

Assignment 3

Multivariate Statistical Analysis



August 28, 2019

PW Janse van rensburg

15338673

Question 6

7.18

**data** ex7\_18;

input X1 X2 X3;

datalines;

108.28 17.05 1484.10

152.36 16.59 750.33

95.04 10.91 766.42

65.45 14.14 1110.46

62.97 9.52 1031.29

263.99 25.33 195.26

265.19 18.54 193.83

285.06 15.73 191.11

92.01 8.10 1175.16

165.68 11.13 211.15

;

**proc** **reg** data = ex7\_18;

title 'Variable selection';

model X1 = X2 X3 / selection= cp aic;

plot cp.\*np. / cmallows=green;

model X2 = X1 X3 / selection= cp aic;

plot cp.\*np. / cmallows=green;

model X3 = X1 X2 / selection= cp aic;

plot cp.\*np. / cmallows=green;

**run**;

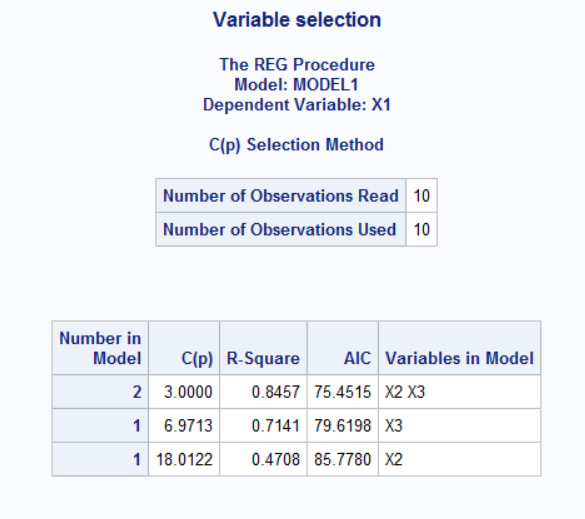
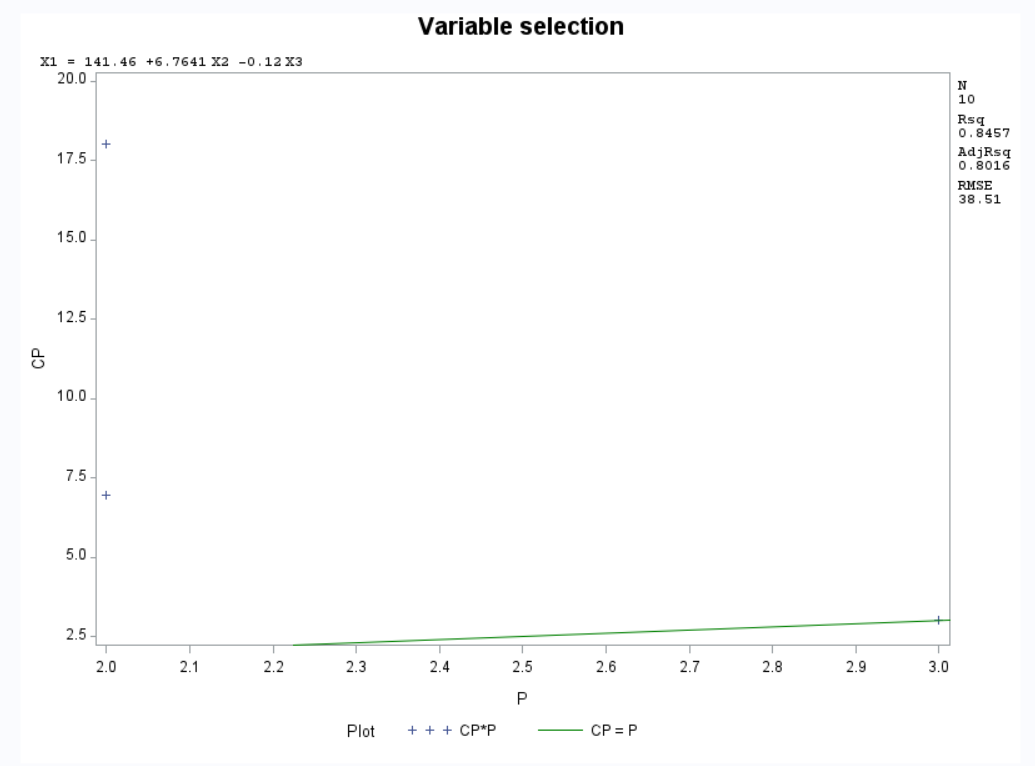


Figure 1



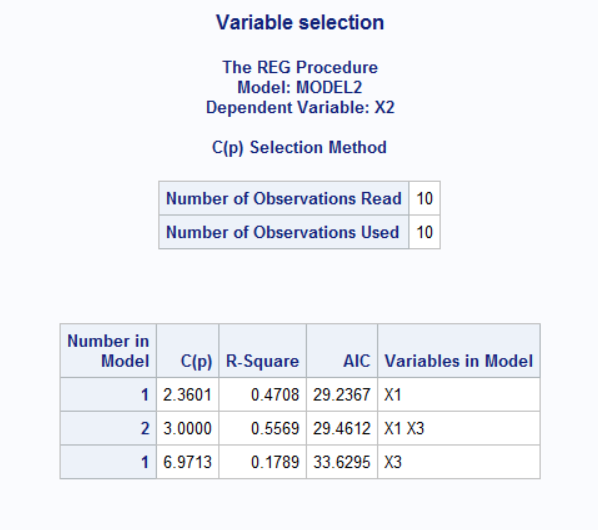
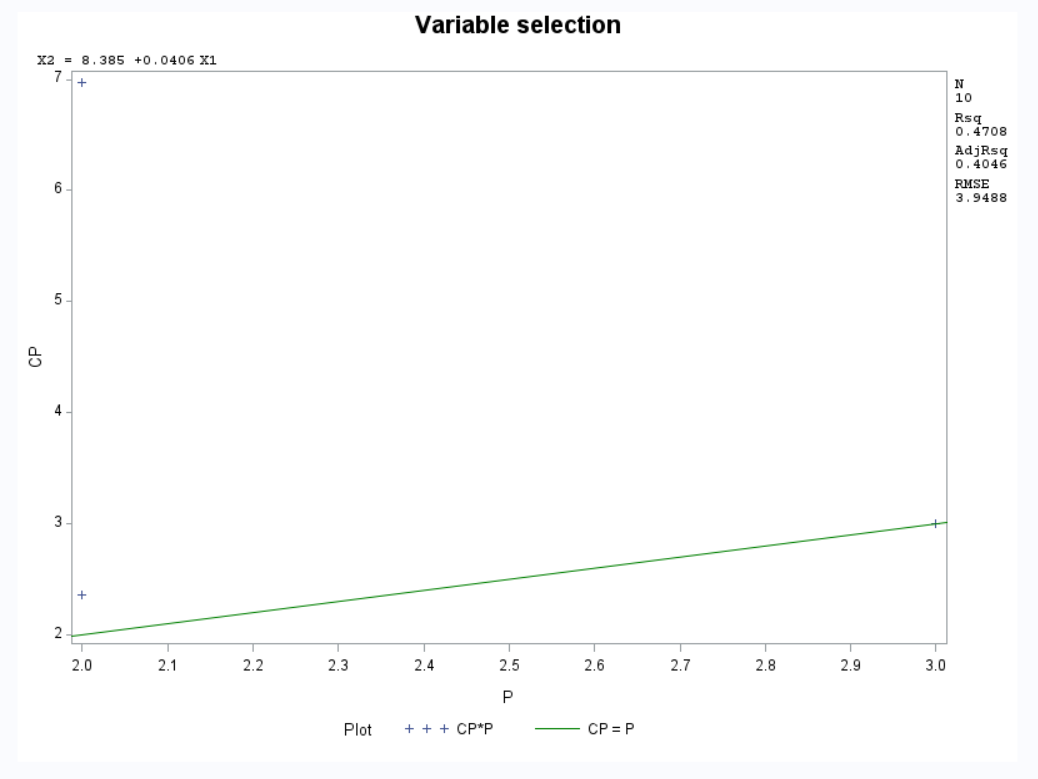


Figure 2



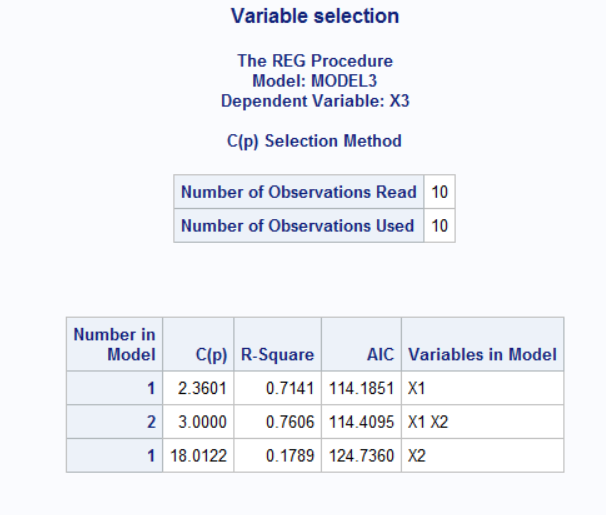
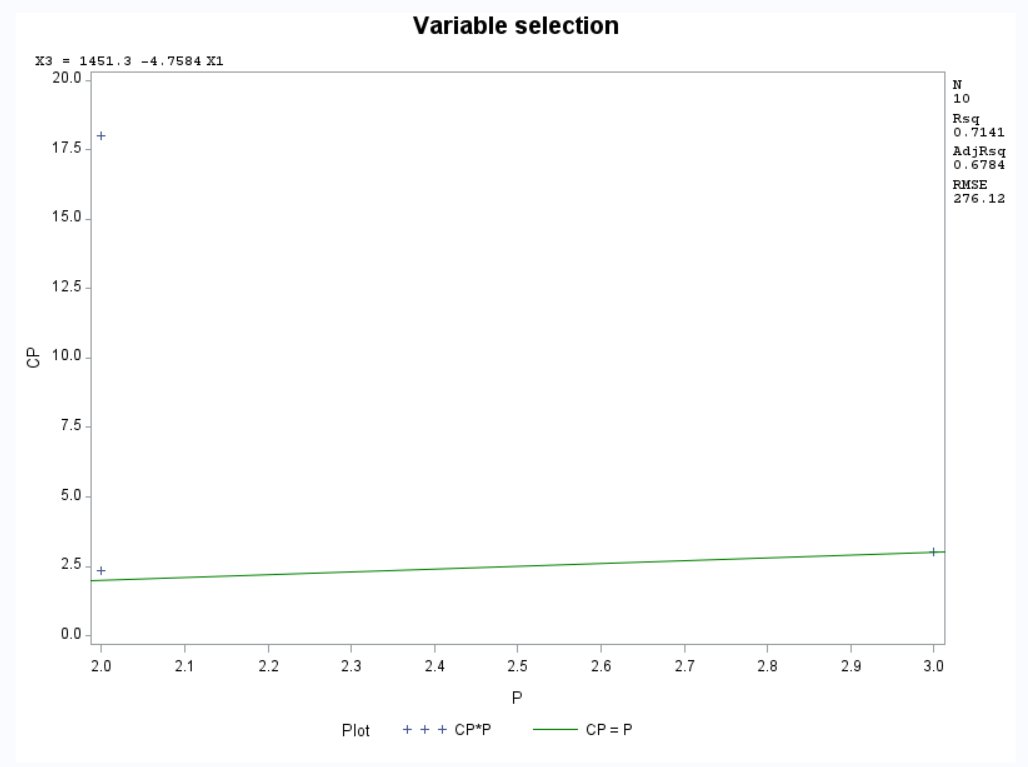


Figure 3



7.19.a.

**data** ex7\_19;

input Z1 Z2 Z3 Z4 Z5 Y;

datalines;

0.375 3.13 60.0 40 2.00 101

1.000 3.13 76.8 30 1.99 141

1.000 3.13 60.0 20 2.00 96

1.000 3.13 60.0 20 1.98 125

1.625 3.13 43.2 10 2.01 43

1.625 3.13 60.0 20 2.00 16

1.625 3.13 60.0 20 2.02 188

0.375 5.00 76.8 10 2.01 10

1.000 5.00 43.2 10 1.99 3

1.000 5.00 43.2 30 2.01 386

1.000 5.00 100.0 20 2.00 45

1.625 5.00 76.8 10 1.99 2

0.375 1.25 76.8 10 2.01 76

1.000 1.25 43.2 10 1.99 78

1.000 1.25 76.8 30 2.00 160

1.000 1.25 60.0 0 2.00 3

1.625 1.25 43.2 30 1.99 216

1.625 1.25 60.0 20 2.00 73

0.375 3.13 76.8 30 1.99 314

0.375 3.13 60.0 20 2.00 170

;

**data** ex7\_19;

set ex7\_19;

A = log10(Y);

**run**;

ods graphics on;

**proc** **reg** data=ex7\_19;

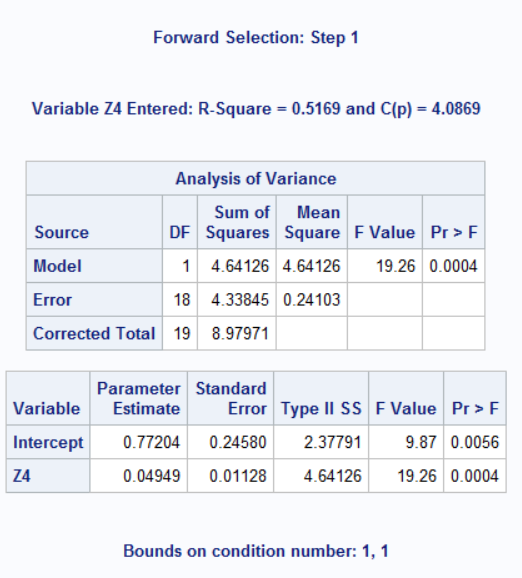
model A = Z1 Z2 Z3 Z4 Z5 / selection=cp aic;

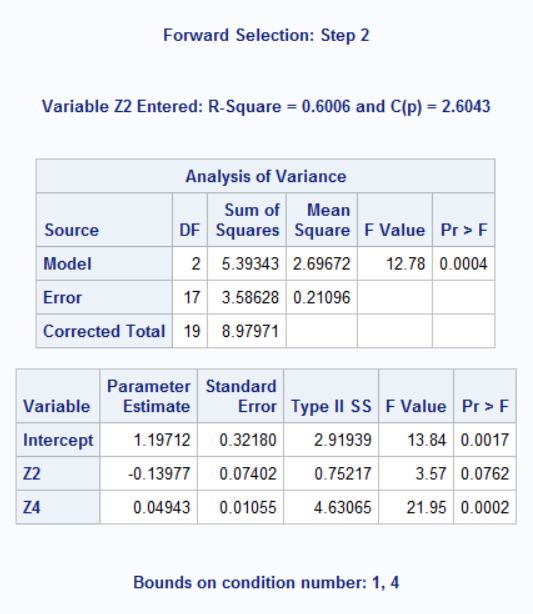
**run**;

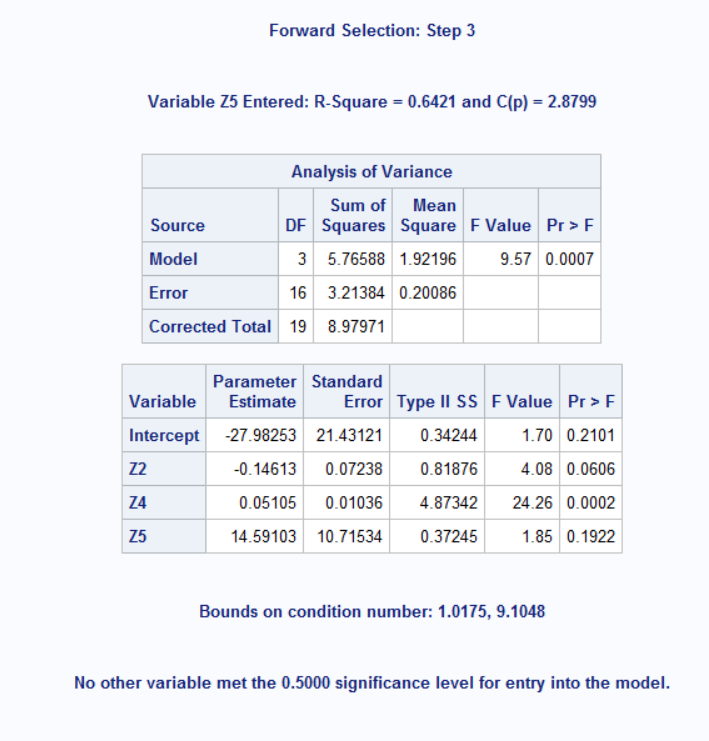
**proc** **reg** data=ex7\_19;

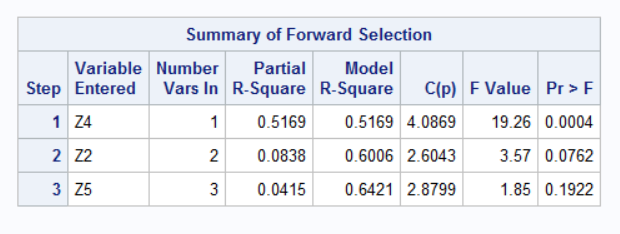
model A = Z1 Z2 Z3 Z4 Z5 / selection=forward;

**run**;









7.19.b

**proc** **reg** data=ex7\_19;

model A = Z2 Z4 Z5;

plot rstudent.\*p.;

plot nqq.\* rstudent.;

output out=B student=resid;

**run**;

Figure 4

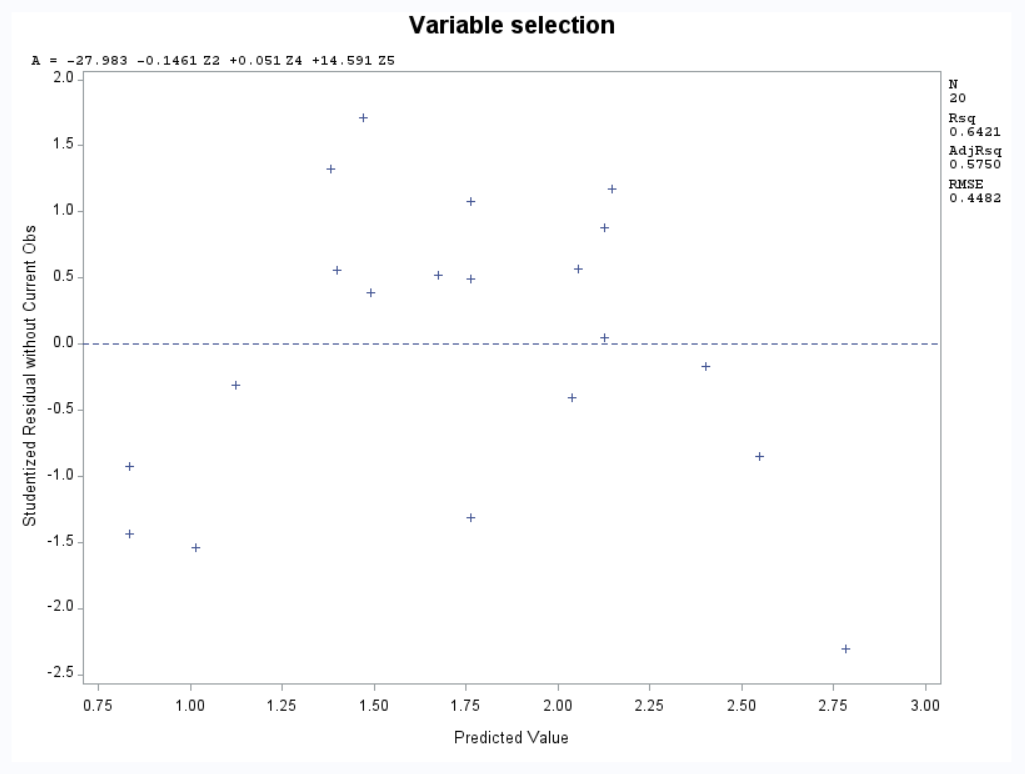
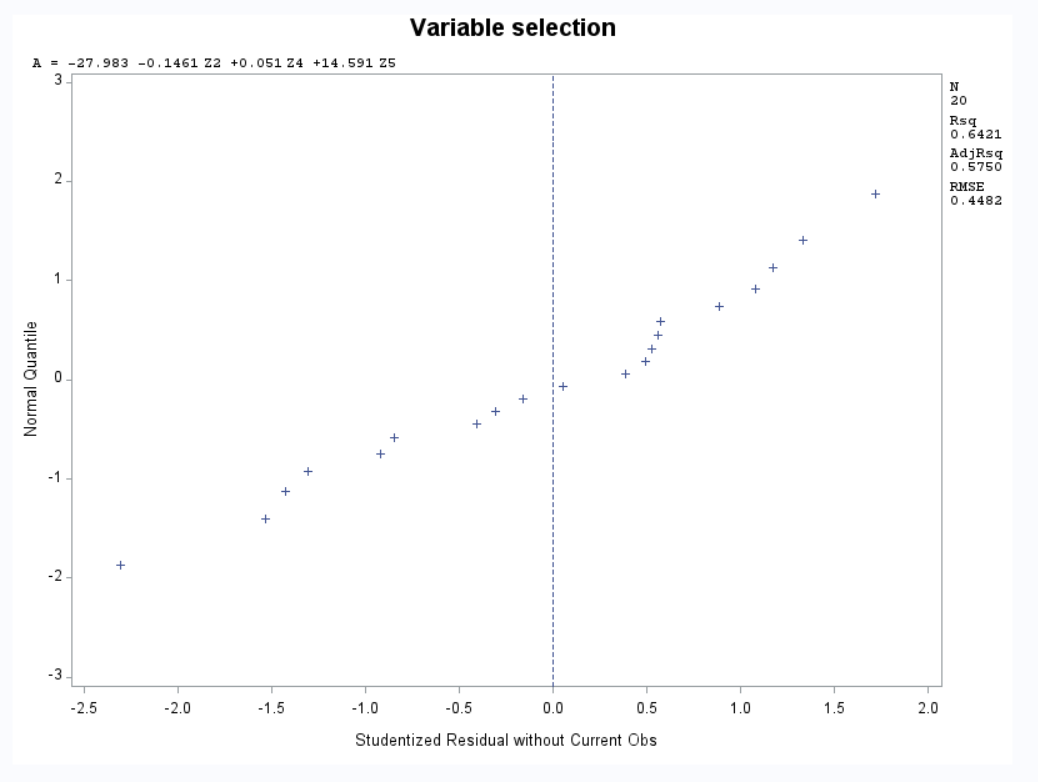


Figure 5



From Figure 5 we can discern a linear diagonal line from the QQ-plot, which indicates that the data is approximately normally distributed.