NORTHWESTERN UNIVERSITY

SCHOOL OF CONTINUING STUDIES

Handout: Problem Set #5 Solutions

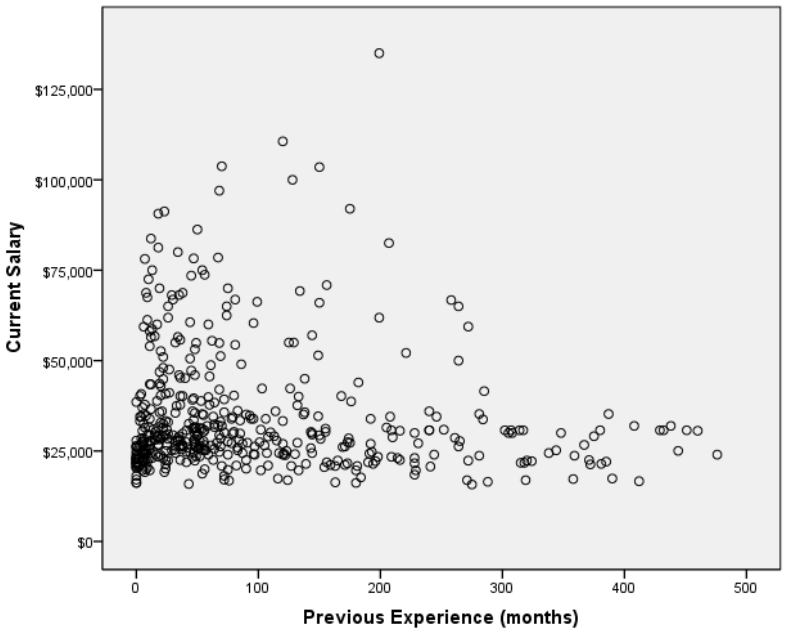
PREDICT 401: Introduction to Statistical Analysis

These problem sets are meant to allow you to practice and check the accuracy of your work. Please do not review the solutions until you have finalized your work. Although these problem sets are not submitted and graded, treat them as if they were. It is to your great benefit to work on and even struggle with the problem sets. Looking at the solutions before finalizing your work will, quite simply, make for a less meaningful learning experience.

For this problem set, use the "Employee data" dataset provided to you.

1. In this problem, we will address whether an employee's salary is related to the length of previous experience?

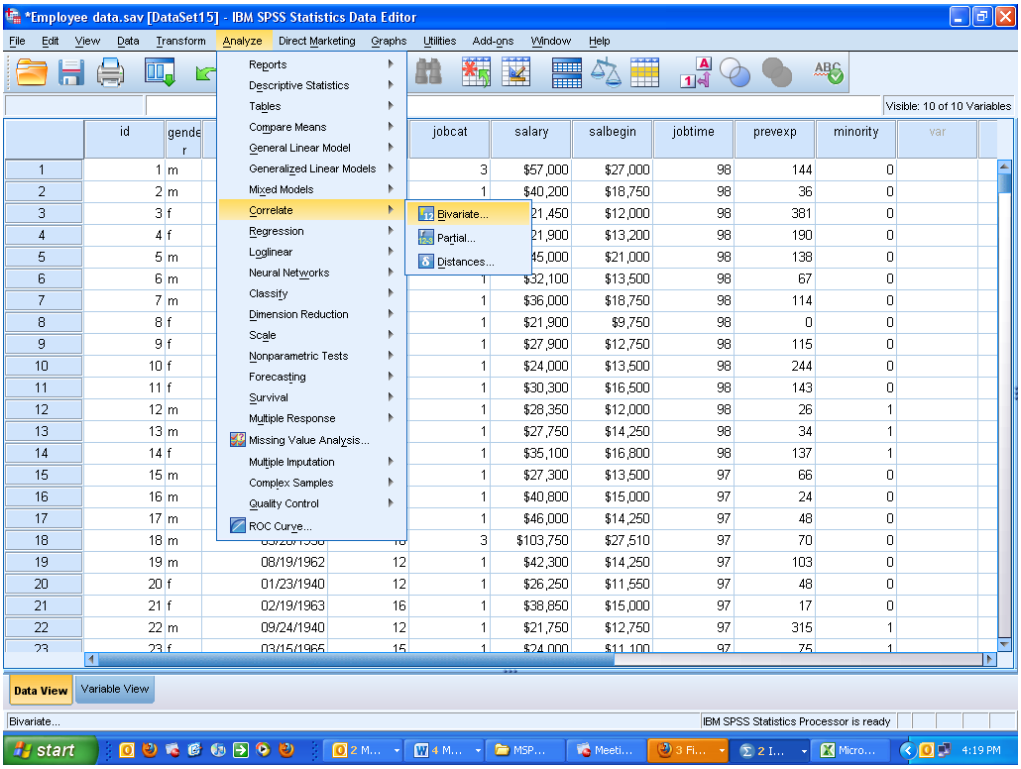
    a. Using SPSS, create a scatterplot with salary on the vertical axis and length of previous experience on the horizontal axis. Does this graph tell you anything about the relationship?



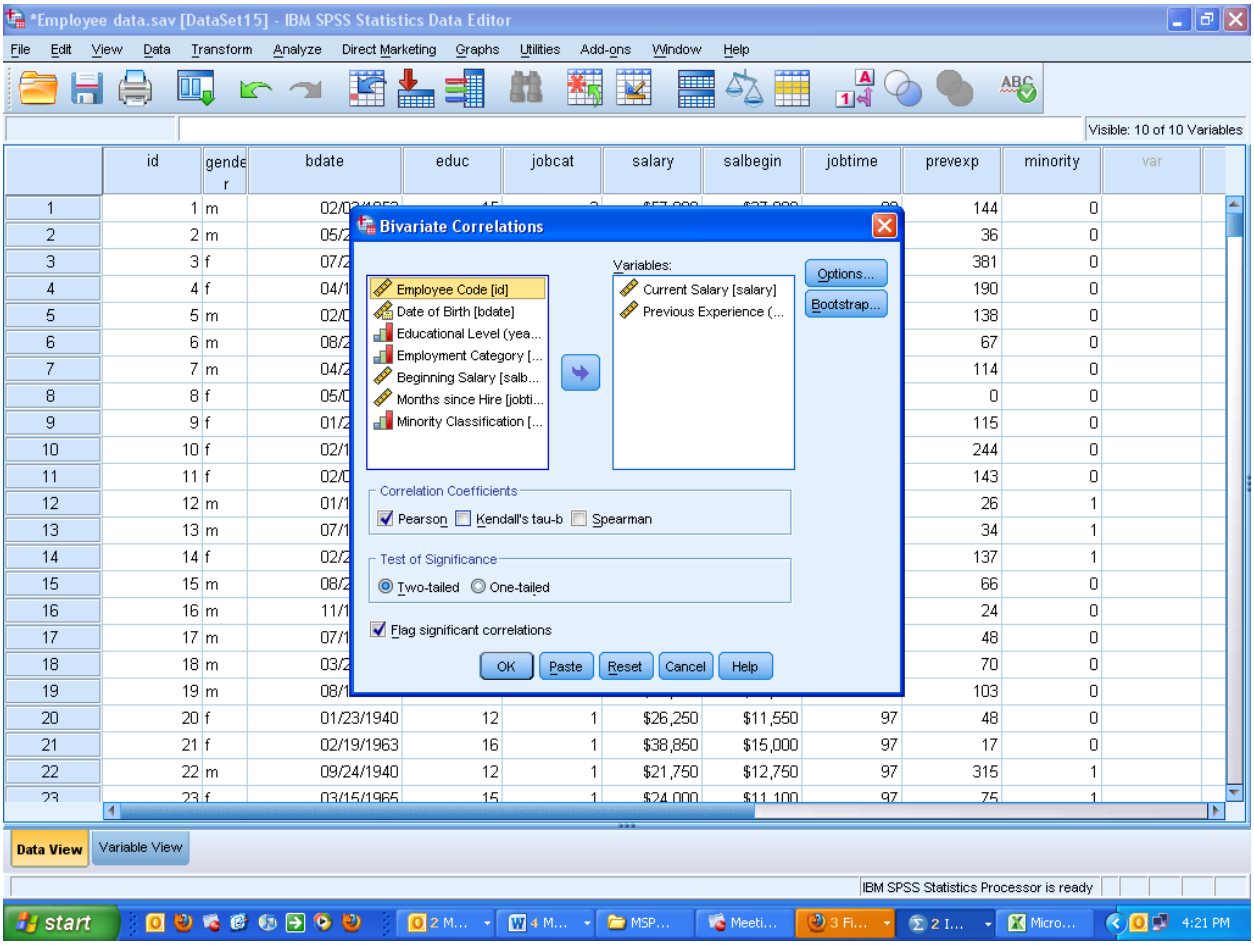
This graph does not easily allow me to determine whether there's any relationship between current salary and previous experience. There is no obvious trend that jumps out at me.

    b. What is the Pearson's correlation between these two variables?

We can find the correlation using the bivariate correlation command, as shown below.



Then, we select the two variables to correlate and ensure we've chosen the "Pearson's" option, as shown below.



The output provided is:

Correlations

		Current Salary	Previous Experience (months)
Current Salary	Pearson Correlation	1	-.097*
	Sig. (2-tailed)		.034
	N	474	474
Previous Experience (months)	Pearson Correlation	-.097*	1
	Sig. (2-tailed)	.034	
	N	474	474

***. Correlation is significant at the 0.05 level (2-tailed).**

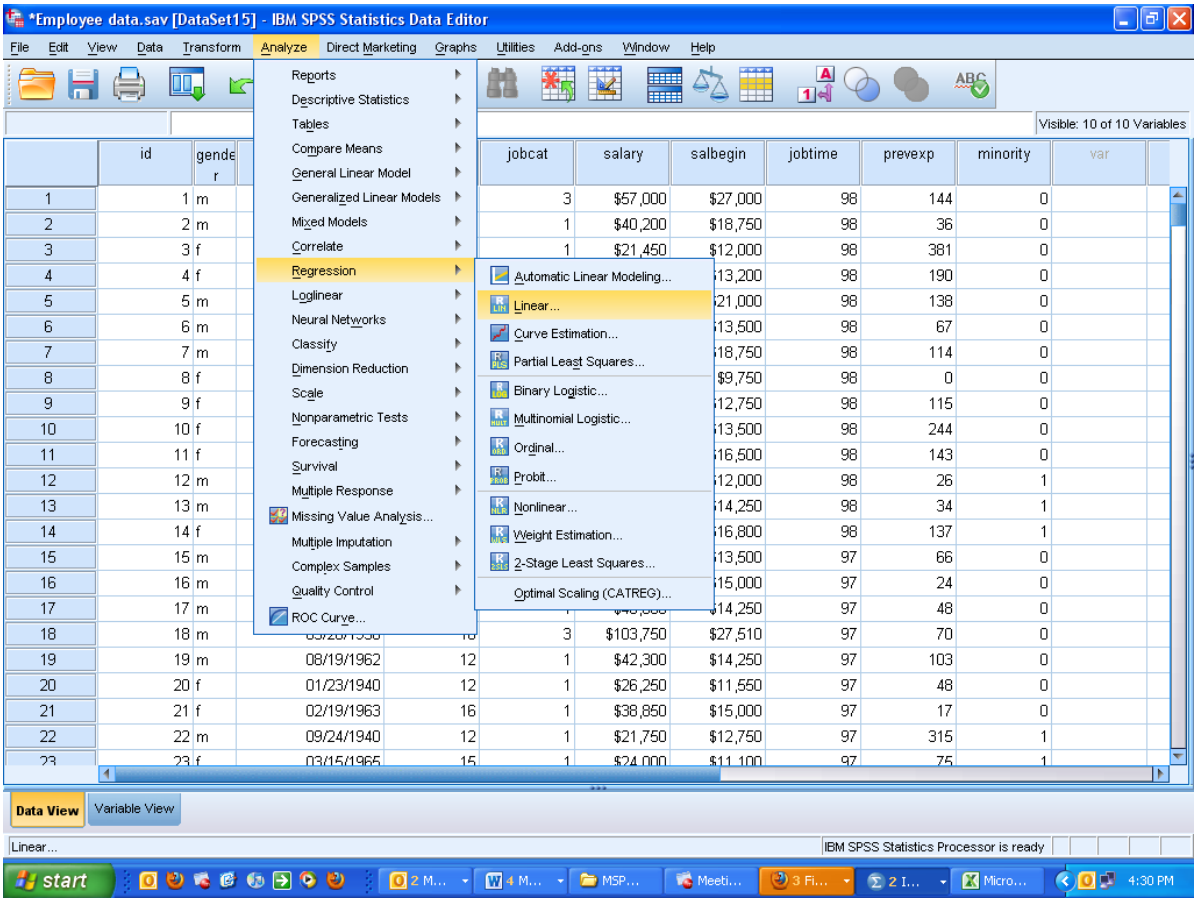
Thus, the correlation between current salary and length of previous experience equals -0.097. In other words, there is a very slight negative linear relationship between salary and length prior experience. We also note

that we are confident that there is a linear relationship at the 95% level, since the significance level calculated (0.034) is less than 0.05. That is, we can reject the null hypothesis that there is no linear relationship between the two variables.

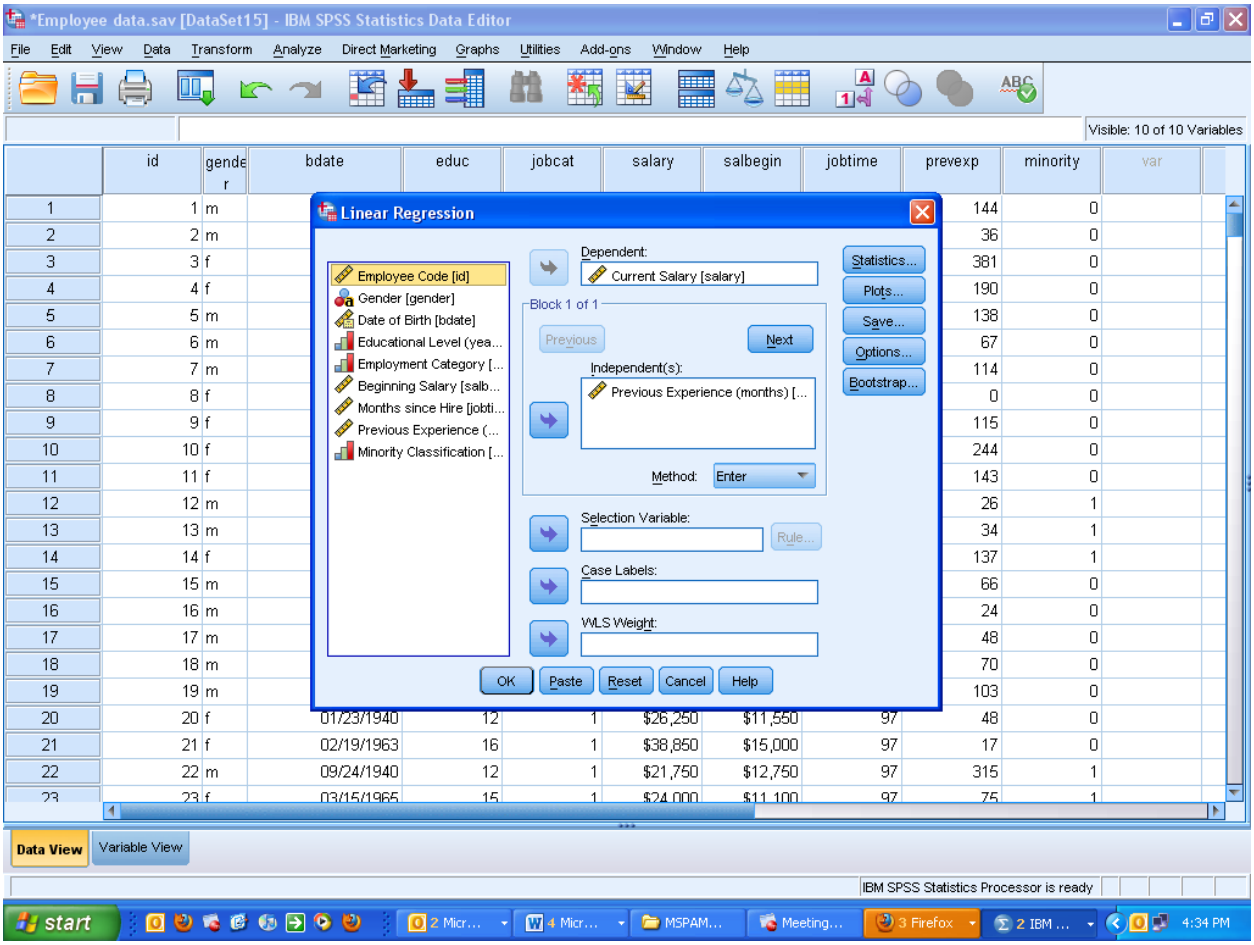
c. Run a regression of "salary" as the dependent variable on "length of experience" as the independent variable. That is, use SPSS to find alpha and beta in the equation below:

$$\text{Salary} = \alpha + \beta * \text{experience} + \varepsilon$$

We can run a linear regression with the "analyze" command:



We have to include salary as the dependent variable and length of experience as the independent variable as shown here:



The results of the regression are shown here:

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	35945.029	1060.488		33.895	.000
	Previous Experience (months)	-15.913	7.479	-.097	-2.128	.034

a. Dependent Variable: Current Salary

In this model, value of alpha = 35945.029 and the value of beta = -15.913. In other words, the equation described earlier becomes:

$$\text{Salary} = 35{,}945.029 + -15.913 * \text{experience} + \varepsilon$$

How do we interpret these coefficients?

- **Our value of beta tells us that for every additional month of experience, salary actually decreases by $15.91.**
- **Our value of alpha tells us that at 0 months of experience, we would expect a salary of $35,945.03.**

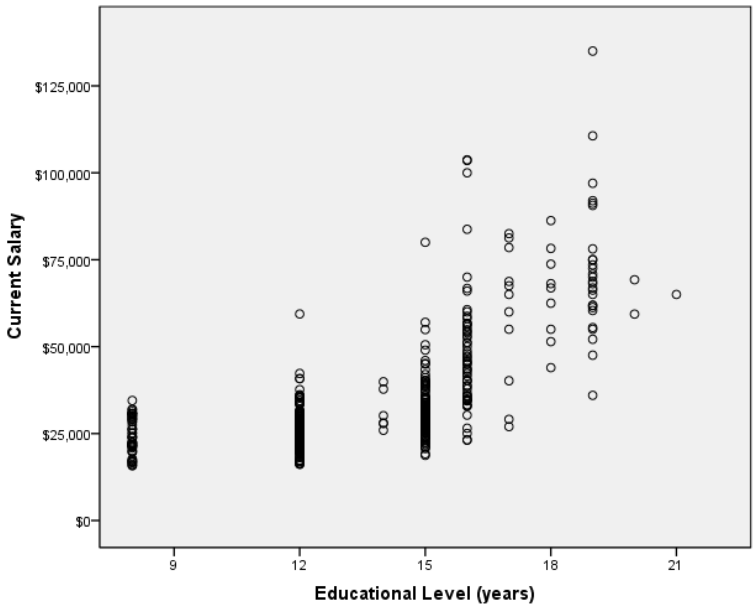
We can also reject at the 95% level the null hypothesis that there is no relationship between length of prior experience and salary. We know this because the significance of beta is less than 0.05.

d. So, what's really going on here?

Salary and length of prior experience are only very slightly related. After all, every additional year of experience results in about $16 less of salary. That's a really small effect. So why is it statistically significant? Well, we get statistical significance with greater effects, small standard deviations, and lots of observations. Here, the relatively small effect size is countered by a moderately large data base. So, while the effect isn't big, it is statistically significant.

2. What is the relationship between salary and years of education? Use the same tools and processes as described in problem 1.

a. According to the scatterplot below, I suspect that there is a positive relationship between salary and educational level. Salary seems to be increasing as education increases.



b. Pearson's correlation output, below, shows a strong positive linear relationship that is statistically significant at 95% (the significant shown is rounded down to 0.000)

Correlations

		Current Salary	Educational Level (years)
Current Salary	**Pearson Correlation**	**1**	.661**
	Sig. (2-tailed)		.000
	N	**474**	474
Educational Level (years)	**Pearson Correlation**	**.661****	1
	Sig. (2-tailed)	**.000**	
	N	**474**	474

****. Correlation is significant at the 0.01 level (2-tailed).**

c. To find the actual relationship, we can model the equation

$$\text{Salary} = \alpha + \beta * \text{education} + \varepsilon$$

OLS regression output shows:

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	**(Constant)**	**-18331.178**	**2821.912**		**-6.496**	**.000**
	Educational Level (years)	**3909.907**	**204.547**	**.661**	**19.115**	**.000**

a. Dependent Variable: Current Salary

So, the OLS equation becomes

$$\text{Salary} = -18{,}331.178 + 3909.907 * \text{education} + \varepsilon$$

In other words, every additional year of experience yields, on average, $3,909,91 more in current salary. At 0 years of education, we would infer a *negative* salary. Obviously, there are no observations with those actual values, but that point is suggested by the shape of the OLS line. Given the significance calculations, we are highly confident that the relationship between salary and education is not zero.

3. Given your answers to problems 1 and 2 above, does prior experience cause a person to earn less money at this company? Does more education cause a person to earn more money?

We cannot answer either of these questions with any certainty at this point. These analyses have simply told us the statistical relationship between the variables. It doesn't allow us to infer any sort of causation. For instance, in question 2, it may well be that this company values education and explicitly offers compensation that is related to years of education. But perhaps not. Perhaps the company offers compensation based on another factor entirely, such as quality of employee's past performance. If those who earn higher levels of

education also tend to be those who are high performers, then perhaps education is merely a proxy for performance. That is, education may be related to higher pay, but it might not directly cause the higher pay, except through quality of job performance.

