TUTORIAL Chapter 14.

EXAMPLE 1(28). The following data is given:

Y	X1	X2	X3
33	9	192	40
44	11	397	47
34	10	235	37
60	13	345	61
20	11	245	23
30	7	235	35
45	12	296	52
25	9	235	27
53	10	295	57
45	13	335	50
37	11	243	41
44	13	413	51

a) Construct the correlation matrix. Which variable will be entered in the first step of a stepwise regression model?
b) Use forward stepwise regression to construct a regression equation, entering all significance variables.
c) Construct and estimate of the full regression model. Which equation explains the most variation in the dependent variable?

Solution:

a.

	y	x1	x2
x1	0.582		
	0.047		
x2	0.645	0.709	
	0.024	0.010	
x3	0.985	0.578	0.640
	0.000	0.049	0.025

Cell Contents: Pearson correlation
P-Value

%Which variable must be entered first?

Answer:

x_3 is the most highly correlated with the dependent variable. It should enter the model first.

%Why?

b.

```
Forward selection.  Alpha-to-Enter: 0.25

 Response is    y     on  3 predictors, with N =   12


    Step          1
Constant     -3.097

x3            0.973
T-Value       18.12
P-Value       0.000

S              2.08
R-Sq          97.04
R-Sq(adj)     96.75
C-p             0.1
```

The estimated equation is $\hat{y}$ = -3.097 + 0.973x_3.

c.

```
The regression equation is
y = - 3.84 + 0.053 x1 + 0.0032 x2 + 0.956 x3

Predictor        Coef     SE Coef          T        P
Constant       -3.842       4.083      -0.94    0.374
x1             0.0531      0.5437       0.10    0.925
x2            0.00320     0.01529       0.21    0.840
x3            0.95618     0.07976      11.99    0.000

S = 2.308       R-Sq = 97.1%     R-Sq(adj) = 96.0%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         3     1419.04      473.01     88.77    0.000
Residual Error     8       42.63        5.33
Total             11     1461.67
```

The difference between the two models is:

- **that the full model has two additional independent variables.**
- **Viewing the p-values and the adjusted R^2, however, it appears that x_1 and x_2 are not significant additions to the model.**
- **Using the adjusted R^2: it is apparent that the equation developed using Forward Selection explains the most variation in the dependent variable adjusting for the number of independent variables in the model.**