Homework5

Yuki Joyama

2023-12-12

Problem 1

a) The following table shows the descriptive statistics for all variables of interest in 50 States.

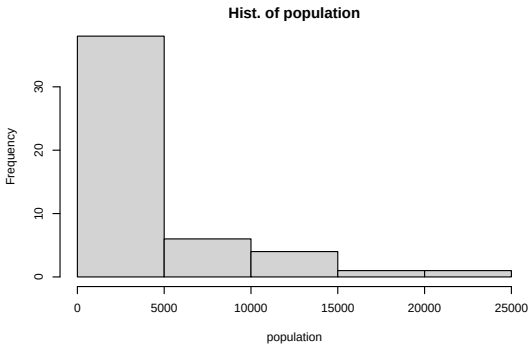
Characteristic	**N = 50**[1]
Population	4,246.4 / 2,838.5 (4,464.5)
Income per capita	4,435.8 / 4,519.0 (614.5)
Illiteracy (%)	1.2 / 1.0 (0.6)
Life Expectancy (years)	70.9 / 70.7 (1.3)
Murder rate (per 100,000)	7.4 / 6.9 (3.7)
High graduates (%)	53.1 / 53.3 (8.1)
Number of days below freezing	104.5 / 114.5 (52.0)
Land area (mile ^2)	70,735.9 / 54,277.0 (85,327.3)

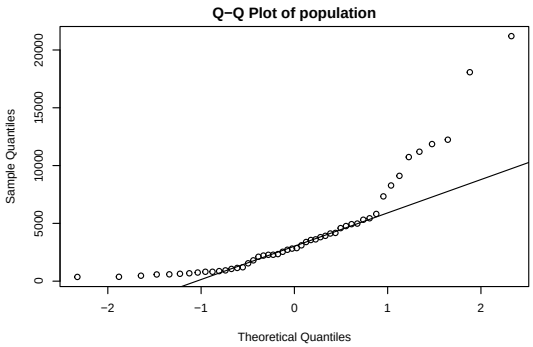
[1]Mean / Median (SD)

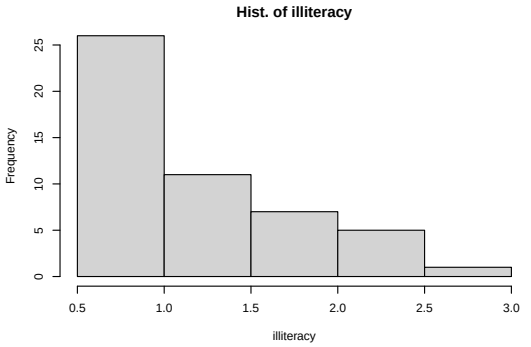
b)

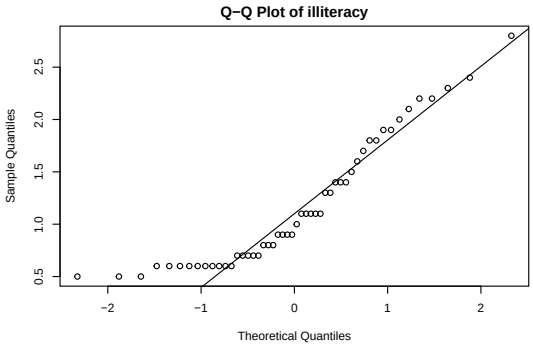
var	statistic	p.value
population	0.770	<0.001
income	0.977	0.43
illiteracy	0.883	<0.001
life_exp	0.977	0.442
murder	0.953	0.047
hs_grad	0.953	0.046
frost	0.955	0.053
area	0.572	<0.001

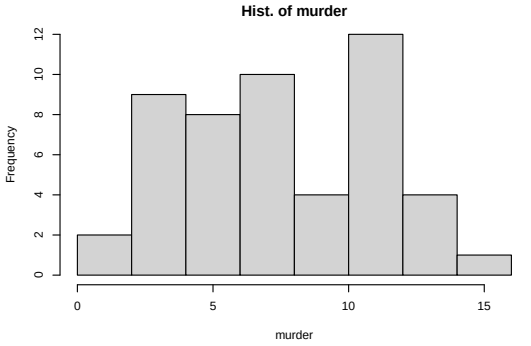
The results of Shapiro-Wilk test indicates that variable `population`, `illiteracy`, `murder`, `hs_grad`, and `area` is not normally distributed.
The histogram and Q-Q plots for these variables are as follows:

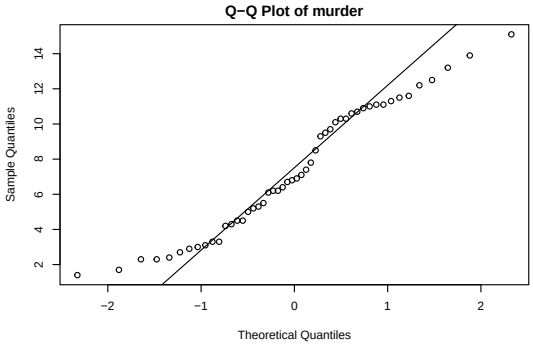

Hist. of population
Frequency
0
10
20
30
0
5000
10000
15000
20000
25000
population

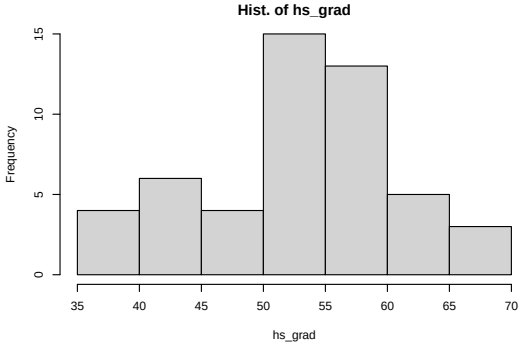

Q−Q Plot of population
Sample Quantiles
0
5000
10000
15000
20000
−2
−1
0
1
2
Theoretical Quantiles

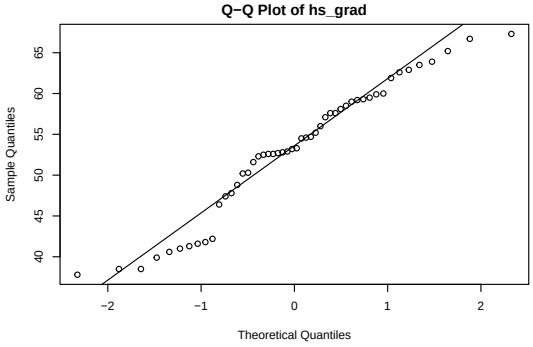

Hist. of illiteracy
Frequency
0
5
10
15
20
25
0.5
1.0
1.5
2.0
2.5
3.0
illiteracy

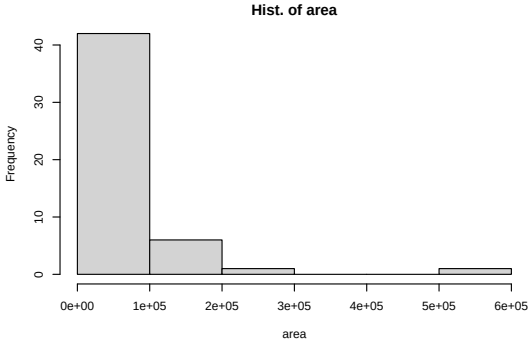

Q−Q Plot of illiteracy
Sample Quantiles
0.5
1.0
1.5
2.0
2.5
−2
−1
0
1
2
Theoretical Quantiles

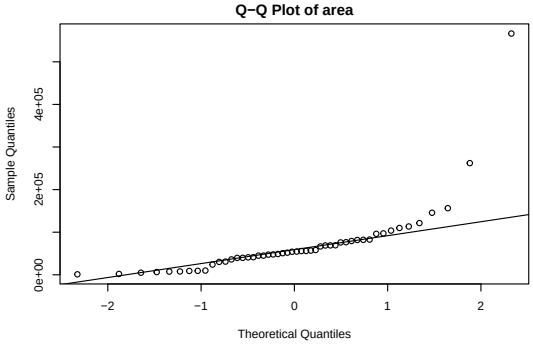

Hist. of murder
Frequency
0
2
4
6
8
10
12
0
5
10
15
murder


Q−Q Plot of murder
Sample Quantiles
2
4
6
8
10
12
14
−2
−1
0
1
2
Theoretical Quantiles


Hist. of hs_grad
Frequency
0
5
10
15
35
40
45
50
55
60
65
70
hs_grad

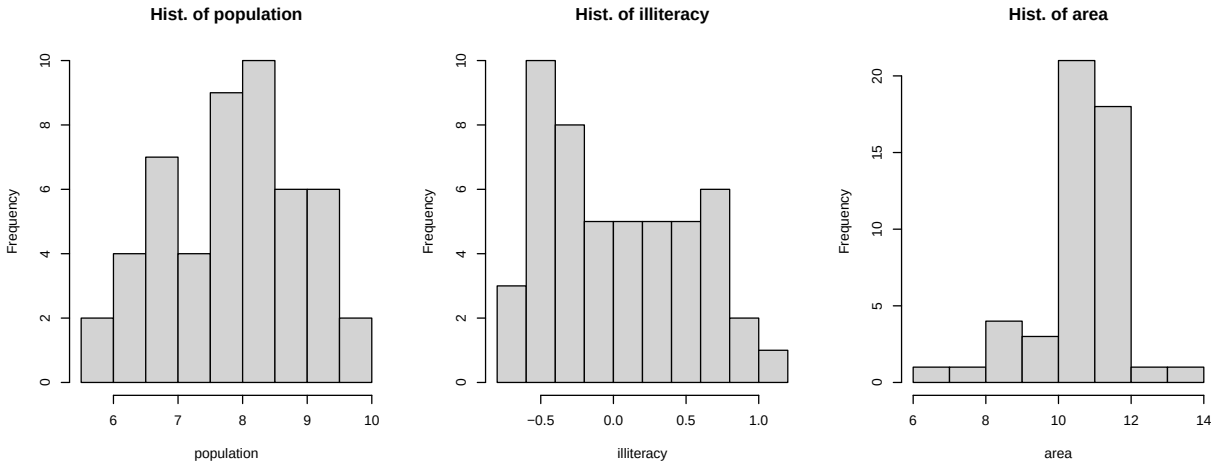

Q−Q Plot of hs_grad
Sample Quantiles
40
45
50
55
60
65
−2
−1
0
1
2
Theoretical Quantiles


Hist. of area
Frequency
0
10
20
30
40
0e+00
1e+05
2e+05
3e+05
4e+05
5e+05
6e+05
area


Q−Q Plot of area
Sample Quantiles
0e+00
2e+05
4e+05
−2
−1
0
1
2
Theoretical Quantiles

Given the shape of the histograms, I will log-transform `population`, `illiteracy`, and `area`.

Now, let's check these histograms.



c) Automatic procedures

d) Criterion-based procedures

e) The LASSO method

f) Compare the subsets from c, d, and e

g) Findings