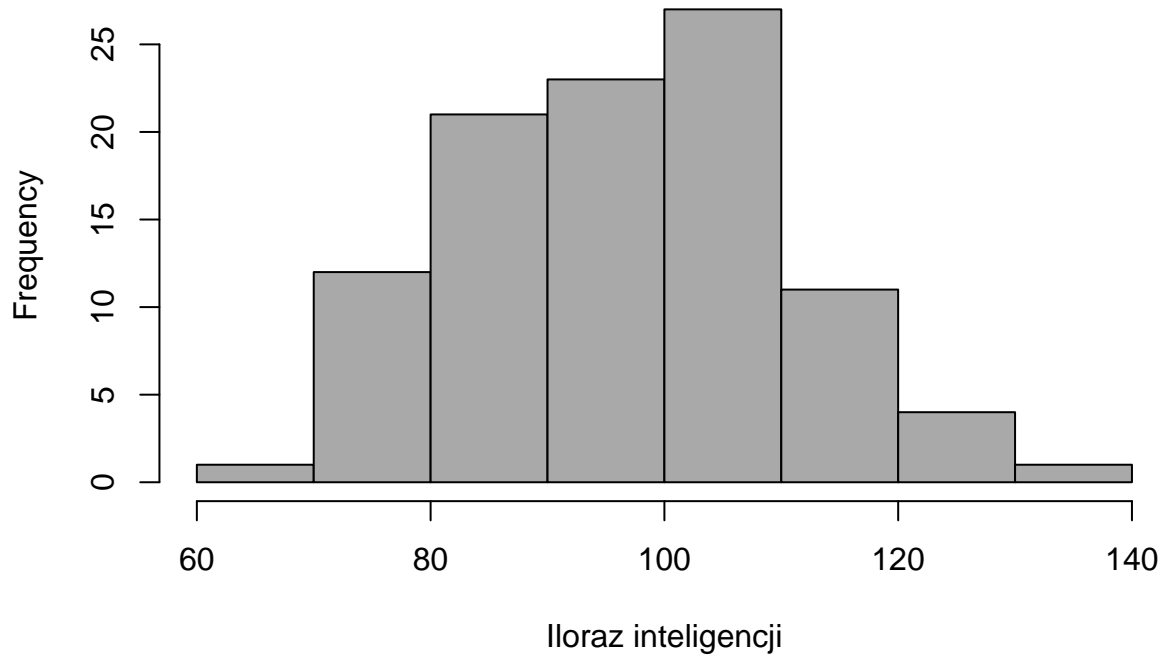
Histogram for earnings with outliers



Histogram for earnings without outliers



Histogram for IQ



```
## Warning in FUN(newX[, i], ...): brak argumentów w min; zwracanie wartości
## Inf

## Warning in FUN(newX[, i], ...): brak argumentów w max; zwracanie wartości -
## Inf

## [[1]]
##      vars  n    mean      sd median trimmed   mad min  max range
## group*    1 80    1.00    0.00      1    1.00   0.00  1    1     0
## grades    2  0    NaN     NA     NA     NaN    NA Inf  -Inf -Inf
## IQ        3 80   96.04   14.39     93   95.48  16.31  67  134    67
## earnings  4  5 8938.00 17365.58  1200 8938.00 444.78 890 40000 39110
##      skew kurtosis      se
## group*   NaN      NaN    0.00
## grades   NA      NA     NA
## IQ       0.31   -0.62    1.61
## earnings 1.07   -0.92  7766.12
##
## $A
##      vars  n    mean      sd median trimmed   mad min  max range skew
## group*    1  8    2.00    0.00    2.00    2.00   0.00  2    2     0  NaN
## grades    2  8    3.31    1.03    3.25    3.31   1.11  2    5     3  0.10
## IQ        3  8   104.88  12.44   106.50  104.88   8.90  80  122    42 -0.64
## earnings  4  8  1897.50  824.30  1890.00  1897.50  578.21 600 3500  2900  0.42
##      kurtosis      se
## group*      NaN    0.00
## grades    -1.41    0.37
```

```
## IQ          -0.51  4.40
## earnings    -0.43 291.44
##
## $B
##      vars  n    mean      sd median trimmed   mad  min  max range
## group*    1 12    3.00    0.00   3.0    3.0   0.00   3    3     0
## grades    2 12    3.67    0.81   3.5    3.7   0.74   2    5     3
## IQ        3 12   98.25   10.45  100.5   99.8  10.38  72   109    37
## earnings  4 12 2755.83 2929.61 2050.0  1982.0 444.78 1250 12000 10750
##      skew kurtosis      se
## group*   NaN      NaN   0.00
## grades  -0.26   -0.59   0.23
## IQ      -1.16    0.60   3.02
## earnings 2.59    5.27  845.70
##
## attr("call")
## by.data.frame(data = x, INDICES = group, FUN = describe, type = type)
```

Boxplot