Problem #1

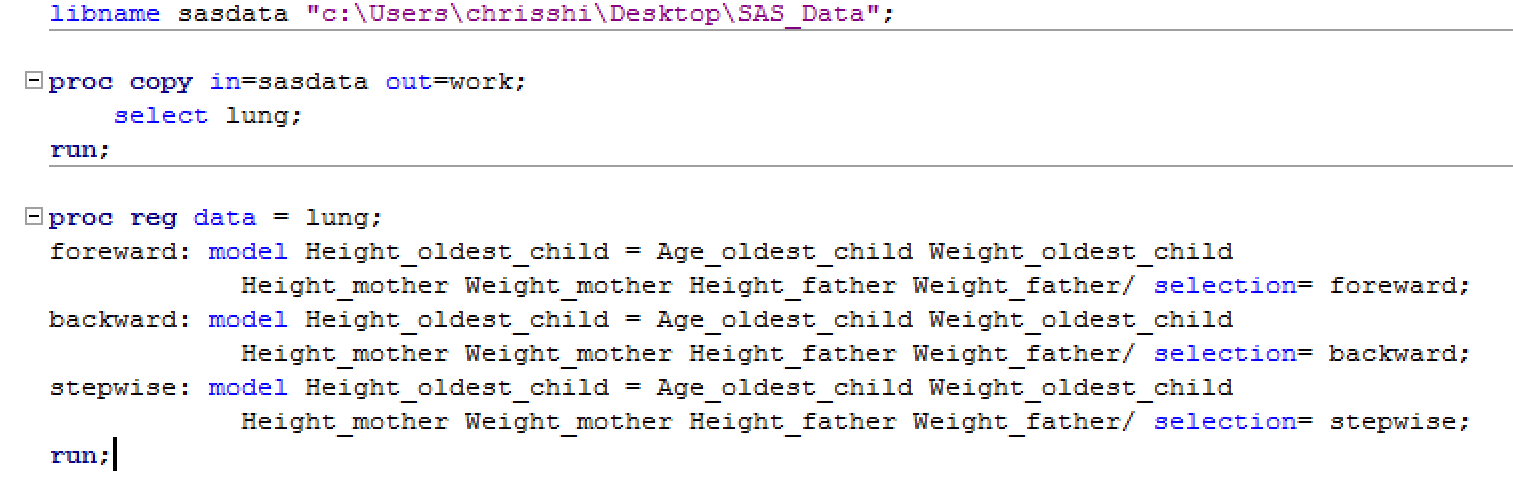
1. 200 students. (199+1 = 200)
2. It’s significant because p-value of this model is less than 0.0001.
3. F = 2385.93019/51.09630 = 46.6947742
4. R-square = 1 – (9963.77926/19508) = 0.4892
5. Science = 12.32529 + 0.38931 \* math – 2.00976 \* female + 0.04984 \* socst + 0.33530 \* read
6. I don’t think it’s a good model because the p-value of female and socst are higher than 0.05.
7. I will try to remove socst variable and analyze the result of new regression model.

Problem #2

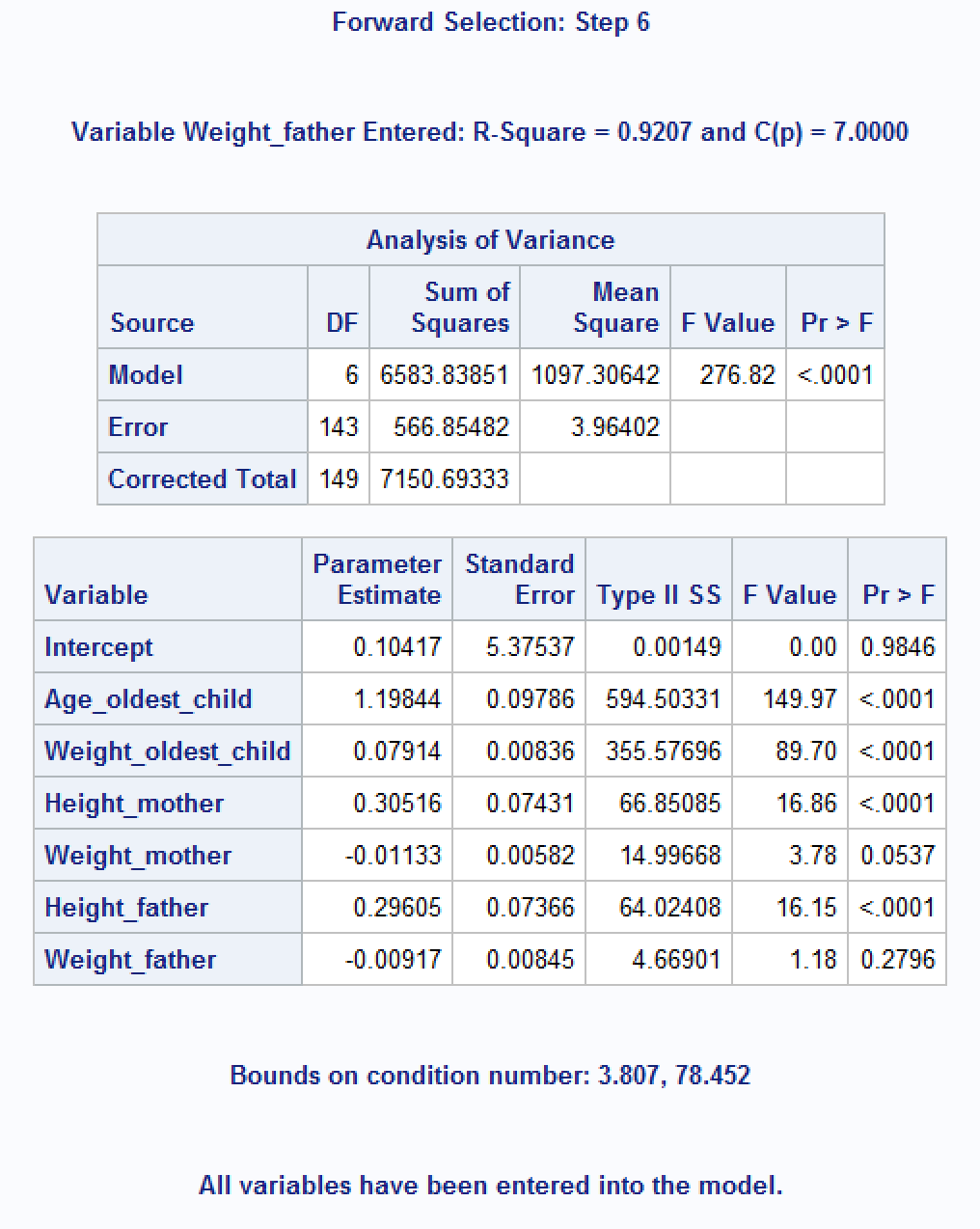
1. II

Problem #3

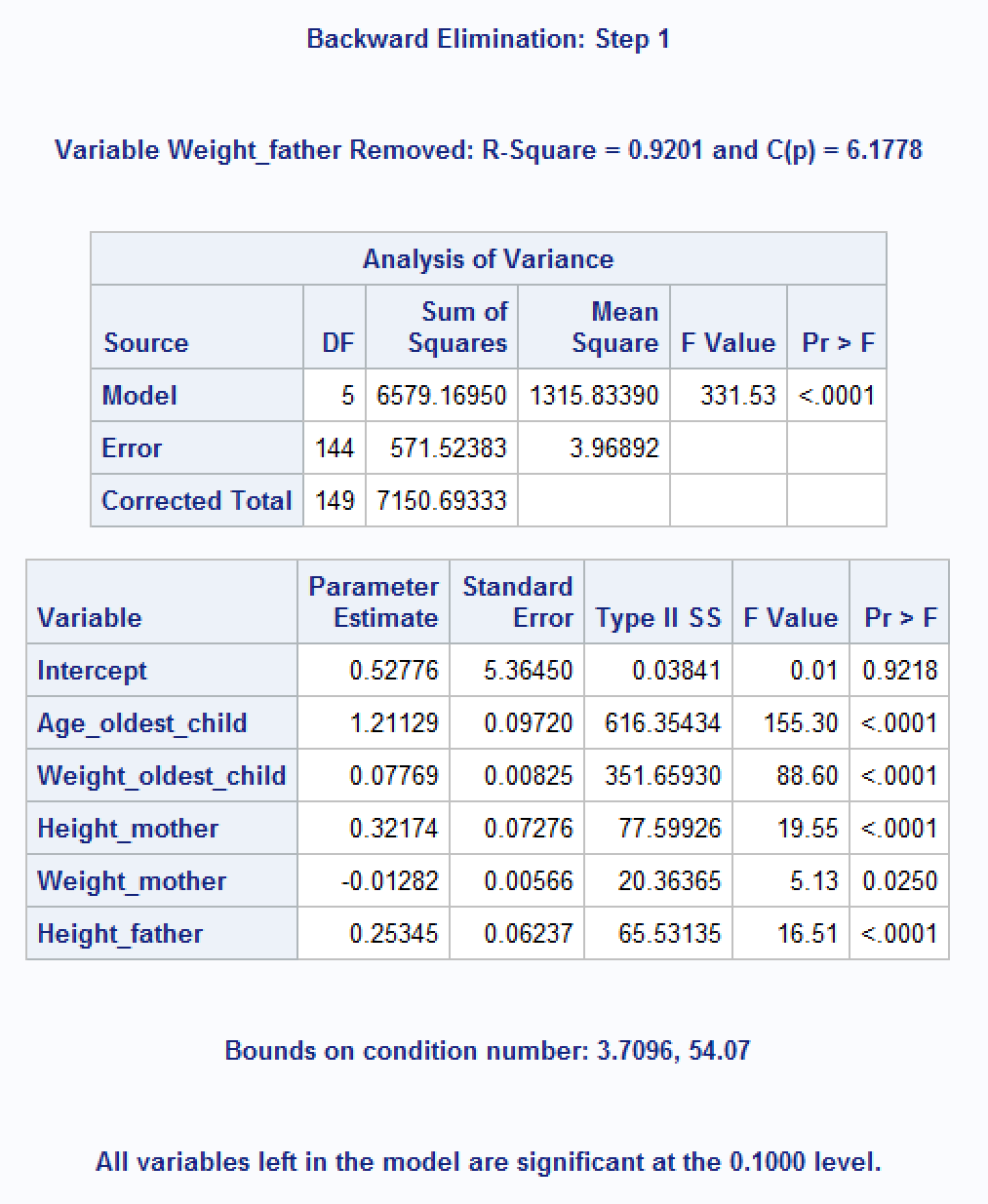
1. Code:



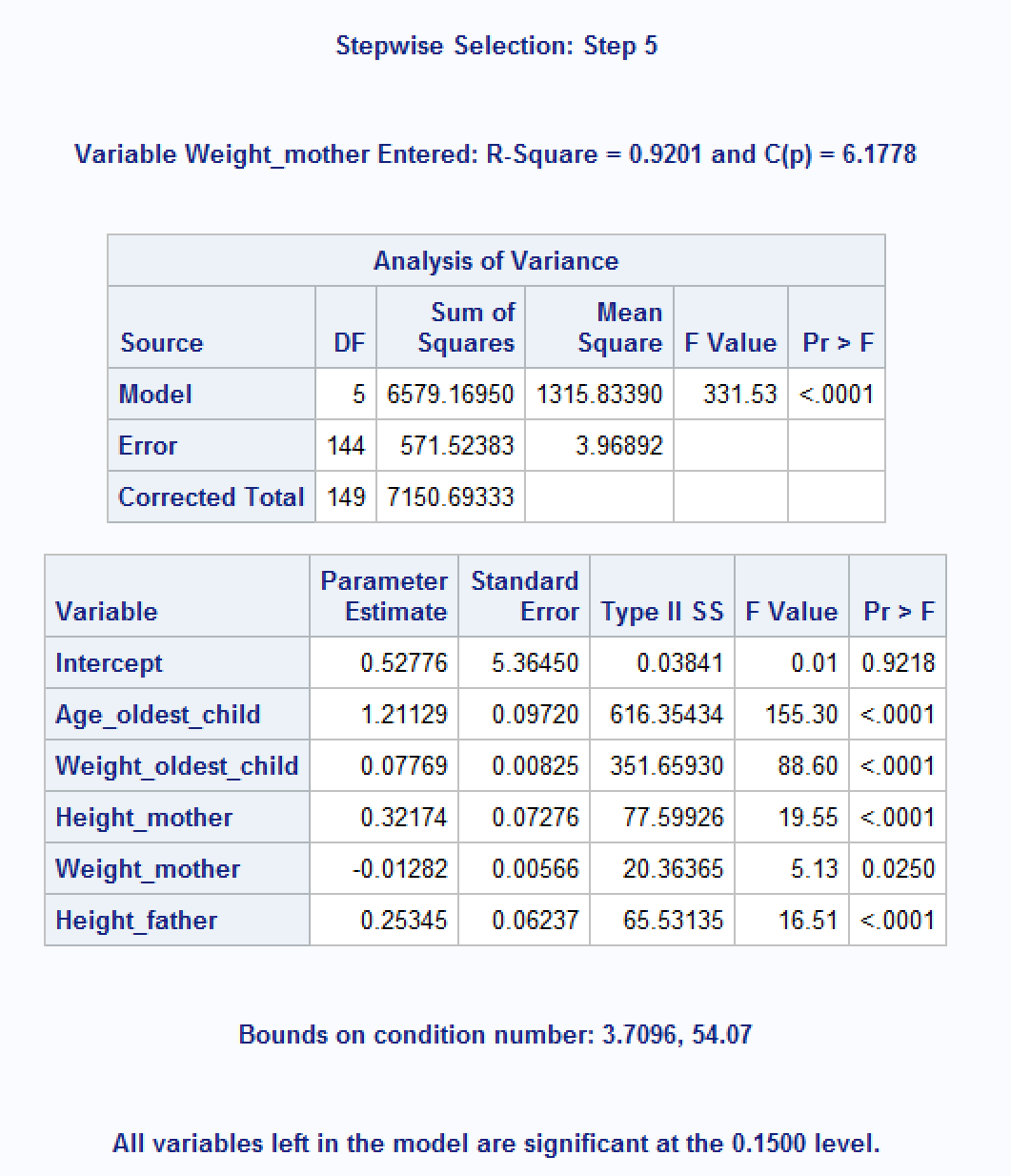
Result of Foreward:



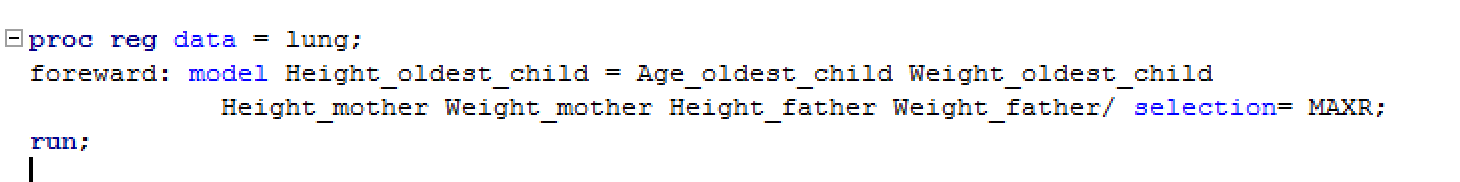
Result of Backword:



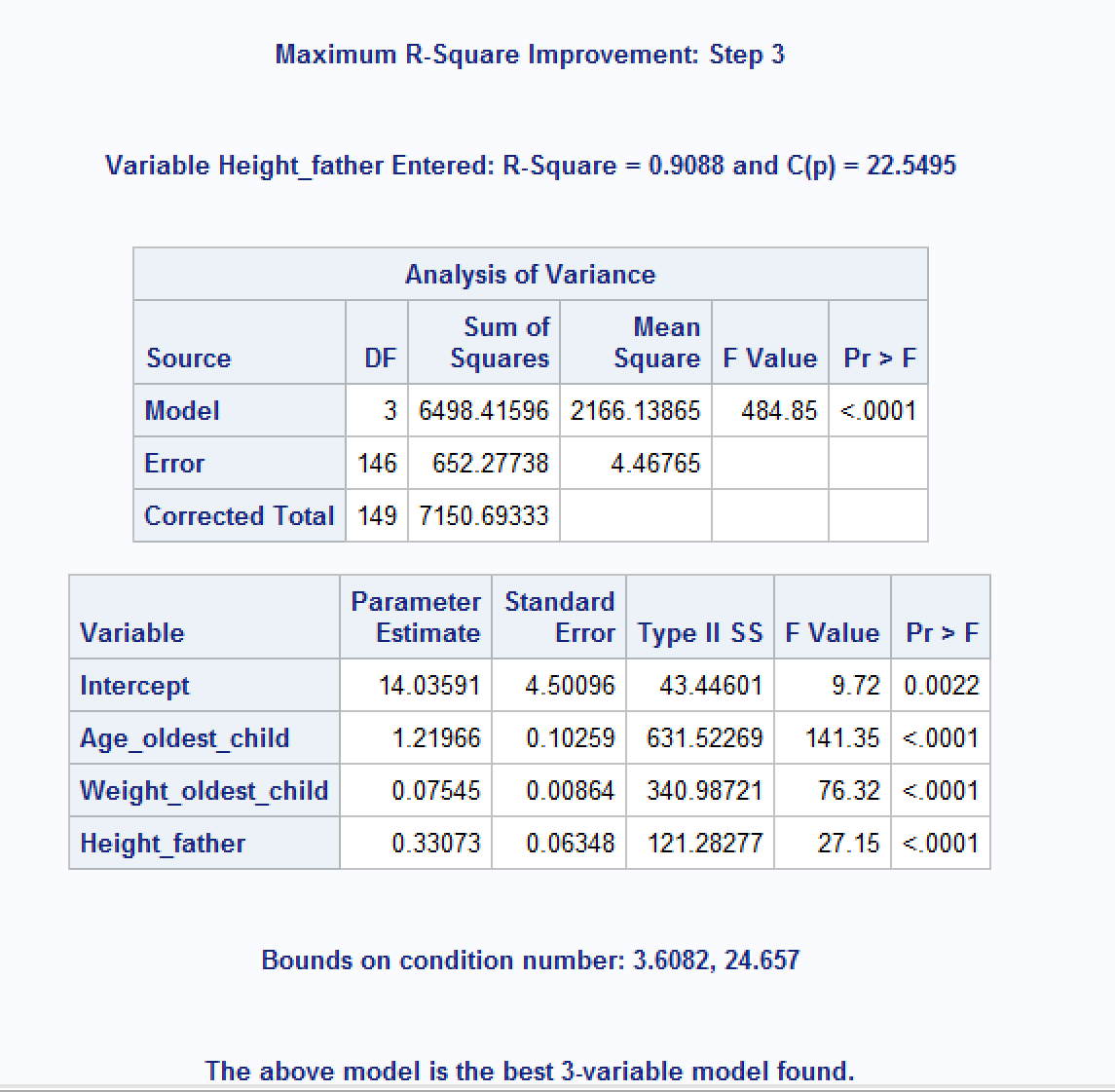
Reuslt of Stepwise:



1. Code:



Result (with 3 independent variables):

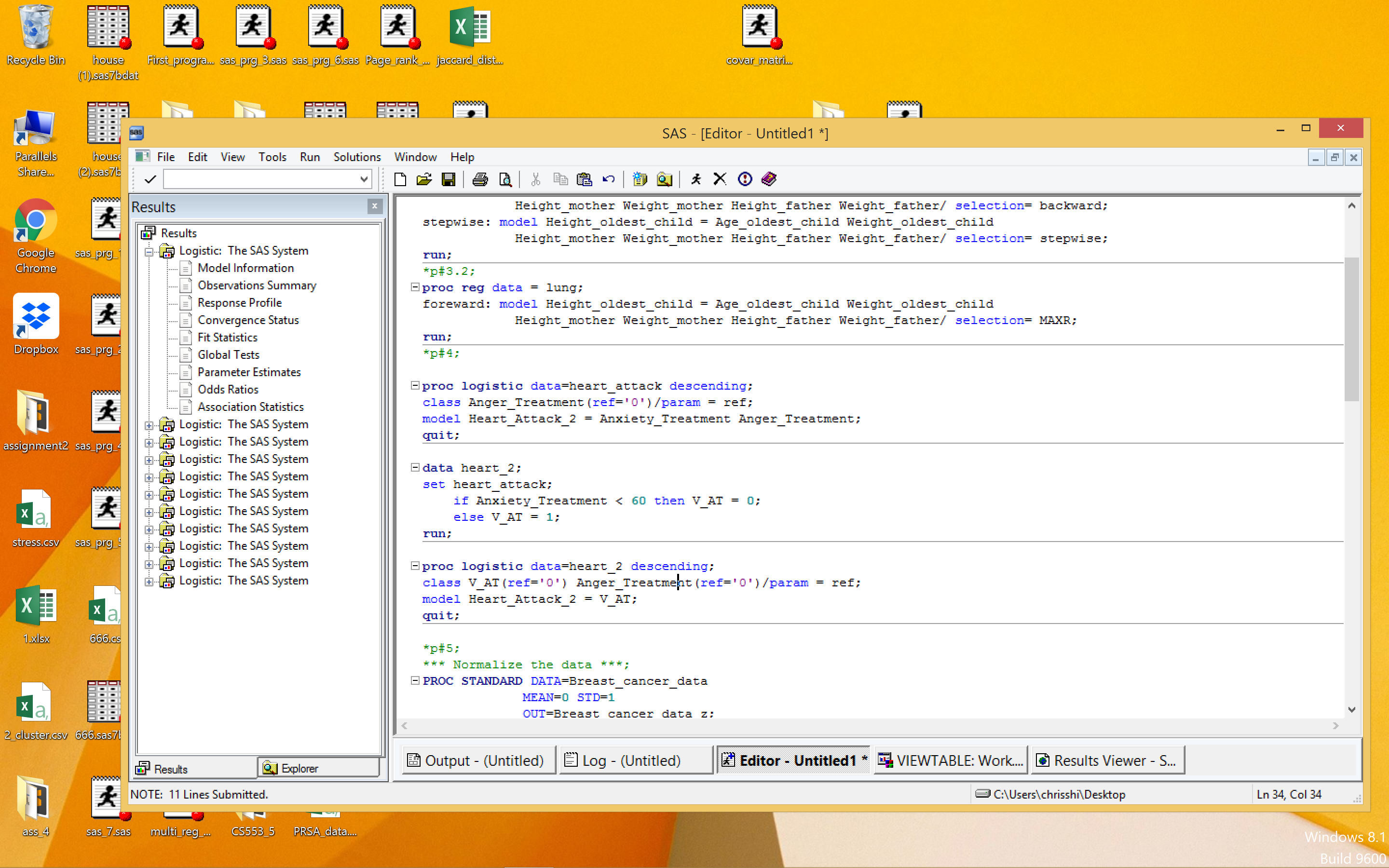


Conclusion:

Best subset includes “AGE of Oldest Child”, “WEIGHT of Oldest Child” and “HEIGHT of Father”.

Problem #4

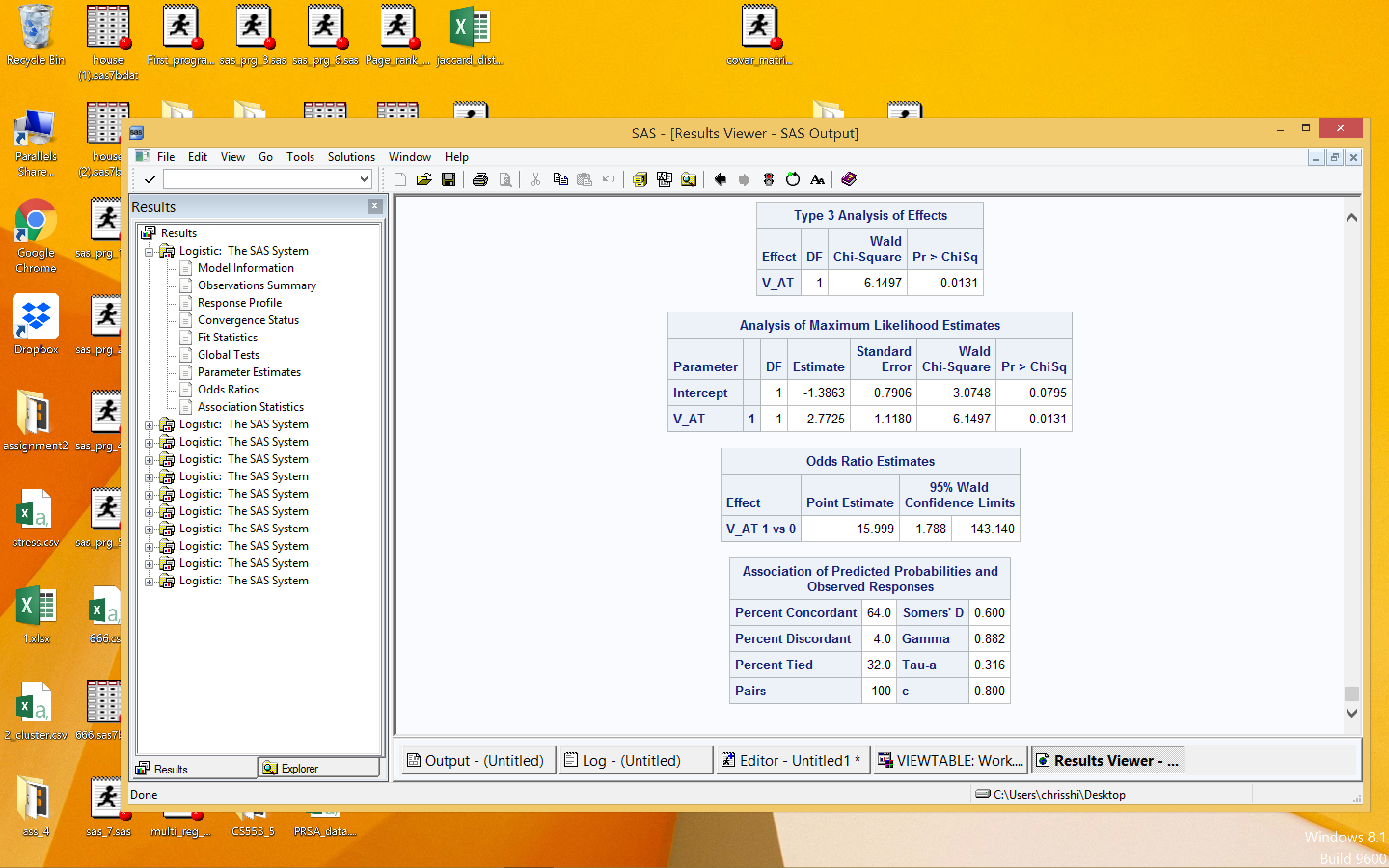
1. Code:



Firstly, try to build a logistic regression on two variables and then get the result below.



Then I removed Anger\_Treatment and let V\_AT = 1 when Anxiety\_Treatment >=60 and got a better model.



Log(odd of having a second heart attack) = -1.3863+2.7725\*V\_AT

1. 1. Log(odd(A)) = -1.3863 odd(A) = exp(-1.3863)= 0.25

P(A) = 0.25/(1+0.25) = 0.2

Log(odd(B)) = -1.3863+2.7725 = 1.3862

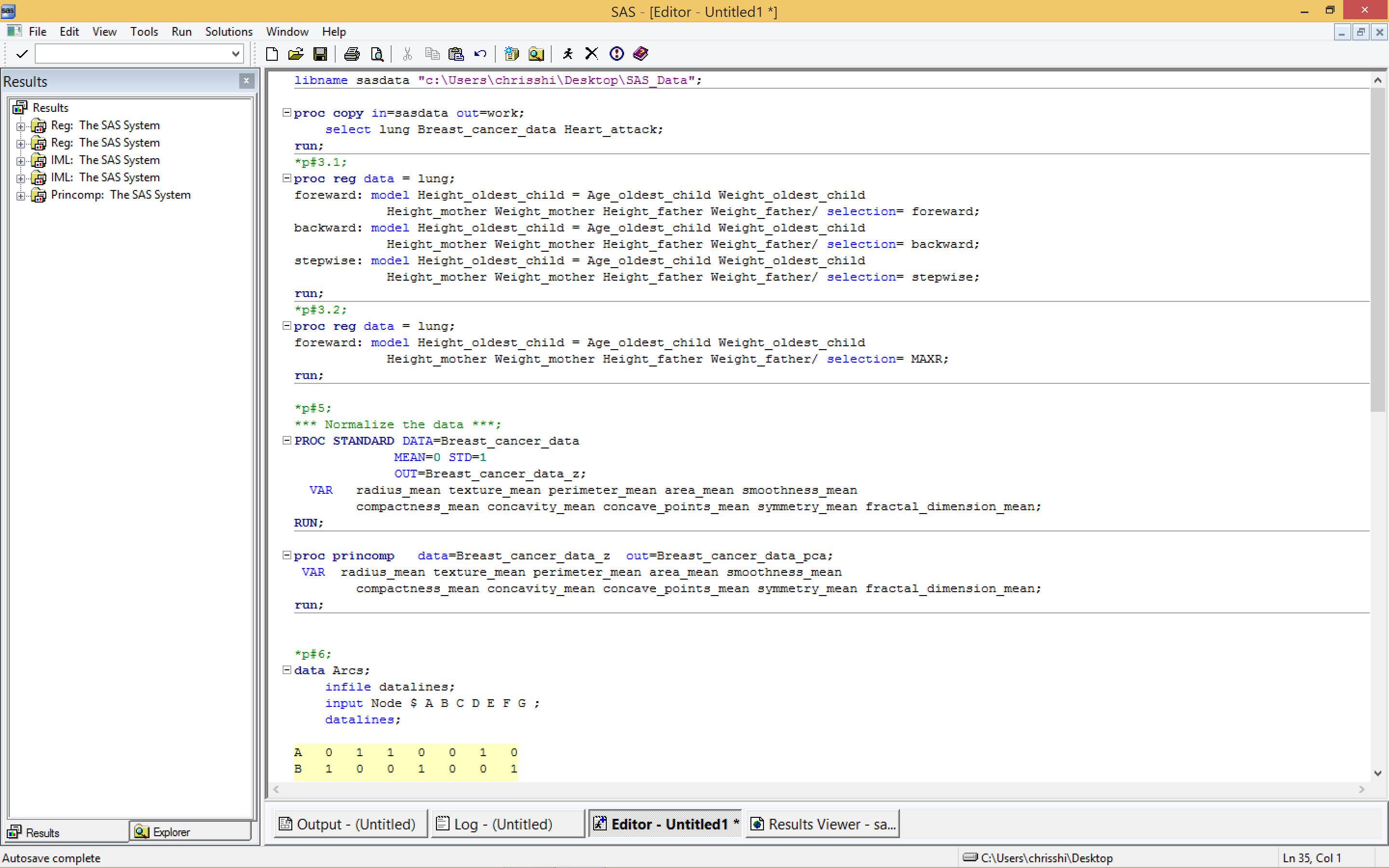
Odd(B) = exp(1.3862) = 3.9996

P(B) = 3.9996/(1+3.9996) = 0.8

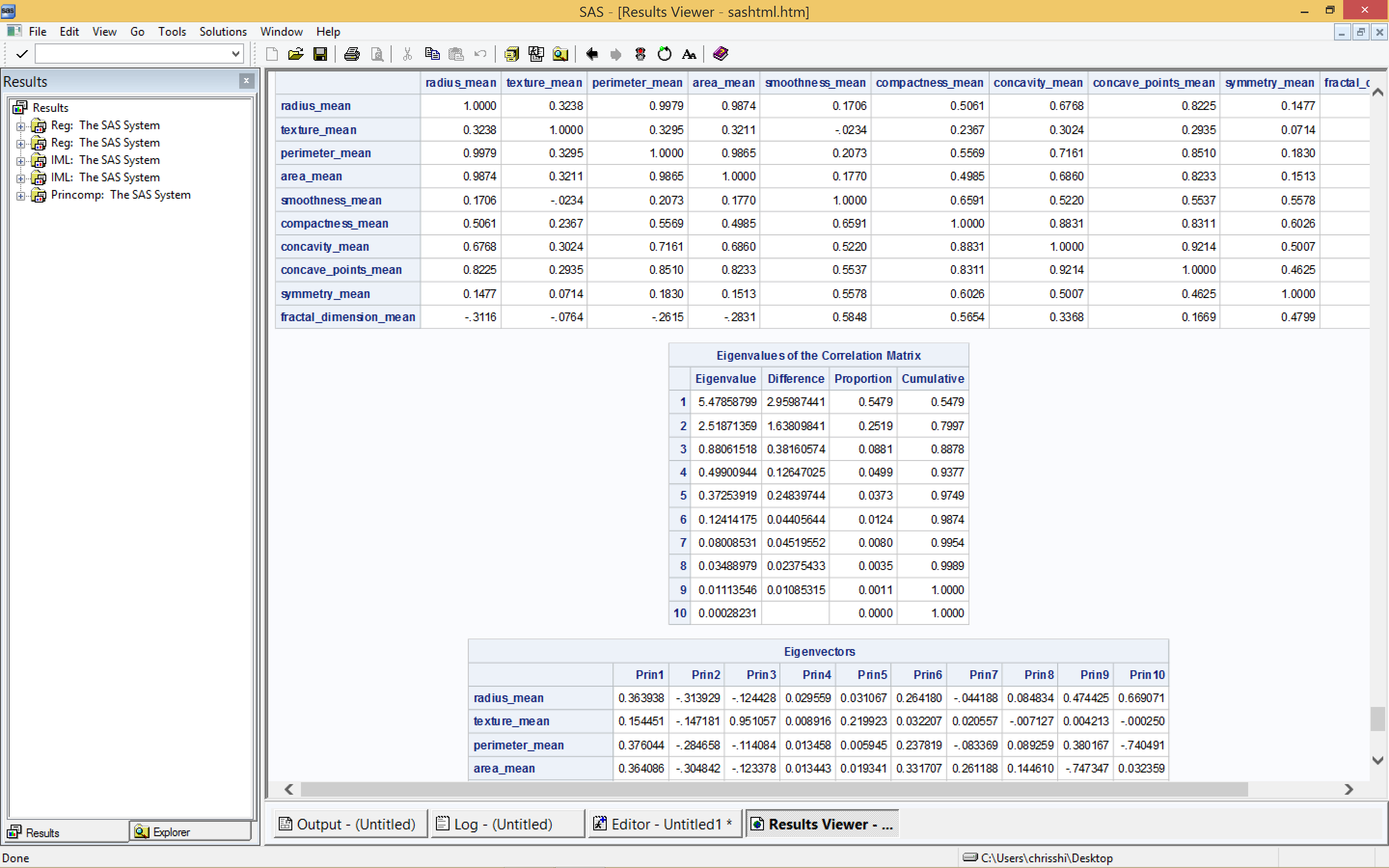
* 1. Odd(A) = 0.25 Odd(B) = 3.996
  2. Odds ratio = 0.25/3.996 = 0.0625

Problem #5

Code:



Result:



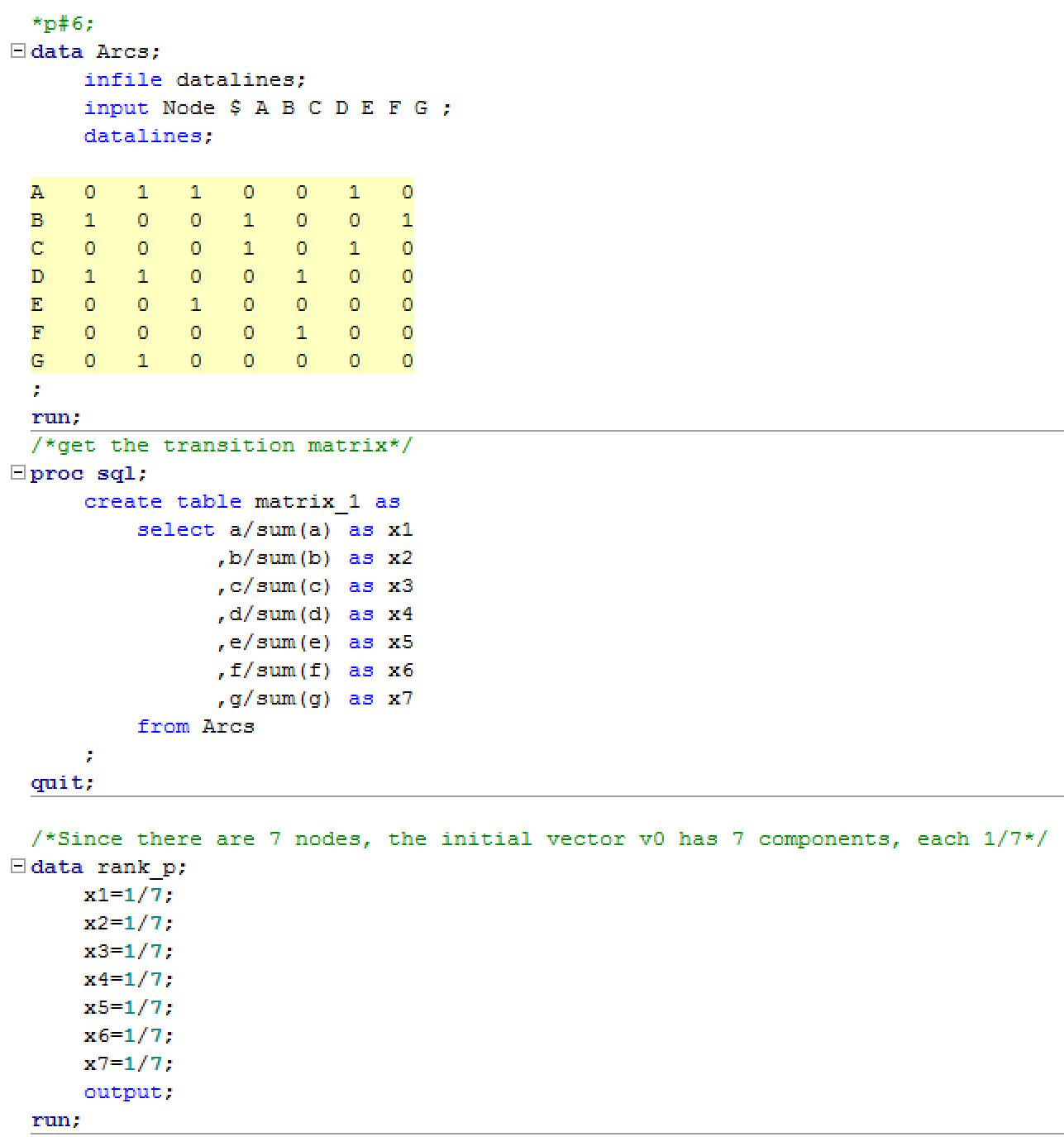
Conclusion:

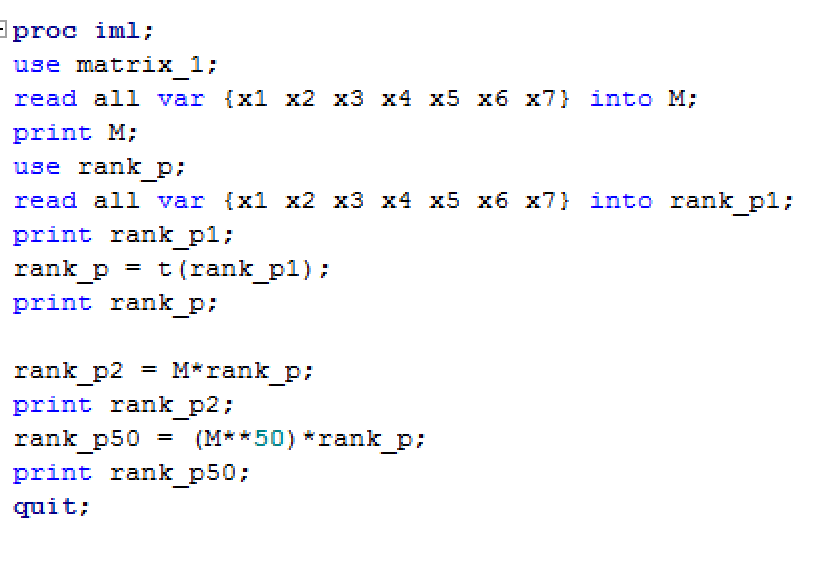
1. To explain at least 85 percent of the variability 3 principal components should be used.
2. To explain at least 95 percent of the variability 5 principal components should be used.

Problem #6

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | E | F | G |
| A | 0 | 1 | 1 | 0 | 0 | 1 | 0 |
| B | 1 | 0 | 0 | 1 | 0 | 0 | 1 |
| C | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| D | 1 | 1 | 0 | 0 | 1 | 0 | 0 |
| E | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| F | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| G | 0 | 1 | 0 | 0 | 0 | 0 | 0 |

Code:





Result:

|  |
| --- |
| **rank\_p2** |
| 0.1904762 |
| 0.2857143 |
| 0.1428571 |
| 0.1904762 |
| 0.0714286 |
| 0.0714286 |
| 0.047619 |