REPORT SERIES WITH DLOOKR

Exploratory Data Analysis Report

Author:
dlookr package

Version:
0.3.12

November 19, 2020

Contents

Chapter 1

Introduction

The EDA Report provides exploratory data analysis information on objects that inherit data.frame and data.frame.

1.1 Information of Dataset

The dataset that generated the EDA Report is an 'data.frame' object. It consists of 4,950 observations and 22 variables.

1.2 Information of Variables

Table 1.1: Information of Variables

variables	types	missing_count	missing_percent	unique_count	unique_rate
workerid	numeric	0	0	495	0.1000000
year	numeric	0	0	10	0.0020202
ui	numeric	0	0	495	0.1000000
quarter	numeric	0	0	4	0.0008081
q1	numeric	0	0	2	0.0004040
wage	numeric	0	0	4950	1.0000000
educ	numeric	0	0	2469	0.4987879
exper	numeric	0	0	29	0.0058586
union	numeric	0	0	2	0.0004040
exper2	numeric	0	0	29	0.0058586
lnwage	numeric	0	0	4950	1.0000000
yy1	numeric	0	0	2	0.0004040
yy2	numeric	0	0	2	0.0004040
yy3	numeric	0	0	2	0.0004040
yy4	numeric	0	0	2	0.0004040
yy5	numeric	0	0	2	0.0004040
yy6	numeric	0	0	2	0.0004040
yy7	numeric	0	0	2	0.0004040
yy8	numeric	0	0	2	0.0004040
yy9	numeric	0	0	2	0.0004040
yy10	numeric	0	0	2	0.0004040
lag_lnwage	numeric	495	10	4455	0.9000000

The target variable of the data is 'NULL', and the data type of the variable is NULL(You did not specify a target variable).

1.3 About EDA Report

EDA reports provide information and visualization results that support the EDA process. In particular, it provides a variety of information to understand the relationship between the target variable and the rest of the variables of interest.

Chapter 2

Univariate Analysis

2.1 Descriptive Statistics

edaData
22 Variables 4950 Observations

---

workerid Format:%9.0g

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
4950	0	495	1	248	165	25	50	124	248	372	446	471

```
lowest :   1   2   3   4   5, highest: 491 492 493 494 495
```

---

year Format:%9.0g

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
4950	0	10	0.99	5.5	3.301	1.0	1.9	3.0	5.5	8.0	9.1	10.0

```
lowest :  1  2  3  4  5, highest:  6  7  8  9 10

Value        1   2   3   4   5   6   7   8   9  10
Frequency  495 495 495 495 495 495 495 495 495 495
Proportion 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1
```

---

ui Format:%10.0g

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50
4950	0	495	1	0.4879	0.333	0.04343	0.08252	0.24384	0.47667
.75	.90	.95							
0.74471	0.89674	0.94917							

```
lowest : 0.0004503177 0.0016528674 0.0094674007 0.0109514188 0.0114820479
highest: 0.9912454959 0.9929943224 0.9932442058 0.9942212892 0.9957856431
```

---

quarter Format:%9.0g

n	missing	distinct	Info	Mean	Gmd
4950	0	4	0.936	2.517	1.26

```
Value          1     2     3     4
Frequency   1190  1350  1070  1340
Proportion 0.240 0.273 0.216 0.271
```

---

q1 Format:%9.0g

n	missing	distinct	Info	Sum	Mean	Gmd
4950	0	2	0.548	1190	0.2404	0.3653

---

wage Format:%9.0g

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
4950	0	4950	1	1999	1092	861.0	969.9	1226.8	1718.1	2482.9	3419.9	4112.1

```
lowest :  662.0831  666.1646  666.9563  667.3817  668.4014
highest: 6751.8921 6752.1616 6759.6406 7075.5063 7170.2417
```

---

educ Format:%9.0g

```
      n  missing distinct     Info     Mean      Gmd      .05      .10      .25      .50
   4950        0     2469        1    5.226    4.244   0.0000   0.5168   2.3216   4.6458
    .75      .90      .95
 7.6351  10.6920  12.6225

lowest : 0.000000e+00 9.006355e-04 6.068707e-03 2.190284e-02 2.314014e-02
highest: 1.815874e+01 1.825217e+01 1.830014e+01 1.844627e+01 1.944627e+01
```

---

exper Format:%9.0g

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
4950	0	29	0.998	14.34	7.415	4	5	9	15	19	23	25

```
lowest :  0  1  2  3  4, highest: 24 25 26 27 28
```

---

union Format:%9.0g

n	missing	distinct	Info	Sum	Mean	Gmd
4950	0	2	0.749	2407	0.4863	0.4997

---

exper2 Format:%9.0g

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
4950	0	29	0.998	247.3	210	16	25	81	225	361	529	625

```
lowest :   0   1   4   9  16, highest: 576 625 676 729 784
```

---

lnwage Format:%10.0g

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
4950	0	4950	1	7.483	0.5438	6.758	6.877	7.112	7.449	7.817	8.137	8.322

```
lowest : 6.495391 6.501537 6.502724 6.503362 6.504889, highest: 8.817578 8.817618 8.818725 8.864394 8.877695
```

---

yy1 : year== 1.0000 Format:%8.0g

n	missing	distinct	Info	Sum	Mean	Gmd
4950	0	2	0.27	495	0.1	0.18

---

yy2 : year== 2.0000 Format:%8.0g

n	missing	distinct	Info	Sum	Mean	Gmd
4950	0	2	0.27	495	0.1	0.18

---

yy3 : year== 3.0000 Format:%8.0g

n	missing	distinct	Info	Sum	Mean	Gmd
4950	0	2	0.27	495	0.1	0.18

---

yy4 : year== 4.0000 Format:%8.0g

n	missing	distinct	Info	Sum	Mean	Gmd
4950	0	2	0.27	495	0.1	0.18

---

yy5 : year== 5.0000 Format:%8.0g

n	missing	distinct	Info	Sum	Mean	Gmd
4950	0	2	0.27	495	0.1	0.18

---

yy6 : year== 6.0000 Format:%8.0g

n	missing	distinct	Info	Sum	Mean	Gmd
4950	0	2	0.27	495	0.1	0.18

---

yy7 : year== 7.0000 Format:%8.0g

n	missing	distinct	Info	Sum	Mean	Gmd
4950	0	2	0.27	495	0.1	0.18

---

yy8 : year== 8.0000 Format:%8.0g

n	missing	distinct	Info	Sum	Mean	Gmd
4950	0	2	0.27	495	0.1	0.18

---

yy9 : year== 9.0000 Format:%8.0g

n	missing	distinct	Info	Sum	Mean	Gmd
4950	0	2	0.27	495	0.1	0.18

---

yy10 : year== 10.0000 Format:%8.0g

n	missing	distinct	Info	Sum	Mean	Gmd
4950	0	2	0.27	495	0.1	0.18

---

lag_lnwage Format:%9.0g

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
4455	495	4454	1	7.455	0.5363	6.742	6.860	7.085	7.425	7.784	8.107	8.292

```
lowest : 6.495391 6.501537 6.502725 6.503362 6.504889, highest: 8.707273 8.708329 8.717245 8.723325 8.723684
```

2.2 Normality Test of Numerical Variables

2.2.1 Statistics and Visualization of (Sample) Data

workerid

normality test : Shapiro-Wilk normality test
statistic : 0.95489, p-value : 1.31917E-36

type	skewness	kurtosis
original	0.0000	1.8000
log transformation	-1.7851	6.9849
sqrt transformation	-0.5591	2.3823

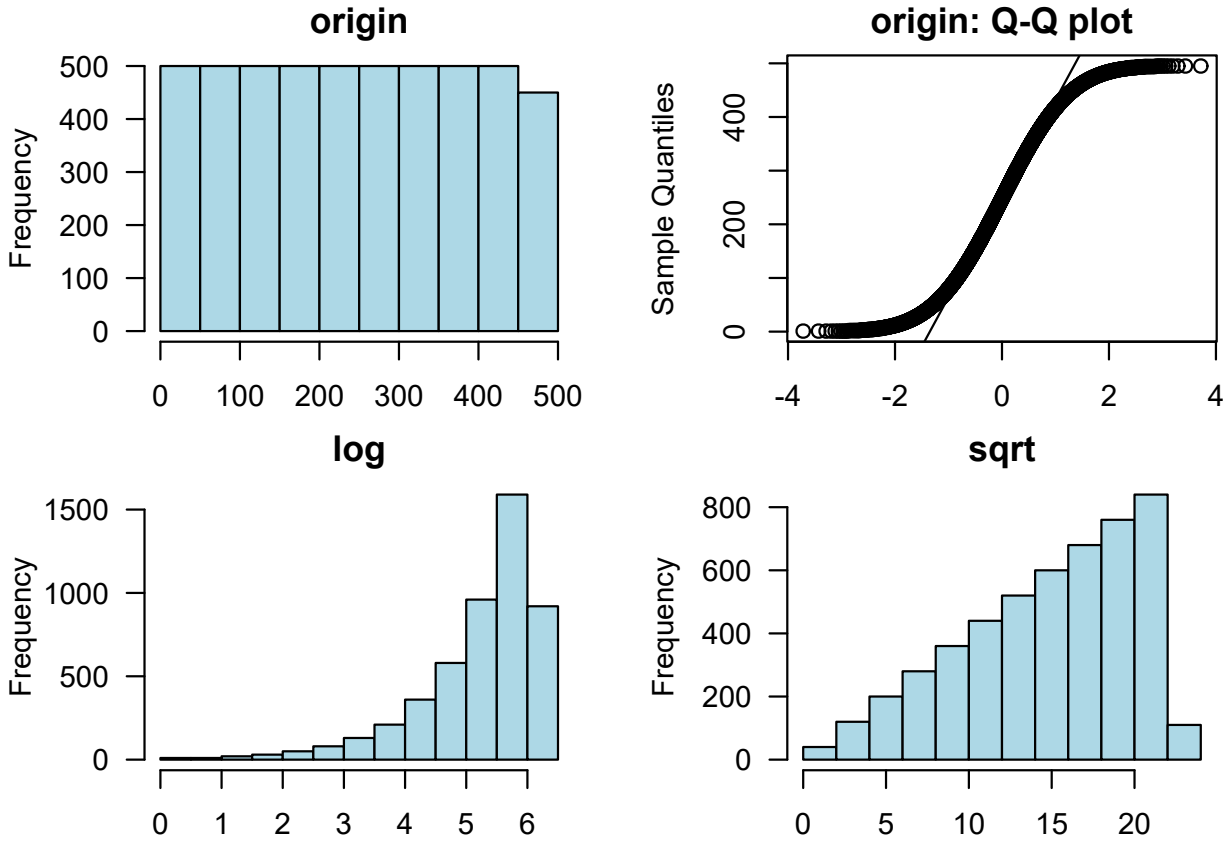

Figure 2.1: workerid

year

normality test : Shapiro-Wilk normality test
statistic : 0.93536, p-value : 5.49024E-42

type	skewness	kurtosis
original	0.0000	1.7758
log transformation	-0.8780	2.7553
sqrt transformation	-0.3965	2.0269

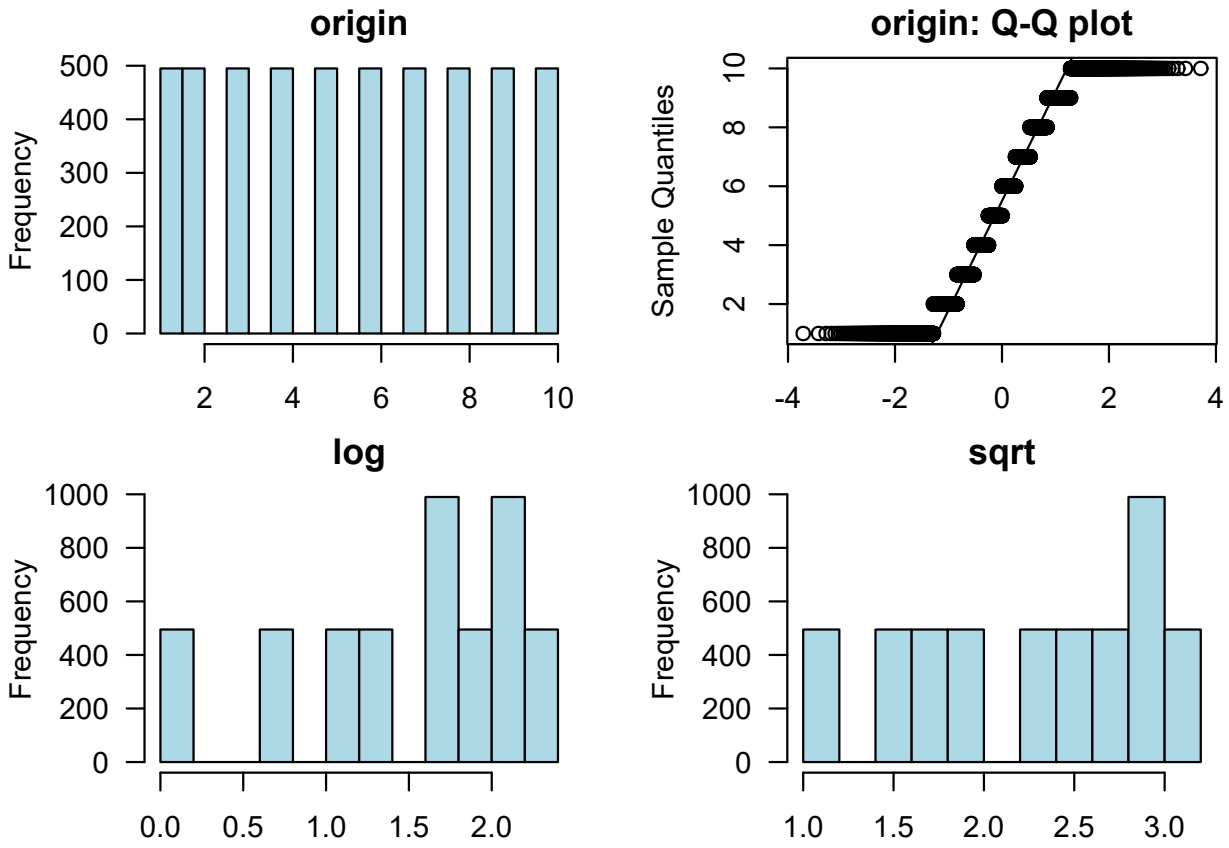

Figure 2.2: year

ui

normality test : Shapiro-Wilk normality test
statistic : 0.95431, p-value : 8.58476E-37

type	skewness	kurtosis
original	0.0512	1.8115
log transformation	-1.9405	8.8040
sqrt transformation	-0.5163	2.3238

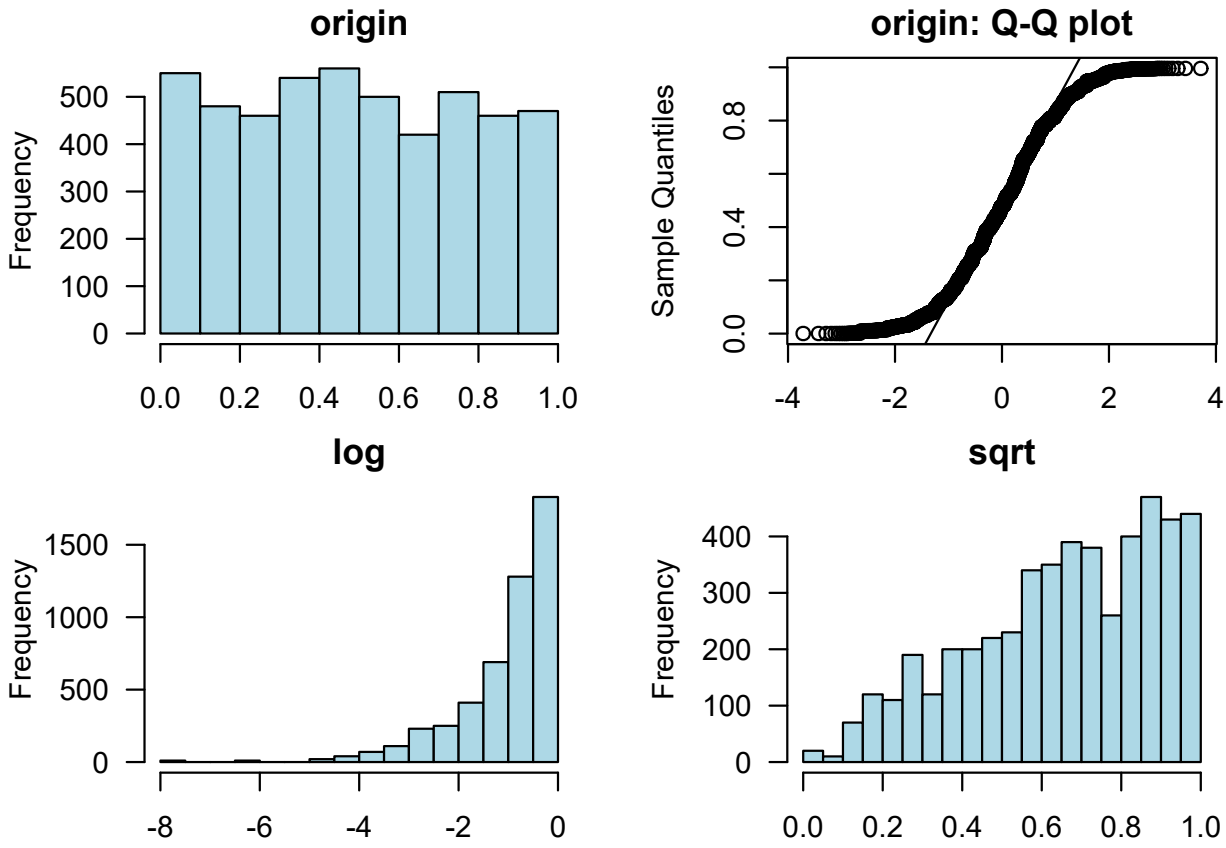

Figure 2.3: ui

quarter

normality test : Shapiro-Wilk normality test
statistic : 0.85193, p-value : 7.53006E-56

type	skewness	kurtosis
original	0.0207	1.6156
log transformation	-0.4620	1.8294
sqrt transformation	-0.2169	1.6921

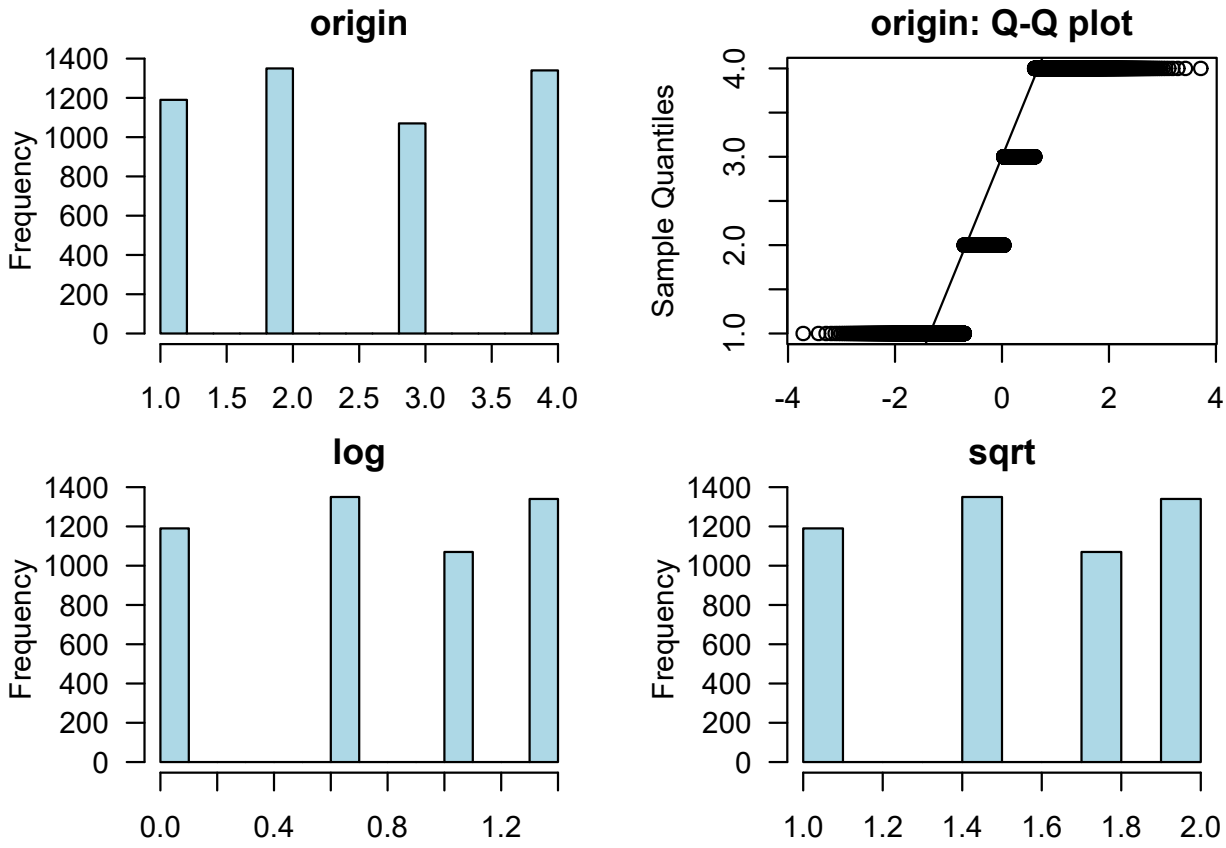

Figure 2.4: quarter

q1

normality test : Shapiro-Wilk normality test
statistic : 0.53019, p-value : 1.43711E-78

type	skewness	kurtosis
original	1.2150	2.4762
log transformation		
sqrt transformation	1.2150	2.4762

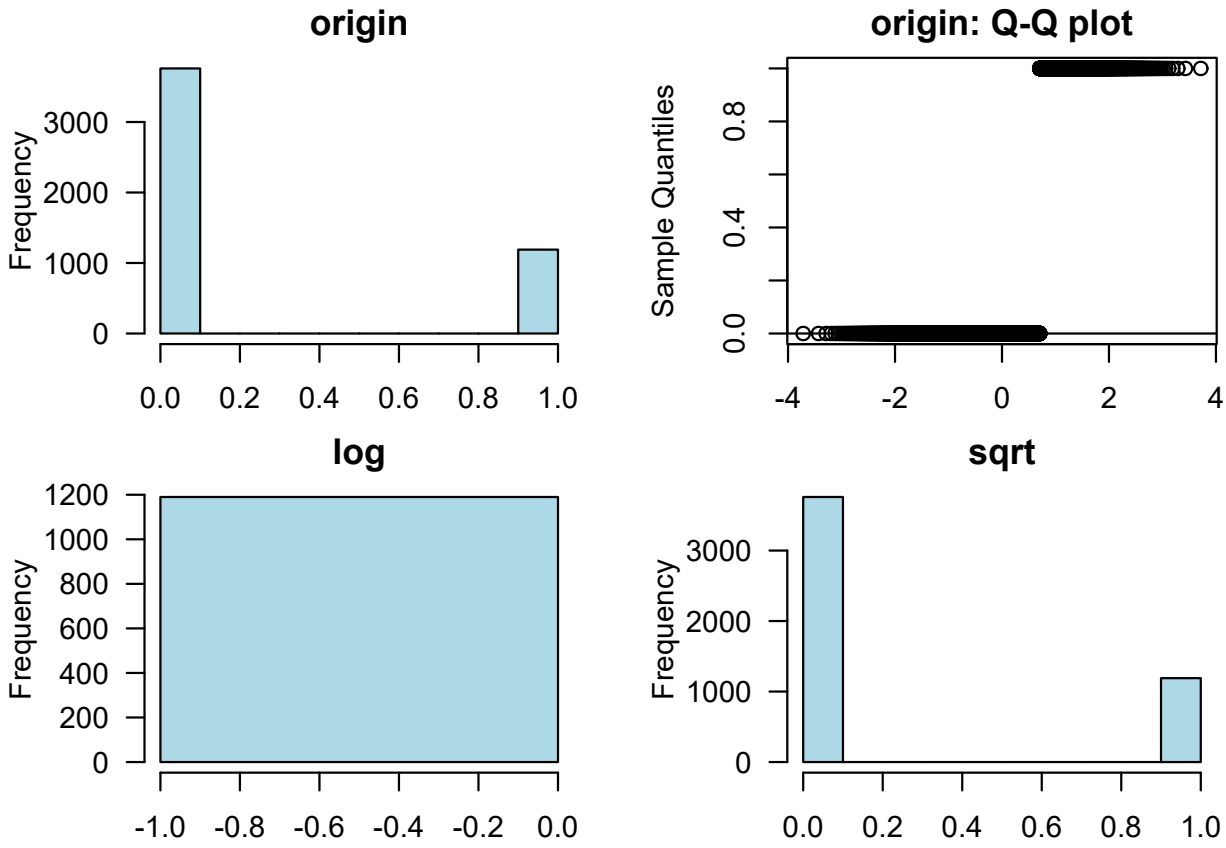

Figure 2.5: q1

wage

normality test : Shapiro-Wilk normality test
statistic : 0.88594, p-value : 2.89085E-51

type	skewness	kurtosis
original	1.3634	4.9938
log transformation	0.2803	2.4234
sqrt transformation	0.7889	3.2384

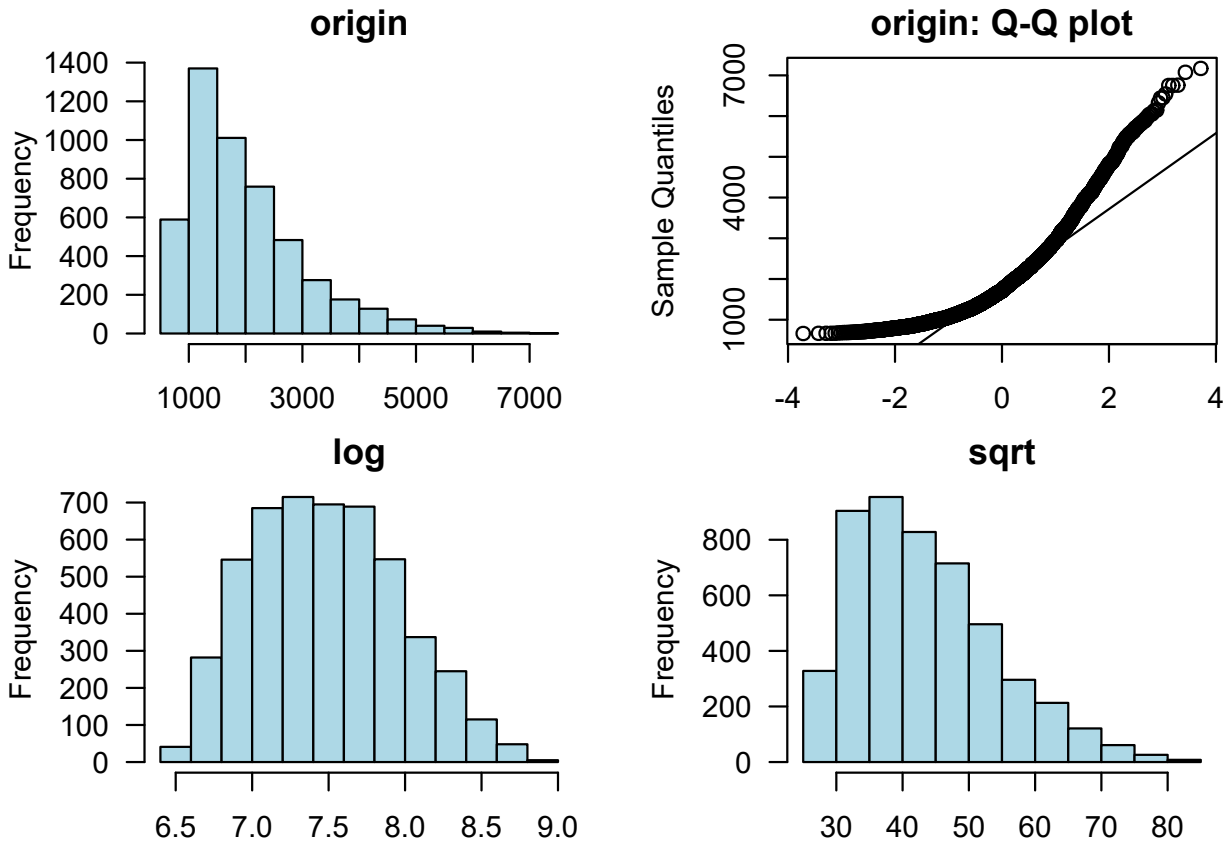

Figure 2.6: wage

educ

normality test : Shapiro-Wilk normality test
statistic : 0.95107, p-value : 8.67426E-38

type	skewness	kurtosis
original	0.7109	3.0196
log transformation		
sqrt transformation	-0.3967	2.7665

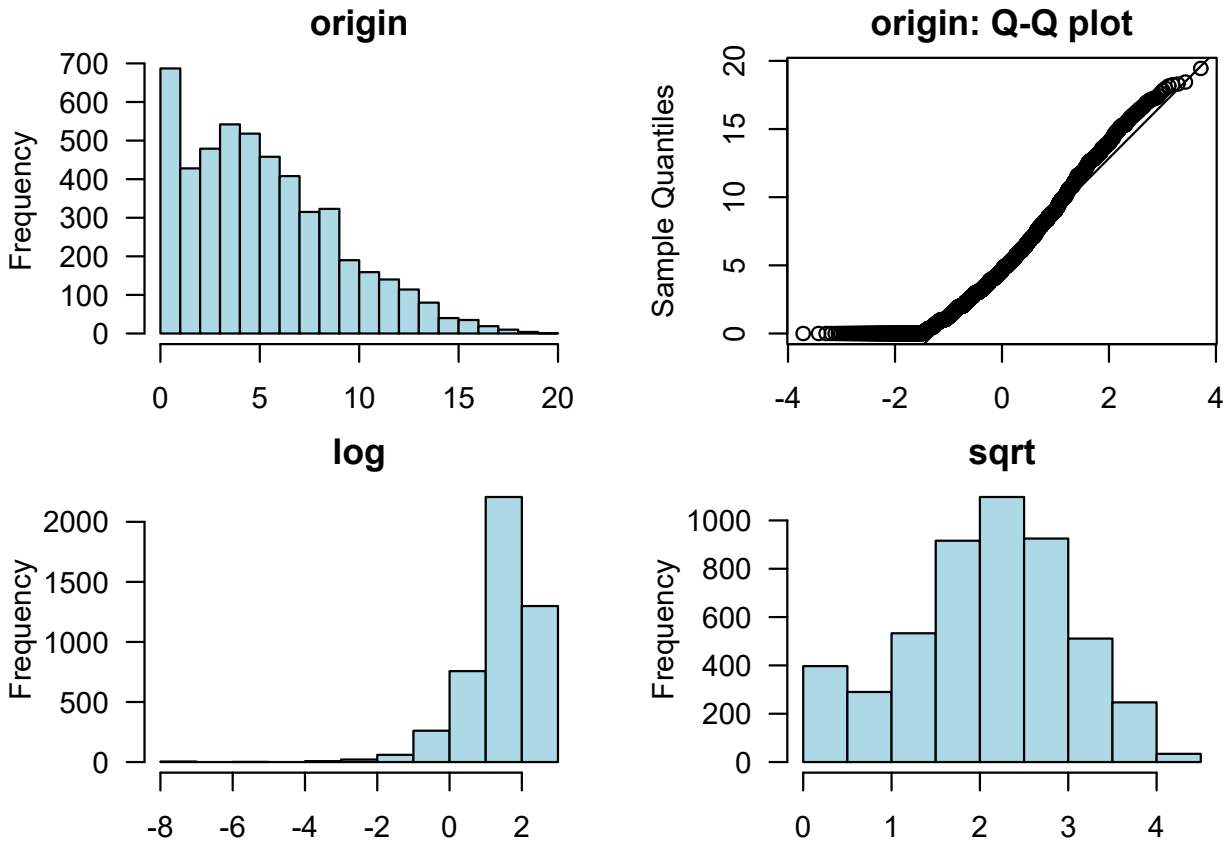

Figure 2.7: educ

exper

normality test : Shapiro-Wilk normality test
statistic : 0.9824, p-value : 2.41918E-24

type	skewness	kurtosis
original	-0.0863	2.1745
log transformation		
sqrt transformation	-0.7659	3.4342

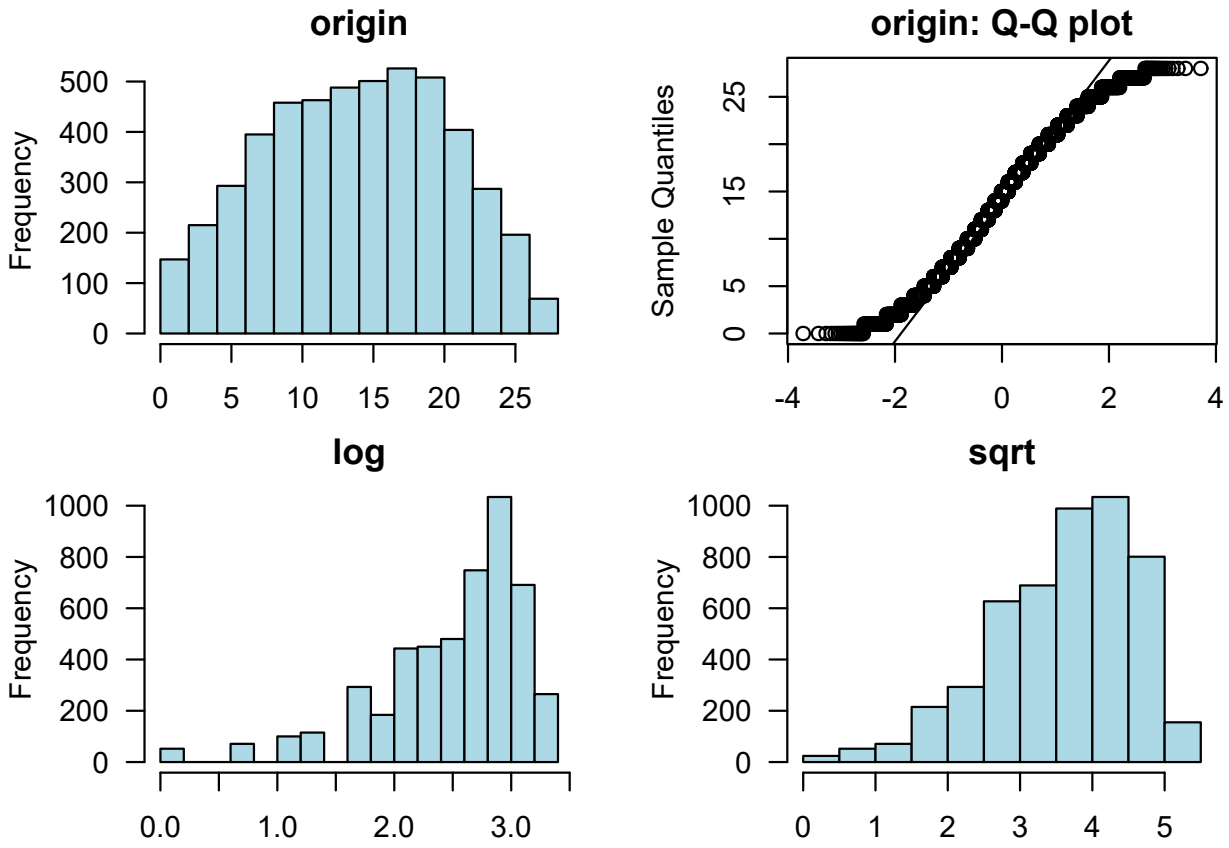

Figure 2.8: exper

union

normality test : Shapiro-Wilk normality test
statistic : 0.63635, p-value : 3.46935E-73

type	skewness	kurtosis
original	0.0550	1.0030
log transformation		
sqrt transformation	0.0550	1.0030

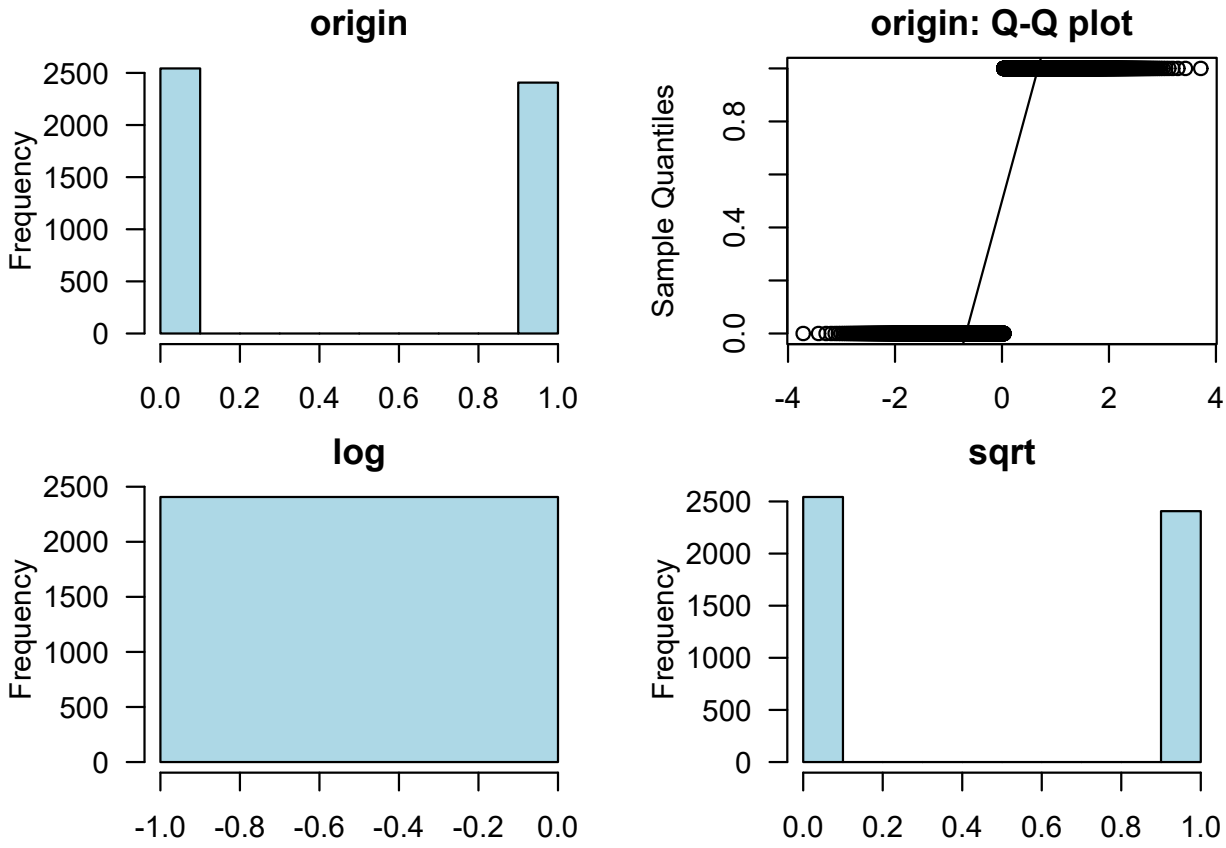

Figure 2.9: union

exper2

normality test : Shapiro-Wilk normality test
statistic : 0.9356, p-value : 6.26713E-42

type	skewness	kurtosis
original	0.6649	2.5948
log transformation		
sqrt transformation	-0.0863	2.1745

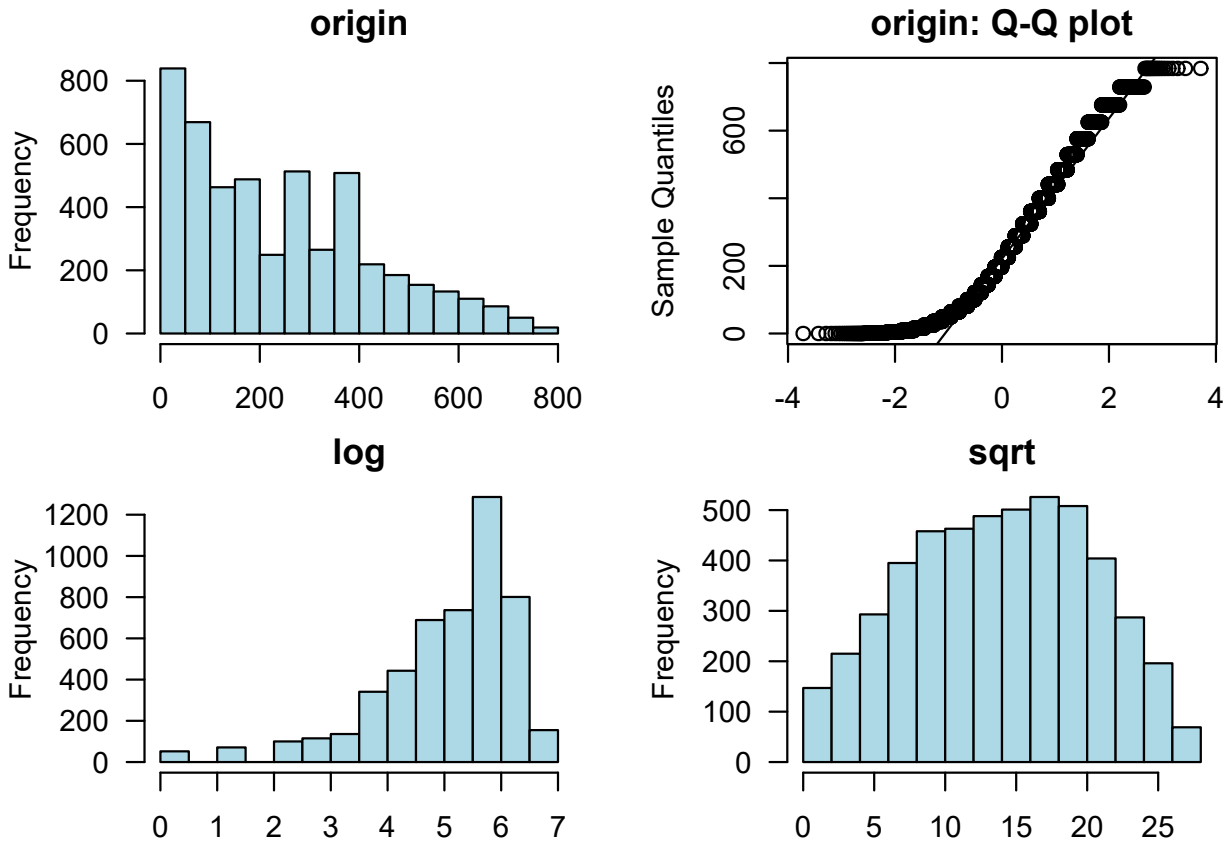

Figure 2.10: exper2

lnwage

normality test : Shapiro-Wilk normality test
statistic : 0.98514, p-value : 2.02874E-22

type	skewness	kurtosis
original	0.2803	2.4234
log transformation	0.1546	2.3344
sqrt transformation	0.2171	2.3734

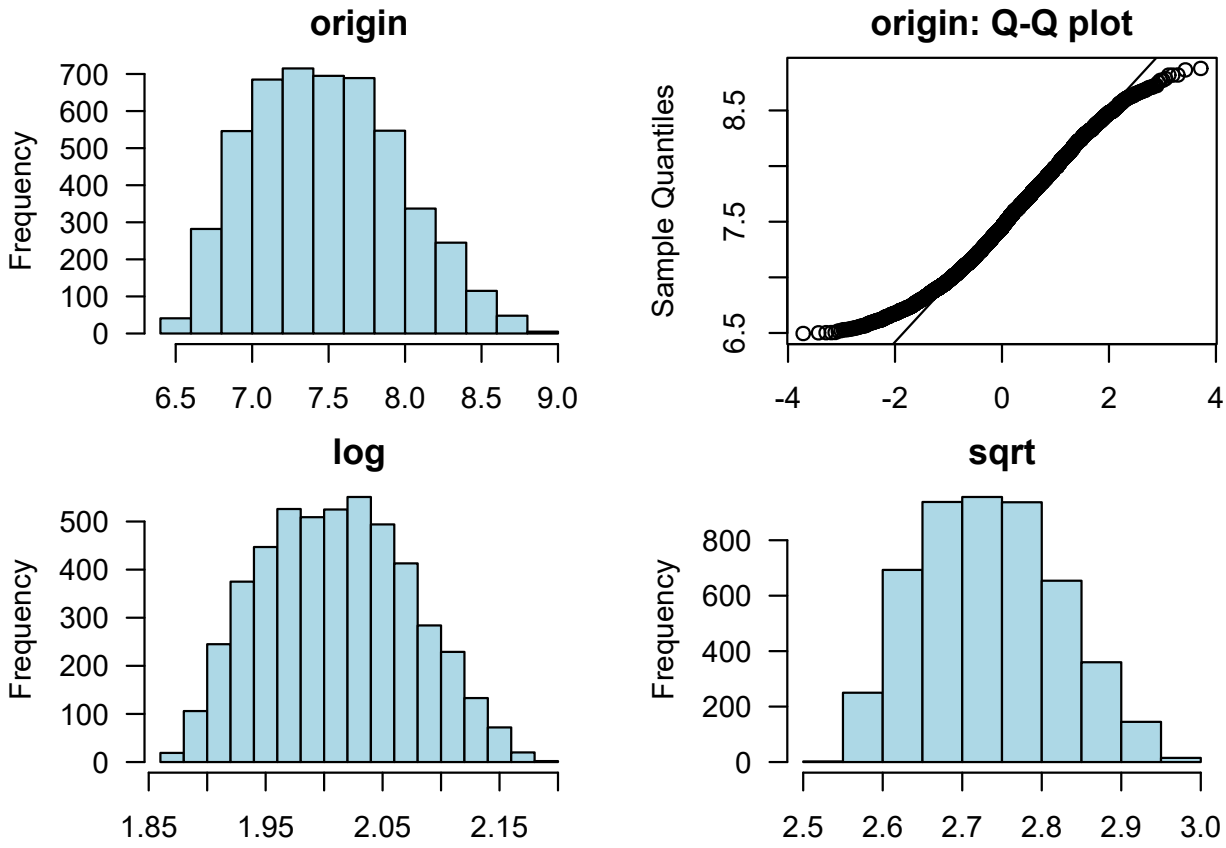

Figure 2.11: lnwage

yy1

normality test : Shapiro-Wilk normality test
statistic : 0.34218, p-value : 6.12319E-86

type	skewness	kurtosis
original	2.6667	8.1111
log transformation		
sqrt transformation	2.6667	8.1111

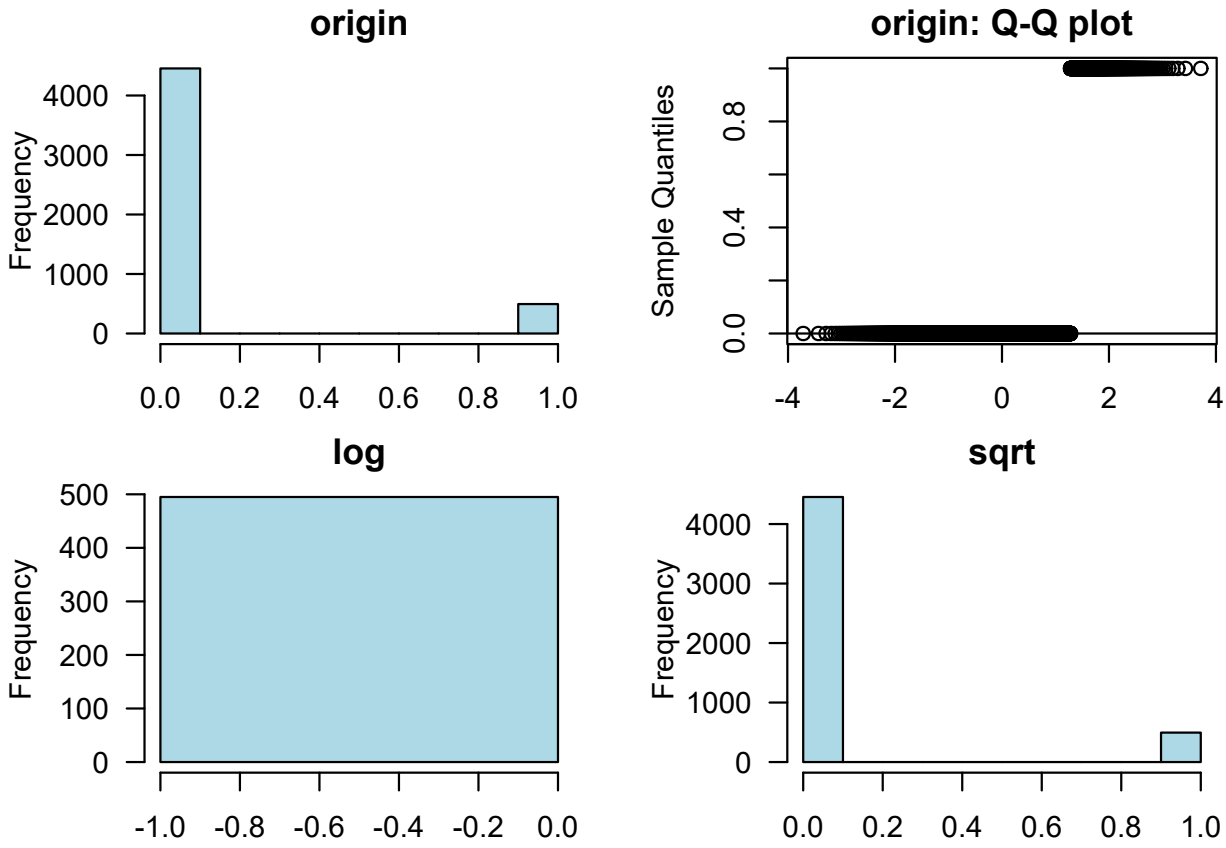

Figure 2.12: yy1

yy2

normality test : Shapiro-Wilk normality test
statistic : 0.34218, p-value : 6.12319E-86

type	skewness	kurtosis
original	2.6667	8.1111
log transformation		
sqrt transformation	2.6667	8.1111

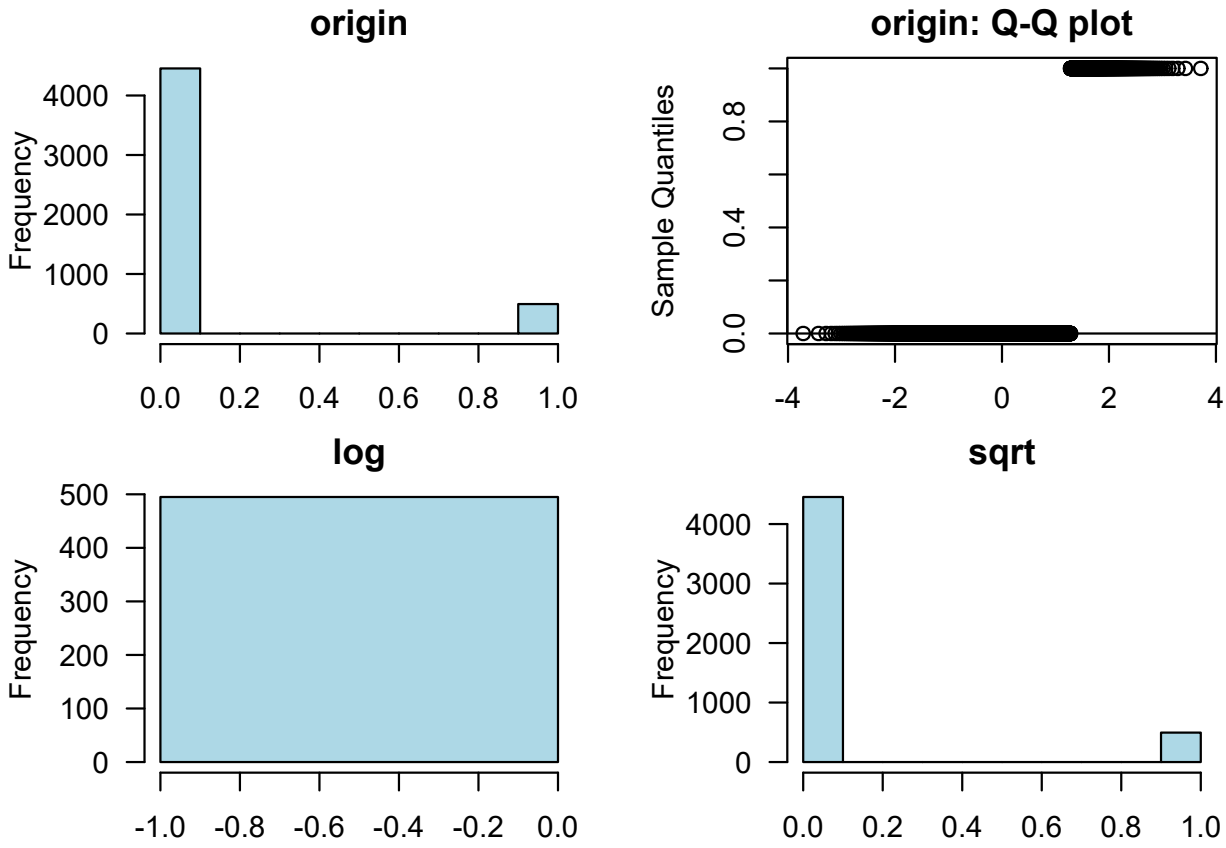

Figure 2.13: yy2

yy3

normality test : Shapiro-Wilk normality test
statistic : 0.34218, p-value : 6.12319E-86

type	skewness	kurtosis
original	2.6667	8.1111
log transformation		
sqrt transformation	2.6667	8.1111

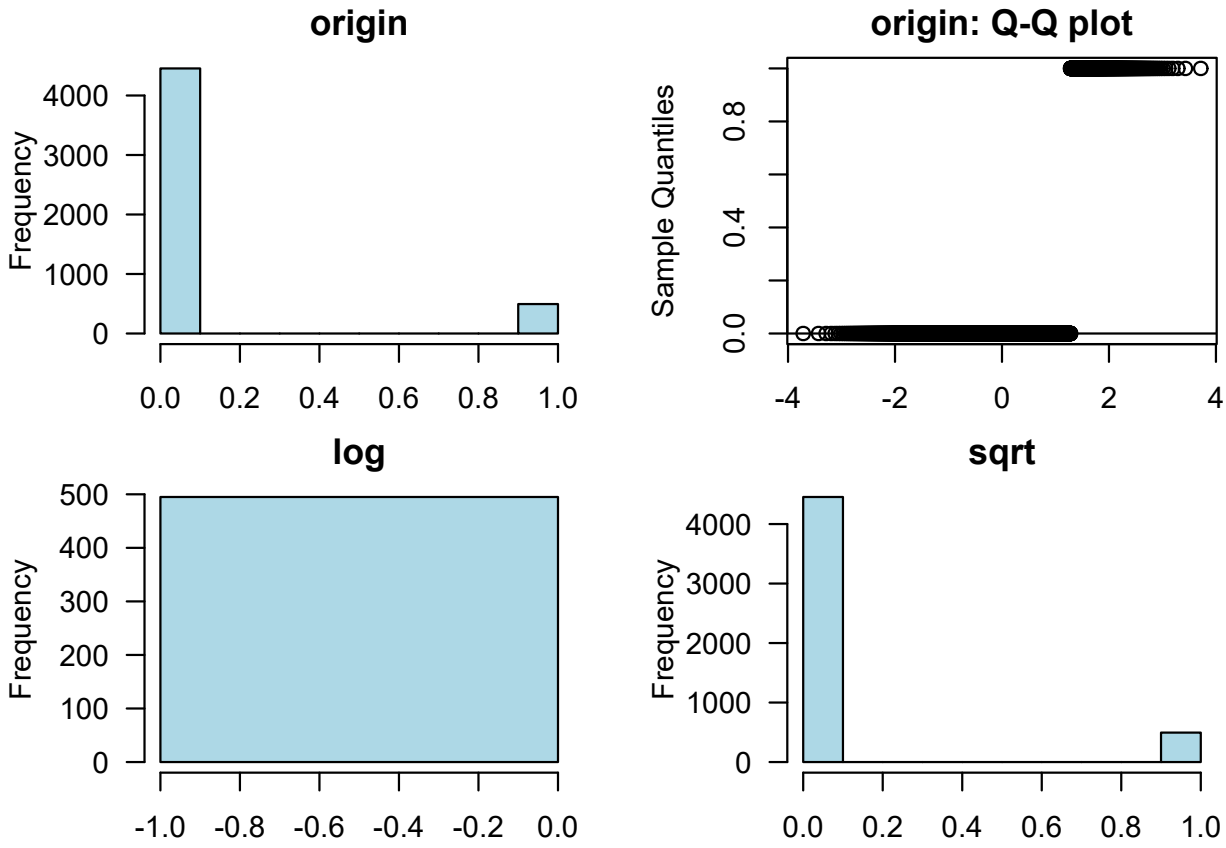

Figure 2.14: yy3

yy4

normality test : Shapiro-Wilk normality test
statistic : 0.34218, p-value : 6.12319E-86

type	skewness	kurtosis
original	2.6667	8.1111
log transformation		
sqrt transformation	2.6667	8.1111

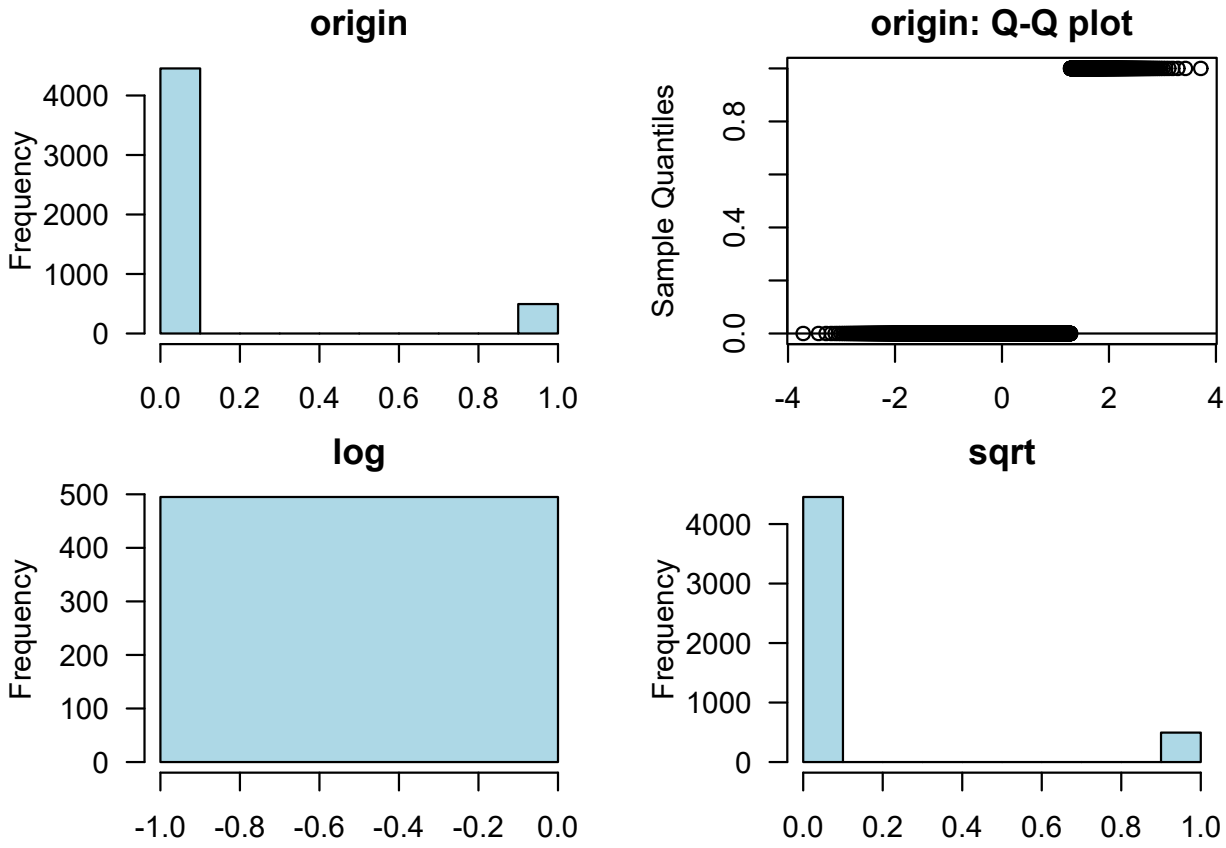

Figure 2.15: yy4

yy5

normality test : Shapiro-Wilk normality test
statistic : 0.34218, p-value : 6.12319E-86

type	skewness	kurtosis
original	2.6667	8.1111
log transformation		
sqrt transformation	2.6667	8.1111

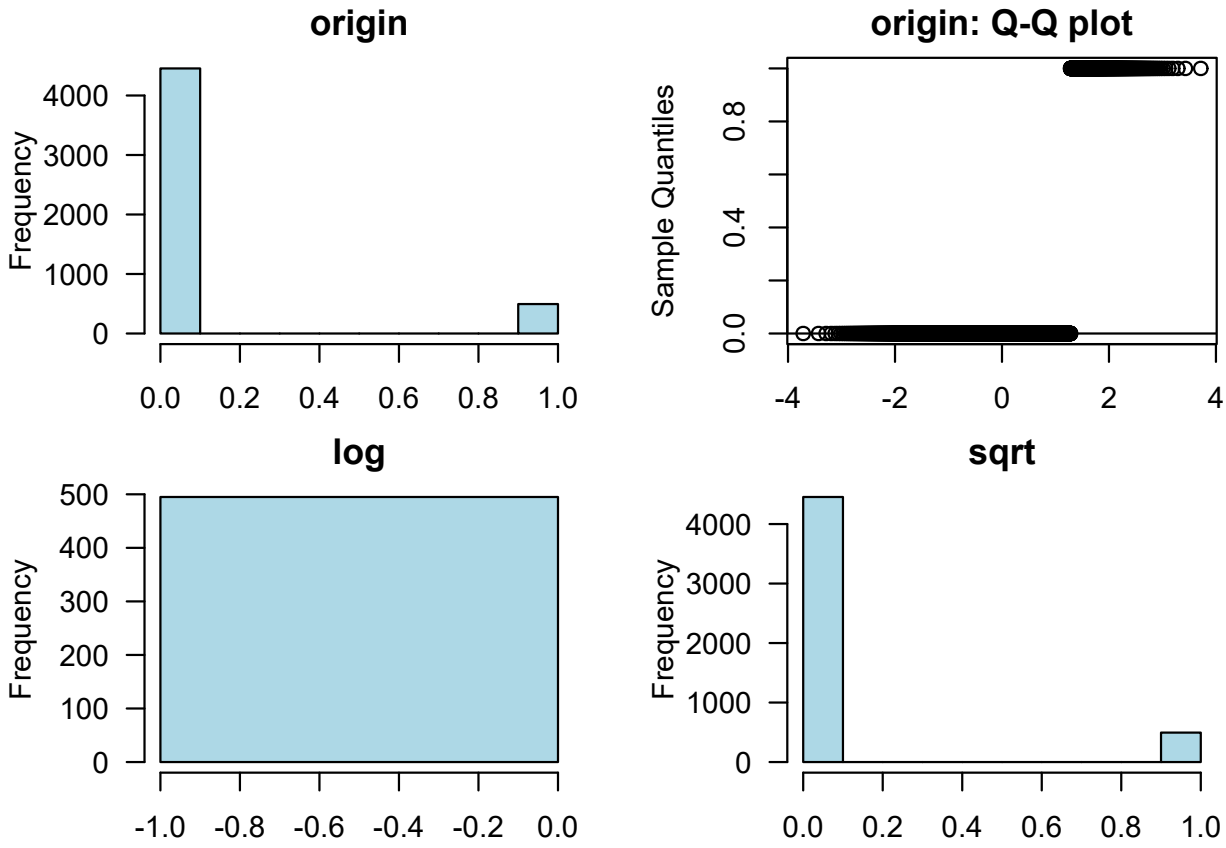

Figure 2.16: yy5

yy6

normality test : Shapiro-Wilk normality test
statistic : 0.34218, p-value : 6.12319E-86

type	skewness	kurtosis
original	2.6667	8.1111
log transformation		
sqrt transformation	2.6667	8.1111

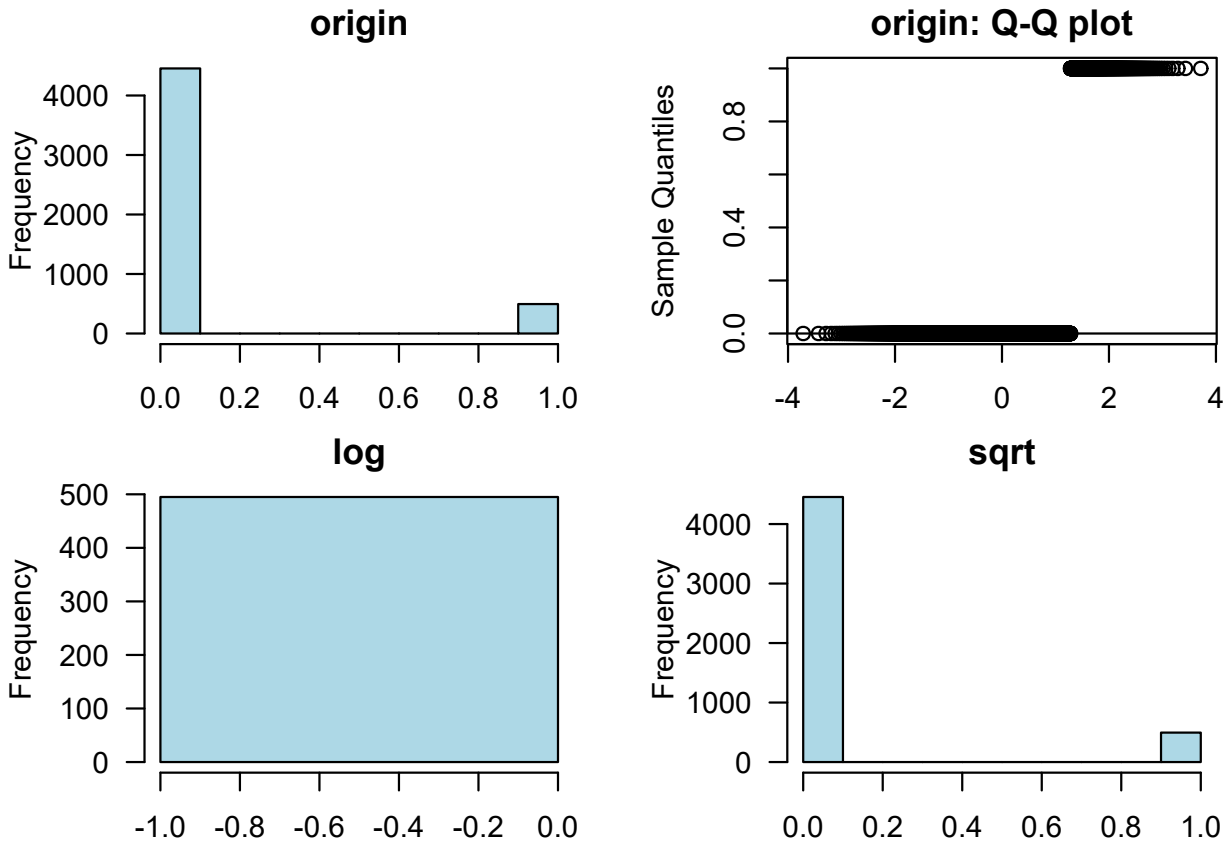

Figure 2.17: yy6

yy7

normality test : Shapiro-Wilk normality test
statistic : 0.34218, p-value : 6.12319E-86

type	skewness	kurtosis
original	2.6667	8.1111
log transformation		
sqrt transformation	2.6667	8.1111

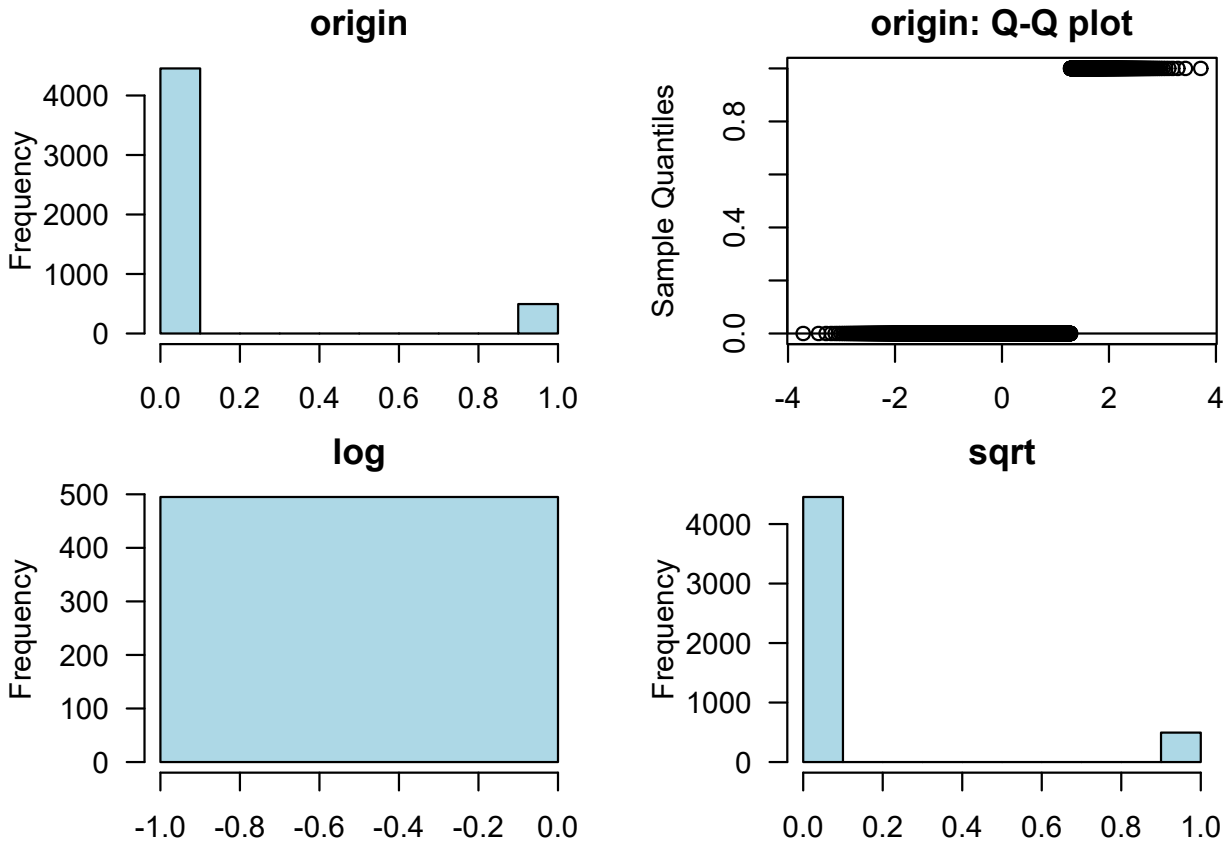

Figure 2.18: yy7

yy8

normality test : Shapiro-Wilk normality test
statistic : 0.34218, p-value : 6.12319E-86

type	skewness	kurtosis
original	2.6667	8.1111
log transformation		
sqrt transformation	2.6667	8.1111

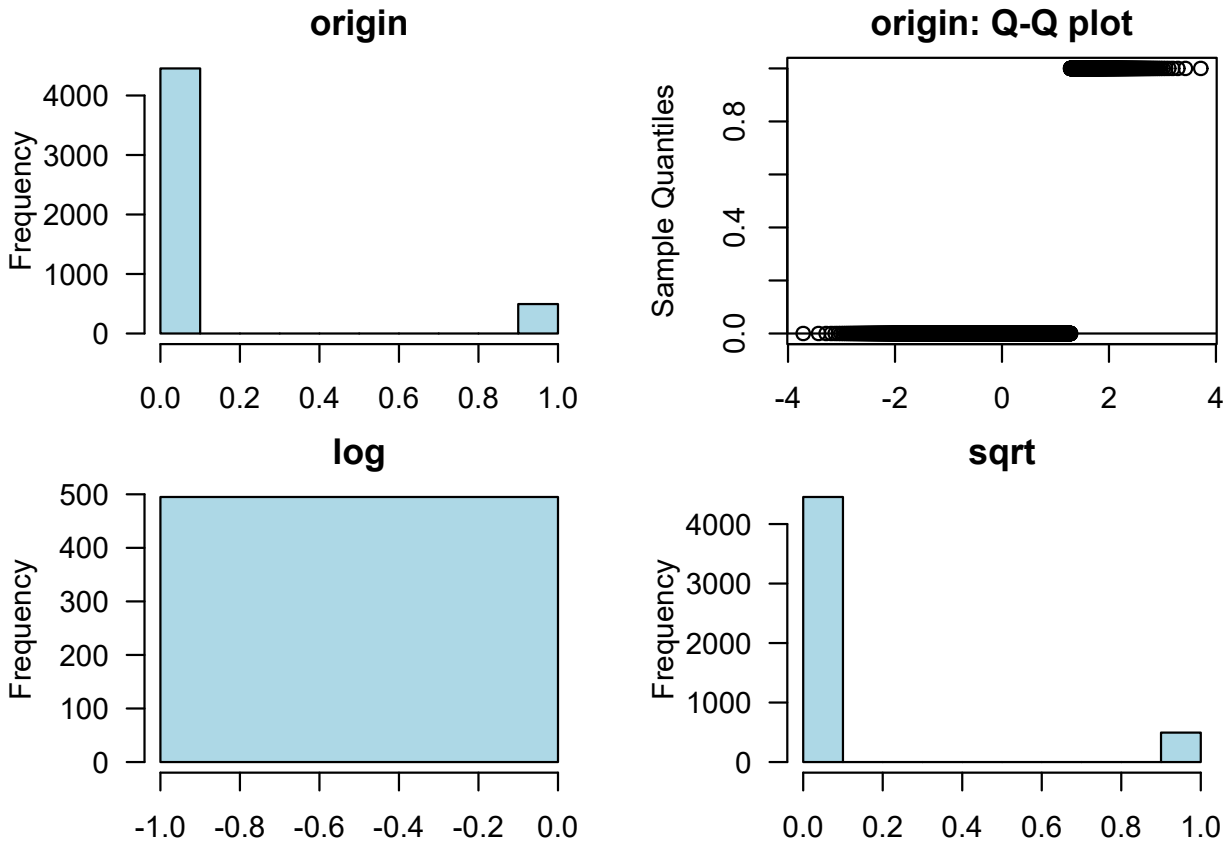

Figure 2.19: yy8

yy9

normality test : Shapiro-Wilk normality test
statistic : 0.34218, p-value : 6.12319E-86

type	skewness	kurtosis
original	2.6667	8.1111
log transformation		
sqrt transformation	2.6667	8.1111

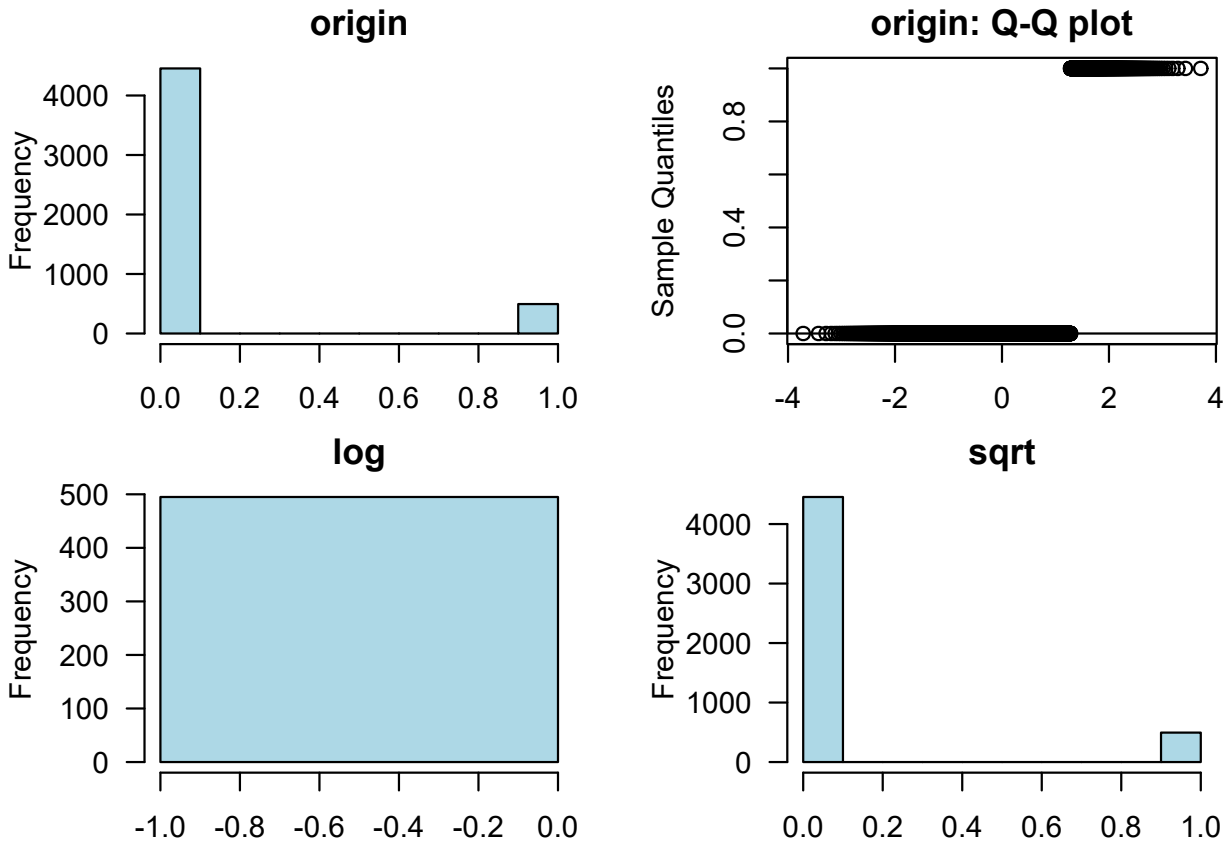

Figure 2.20: yy9

yy10

normality test : Shapiro-Wilk normality test
statistic : 0.34218, p-value : 6.12319E-86

type	skewness	kurtosis
original	2.6667	8.1111
log transformation		
sqrt transformation	2.6667	8.1111

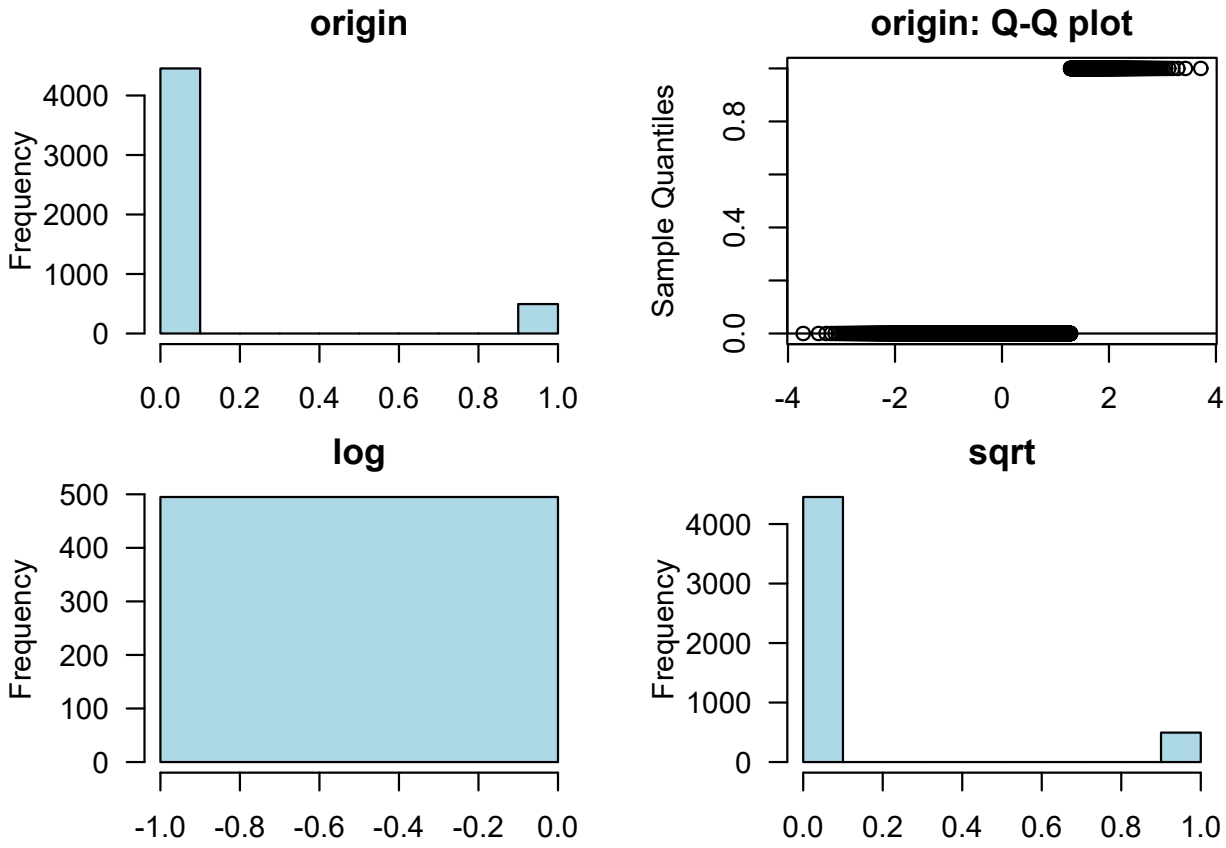

Figure 2.21: yy10

lag_lnwage

normality test : Shapiro-Wilk normality test
statistic : 0.98365, p-value : 2.66633E-22

type	skewness	kurtosis
original	0.2925	2.4111
log transformation	0.1699	2.3204
sqrt transformation	0.2309	2.3606

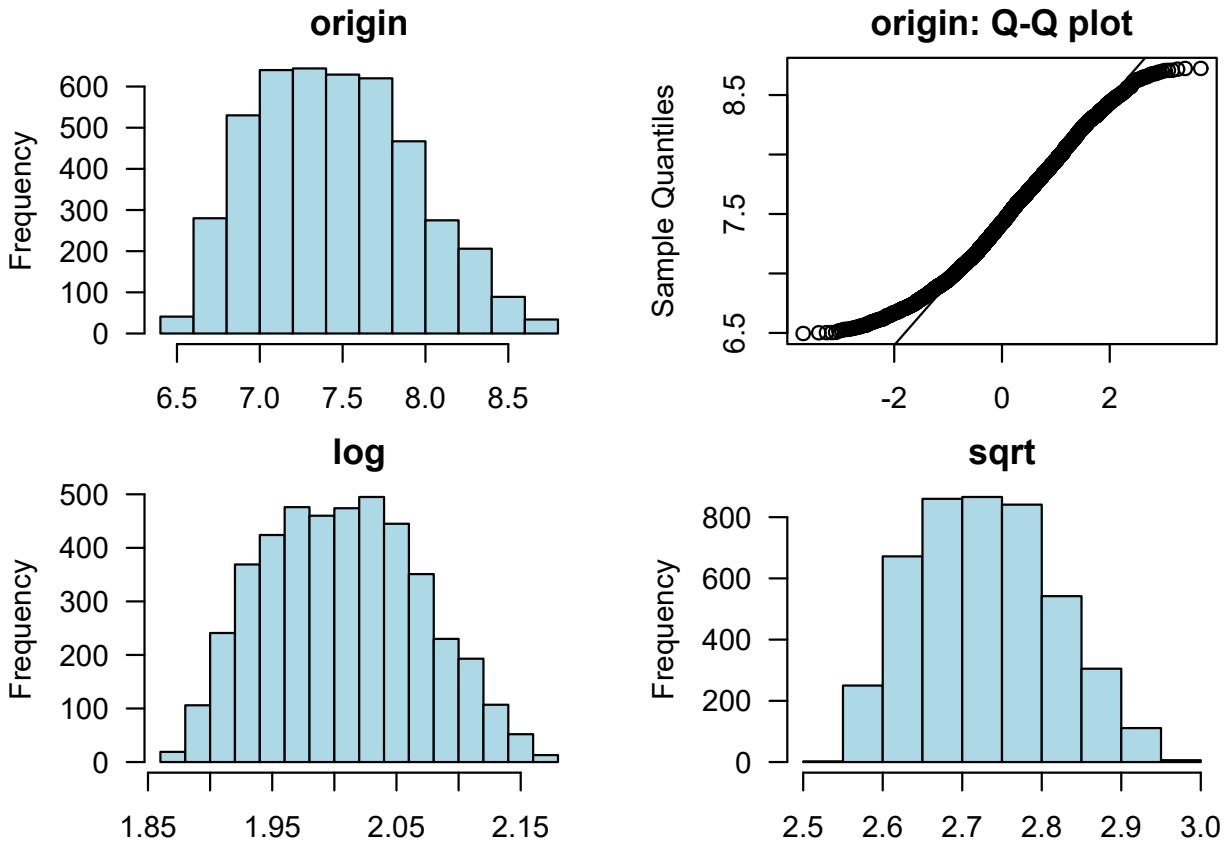

Figure 2.22: lag_lnwage

Chapter 3

Relationship Between Variables

3.1 Correlation Coefficient

3.1.1 Correlation Coefficient by Variable Combination

Table 3.1: The correlation coefficients (0.5 or more)

Variable1	Variable2	Correlation Coefficient
lag_lnwage	lnwage	0.989
exper2	exper	0.971
lnwage	wage	0.961
lag_lnwage	wage	0.952
educ	wage	0.871
lag_lnwage	ui	0.870
lnwage	educ	0.860
lnwage	ui	0.859
lag_lnwage	educ	0.849
wage	ui	0.794
q1	quarter	-0.757
educ	ui	0.533
yy1	year	-0.522
yy10	year	0.522

3.1.2 Correlation Plot of Numerical Variables

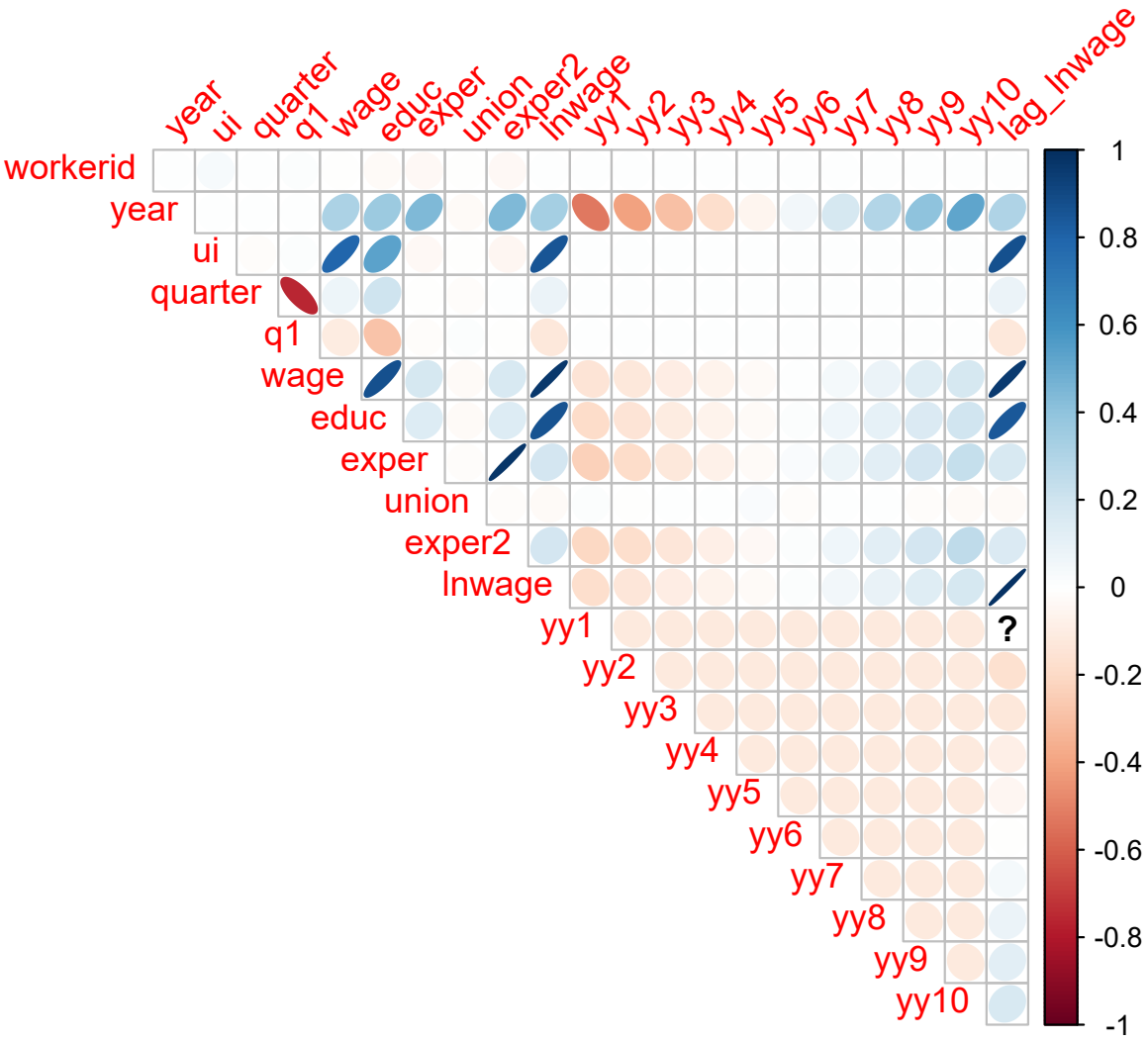

Figure 3.1: The correlation coefficient of numerical variables

Chapter 4

Target based Analysis

4.1 Grouped Descriptive Statistics

4.1.1 Grouped Numerical Variables

There is no target variable.

4.1.2 Grouped Categorical Variables

There is no target variable.

4.2 Grouped Relationship Between Variables

4.2.1 Grouped Correlation Coefficient

There is no target variable.

4.2.2 Grouped Correlation Plot of Numerical Variables

There is no target variable.