**Name:** LAAVANYA GANESH

**UIN:** 6543324917

**IDS 462:** Statistical Software for Business

**ASSIGNMENT 4**

**PROGRAM:**

/\* DATA INPUT \*/

DATA LIBRARY;

INPUT BOOKS ENROLL DEGREE AREA;

DATALINES;

4 5 3 20

5 8 3 40

10 40 3 100

1 4 2 50

.5 2 1 300

2 8 1 400

7 30 3 40

4 20 2 200

1 10 2 5

1 12 1 100

;

RUN;

/\* DUMMY VARIABLES \*/

DATA DUMMY;

SET LIBRARY;

LOGAREA = LOG(AREA);

IF DEGREE = 2

THEN MA = 1;

ELSE MA = 0;

IF DEGREE > 2

THEN PhD = 1;

ELSE PhD = 0;

RUN;

/\* QUESTION 9.3 \*/

PROC REG DATA = DUMMY;

MODEL BOOKS = ENROLL DEGREE AREA / SELECTION = STEPWISE;

OUTPUT OUT=OUTPUT\_DATA R= R P = P;

RUN;

QUIT;

/\* QUESTION 9.4 \*/

PROC REG DATA = DUMMY;

MODEL BOOKS = ENROLL LOGAREA MA PhD / SELECTION = STEPWISE;

OUTPUT OUT=OUTPUT\_DATA\_WITH\_DUMMYVAR R= R P = P;

RUN;

QUIT;

**LOG**

1 OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;

61

62 /\* DATA INPUT \*/

63 DATA LIBRARY;

64 INPUT BOOKS ENROLL DEGREE AREA;

65 DATALINES;

NOTE: The data set WORK.LIBRARY has 10 observations and 4 variables.

NOTE: DATA statement used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

76 ;

77 RUN;

78

79 /\* DUMMY VARIABLES \*/

80

81 DATA DUMMY;

82 SET LIBRARY;

83 LOGAREA = LOG(AREA);

84 IF DEGREE = 2

85 THEN MA = 1;

86 ELSE MA = 0;

87

88 IF DEGREE > 2

89 THEN PhD = 1;

90 ELSE PhD = 0;

91 RUN;

NOTE: There were 10 observations read from the data set WORK.LIBRARY.

NOTE: The data set WORK.DUMMY has 10 observations and 7 variables.

NOTE: DATA statement used (Total process time):

real time 0.00 seconds

cpu time 0.00 seconds

92

93 /\* QUESTION 9.3 \*/

94 PROC REG DATA = DUMMY;

95 MODEL BOOKS = ENROLL DEGREE AREA / SELECTION = STEPWISE;

96 OUTPUT OUT=OUTPUT\_DATA R= R P = P;

97 RUN;

98 QUIT;

NOTE: The data set WORK.OUTPUT\_DATA has 10 observations and 9 variables.

NOTE: PROCEDURE REG used (Total process time):

real time 1.09 seconds

cpu time 0.42 seconds

99

100

101 /\* QUESTION 9.4 \*/

102

103 PROC REG DATA = DUMMY;

104 MODEL BOOKS = ENROLL LOGAREA MA PhD / SELECTION = STEPWISE;

105 OUTPUT OUT=OUTPUT\_DATA\_WITH\_DUMMYVAR R= R P = P;

106 RUN;

107 QUIT;

NOTE: The data set WORK.OUTPUT\_DATA\_WITH\_DUMMYVAR has 10 observations and 9 variables.

NOTE: PROCEDURE REG used (Total process time):

real time 0.87 seconds

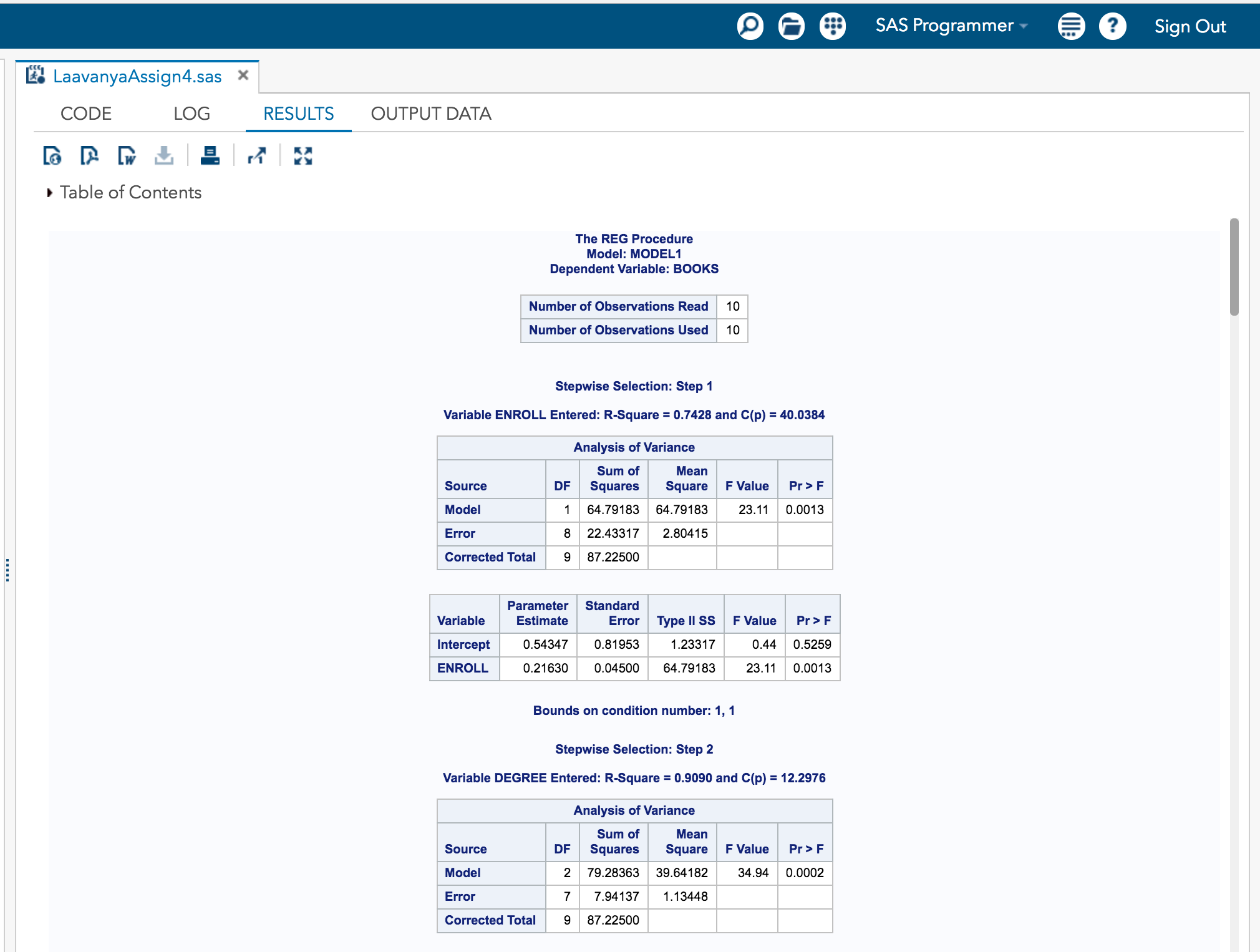
cpu time 0.34 seconds

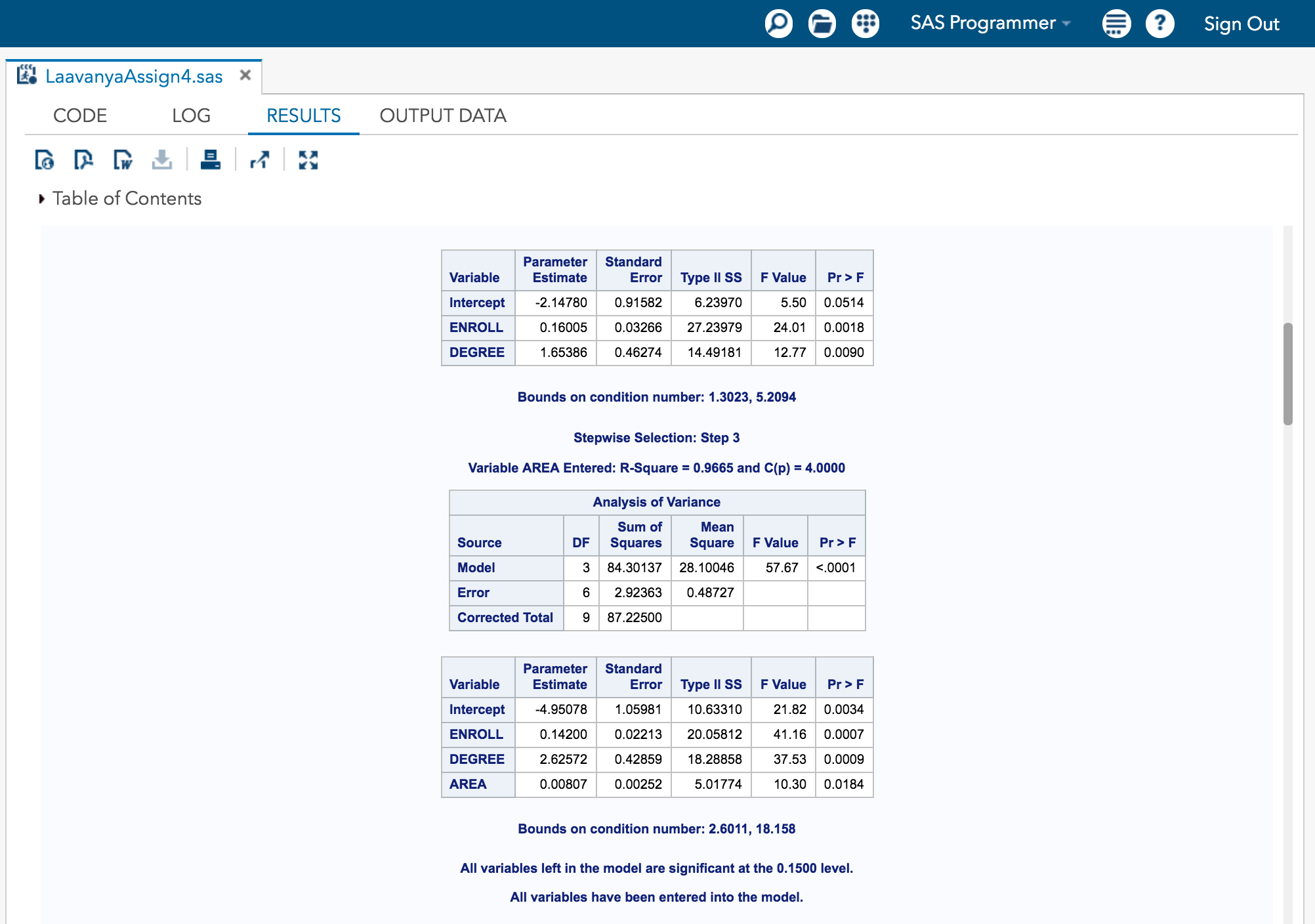
108

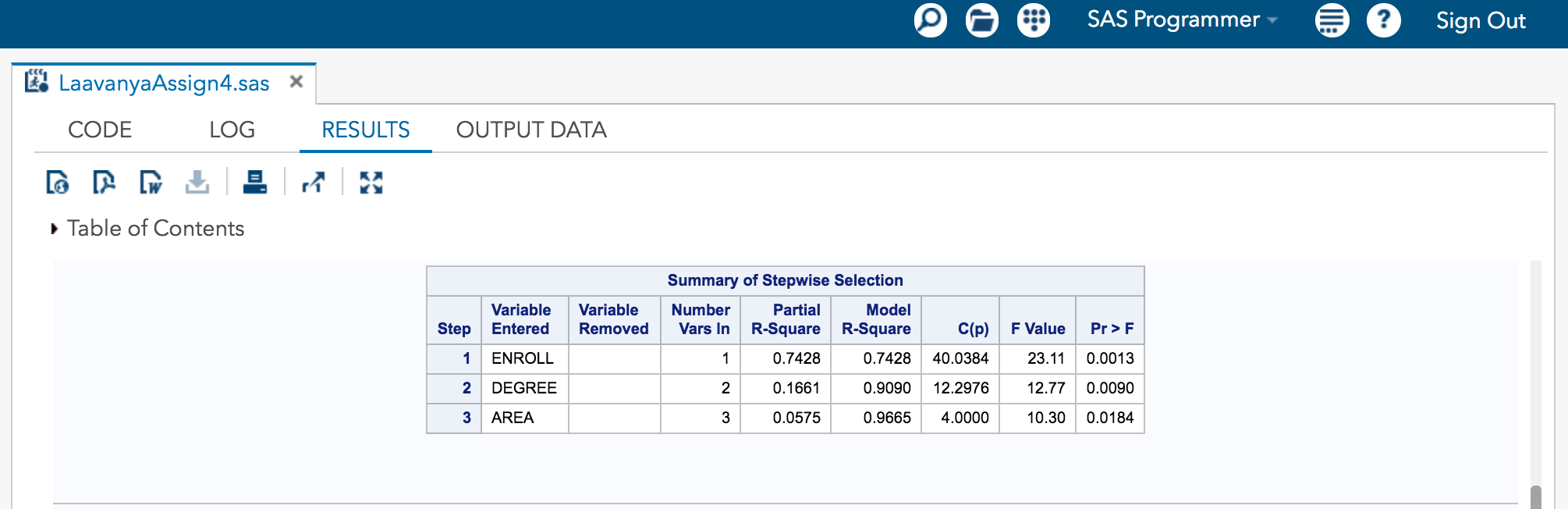
109 OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;

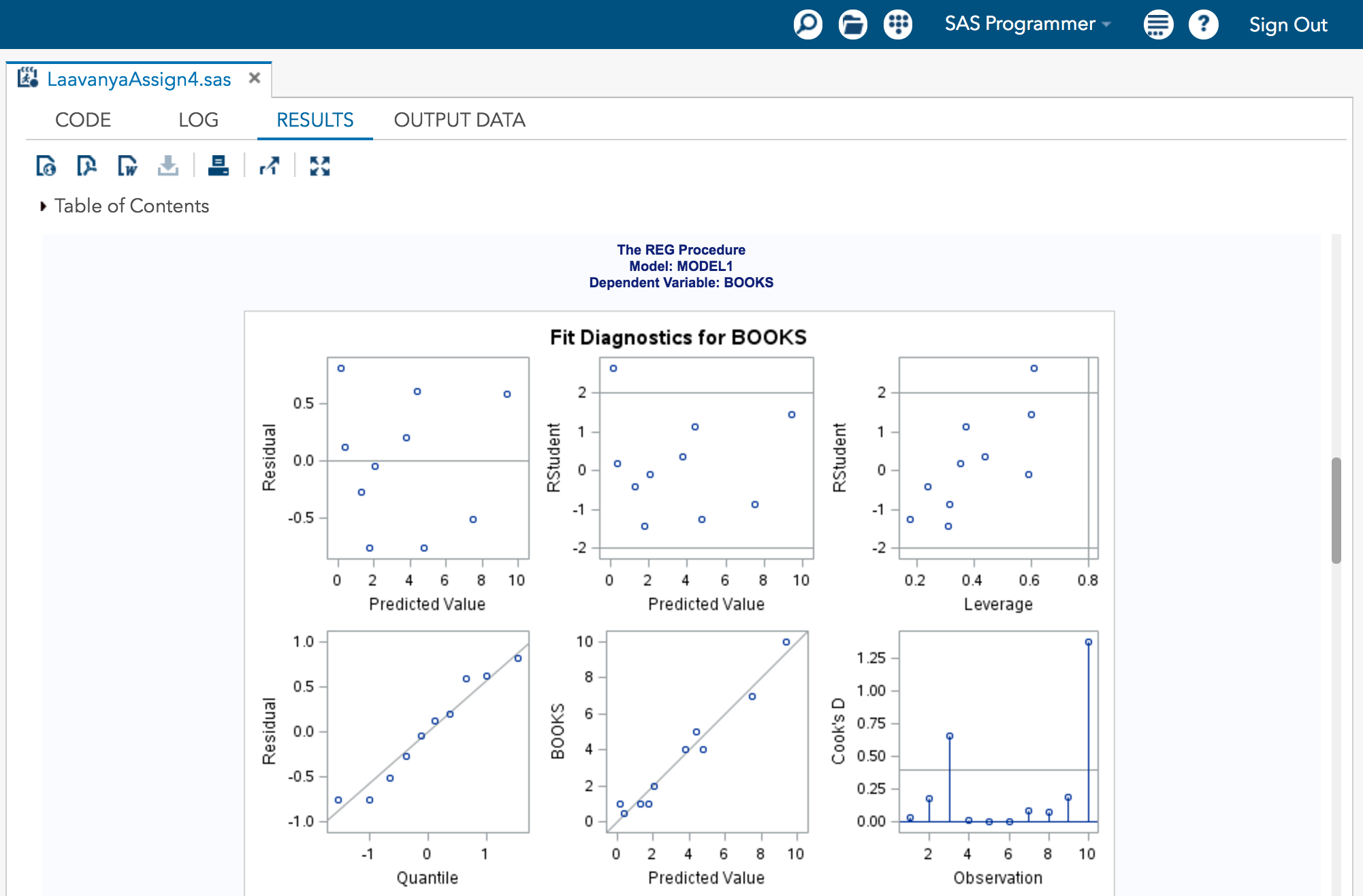
122

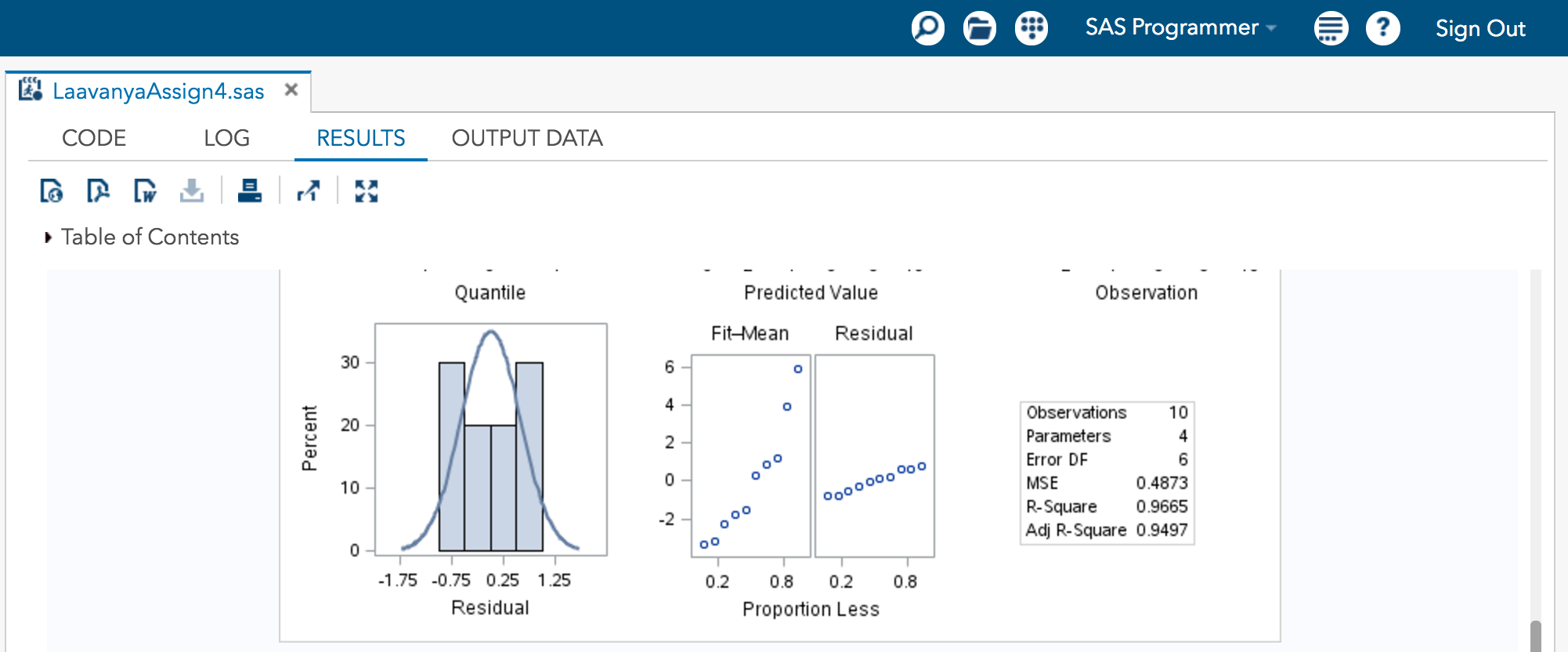
**SCREENSHOTS of RESULTS:**

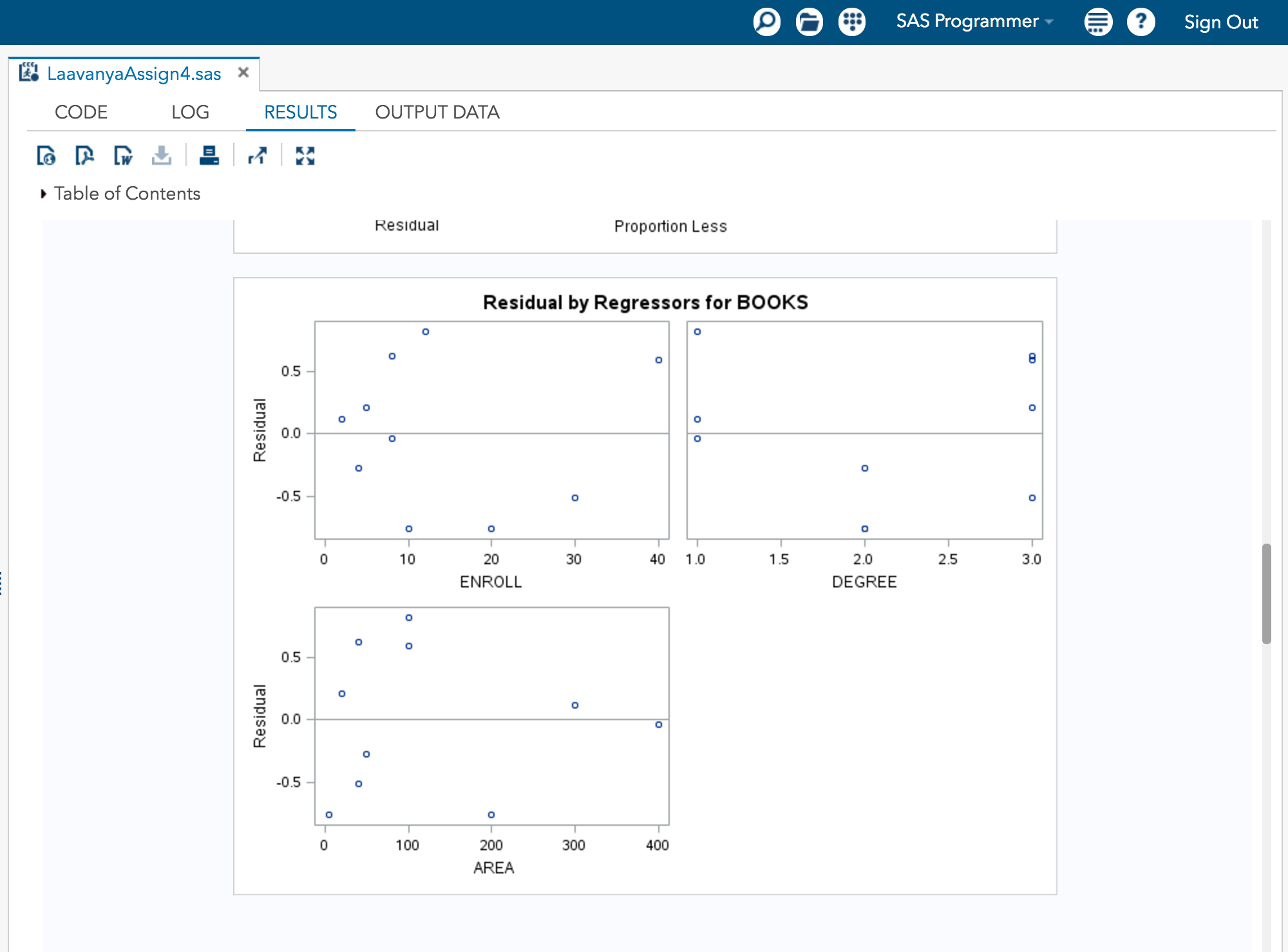
****

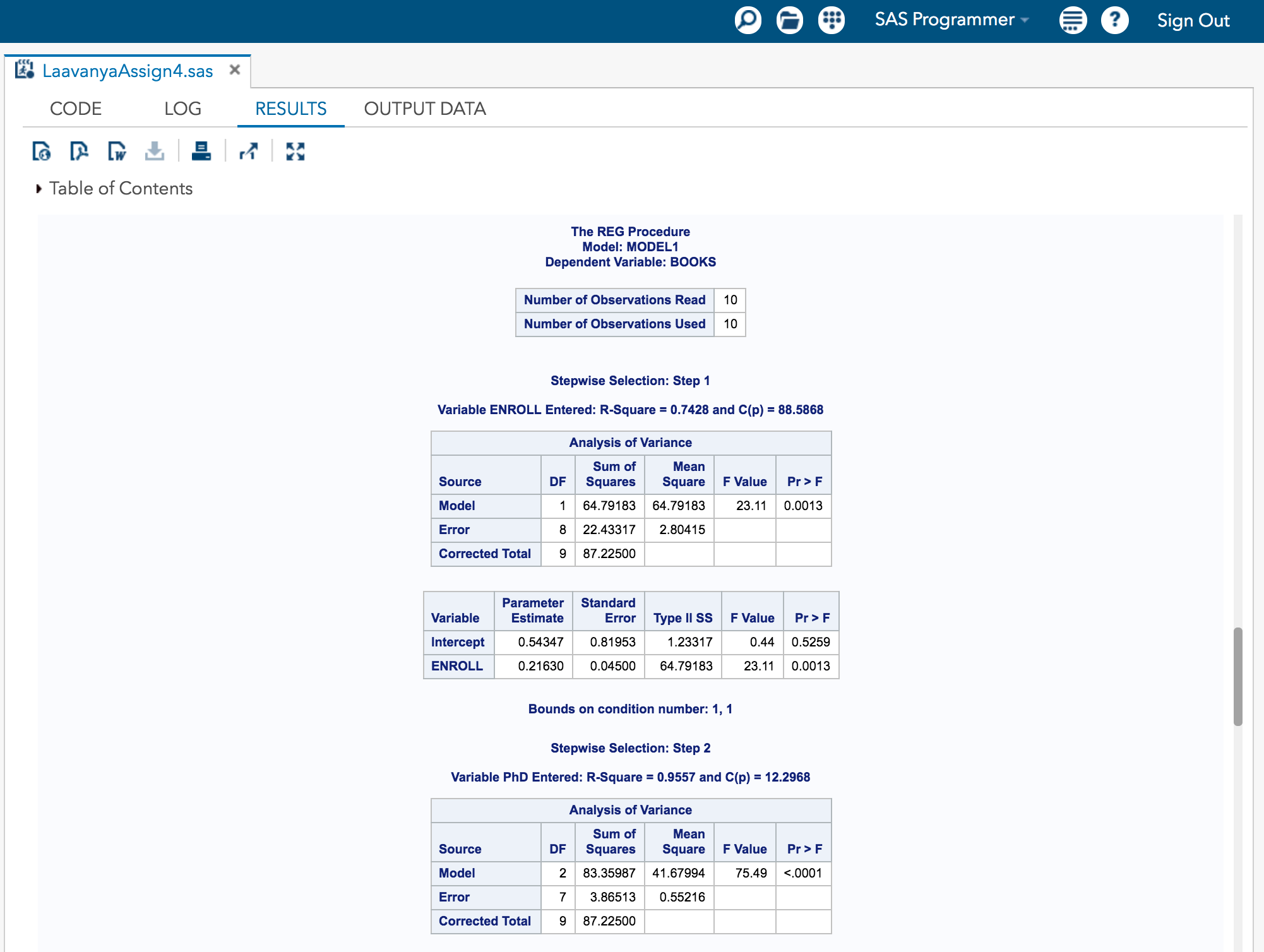
****

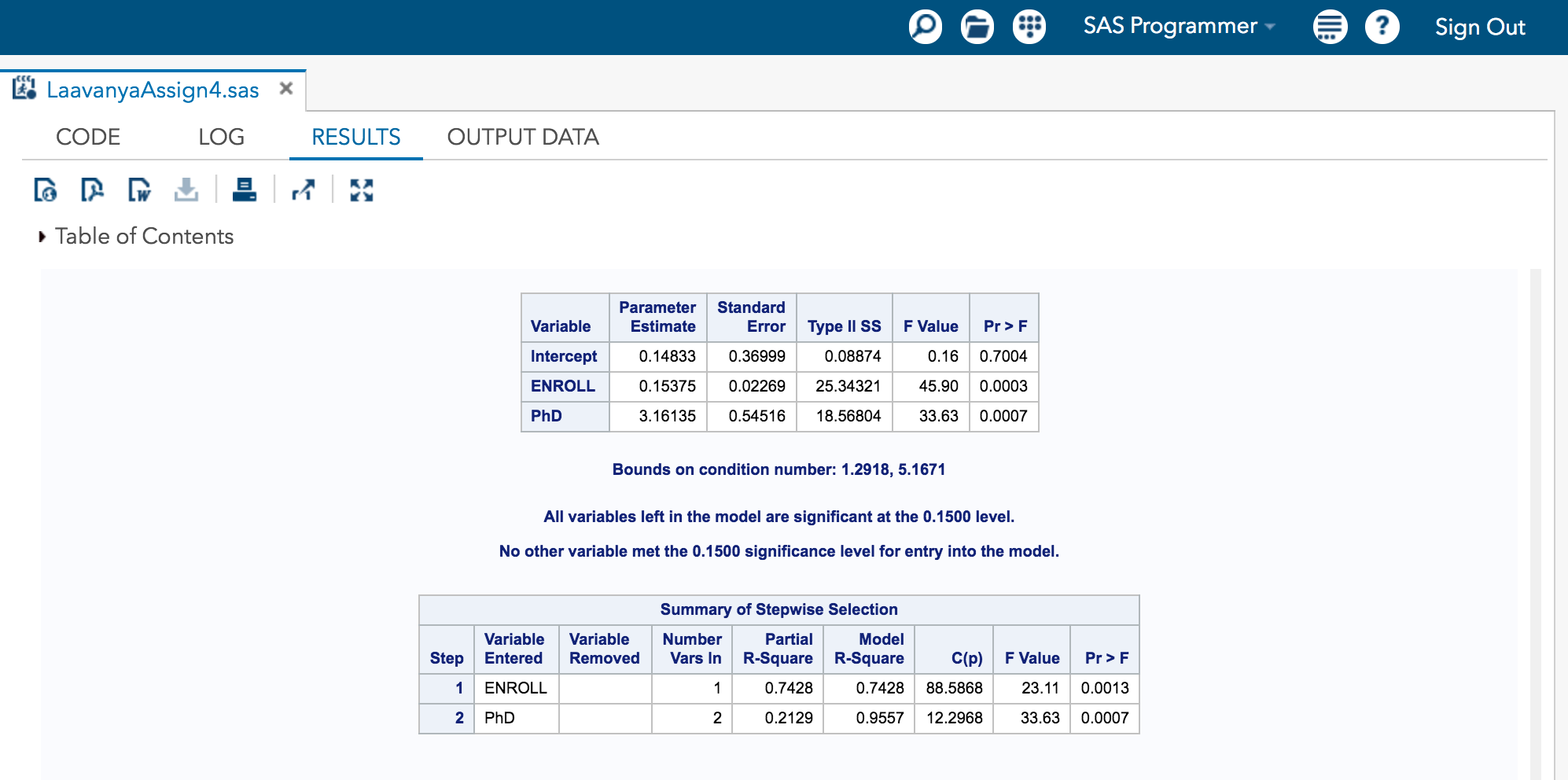


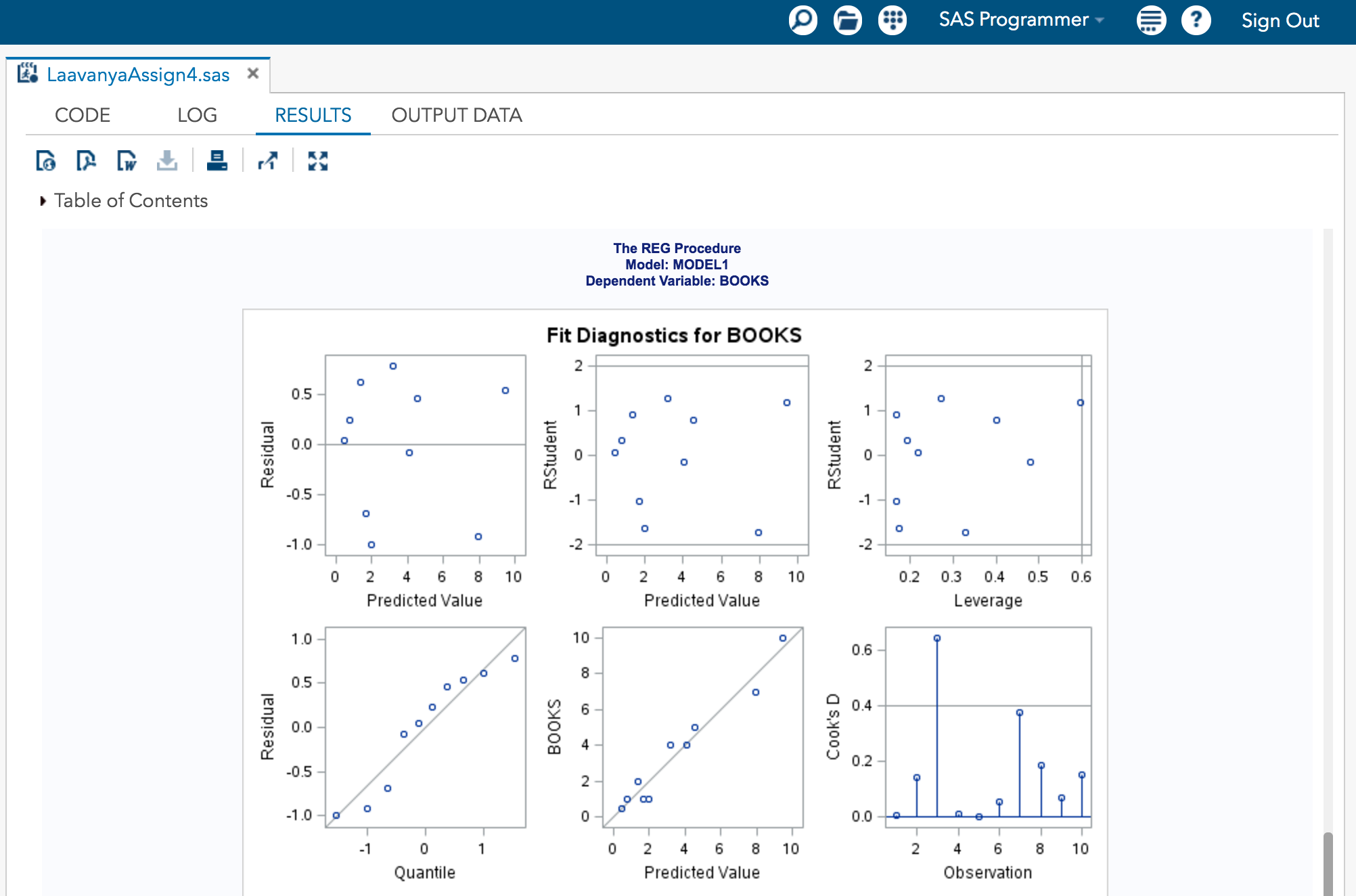


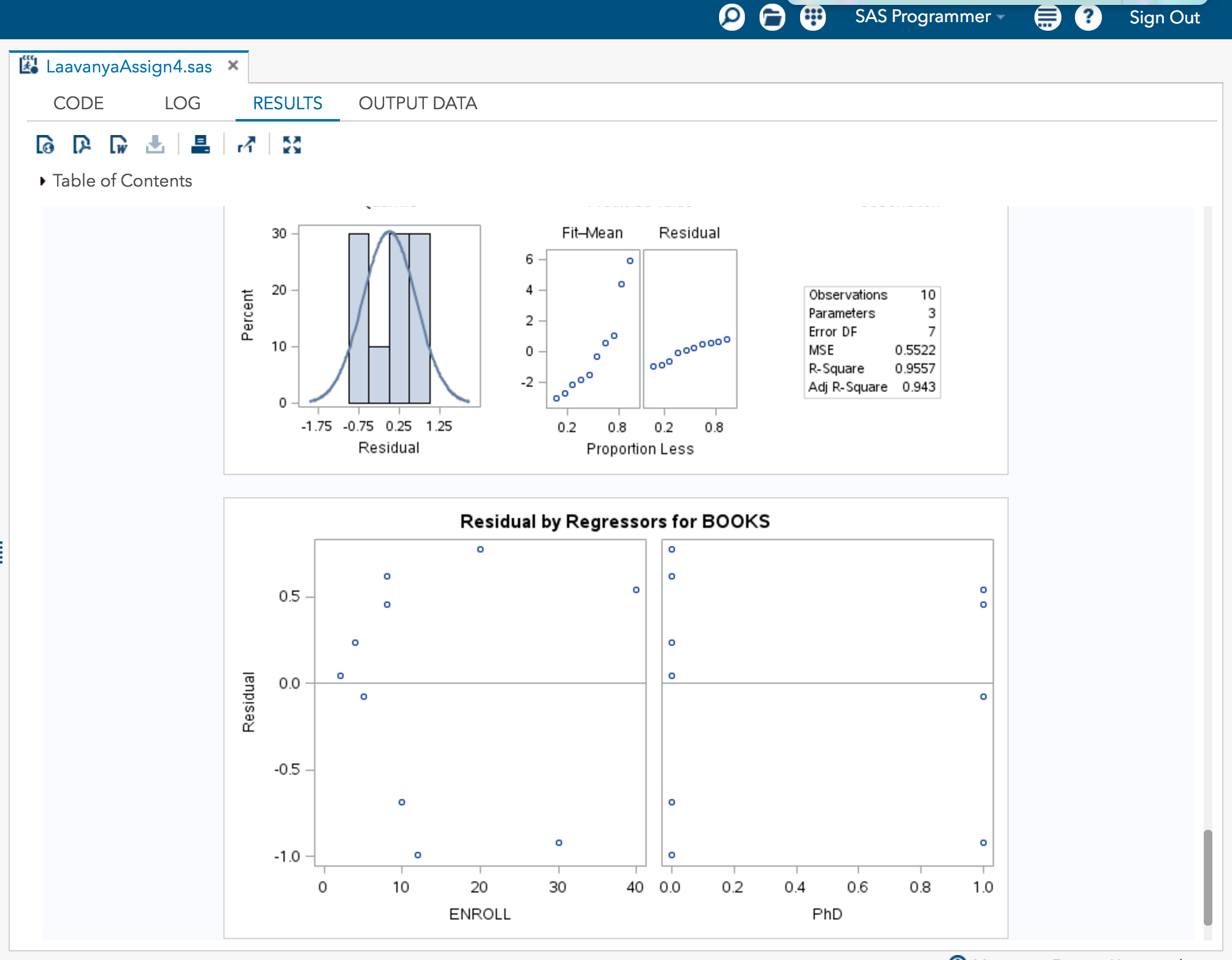












**REVIEW QUESTIONS**

1. What is the dependent variable(s) in this problem? That is, what are we trying to predict (or estimate)?

**ANSWER**

The dependent variable in both the questions 9.3 and 9.4 is number of volume of books (BOOKS) across the country.

1. What are the independent variable(s) in this problem? That is, what variables are we using to build the prediction?

**ANSWER**

The independent variables in 9.3 are Student Enrollment (ENROLL) , the highest degree offered (DEGREE) and the size of campus (AREA).

The independent variables in 9.4 are Student Enrollment (ENROLL) , the dummy variable Phd , the dummy variable MA and natural logarithm of the size of campus (LOGAREA).

1. Within the full regression model for problem 9.3, what is the slope of the regression line between the dependent variable and area?

**ANSWER**

The slope of the regression line between the dependent variable and area is 0.0.00807.

1. Interpret the slope from 3 above. That is, as area increases by 1, what happens?

**ANSWER**

As the area increases by 1, the volume of books across the country increases by 8070.

1. What is the unit of measurement of the area variable?

**ANSWER**

The unit of measurement of the area variable is acres.

1. In which problem (9.3 or 9.4) is highest degree treated as quantitative? In which problem is it treated as categorical?

**ANSWER**

In 9.3 the highest degree treated as quantitative. In 9.4 the highest degree treated as categorical.

1. How many more (or less) books is the library estimated to have if the university awards Ph.D.s as opposed to only Bachelor’s degrees. The answer to this question is a single number.

**ANSWER**

The library is estimated to have 3.16135 million more books if the university awards Ph.D.s as opposed to only Bachelor’s degrees.

1. Provide the full interpretation of the Ph.D. variable: If a university award’s Ph.D.s…The answer to this question is a full sentence.

**ANSWER**

If a university award’s Ph.D.s, the number of volume of books across the country increases by 3.16135 million.