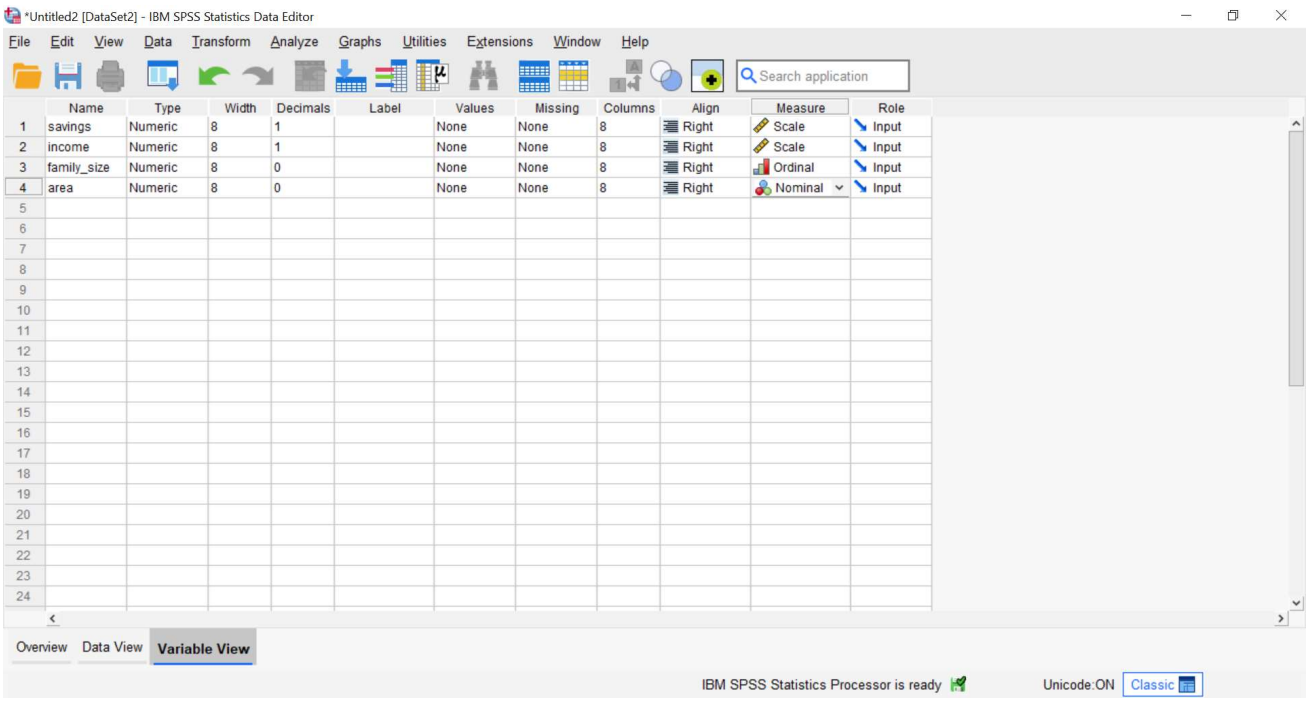
Lab #6.

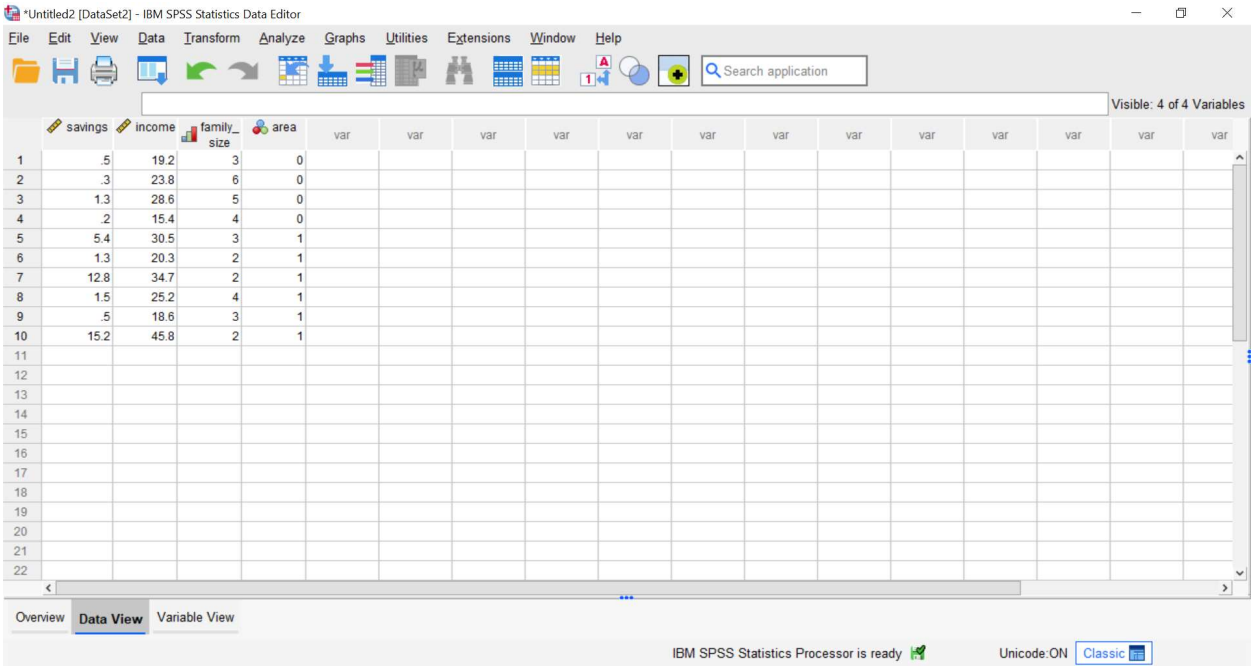
Family Savings_Model Selection MLR Ex.

A particular savings and loan corporation is interested in determining how well the amount of money in family savings accounts (y) can be predicted using annual income (x_1), number of members in a family unit (x_2) and the area in which the family unit lives (x_3).

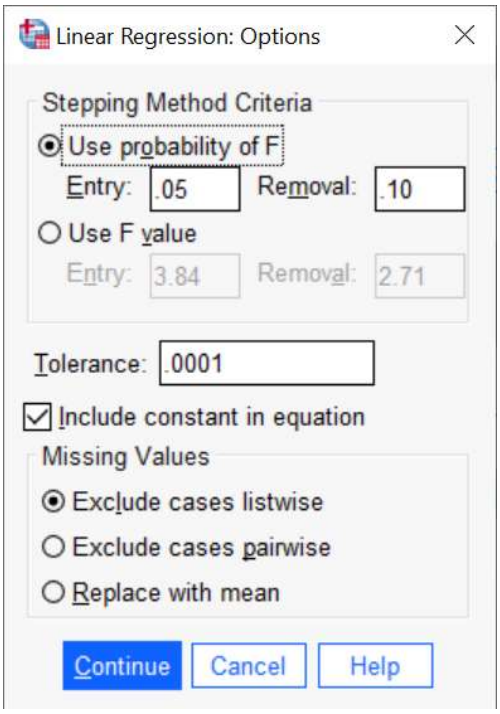
y	X1	X2	X3
0.5	19.2	3	0
0.3	23.8	6	0
1.3	28.6	5	0
0.2	15.4	4	0
5.4	30.5	3	1
1.3	20.3	2	1
12.8	34.7	2	1
1.5	25.2	4	1
0.5	18.6	3	1
15.2	45.8	2	1

Enter the data





To perform model selection procedures, go to “Analyse” → “Regression” → “Linear” , then select ‘savings’ as Dependent variable and the other 3 variables as Independent ones. Then click on ‘Options’ and select 0.05 for the Entry and 0.10 for the Removal. Click on ‘Continue’.



After that, click on ‘Method” and select ‘**Forward**’, then click on ‘OK’.

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	income	.	Forward (Criterion: Probability-of-F-to-enter <= .050)
2	family_size	.	Forward (Criterion: Probability-of-F-to-enter <= .050)

a. Dependent Variable: savings

Step 1 variable x_1 is entered

Step 2 variable x_2 is entered

Excluded Variables[a]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics Tolerance
1	family_size	-.324[b]	-2.809	.026	-.728	.891
	area	.161[b]	.978	.361	.347	.822
2	area	-.107[c]	-.645	.543	-.255	.467

a. Dependent Variable: savings

b. Predictors in the Model: (Constant), income

c. Predictors in the Model: (Constant), income, family_size

variable x_3 was not entered because its p-value > 0.05 (probability to "enter" should be ≤ 0.05)

Hence, the best model based on the **Forward Selection Procedure** has variables X1 and X2 only.

- Now repeat the steps: go to “Analyse” → “Regression” → “Linear” , then select ‘savings’ as Dependent variable and the other 3 variables as Independent ones. After that, click on ‘Method” and select ‘**Backward**’, then click on ‘OK’.

ANOVA for Full model, before we do step 1

First, we verify, whether the full model makes sense (i.e. whether it can be used):

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	256.621	3	85.540	23.784	<.001[b]
	Residual	21.579	6	3.597		
	Total	278.200	9			
2	Regression	255.125	2	127.562	38.697	<.001[c]
	Residual	23.075	7	3.296		
	Total	278.200	9			

a. Dependent Variable: savings
b. Predictors: (Constant), area, income, family_size
c. Predictors: (Constant), income, family_size

p-value < α : full model is OK

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	area, income, family_size[b]	.	Enter
2	.	area	Backward (criterion: Probability of F-to-remove >= .100).

a. Dependent Variable: savings
b. All requested variables entered.

step 1 / we start with all 3 variables

step 2 – we remove/delete X_3

Excluded Variables[a]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics Tolerance
2	area	-.107b	-.645	.543	-.255	.467

a. Dependent Variable: savings
b. Predictors in the Model: (Constant), income, family_size

p-value > 0,10 (probability to "remove")

Hence, the best model based on the **Backward Elimination Procedure** has variables X1 and X2 only.

- Now repeat the steps: go to “Analyse” → “Regression” → “Linear” , then select ‘savings’ as Dependent variable and the other 3 variables as Independent ones. After that, click on ‘Method” and select ‘**Stepwise**’, then click on ‘OK’.

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	income	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
2	family_size	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

a. Dependent Variable: savings

Step 1 variable X_1 is entered

Step 2 variable

Excluded Variables[a]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics Tolerance
1	family_size	-.324[b]	-2.809	.026	-.728	.891
	area	.161[b]	.978	.361	.347	.822
2	area	-.107[c]	-.645	.543	-.255	.467

a. Dependent Variable: savings
b. Predictors in the Model: (Constant), income
c. Predictors in the Model: (Constant), income, family_size

variable X_3 was not entered because its p-value > 0.05 (probability to "enter") and also is > 0.10 (probability to "remove")

Hence, the best model based on the **Stepwise Regression** has variables X1 and X2 only.