What are the principles behind performing multiple regressions in SPSS?

Authored by

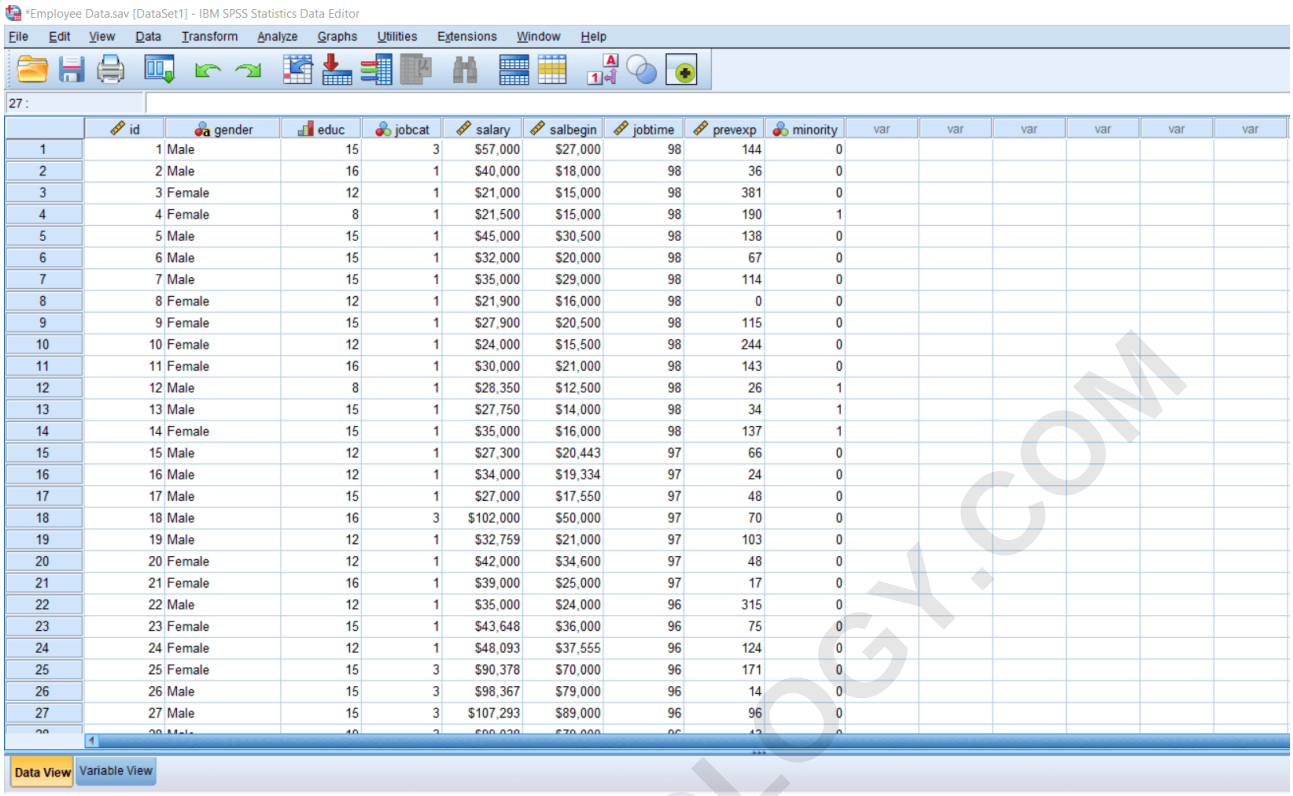
stats writer

June 24, 2024

Multiple regression is a statistical technique used to examine the relationship between multiple independent variables and a single dependent variable. In SPSS, this technique is performed by using the principles of linear regression, where the dependent variable is predicted based on the values of the independent variables. The principles behind performing multiple regressions in SPSS include identifying the appropriate variables to include in the analysis, checking for multicollinearity among the independent variables, assessing the linearity and normality of the data, and interpreting the results to determine the strength and direction of the relationships between the variables. Additionally, SPSS provides tools for evaluating the overall model fit and identifying any outliers or influential data points that may affect the results. By following these principles, SPSS allows for a thorough and accurate analysis of multiple regression models.

Multiple Regressions of SPSS

In this section, we are going to learn about Multiple Regression. Multiple Regression is a regression analysis method in which we see the effect of multiple independent variables on one dependent variable. For this, we will take the Employee data set. This data set is arranged according to their ID, gender, education, job category, salary, salary at the beginning, job time, previous experience, and whether they belong to a minority community or not.


*Employee Data.sav [DataSet1] - IBM SPSS Statistics Data Editor

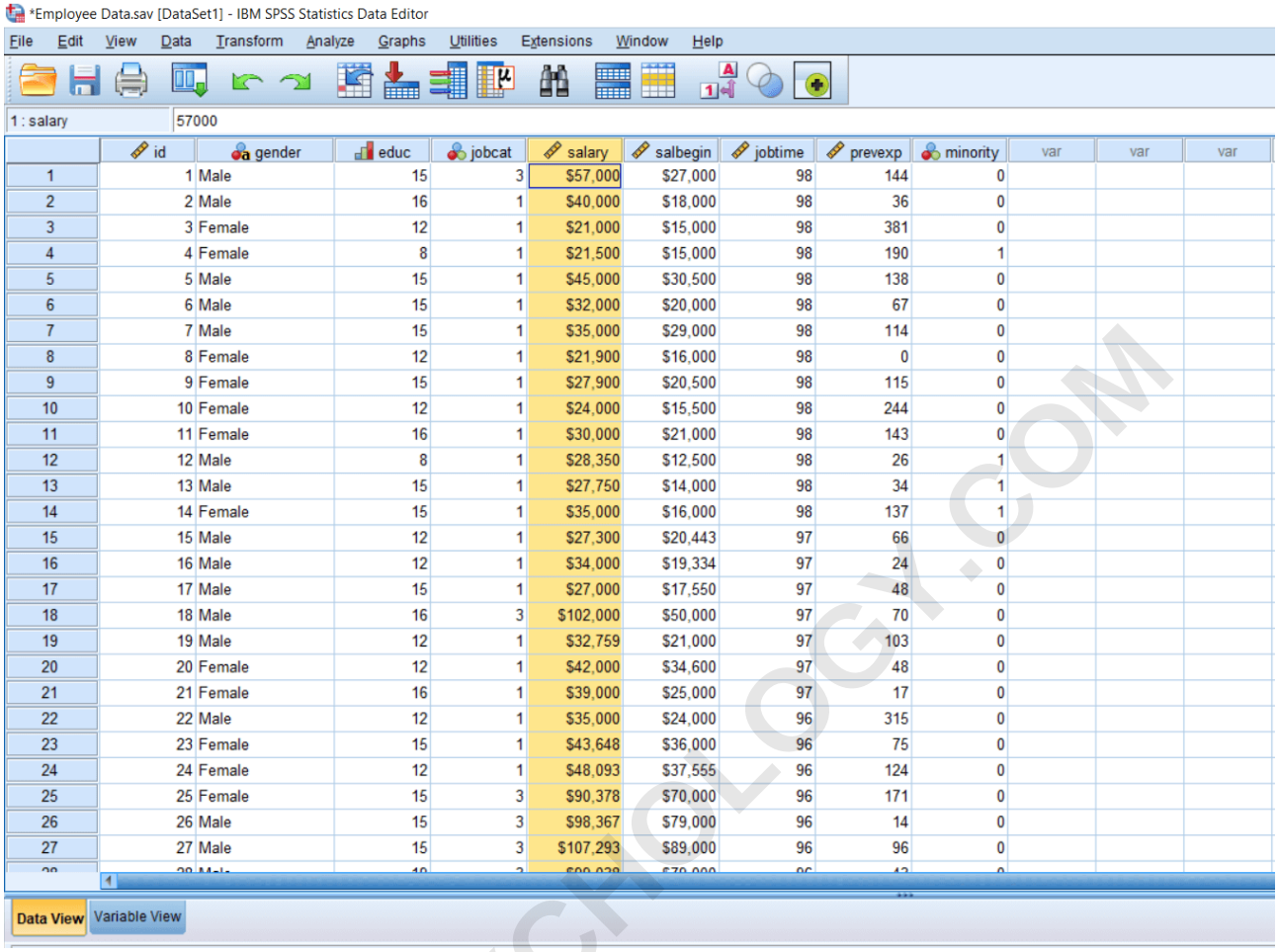
File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

27 :

	id	gender	educ	jobcat	salary	salbegin	jobtime	prevexp	minority	var	var	var	var	var	var
1	1	Male	15	3	$57,000	$27,000	98	144	0						
2	2	Male	16	1	$40,000	$18,000	98	36	0						
3	3	Female	12	1	$21,000	$15,000	98	381	0						
4	4	Female	8	1	$21,500	$15,000	98	190	1						
5	5	Male	15	1	$45,000	$30,500	98	138	0						
6	6	Male	15	1	$32,000	$20,000	98	67	0						
7	7	Male	15	1	$35,000	$29,000	98	114	0						
8	8	Female	12	1	$21,900	$16,000	98	0	0						
9	9	Female	15	1	$27,900	$20,500	98	115	0						
10	10	Female	12	1	$24,000	$15,500	98	244	0						
11	11	Female	16	1	$30,000	$21,000	98	143	0						
12	12	Male	8	1	$28,350	$12,500	98	26	1						
13	13	Male	15	1	$27,750	$14,000	98	34	1						
14	14	Female	15	1	$35,000	$16,000	98	137	1						
15	15	Male	12	1	$27,300	$20,443	97	66	0						
16	16	Male	12	1	$34,000	$19,334	97	24	0						
17	17	Male	15	1	$27,000	$17,550	97	48	0						
18	18	Male	16	3	$102,000	$50,000	97	70	0						
19	19	Male	12	1	$32,759	$21,000	97	103	0						
20	20	Female	12	1	$42,000	$34,600	97	48	0						
21	21	Female	16	1	$39,000	$25,000	97	17	0						
22	22	Male	12	1	$35,000	$24,000	96	315	0						
23	23	Female	15	1	$43,648	$36,000	96	75	0						
24	24	Female	12	1	$48,093	$37,555	96	124	0						
25	25	Female	15	3	$90,378	$70,000	96	171	0						
26	26	Male	15	3	$98,367	$79,000	96	14	0						
27	27	Male	15	3	$107,293	$89,000	96	96	0						

Data View Variable View

Now suppose in this data set, we want to find out what exactly determines the Current Salary drawn by the employees.


*Employee Data.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

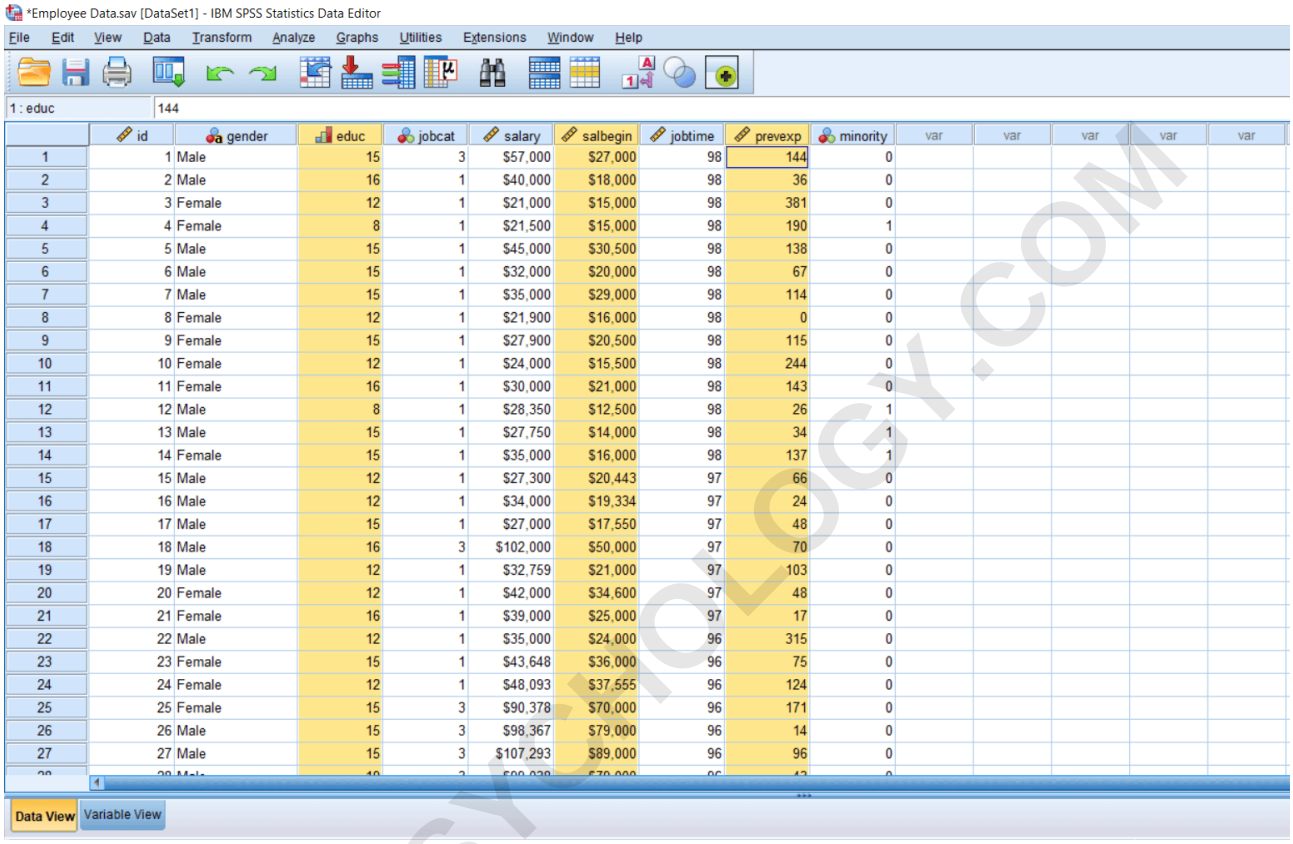
1 : salary 57000

	id	gender	educ	jobcat	salary	salbegin	jobtime	prevexp	minority	var	var	var
1	1	Male	15	3	$57,000	$27,000	98	144	0			
2	2	Male	16	1	$40,000	$18,000	98	36	0			
3	3	Female	12	1	$21,000	$15,000	98	381	0			
4	4	Female	8	1	$21,500	$15,000	98	190	1			
5	5	Male	15	1	$45,000	$30,500	98	138	0			
6	6	Male	15	1	$32,000	$20,000	98	67	0			
7	7	Male	15	1	$35,000	$29,000	98	114	0			
8	8	Female	12	1	$21,900	$16,000	98	0	0			
9	9	Female	15	1	$27,900	$20,500	98	115	0			
10	10	Female	12	1	$24,000	$15,500	98	244	0			
11	11	Female	16	1	$30,000	$21,000	98	143	0			
12	12	Male	8	1	$28,350	$12,500	98	26	1			
13	13	Male	15	1	$27,750	$14,000	98	34	1			
14	14	Female	15	1	$35,000	$16,000	98	137	1			
15	15	Male	12	1	$27,300	$20,443	97	66	0			
16	16	Male	12	1	$34,000	$19,334	97	24	0			
17	17	Male	15	1	$27,000	$17,550	97	48	0			
18	18	Male	16	3	$102,000	$50,000	97	70	0			
19	19	Male	12	1	$32,759	$21,000	97	103	0			
20	20	Female	12	1	$42,000	$34,600	97	48	0			
21	21	Female	16	1	$39,000	$25,000	97	17	0			
22	22	Male	12	1	$35,000	$24,000	96	315	0			
23	23	Female	15	1	$43,648	$36,000	96	75	0			
24	24	Female	12	1	$48,093	$37,555	96	124	0			
25	25	Female	15	3	$90,378	$70,000	96	171	0			
26	26	Male	15	3	$98,367	$79,000	96	14	0			
27	27	Male	15	3	$107,293	$89,000	96	96	0			

Data View Variable View

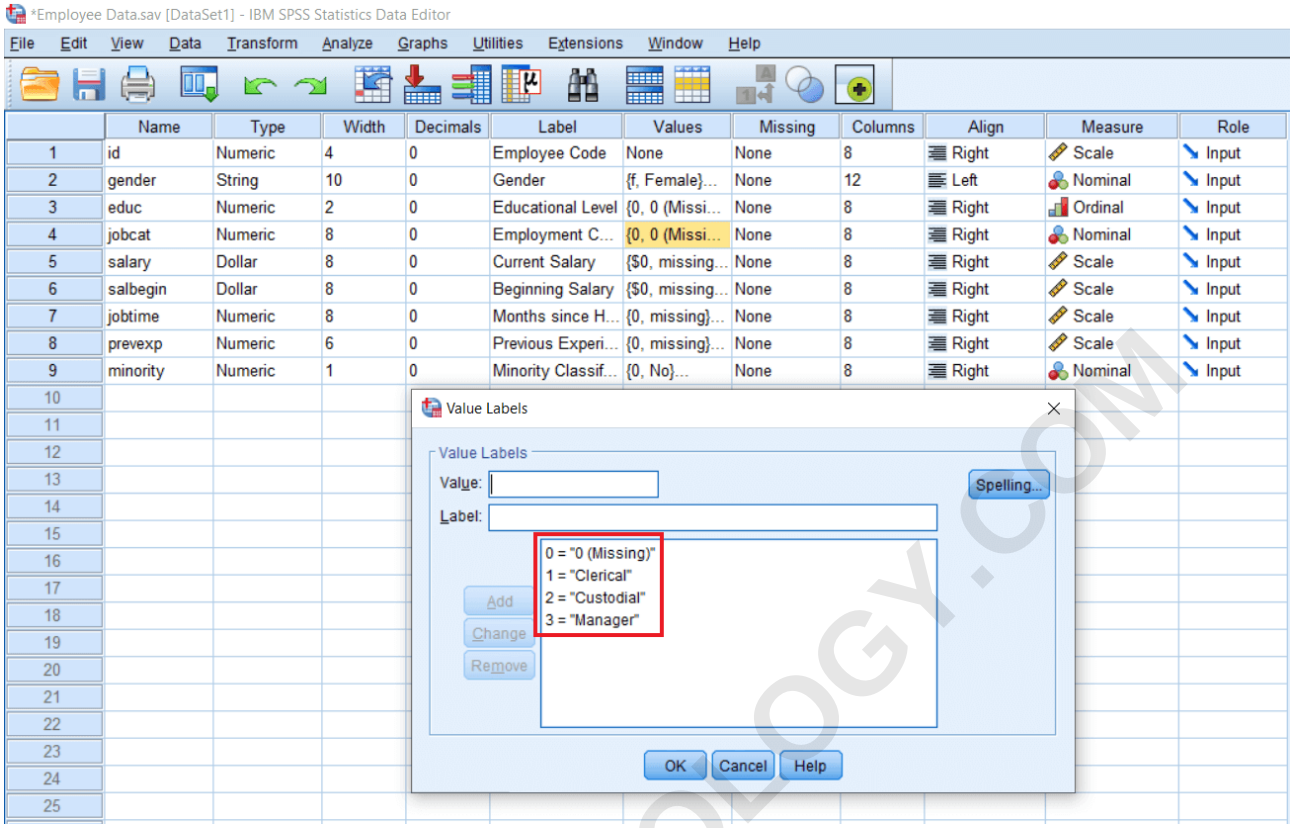
The current salary can be determined by their salary at the time of joining because it's logical to assume that those employees who are drawing a higher salary at the time of joining they will also draw a higher salary currently. We can also guess that the employee's experience will also contribute to the salary they are drawing. Apart from this, we also have an education. So education is again an important criterion for determining the salary. Those highly educated employees can presume that they are drawing a higher

amount of salary compared to those who are less educated.



	id	gender	educ	jobcat	salary	salbegin	jobtime	prevexp	minority	var	var	var	var	var
1	1	Male	15	3	$57,000	$27,000	98	144	0					
2	2	Male	16	1	$40,000	$18,000	98	36	0					
3	3	Female	12	1	$21,000	$15,000	98	381	0					
4	4	Female	8	1	$21,500	$15,000	98	190	1					
5	5	Male	15	1	$45,000	$30,500	98	138	0					
6	6	Male	15	1	$32,000	$20,000	98	67	0					
7	7	Male	15	1	$35,000	$29,000	98	114	0					
8	8	Female	12	1	$21,900	$16,000	98	0	0					
9	9	Female	15	1	$27,900	$20,500	98	115	0					
10	10	Female	12	1	$24,000	$15,500	98	244	0					
11	11	Female	16	1	$30,000	$21,000	98	143	0					
12	12	Male	8	1	$28,350	$12,500	98	26	1					
13	13	Male	15	1	$27,750	$14,000	98	34	1					
14	14	Female	15	1	$35,000	$16,000	98	137	1					
15	15	Male	12	1	$27,300	$20,443	97	66	0					
16	16	Male	12	1	$34,000	$19,334	97	24	0					
17	17	Male	15	1	$27,000	$17,550	97	48	0					
18	18	Male	16	3	$102,000	$50,000	97	70	0					
19	19	Male	12	1	$32,759	$21,000	97	103	0					
20	20	Female	12	1	$42,000	$34,600	97	48	0					
21	21	Female	16	1	$39,000	$25,000	97	17	0					
22	22	Male	12	1	$35,000	$24,000	96	315	0					
23	23	Female	15	1	$43,648	$36,000	96	75	0					
24	24	Female	12	1	$48,093	$37,555	96	124	0					
25	25	Female	15	3	$90,378	$70,000	96	171	0					
26	26	Male	15	3	$98,367	$79,000	96	14	0					
27	27	Male	15	3	$107,293	$89,000	96	96	0					

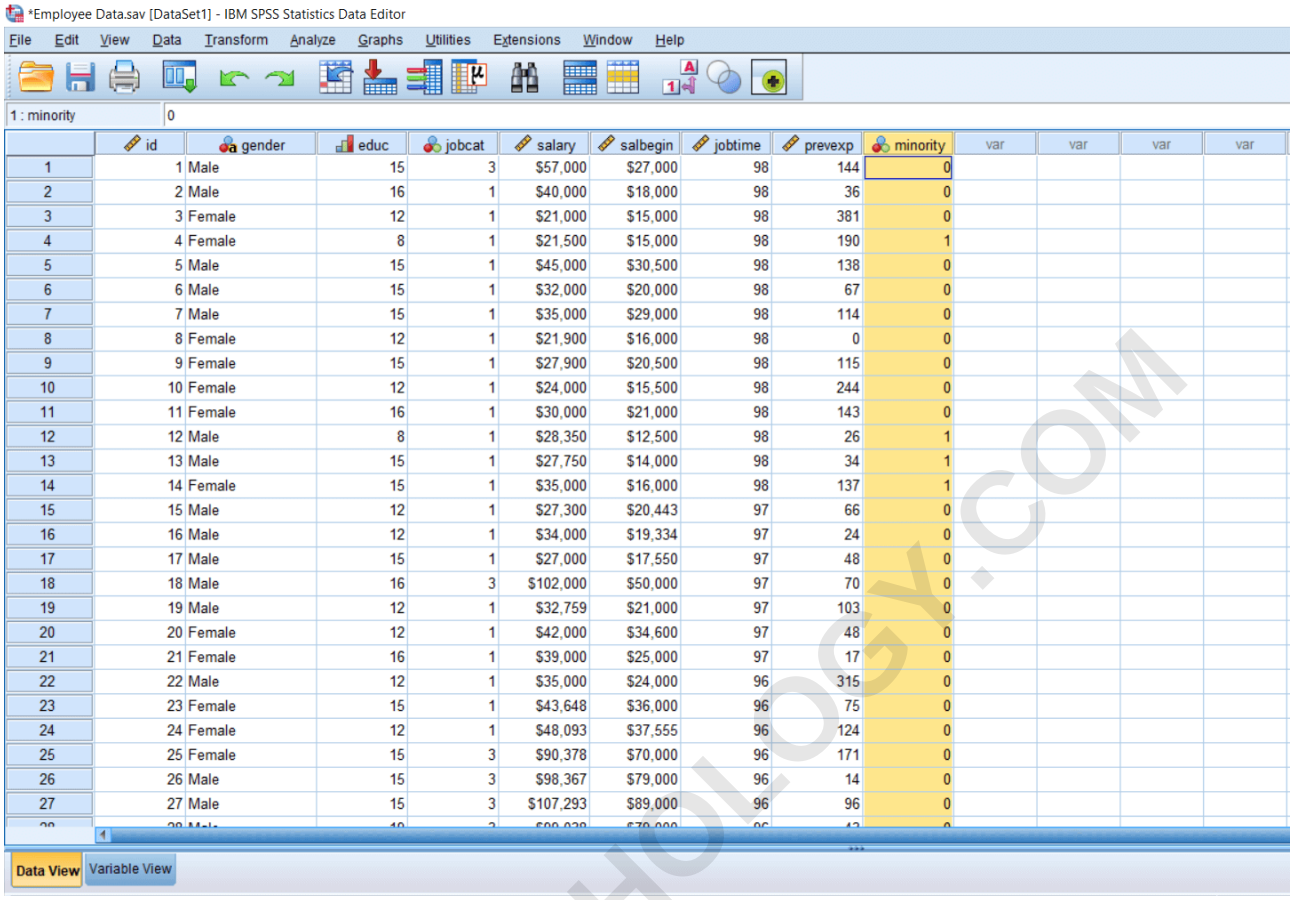
Similarly, we can also guess that salary drawn will also be affected by the employee's position. In our case, we have three categories of employees, i.e., Clerical, Custodial, Manager.



We will guess that Managers are drawing a higher salary as compared to Clerical or Custodial employees. Suppose we want to test these assumptions that job category, education, or position of the employee or the salary at the time of joining the organization are the influence factors in the employee's current salary. In that case, we have to run the multiple regression analysis. The idea of multiple regression analysis is very clear. When we want to predict one dependent variable in our case, it's a Current salary by many independent variables like education, job category,

beginning salary, job timing, previous experience, then we can perform a multiple regression analysis.

When we perform a multiple regression analysis, our variables must be logically selected. For example, do we believe that being in a minority status does affect the salary of a person. Well, it may or may not. So, it's interesting to say this, but theoretically, if we find any justification that minority affiliation of the person may affect his or her salary, then we can include that variable as well in the multiple regression analysis, or if we want to include all the variables in our multiple regression analysis, we can do that. SPSS is going to tell whether this variable exercises a significant influence on the dependent variable or not.


*Employee Data.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

1 : minority 0

	id	gender	educ	jobcat	salary	salbegin	jobtime	prevexp	minority	var	var	var	var
1	1	Male	15	3	$57,000	$27,000	98	144	0				
2	2	Male	16	1	$40,000	$18,000	98	36	0				
3	3	Female	12	1	$21,000	$15,000	98	381	0				
4	4	Female	8	1	$21,500	$15,000	98	190	1				
5	5	Male	15	1	$45,000	$30,500	98	138	0				
6	6	Male	15	1	$32,000	$20,000	98	67	0				
7	7	Male	15	1	$35,000	$29,000	98	114	0				
8	8	Female	12	1	$21,900	$16,000	98	0	0				
9	9	Female	15	1	$27,900	$20,500	98	115	0				
10	10	Female	12	1	$24,000	$15,500	98	244	0				
11	11	Female	16	1	$30,000	$21,000	98	143	0				
12	12	Male	8	1	$28,350	$12,500	98	26	1				
13	13	Male	15	1	$27,750	$14,000	98	34	1				
14	14	Female	15	1	$35,000	$16,000	98	137	1				
15	15	Male	12	1	$27,300	$20,443	97	66	0				
16	16	Male	12	1	$34,000	$19,334	97	24	0				
17	17	Male	15	1	$27,000	$17,550	97	48	0				
18	18	Male	16	3	$102,000	$50,000	97	70	0				
19	19	Male	12	1	$32,759	$21,000	97	103	0				
20	20	Female	12	1	$42,000	$34,600	97	48	0				
21	21	Female	16	1	$39,000	$25,000	97	17	0				
22	22	Male	12	1	$35,000	$24,000	96	315	0				
23	23	Female	15	1	$43,648	$36,000	96	75	0				
24	24	Female	12	1	$48,093	$37,555	96	124	0				
25	25	Female	15	3	$90,378	$70,000	96	171	0				
26	26	Male	15	3	$98,367	$79,000	96	14	0				
27	27	Male	15	3	$107,293	$89,000	96	96	0				

Data View Variable View

The model of Multiple Regression is very simple. We have to select a dependent variable. Generally, we denote our dependent variable by the symbol y, and then we have many independent variables, and we can call them x1, x2, x3 till we can have xn.

(y = x1 x2 x3 + ----- + xn

Now we are going to get the coefficient by applying the multiple regression analysis. So suppose those constant or coefficient is α1x1 + α2x2 + α3x3 till αnxn.

Now we will make a prediction. Now we are predicting y based on these x variables from x1 till xn. We will make some error because we cannot always find all those variables that will completely predict y. So these are bound to some error term. Again, we are going to find that out. Apart from that, we are going to have a constant as well in our regression equation.

$y = \alpha 1x1 + \alpha 2x2 + \alpha 3x3 + \text{----- } + \alpha nxn + \text{error} + \text{constant}$

So that's our typical theoretical regression model. Now we have to use the word α, but most typically, people use the word β, so we can again rewrite the equation as $\beta 1x1 + \beta 2x2 + \beta 3x3$ till we have βnxn then our error term plus constant term.

$y = \beta 1x1 + \beta 2x2 + \beta 3x3 + \text{----- } + \beta nxn + \text{error} + \text{constant}$

This β is the standardized regression weights that we are going to get after the regression analysis. In our case, we wanted to predict the Current Salary of the employee. So, we will write our multiple regression equation as Current Salary = $\beta 1$. Now take one variable as the employee's beginning salary, so Current Salary =

β1* Beginning salary.

Current Salary = β1* Beginning salary

Now we can take our second variable as education. So for that, we are going to get a second coefficient that is β2* education category.

Current Salary = β1* Beginning salary + β2* education category

Then the third variable we may take as an experience. So for that, we are going to get a third coefficient that is β3* experience. So we have built a sample model by taking into three variables.

Current Salary = β1* Beginning salary + β2* education category + β3* experience

If we want to take more variables, we can do that. We can make a lengthy or complex regression model. Now we are going to add our error term and then constant. So that makes our regression model clear.

Current Salary = β1* Beginning salary + β2* education

category + β3* experience + error + constant

We can see, in the case of multiple regression analysis, we can take our independent variable as either a nominal variable or a metric variable. The independent variable could be metric or non-metric. But our dependent variable in case of multiple regression analysis or linear regression analysis or hierarchy regression analysis should always be metric.

Current Salary = β1* Beginning salary + β2* education category + β3* experience + error + constant