

**Research Methods**

# **SPSS and Statistics**

**Final Exam for PA 706 at SFSU**

**Spring 2006**



**Nikhila Pai**

**PART A.**

1a.

**Statistics**

Gov11 For my next series of questions I'd like you to think about the LAST time you made contact with the government in the past year, for any reason. It could have been over the phone, online, in person, or by mail. What was the purpose of that contact

N	Valid	2942
	Missing	1991
Mode		4

**Gov11 For my next series of questions I'd like you to think about the LAST time you made contact with the government in the past year, for any reason. It could have been over the phone, online, in person, or by mail. What was the purpose of that contact**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	To get information or an answer to a specific question	686	13.9	23.3	23.3
	To express your opinion	497	10.1	16.9	40.2
	To get help solving a problem	295	6.0	10.0	50.2
	To carry out a transaction, such as filing taxes or register	879	17.8	29.9	80.1
	Was it for some other purpose I haven't mentioned? (SPECIFY)	186	3.8	6.3	86.4
	(VOL) Combination of above	147	3.0	5.0	91.4
	(VOL) Didn't make contact with government in past year (SKIP	171	3.5	5.8	97.3
	Don't know/Refused	81	1.6	2.7	100.0
	Total	2942	59.6	100.0	
Missing	System	1991	40.4		
Total		4933	100.0		

Gov 11 is nominal because the values are categorical variables with no numerical meaning or ranking. The only appropriate measure of central tendency is the mode on nominal level data as it is impossible to calculate the median and mean (as it is not possible to assign numerical scores to the cases). According to the *Social Statistics* SPSS textbook, there is no dispersion statistic for nominal level data, however, a frequency distribution can help in assessing the variability as it will tell us the percentage of respondents in each category.

**Statistics**

Gov12 Is the government the FIRST place you went for this purpose, or did you go someplace else first?

N	Valid	2771
	Missing	2162
Mode		1

**Gov12 Is the government the FIRST place you went for this purpose, or did you go someplace else first?**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Government first	2164	43.9	78.1	78.1
	Someplace else first	533	10.8	19.2	97.3
	Don't know/Refused	74	1.5	2.7	100.0
	Total	2771	56.2	100.0	
Missing	System	2162	43.8		
Total		4933	100.0		

Gov 12 is nominal because the values are categorical variables with no numerical meaning or ranking. The only appropriate measure of central tendency is the mode on nominal level data as it is impossible to calculate the median and mean (as it is not possible to assign numerical scores to the cases). According to the *Social Statistics SPSS* textbook, there is no dispersion statistic for nominal level data, however, a frequency distribution can help in assessing the variability as it will tell us the percentage of respondents in each category.

**Statistics**

Gov13 At the time you contacted the government, how complicated did you think your question, transaction, or problem was? Did you think it was... (READ 1-3)

N	Valid	2771
	Missing	2162
Median		3.00
Mode		3
Range		8
Percentiles	25	2.00
	50	3.00
	75	3.00

**Gov13 At the time you contacted the government, how complicated did you think your question, transaction, or problem was? Did you think it was... (READ 1-3)**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	VERY COMPLICATED, meaning it would require substantial effort	240	4.9	8.7	8.7

	SOMEWHAT COMPLICATED, requiring moderate effort and only som	606	12.3	21.9	30.5
	NOT REALLY COMPLICATED AT ALL, requiring minimal effort and	1879	38.1	67.8	98.4
	Don't know/Refused	45	.9	1.6	100.0
	Total	2771	56.2	100.0	
Missing	System	2162	43.8		
Total		4933	100.0		

Gov 13 is ordinal because the values are categorical and ranked. The median and the mode are considered appropriate measures of central tendency for ordinal level data. In this case, it seems the mode would be more appropriate because the values are ranked and the middle value (median) will always be the same no matter how many cases fall outside the middle value. According to the *Social Statistics* SPSS textbook, the range and the quartiles are useful measure of dispersion for ordinal level data. The range is the distance between the smallest and largest values in a set of data while providing a context for interpreting other measures of dispersion like the interquartile range and standard deviation. The IQR tells us about the distribution of responses in relation to the median; where the smaller the IQR the more homogenous the distribution and the larger the IQR the more heterogeneous the distribution.

### Statistics

Gov14 How URGENT was your last contact with the government? Was it... (READ 1-3)

N	Valid	2771
	Missing	2162
Median		3.00
Mode		3
Range		8
Percentiles	25	2.00
	50	3.00
	75	3.00

Gov14 How URGENT was your last contact with the government? Was it... (READ 1-3)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	VERY urgent, meaning you needed a response within 24 hours	313	6.3	11.3	11.3
	Somewhat urgent, but not requiring an immediate response	1068	21.6	38.5	49.8
	Not really urgent at all	1356	27.5	48.9	98.8
	Don't know/Refused	34	.7	1.2	100.0
	Total	2771	56.2	100.0	
Missing	System	2162	43.8		

Total	4933	100.0		
-------	------	-------	--	--

Gov 14 is ordinal because the values are categorical and ranked. The median and the mode are considered appropriate measures of central tendency for ordinal level data. In this case, it seems the mode would be more appropriate because the values are ranked and the middle value (median) will always be the same no matter how many cases fall outside the middle value. According to the *Social Statistics* SPSS textbook, the range and the quartiles are useful measure of dispersion for ordinal level data. The range is the distance between the smallest and largest values in a set of data while providing a context for interpreting other measures of dispersion like the interquartile range and standard deviation. The IQR tells us about the distribution of responses in relation to the median; where the smaller the IQR the more homogenous the distribution and the larger the IQR the more heterogeneous the distribution.

**Statistics**

Gov15 Were you contacting the government for PERSONAL reasons or for BUSINESS reasons?

N	Valid	2771
	Missing	2162
Mode		1

Gov15 Were you contacting the government for PERSONAL reasons or for BUSINESS reasons?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Personal	1969	39.9	71.1	71.1
	Business	540	10.9	19.5	90.6
	(VOL) Both	198	4.0	7.1	97.7
	(VOL) Neither/Something else (SPECIFY)	52	1.1	1.9	99.6
	Don't know/Refused	12	.2	.4	100.0
	Total	2771	56.2	100.0	
Missing	System	2162	43.8		
Total		4933	100.0		

Gov 15 is nominal because the values are categorical variables with no numerical meaning or ranking. The only appropriate measure of central tendency is the mode on nominal level data as it is impossible to calculate the median and mean (as it is not possible to assign numerical scores to the cases). According to the *Social Statistics* SPSS textbook, there is no dispersion statistic for nominal level data, however, a frequency distribution can help in assessing the variability as it will tell us the percentage of respondents in each category.

**Statistics**

GOV16 Do you happen to recall what LEVEL of government—local, state, or federal—you were trying to reach?

N	Valid	2771
	Missing	2162
Mode		2

**GOV16 Do you happen to recall what LEVEL of government—local, state, or federal—you were trying to reach?**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Local	523	10.6	18.9	18.9
	State	962	19.5	34.7	53.6
	Federal	887	18.0	32.0	85.6
	(VOL) Combination of above	198	4.0	7.1	92.8
	(VOL) Don't remember	154	3.1	5.6	98.3
	Don't know/Refused	46	.9	1.7	100.0
	Total	2771	56.2	100.0	
Missing	System	2162	43.8		
Total		4933	100.0		

Gov 16 is nominal because the values are categorical variables with no numerical meaning or ranking. The only appropriate measure of central tendency is the mode on nominal level data as it is impossible to calculate the median and mean (as it is not possible to assign numerical scores to the cases). According to the *Social Statistics SPSS* textbook, there is no dispersion statistic for nominal level data, however, a frequency distribution can help in assessing the variability as it will tell us the percentage of respondents in each category.

**Statistics**

**GOV17 Was your last contact with the government related to filing taxes, or not?**

N	Valid	2771
	Missing	2162
Mode		2

**GOV17 Was your last contact with the government related to filing taxes, or not?**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	662	13.4	23.9	23.9
	No	2093	42.4	75.5	99.4
	Don't know/Refused	16	.3	.6	100.0
	Total	2771	56.2	100.0	
Missing	System	2162	43.8		
Total		4933	100.0		

Gov 17 is nominal because the values are categorical variables with no numerical meaning or ranking. The only appropriate measure of central tendency is the mode on nominal level data as it is impossible to calculate the median and mean (as it is not possible to assign numerical scores to the cases). According to the *Social Statistics SPSS* textbook, there is no dispersion statistic for nominal level data, however, a frequency distribution can help in assessing the variability as it will tell us the percentage of respondents in each category.

1b.

**Statistics**

GOV23 What caused you to switch? Did you switch... (READ 1-3; ROTATE 1-2)?

N	Valid	615
	Missing	4318
Mode		1

GOV23 What caused you to switch? Did you switch... (READ 1-3; ROTATE 1-2)?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Because you weren't getting the response you needed	248	5.0	40.3	40.3
	Because someone instructed you to, or referred you to a diff	136	2.8	22.1	62.4
	Because of some other reason I haven't mentioned? (SPECIFY)	77	1.6	12.5	74.8
	(VOL) Both/All of the above	16	.3	2.7	77.5
	Easier/Faster/More convenient	29	.6	4.8	82.2
	Wanted to speak to person/Do it personally	48	1.0	7.7	90.0
	Needed more information	35	.7	5.6	95.6
	Don't know/Refused	27	.5	4.4	100.0
	Total	615	12.5	100.0	
Missing	System	4318	87.5		
Total		4933	100.0		

Gov 23 is nominal because the values are categorical variables with no numerical meaning or ranking. The only appropriate measure of central tendency is the mode on nominal level data as it is impossible to calculate the median and mean (as it is not possible to assign numerical scores to the cases). According to the *Social Statistics SPSS* textbook, there is no dispersion statistic for nominal level data, however, a frequency distribution can help in assessing the variability as it will tell us the percentage of respondents in each category.

**Statistics**

GOV24 In general, how satisfied were you with your LAST experience contacting the government? Were you...(READ 1-4)

N	Valid	2653
	Missing	2279
Median		2.00
Mode		2
Range		8
Percentiles	25	1.00
	50	2.00

75	2.00
----	------

**GOV24 In general, how satisfied were you with your LAST experience contacting the government?  
Were you...(READ 1-4)**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very satisfied	928	18.8	35.0	35.0
	Somewhat satisfied	1096	22.2	41.3	76.3
	Somewhat DISsatisfied	313	6.4	11.8	88.1
	Very DISsatisfied	261	5.3	9.8	97.9
	Don't know/Refused	56	1.1	2.1	100.0
	Total	2653	53.8	100.0	
Missing	System	2279	46.2		
Total		4933	100.0		

Gov 24 is ordinal because the values are categorical and ranked. The median and the mode are considered appropriate measures of central tendency for ordinal level data. In this case, it seems the mode would be more appropriate because the values are ranked and the middle value (median) will always be the same no matter how many cases fall outside the middle value. According to the *Social Statistics SPSS* textbook, the range and the quartiles are useful measure of dispersion for ordinal level data. The range is the distance between the smallest and largest values in a set of data while providing a context for interpreting other measures of dispersion like the interquartile range and standard deviation. The IQR tells us about the distribution of responses in relation to the median; where the smaller the IQR the more homogenous the distribution and the larger the IQR the more heterogeneous the distribution.

### Statistics

Gov25 Which of the following statements best describes the outcome of your last contact with the government... (READ 1-3)

N	Valid	2653
	Missing	2279
Median		1.00
Mode		1
Range		8
Percentiles	25	1.00
	50	1.00
	75	2.00

**Gov25 Which of the following statements best describes the outcome of your last contact with the government... (READ 1-3)**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	I was successful and accomplished what I wanted to	1673	33.9	63.0	63.0

	I have not yet been successful but I'm still working on it	657	13.3	24.8	87.8
	I was unsuccessful and have stopped trying	257	5.2	9.7	97.5
	(VOL) None of these	41	.8	1.5	99.0
	Don't know/Refused	26	.5	1.0	100.0
	Total	2653	53.8	100.0	
Missing	System	2279	46.2		
Total		4933	100.0		

Gov 25 is ordinal because the values are categorical and ranked. The median and the mode are considered appropriate measures of central tendency for ordinal level data. In this case, it seems the mode would be more appropriate because the values are ranked and the middle value (median) will always be the same no matter how many cases fall outside the middle value. According to the *Social Statistics SPSS* textbook, the range and the quartiles are useful measure of dispersion for ordinal level data. The range is the distance between the smallest and largest values in a set of data while providing a context for interpreting other measures of dispersion like the interquartile range and standard deviation. The IQR tells us about the distribution of responses in relation to the median; where the smaller the IQR the more homogenous the distribution and the larger the IQR the more heterogeneous the distribution.

### Statistics

Gov26 Altogether, did your last contact with the government take about the amount of time you expected, more time than you expected, or less time than you expected?

N	Valid	2653
	Missing	2279
Median		2.00
Mode		1
Range		8
Percentiles	25	1.00
	50	2.00
	75	3.00

Gov26 Altogether, did your last contact with the government take about the amount of time you expected, more time than you expected, or less time than you expected?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	About the time I expected	1232	25.0	46.4	46.4
	More time than I expected	753	15.3	28.4	74.8
	Less time than I expected	627	12.7	23.6	98.4
	Don't know/Refused	42	.8	1.6	100.0
	Total	2653	53.8	100.0	
Missing	System	2279	46.2		
Total		4933	100.0		

Gov 26 is ordinal because the values are categorical and ranked. The median and the mode are considered appropriate measures of central tendency for ordinal level data. In

this case, it seems the mode would be more appropriate because the values are ranked and the middle value (median) will always be the same no matter how many cases fall outside the middle value. According to the *Social Statistics* SPSS textbook, the range and the quartiles are useful measure of dispersion for ordinal level data. The range is the distance between the smallest and largest values in a set of data while providing a context for interpreting other measures of dispersion like the interquartile range and standard deviation. The IQR tells us about the distribution of responses in relation to the median; where the smaller the IQR the more homogenous the distribution and the larger the IQR the more heterogeneous the distribution.

**Statistics**

GOV27 You said that you visited a government website the last time you contacted the government. Did you do that from home or from work?

N	Valid	766
	Missing	4167
Mode		1

**GOV27 You said that you visited a government website the last time you contacted the government. Did you do that from home or from work?**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Home	555	11.2	72.5	72.5
	Work	185	3.8	24.2	96.6
	Don't know/Refused	26	.5	3.4	100.0
	Total	766	15.5	100.0	
Missing	System	4167	84.5		
Total		4933	100.0		

Gov 27 is nominal because the values are categorical variables with no numerical meaning or ranking. The only appropriate measure of central tendency is the mode on nominal level data as it is impossible to calculate the median and mean (as it is not possible to assign numerical scores to the cases). According to the *Social Statistics* SPSS textbook, there is no dispersion statistic for nominal level data, however, a frequency distribution can help in assessing the variability as it will tell us the percentage of respondents in each category.

1c.

**Statistics**

GOV39 For my next few questions, I'd like you to think about ALL of the times you contacted the government in the past year, not just the last time. About what percentage of your contact with the government in the past year was related to filing taxes?

N	Valid	2771
	Missing	2162
Mean		2.72
Std. Error of Mean		.047
Median		1.00
Mode		1
Std. Deviation		2.469

Variance	6.094
Skewness	1.281
Std. Error of Skewness	.047
Kurtosis	.207
Std. Error of Kurtosis	.093
Range	8
Minimum	1
Maximum	9

**GOV39 For my next few questions, I'd like you to think about ALL of the times you contacted the government in the past year, not just the last time. About what percentage of your contact with the government in the past year was related to filing taxes?**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 percent (None)	1473	29.9	53.1	53.1
	Less than 10 percent	363	7.4	13.1	66.2
	10-25 percent (about a quarter)	243	4.9	8.8	75.0
	26-39 percent (about a third)	49	1.0	1.8	76.8
	40-59 percent (about half)	195	4.0	7.0	83.8
	60-79 percent (about two-thirds)	55	1.1	2.0	85.8
	80-99 percent	105	2.1	3.8	89.6
	100 percent (all)	231	4.7	8.3	97.9
	Don't know/Refused	57	1.2	2.1	100.0
	Total	2771	56.2	100.0	
Missing	System	2162	43.8		
Total		4933	100.0		

Gov 39 is ordinal because the values are categorical and ranked. The median and the mode are considered appropriate measures of central tendency for ordinal level data. In this case, it seems the mode would be more appropriate because the values are ranked and the middle value (median) will always be the same no matter how many cases fall outside the middle value. According to the *Social Statistics SPSS* textbook, the range and the quartiles are useful measure of dispersion for ordinal level data. The range is the distance between the smallest and largest values in a set of data while providing a context for interpreting other measures of dispersion like the interquartile range and standard deviation. The IQR tells us about the distribution of responses in relation to the median; where the smaller the IQR the more homogenous the distribution and the larger the IQR the more heterogeneous the distribution.

1d.

### Statistics

		GOV47a. the federal government ... How much, if at all, has the Internet improved the way you interact with... (INSERT; ROTATE) – a lot, some, only a little, or not at all?	GOV47b. your state government ... How much, if at all, has the Internet improved the way you interact with... (INSERT; ROTATE) – a lot, some, only a little, or not at all?	GOV47c. your local government ... How much, if at all, has the Internet improved the way you interact with... (INSERT; ROTATE) – a lot, some, only a little, or not at all?
N	Valid	3063	3063	3063
	Missing	1870	1870	1870
Median		2.00	3.00	3.00
Mode		1	4	4
Range		8	8	8
Percentiles	25	1.00	1.00	2.00
	50	2.00	3.00	3.00
	75	4.00	4.00	4.00

**GOV47a. the federal government ...How much, if at all, has the Internet improved the way you interact with... (INSERT; ROTATE) – a lot, some, only a little, or not at all?**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	A lot	852	17.3	27.8	27.8
	Some	720	14.6	23.5	51.3
	Only a little	498	10.1	16.3	67.6
	Not at all	837	17.0	27.3	94.9
	(VOL) Don't use Internet to contact government	122	2.5	4.0	98.9
	Don't know/Refused	33	.7	1.1	100.0
	Total	3063	62.1	100.0	
Missing	System	1870	37.9		
Total		4933	100.0		

Gov 47a is ordinal because the values are categorical and ranked. The median and the mode are considered appropriate measures of central tendency for ordinal level data. In this case, it seems the mode would be more appropriate because the values are ranked and the middle value (median) will always be the same no matter how many cases fall outside the middle value. According to the *Social Statistics* SPSS textbook, the range and the quartiles are useful measure of dispersion for ordinal level data. The range is the distance between the smallest and largest values in a set of data while providing a context for interpreting other measures of dispersion like the interquartile range and standard deviation. The IQR tells us about the distribution of responses in relation to the median;

where the smaller the IQR the more homogenous the distribution and the larger the IQR the more heterogeneous the distribution.

**GOV47b. your state government ...How much, if at all, has the Internet improved the way you interact with... (INSERT; ROTATE) – a lot, some, only a little, or not at all?**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	A lot	770	15.6	25.1	25.1
	Some	741	15.0	24.2	49.3
	Only a little	576	11.7	18.8	68.1
	Not at all	811	16.5	26.5	94.6
	(VOL) Don't use Internet to contact government	125	2.5	4.1	98.7
	Don't know/Refused	39	.8	1.3	100.0
	Total	3063	62.1	100.0	
Missing	System	1870	37.9		
Total		4933	100.0		

Gov 47b is ordinal because the values are categorical and ranked. The median and the mode are considered appropriate measures of central tendency for ordinal level data. In this case, it seems the mode would be more appropriate because the values are ranked and the middle value (median) will always be the same no matter how many cases fall outside the middle value. According to the *Social Statistics* SPSS textbook, the range and the quartiles are useful measure of dispersion for ordinal level data. The range is the distance between the smallest and largest values in a set of data while providing a context for interpreting other measures of dispersion like the interquartile range and standard deviation. The IQR tells us about the distribution of responses in relation to the median; where the smaller the IQR the more homogenous the distribution and the larger the IQR the more heterogeneous the distribution.

**GOV47c. your local government ...How much, if at all, has the Internet improved the way you interact with... (INSERT; ROTATE) – a lot, some, only a little, or not at all?**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	A lot	549	11.1	17.9	17.9
	Some	612	12.4	20.0	37.9
	Only a little	551	11.2	18.0	55.9
	Not at all	1164	23.6	38.0	93.9
	(VOL) Don't use Internet to contact government	152	3.1	5.0	98.9
	Don't know/Refused	35	.7	1.1	100.0
	Total	3063	62.1	100.0	
Missing	System	1870	37.9		
Total		4933	100.0		

Gov 47c is ordinal because the values are categorical and ranked. The median and the mode are considered appropriate measures of central tendency for ordinal level data. In this case, it seems the mode would be more appropriate because the values are ranked and

the middle value (median) will always be the same no matter how many cases fall outside the middle value. According to the *Social Statistics* SPSS textbook, the range and the quartiles are useful measure of dispersion for ordinal level data. The range is the distance between the smallest and largest values in a set of data while providing a context for interpreting other measures of dispersion like the interquartile range and standard deviation. The IQR tells us about the distribution of responses in relation to the median; where the smaller the IQR the more homogenous the distribution and the larger the IQR the more heterogeneous the distribution.

1e.  
**Statistics**

AGE What is your age?

N	Valid	4933
	Missing	0
Mean		46.57
Std. Error of Mean		.268
Median		44.00
Mode		40
Std. Deviation		18.816
Skewness		.566
Std. Error of Skewness		.035
Kurtosis		-.223
Std. Error of Kurtosis		.070
Range		81
Percentiles	25	32.00
	50	44.00
	75	59.00

AGE What is your age?

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 18	150	3.0	3.0	3.0
19	87	1.8	1.8	4.8
20	81	1.6	1.6	6.4
21	73	1.5	1.5	7.9
22	66	1.3	1.3	9.3
23	72	1.5	1.5	10.7
24	69	1.4	1.4	12.1
25	91	1.8	1.8	14.0
26	82	1.7	1.7	15.6
27	90	1.8	1.8	17.4
28	67	1.4	1.4	18.8
29	82	1.7	1.7	20.5
30	114	2.3	2.3	22.8
31	88	1.8	1.8	24.6
32	88	1.8	1.8	26.3

33	76	1.5	1.5	27.9
34	87	1.8	1.8	29.6
35	87	1.8	1.8	31.4
36	88	1.8	1.8	33.2
37	94	1.9	1.9	35.1
38	99	2.0	2.0	37.1
39	105	2.1	2.1	39.2
40	176	3.6	3.6	42.8
41	98	2.0	2.0	44.8
42	112	2.3	2.3	47.1
43	97	2.0	2.0	49.0
44	71	1.4	1.4	50.5
45	99	2.0	2.0	52.5
46	84	1.7	1.7	54.2
47	85	1.7	1.7	55.9
48	109	2.2	2.2	58.1
49	85	1.7	1.7	59.8
50	134	2.7	2.7	62.5
51	63	1.3	1.3	63.8
52	104	2.1	2.1	65.9
53	84	1.7	1.7	67.6
54	83	1.7	1.7	69.3
55	80	1.6	1.6	70.9
56	72	1.5	1.5	72.4
57	59	1.2	1.2	73.6
58	68	1.4	1.4	75.0
59	48	1.0	1.0	75.9
60	97	2.0	2.0	77.9
61	32	.6	.6	78.5
62	68	1.4	1.4	79.9
63	42	.9	.9	80.8
64	56	1.1	1.1	81.9
65	41	.8	.8	82.7
66	46	.9	.9	83.7
67	48	1.0	1.0	84.6
68	39	.8	.8	85.4
69	37	.8	.8	86.2
70	49	1.0	1.0	87.2
71	44	.9	.9	88.0
72	42	.9	.9	88.9
73	34	.7	.7	89.6
74	52	1.1	1.1	90.7
75	40	.8	.8	91.5
76	28	.6	.6	92.0
77	34	.7	.7	92.7
78	43	.9	.9	93.6
79	29	.6	.6	94.2

80	51	1.0	1.0	95.2
81	29	.6	.6	95.8
82	26	.5	.5	96.3
83	22	.4	.4	96.8
84	12	.2	.2	97.0
85	13	.3	.3	97.3
86	11	.2	.2	97.5
87	8	.2	.2	97.7
88	10	.2	.2	97.9
89	3	.1	.1	97.9
90	6	.1	.1	98.0
91	4	.1	.1	98.1
92	3	.1	.1	98.2
93	1	.0	.0	98.2
94	1	.0	.0	98.2
95	3	.1	.1	98.3
Don't know	1	.0	.0	98.3
Refused	83	1.7	1.7	100.0
Total	4933	100.0	100.0	

Age is a ratio because it is continuous with equal distances between each value that are proportional to one another. The mean, median, and mode can apply as appropriate measures of central tendency for numerical, interval/ratio data. All 3 can be employed for a full picture of the data, but the mean would probably be the most appropriate. The standard deviation is the most appropriate measure of dispersion as it is an important complement to the mean. A mean without the SD has limited value as we can't gauge the volatility of the value.

1f.

**Statistics**

EDUC What is the last grade or class you completed in school?

N	Valid	4933
	Missing	0
Median		4.00
Mode		3
Skewness		.119
Std. Error of Skewness		.035
Kurtosis		-.946
Std. Error of Kurtosis		.070
Range		8
Percentiles	25	3.00
	50	4.00
	75	6.00

**EDUC What is the last grade or class you completed in school?**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid None, or grades 1-8	167	3.4	3.4	3.4
High school incomplete (grades 9-11)	546	11.1	11.1	14.5
High school graduate (grade 12 or GED certificate)	1544	31.3	31.3	45.8
Business, Technical, or vocational school AFTER high school	233	4.7	4.7	50.5
Some college, no 4-year degree	1151	23.3	23.3	73.8
College graduate (B.S., B. A., or other 4-year degree)	801	16.2	16.2	90.1
Post-graduate training/professional school after college (Ma	467	9.5	9.5	99.5
Don't know/Refused	23	.5	.5	100.0
Total	4933	100.0	100.0	

Education is ordinal because the values are categorical and ranked. The median and the mode are considered appropriate measures of central tendency for ordinal level data. In this case, it seems the mode would be more appropriate because the values are ranked and the middle value (median) will always be the same no matter how many cases fall outside the middle value. According to the *Social Statistics* SPSS textbook, the range and the quartiles are useful measure of dispersion for ordinal level data. The range is the distance between the smallest and largest values in a set of data while providing a context for interpreting other measures of dispersion like the interquartile range and standard deviation. The IQR tells us about the distribution of responses in relation to the median; where the smaller the IQR the more homogenous the distribution and the larger the IQR the more heterogeneous the distribution.

1g.  
**Statistics**

MODEM Does the modem you use at home connect through a standard telephone line, or do you also have some other type of connection?

N	Valid	2712
	Missing	2220
Mode		1

**MODEM Does the modem you use at home connect through a standard telephone line, or do you also have some other type of connection?**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Standard telephone line	1803	36.6	66.5	66.5
DSL-enabled phone line	297	6.0	10.9	77.4

	Cable modem	516	10.5	19.0	96.5
	Wireless connection	24	.5	.9	97.4
	T-1 or fiber optic connection	9	.2	.3	97.7
	Other (MAKE SURE NOT ONE OF ABOVE)	27	.6	1.0	98.7
	Don't know/Refused	35	.7	1.3	100.0
	Total	2712	55.0	100.0	
Missing	System	2220	45.0		
Total		4933	100.0		

MODEM is nominal because the values are categorical variables with no numerical meaning or ranking. The only appropriate measure of central tendency is the mode on nominal level data as it is impossible to calculate the median and mean (as it is not possible to assign numerical scores to the cases). According to the *Social Statistics SPSS* textbook, there is no dispersion statistic for nominal level data, however, a frequency distribution can help in assessing the variability as it will tell us the percentage of respondents in each category.

1h.  
**Statistics**

POLAF In politics TODAY, do you consider yourself a Republican, Democrat, or Independent?

N	Valid	4933
	Missing	0
Mode		1

**POLAF In politics TODAY, do you consider yourself a Republican, Democrat, or Independent?**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Republican	1505	30.5	30.5	30.5
	Democrat	1457	29.5	29.5	60.0
	Independent	1476	29.9	29.9	90.0
	No party/Not interested in politics (VOL.)	273	5.5	5.5	95.5
	Other party (VOL.)	22	.4	.4	95.9
	Don't know/Refused	201	4.1	4.1	100.0
	Total	4933	100.0	100.0	

POLAF is nominal because the values are categorical variables with no numerical meaning or ranking. The only appropriate measure of central tendency is the mode on nominal level data as it is impossible to calculate the median and mean (as it is not possible to assign numerical scores to the cases). According to the *Social Statistics SPSS* textbook, there is no dispersion statistic for nominal level data, however, a frequency distribution can help in assessing the variability as it will tell us the percentage of respondents in each category.

1i.  
**Statistics**

Race/Ethnicity

N	Valid	4851
	Missing	82
Mode		1

Race/Ethnicity

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	White~H	3602	73.0	74.2	74.2
	Black~H	513	10.4	10.6	84.8
	Hispanic	495	10.0	10.2	95.0
	other	242	4.9	5.0	100.0
	Total	4851	98.3	100.0	
Missing	DK/ref	82	1.7		
Total		4933	100.0		

RACE is nominal because the values are categorical variables with no numerical meaning or ranking. The only appropriate measure of central tendency is the mode on nominal level data as it is impossible to calculate the median and mean (as it is not possible to assign numerical scores to the cases). According to the *Social Statistics* SPSS textbook, there is no dispersion statistic for nominal level data, however, a frequency distribution can help in assessing the variability as it will tell us the percentage of respondents in each category.

1j.

Statistics

DIS Does any disability, handicap, or chronic disease keep you from participating fully in work, school, housework, or other activities, or not?

N	Valid	4933
	Missing	0
Mode		2

**DIS Does any disability, handicap, or chronic disease keep you from participating fully in work, school, housework, or other activities, or not?**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	687	13.9	13.9	13.9
	No	4229	85.7	85.7	99.7
	Don't Know/Refused	16	.3	.3	100.0
	Total	4933	100.0	100.0	

DIS is nominal because the values are categorical variables with no numerical meaning or ranking. The only appropriate measure of central tendency is the mode on nominal level data as it is impossible to calculate the median and mean (as it is not possible to assign numerical scores to the cases). According to the *Social Statistics* SPSS textbook, there is no dispersion statistic for nominal level data, however, a frequency distribution can help in assessing the variability as it will tell us the percentage of respondents in each category.

2.

**Statistics**

Gov28 Recoded

N	Valid	4933
	Missing	0
Mean		.16
Median		.00
Mode		0
Range		3
Percentiles	25	.00
	50	.00
	75	.00

**Gov28 Recoded**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 0	4167	84.5	84.5	84.5
1	741	15.0	15.0	99.5
2	19	.4	.4	99.9
3	5	.1	.1	100.0
Total	4933	100.0	100.0	

**Descriptive Statistics**

	N	Minimum	Maximum	Mean	Std. Deviation
Gov28 Recoded	4933	0	3	.16	.386
Valid N (listwise)	4933				

Recoding gov28 1 through gov28 7 gives very specific data on the types of access respondents have to the internet. If I want to examine the data in terms of whether or not people are using the internet, I can recode the data to help us see patterns more clearly.

3.

Contingency tables or cross-tabulations are the primary method for examining relationships between variables measured at the nominal and ordinal levels. (*Applied Statistics for Public and Nonprofit Administration*, 243-44) In order to conduct the cross-tabulations, I need to recode age as the ratio level data would be unwieldy and it becomes harder to see the patterns in the responses. This allows us to see if there is a statistical relationship between any of the variables. (*Social Statistics*, 119)

Measures of association are useful for examining statistical significance mainly for these relationships. These tests assess whether the relationship observed in the cross-tabulation in the sample of data is sufficiently strong enough to infer that a relationship likely exists in the full population. The test procedure assumes that there is no relationship between the variables in the population and determines whether any apparent relationship is

obtained in a sample cross-tabulation is attributable to chance. (*Applied Statistics for Public and Nonprofit Administration*, 259-60)

*Gov 11 with AGE, EDUC, Recoded EDUC, MODEM, POLAF, RACE and DIS*

The cross-tabulation between GOV 11 and AGE, EDUC, Recoded EDUC, and RACE does not seem significant upon examining the percentages. This conclusion is supported by the low numbers produced by examining the Lambda and Gamma measures of association for the ordinal level variables (Gov 14, AGE, EDUC, Recoded EDUC) and Phi and Cramer's V for the nominal level variables (RACE).

Gov 11 and MODEM, POLAF and DIS upon first glance, appeared to have significant percentage differences. However, upon examining Phi and Cramer's V for each, it was clear measures reflect relatively small numbers meaning no relationship can be applied from the sample to the larger population for:

**MODEM**

**Symmetric Measures**

		Value	Asymp. Std. Error(a)	Approx. T(b)	Approx. Sig.
Nominal by Nominal	Phi	.218			.000
	Cramer's V	.089			.000
	Contingency Coefficient	.213			.000
Ordinal by Ordinal	Kendall's tau-b	-.033	.018	-1.819	.069
	Kendall's tau-c	-.025	.014	-1.819	.069
	Gamma	-.051	.028	-1.819	.069
N of Valid Cases		2197			

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

**POLAF**

**Symmetric Measures**

		Value	Asymp. Std. Error(a)	Approx. T(b)	Approx. Sig.
Nominal by Nominal	Phi	.178			.000
	Cramer's V	.079			.000
	Contingency Coefficient	.175			.000
Ordinal by Ordinal	Kendall's tau-b	-.010	.015	-.628	.530
	Kendall's tau-c	-.009	.014	-.628	.530
	Gamma	-.013	.020	-.628	.530
N of Valid Cases		2946			

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

and DIS.

**Symmetric Measures**

		Value	Asymp. Std. Error(a)	Approx. T(b)	Approx. Sig.
Nominal by Nominal	Phi	.152			.000
	Cramer's V	.108			.000
	Contingency Coefficient	.151			.000
Ordinal by Ordinal	Kendall's tau-b	-.054	.017	-3.165	.002
	Kendall's tau-c	-.034	.011	-3.165	.002
	Gamma	-.123	.038	-3.165	.002
N of Valid Cases		2942			

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

*Gov 14 with AGE, EDUC, Recoded EDUC, MODEM, POLAF, RACE and DIS*

The cross-tabulation between GOV 14 and AGE, EDUC, Recoded EDUC, MODEM, and POLAF does not seem significant upon examining the percentages. This conclusion is supported by the low numbers produced by examining the Lambda and Gamma measures of association for the ordinal level variables (Gov 14, AGE, EDUC, Recoded EDUC) and Phi and Cramer's V for the nominal level variables (MODEM, POLAF).

Gov 14 and RACE and DIS upon first glance, appeared to have significant percentage differences. However, upon examining Phi and Cramer's V for RACE

**Symmetric Measures**

		Value	Asymp. Std. Error(a)	Approx. T(b)	Approx. Sig.
Nominal by Nominal	Phi	.129			.000
	Cramer's V	.074			.000
Ordinal by Ordinal	Gamma	-.043	.039	-1.106	.269
N of Valid Cases		2770			

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

and for DIS

**Symmetric Measures**

		Value	Asymp. Std. Error(a)	Approx. T(b)	Approx. Sig.
Nominal by Nominal	Phi	.126			.000
	Cramer's V	.089			.000
Ordinal by Ordinal	Gamma	.219	.048	4.290	.000
N of Valid Cases		2771			

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

(both nominal level data) it's clear there is little association between the variables.

*Gov 17 with AGE, EDUC, Recoded EDUC, MODEM, POLAF, RACE and DIS*

The cross-tabulation between GOV 17 and AGE, EDUC, Recoded EDUC, MODEM, POLAF and RACE does not seem significant upon examining the percentages. In fact, I was surprised that the percentages within each category were almost uniformly distributed with almost no percentage differences. This conclusion is supported by the low numbers produced by examining the Lambda and Gamma measures of association for the ordinal level variables (Gov 14, AGE, EDUC, Recoded EDUC) and Phi and Cramer's V for the nominal level variables (MODEM, POLAF, RACE).

Gov 17 and DIS upon first glance, appeared to have significant percentage differences. However, upon examining Phi and Cramer's V for DIS, it's clear the difference is meaningless.

**Symmetric Measures**

		Value	Asymp. Std. Error(a)	Approx. T(b)	Approx. Sig.
Nominal by Nominal	Phi	.225			.000
	Cramer's V	.159			.000
	Contingency Coefficient	.220			.000
Ordinal by Ordinal	Kendall's tau-b	-.033	.018	-1.852	.064
	Kendall's tau-c	-.014	.008	-1.852	.064
	Gamma	-.123	.069	-1.852	.064
N of Valid Cases		2771			

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

*Gov 24 with AGE, EDUC, Recoded EDUC, MODEM, POLAF, RACE and DIS*

The cross-tabulation between GOV 11 and AGE, EDUC, Recoded EDUC, and POLAF does not seem significant upon examining the percentages. This conclusion is supported by the low numbers produced by examining the Lambda and Gamma measures of association for the ordinal level variables (Gov 14, AGE, EDUC, Recoded EDUC) and Phi and Cramer's V for the nominal level variables (POLAF).

Gov 24 and MODEM upon first glance, appeared to have significant percentage differences. However, upon examining Phi and Cramer's V for MODEM, it's clear the difference is meaningless.

**Symmetric Measures**

		Value	Asymp. Std. Error(a)	Approx. T(b)	Approx. Sig.
Nominal by Nominal	Phi	.136			.047
	Cramer's V	.068			.047
	Contingency Coefficient	.134			.047
Ordinal by Ordinal	Kendall's tau-b	-.020	.020	-1.001	.317
	Kendall's tau-c	-.015	.015	-1.001	.317
	Gamma	-.034	.034	-1.001	.317
N of Valid Cases		1993			

- a Not assuming the null hypothesis.
- b Using the asymptotic standard error assuming the null hypothesis.

As is the case with RACE,

**Symmetric Measures**

		Value	Asymp. Std. Error(a)	Approx. T(b)	Approx. Sig.
Nominal by Nominal	Phi	.192			.000
	Cramer's V	.111			.000
	Contingency Coefficient	.189			.000
Ordinal by Ordinal	Kendall's tau-b	.131	.018	7.301	.000
	Kendall's tau-c	.091	.012	7.301	.000
	Gamma	.244	.032	7.301	.000
N of Valid Cases		2599			

- a Not assuming the null hypothesis.
- b Using the asymptotic standard error assuming the null hypothesis.

and DIS.

**Symmetric Measures**

		Value	Asymp. Std. Error(a)	Approx. T(b)	Approx. Sig.
Nominal by Nominal	Phi	.139			.000
	Cramer's V	.099			.000
	Contingency Coefficient	.138			.000
Ordinal by Ordinal	Kendall's tau-b	-.085	.019	-4.439	.000
	Kendall's tau-c	-.050	.011	-4.439	.000
	Gamma	-.212	.045	-4.439	.000
N of Valid Cases		2653			

- a Not assuming the null hypothesis.
- b Using the asymptotic standard error assuming the null hypothesis.

B.1. Upon first running the frequency analysis, it is clear that some of the data needs to be recoded in order to be analyzed for any meaning. GENDER is a nominal level variable making the mode the most relevant measure of central tendency and leaving us without an accurate measure of dispersion. However, when the data was coded f=female and m=male, making it impossible to use measures of central tendency or dispersion. We need to recode these values into numbers. DATE OF BIRTH is interval level data, but by using dates rather than age or range of ages I get inaccurate measures of central tendency. Percentile (50% born before 1962; 75% born before 1965) seems to be the only measure of dispersion that seems somewhat accurate, but the rest are inaccurate. I need to recode this into age or year of birth. Once this variable is recoded, I can use the mean as a measure of central tendency (assuming I allow it to remain interval level data) and I can use standard deviation as a measure of dispersion.

EDUCATION LEVEL in years is interval level data with mean (13.49 years) as the best measure of central tendency and standard deviation (2.89 years) for measure of dispersion.

EMPLOYEE CATAGORY is nominal level data leaving us with mode (employment catagory 1=clerical) as a central measure and no accurate measure of dispersion.

CURRENT SALARY and BEGINNING SALARY are both interval level data, but the numbers are so vast it might be easier to look at them after I recode into ordinal level ranges. As interval level data, I can use mean as a measure of central tendency and standard deviation as a measure of dispersion, but if I recode to ordinal level data, then a median or mode might better suit as a measure of central tendency with quartiles and range as a good measure of dispersion. Current salary mean is \$34,419.57 with SD of \$17,075.66. The standard error of the mean is \$784.31 which means it differs by as much from the population mean. The SD is so high because 90% of the salary range falls below \$60,000 while the remaining 10% ranges up to a max of \$135,000—making for a positive skewness of 2.125 indicating that the mean is artificially high. If it was unskewed the value would be closer to 0 (with a normal distribution). The variance is also ridiculously high. Beginning salary mean is \$17,016.09 with SD equaling \$7,870.64. Although the SD seems to be a low number, the skewness is quite high at 2.853 indicating the mean is skewed higher. The standard error is \$361.51 meaning it differs by this amount from the population mean.

MONTHS SINCE HIRE and PREVIOUS EXPERIENCE IN MONTHS are both interval level data and I would need to look at mean for a measure of central tendency and standard deviation as a measure of dispersion. Months since hire mean is 81.11 with SD as 10.06. The skewness is only -.053 meaning the data falls within a normal distribution. Previous experience in months mean is 95.86 with SD equaling 104.59. The standard error is 4.80, which seems quite low given the range of months (0-476). Clearly, there is a problem as shown by the skewness equaling 1.510 and a ridiculously high variance equaling 10938.28. This is because 90% of the respondents have less than 261 months, while 10% range up to 476 months.

MINORITY CLASSIFICATION is nominal level data and I can examine standard deviation for a measure of dispersion. The mode of the respondents is ‘not minority.’

Statistics before recoding:

Statistics											
		Gender	Date of Birth	Educational Level (years)	Employment Category	Current Salary	Beginning Salary	Months since Hire	Previous Experience (months)	Minority Classification	
N	Valid	474	474	474	474	474	474	474	474	474	
	Missing	0	1	0	0	0	0	0	0	0	
Mean			10/08/1956	13.49	1.41	\$34,419.57	\$17,016.09	81.11	95.86	.22	
Std. Error of Mean			197 19:42	.133	.036	\$784.311	\$361.510	.462	4.804	.019	
Median			01/23/1962	12.00	1.00	\$28,875.00	\$15,000.00	81.00	55.00	.00	
Mode			02/04/1934 <sup>a</sup>	12	1	\$30,750	\$15,000	81 <sup>a</sup>	0	0	
Std. Deviation			4302 07:54	2.885	.773	\$17,075.661	\$7,870.638	10.061	104.586	.414	
Variance			1.38E+017	8.322	.598	291578214.5	61946945	101.223	10938.281	.172	
Skewness				-.856	-.114	1.456	2.125	2.853	-.053	1.510	1.360
Std. Error of Skewness				.112	.112	.112	.112	.112	.112	.112	
Kurtosis				-.578	-.265	.268	5.378	12.390	-1.153	1.696	-1.150
Std. Error of Kurtosis				.224	.224	.224	.224	.224	.224	.224	
Range			15340	13	2	\$119,250	\$70,980	35	476	1	
Percentiles	25		12/07/1947	12.00	1.00	\$24,000.00	\$12,450.00	72.00	19.00	.00	
	50		01/23/1962	12.00	1.00	\$28,875.00	\$15,000.00	81.00	55.00	.00	
	75		07/11/1965	15.00	1.00	\$37,162.50	\$17,617.50	90.00	140.00	.00	

a. Multiple modes exist. The smallest value is shown

2. As the means between current salary and beginning salary are so great, it is useful to conduct a t-test in order to see if our pattern in the sample is generalizable to the population from which the sample was drawn. As the variance between current salary and beginning salary are so vast, it seems best to conduct a t-test for unequal variances. Not only are the variances huge, the standard deviations for the two data sets are widely different as well. After conducting the paired sample t-test I gathered the following results:

**Paired Samples Statistics**

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Current Salary	\$34,419.57	474	\$17,075.661	\$784.311
	Beginning Salary	\$17,016.09	474	\$7,870.638	\$361.510

**Paired Samples Correlations**

		N	Correlation	Sig.
Pair 1	Current Salary & Beginning Salary	474	.880	.000

**Paired Samples Test**

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Current Salary - Beginning Salary	\$17403.48	\$10,814.620	\$496.732	\$16,427.407	\$18379.56	35.036	473	.000

The samples seem correlated highly with the value of .880. With a level of significance of .000, this result indicated the probability of the two sample variances are equal is very small therefore the sample variances are not equal, so there is a significant difference between the two samples and that difference is between \$16,427 and \$18,379.

3. A correlational validity establishes when an indicator correlated strongly to other accepted valid indicators. Pearson’s r is one very commonly used statistic for assessing the strength and direction of an association between two variables. It evaluates the extent to which the responses to two variables tend to vary together. In looking at the outputs from the SPSS results of a correlation analysis, we see that the starred (\*\*) results are all significant at the .01 level (two-tailed).

The following is a discussion of the most significant correlations:

When examining the value of Pearson’s r for CURRENT SALARY and BEGINNING SALARY we get a value of .88. This means that BEGINNING SALARY explains 88% of the variance in CURRENT SALARY as the two are positively and highly correlated because as beginning salary increases so does current salary, and vice versa. When examining the value of Pearson’s r for EDUCATIONAL LEVEL and BEGINNING SALARY we get a value of .633. This means that BEGINNING SALARY explains 63% of the variance in EDUCATIONAL LEVEL as the two are positively correlated because

as beginning salary increases so does educational level, and vice versa. When examining the value of Pearson's  $r$  for EMPLOYMENT CATEGORY and BEGINNING SALARY we get a value of .755. This means that BEGINNING SALARY explains 76% of the variance in EMPLOYMENT CATEGORY as the two are positively and highly correlated because as beginning salary increases so does employment category and vice versa.

When examining the value of Pearson's  $r$  for CURRENT SALARY and EDUCATIONAL LEVEL we get a value of .661. This means that CURRENT SALARY explains 66% of the variance in EDUCATIONAL LEVEL as the two are positively correlated because as current salary increases so does educational level, and vice versa. When examining the value of Pearson's  $r$  for CURRENT SALARY and EMPLOYMENT CATEGORY we get a value of .78. This means that CURRENT SALARY explains 78% of the variance in EMPLOYMENT CATEGORY as the two are positively and highly correlated because as current salary increases so does educational level, and vice versa.

When examining the value of Pearson's  $r$  for EDUCATIONAL LEVEL and EMPLOYMENT CATEGORY we get a value of .514. This means that EDUCATIONAL LEVEL explains 66% of the variance in EMPLOYMENT CATEGORY as the two are positively, but only slightly correlated.

No strong correlations:

When examining the value of Pearson's  $r$  for PREVIOUS EXPERIENCE and all the other variables we see that it has weak, not significant correlations at a .01 level (two-tailed). This is surprising because one would expect that previous experience would have a correlation with variables like beginning salary, current salary, and position category. One would expect a correlation as it's easy to assume that previous experience would have some relationship with these variables. Another surprise is MONTHS OF HIRE has no strong correlation in any direction. MINORITY CLASSIFICATION does not seem to be strongly correlated with any of the variables, if only slightly negatively correlated. This indicates that minority classification does have a slight inverse relationship with salary, education or employment category. If one is minority, for example, beginning salary, current salary, educational level and employment category go down.

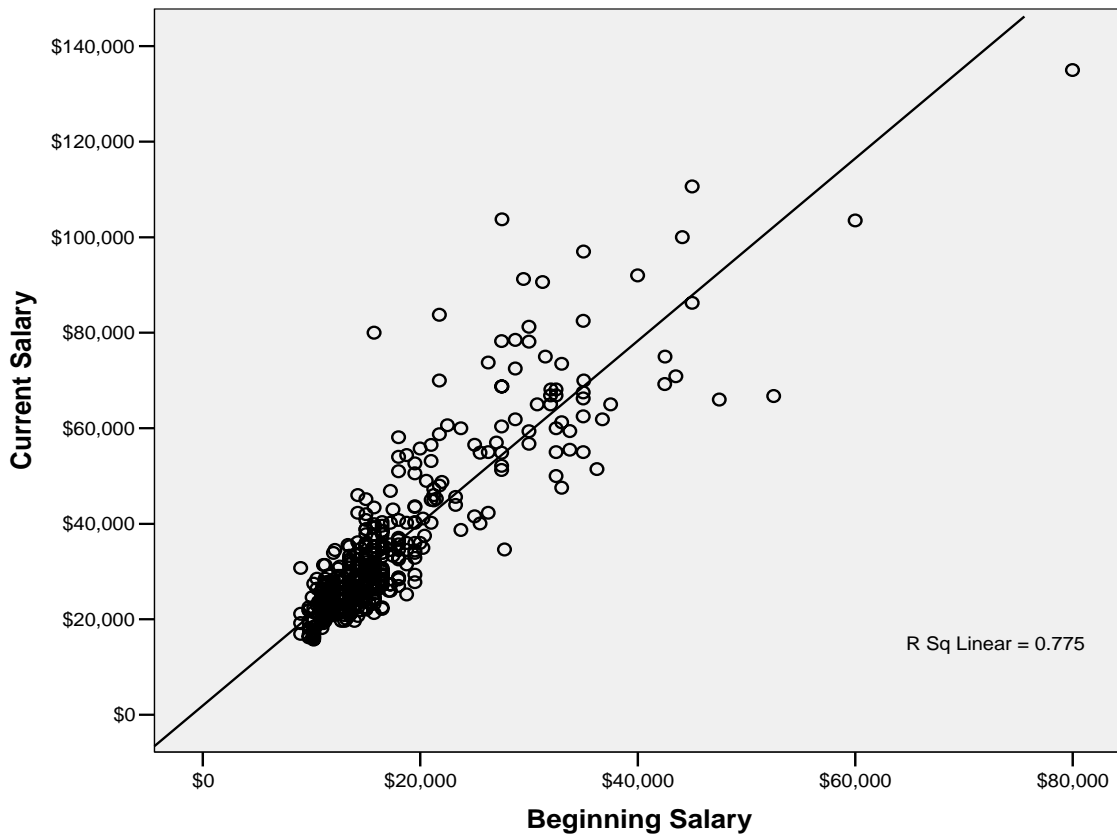
**Correlations**

		Beginning Salary	Current Salary	Previous Experience (months)	Educational Level (years)	Employment Category	Months since Hire	Minority Classification
Beginning Salary	Pearson Correlation	1	.880**	.045	.633**	.755**	-.020	-.158**
	Sig. (2-tailed)		.000	.327	.000	.000	.668	.001
	N	474	474	474	474	474	474	474
Current Salary	Pearson Correlation	.880**	1	-.097*	.661**	.780**	.084	-.177**
	Sig. (2-tailed)	.000		.034	.000	.000	.067	.000
	N	474	474	474	474	474	474	474
Previous Experience (months)	Pearson Correlation	.045	-.097*	1	-.252**	.063	.003	.145**
	Sig. (2-tailed)	.327	.034		.000	.173	.948	.002
	N	474	474	474	474	474	474	474
Educational Level (years)	Pearson Correlation	.633**	.661**	-.252**	1	.514**	.047	-.133**
	Sig. (2-tailed)	.000	.000	.000		.000	.303	.004
	N	474	474	474	474	474	474	474
Employment Category	Pearson Correlation	.755**	.780**	.063	.514**	1	.005	-.144**
	Sig. (2-tailed)	.000	.000	.173	.000		.908	.002
	N	474	474	474	474	474	474	474
Months since Hire	Pearson Correlation	-.020	.084	.003	.047	.005	1	.050
	Sig. (2-tailed)	.668	.067	.948	.303	.908		.282
	N	474	474	474	474	474	474	474
Minority Classification	Pearson Correlation	-.158**	-.177**	.145**	-.133**	-.144**	.050	1
	Sig. (2-tailed)	.001	.000	.002	.004	.002	.282	
	N	474	474	474	474	474	474	474

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

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

4.



### Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.880(a)	.775	.774	\$8,115.356

a Predictors: (Constant), Beginning Salary

From the regression analysis, the initial .880 correlation is confirmed by the .775 R Square correlation as indicated above. Beginning Salary explains current salary to 78%. The scatterplot best fit line demonstrates this relationship graphically indicating that most of the data rests fairly close to the line and the few outliers are in the upper salary ranges.

### Coefficients<sup>a</sup>

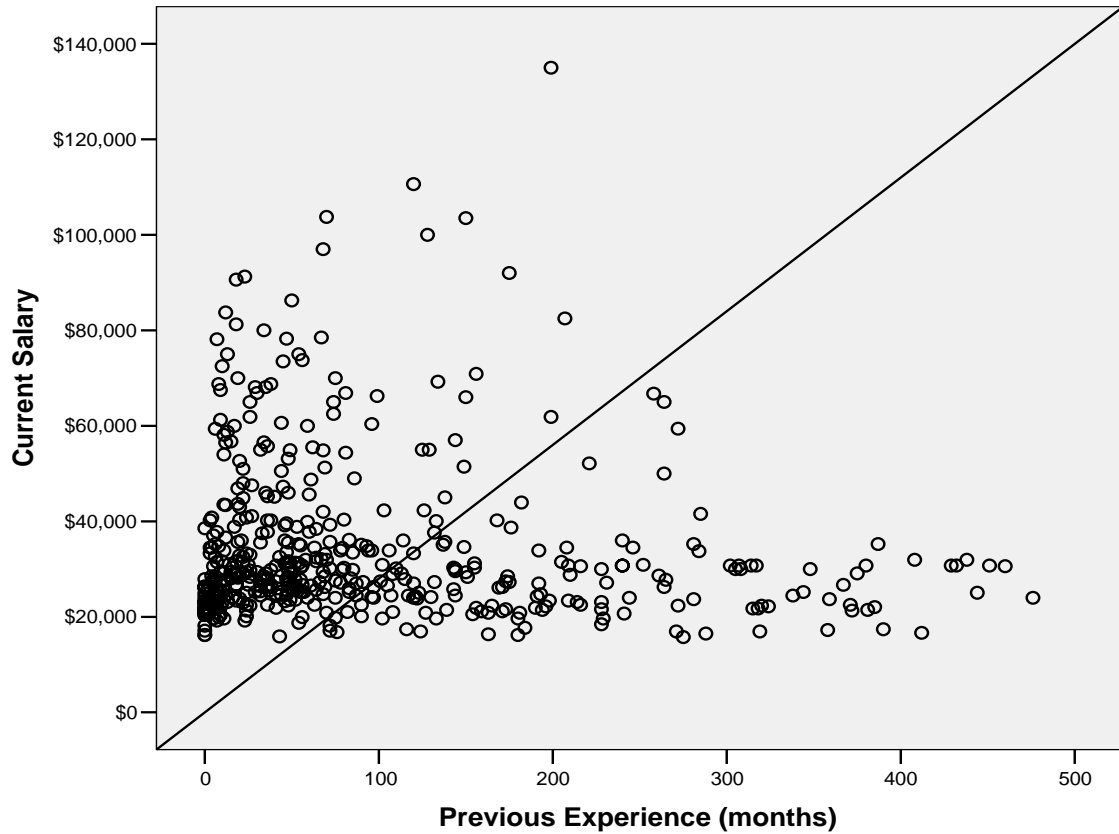
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	1928.206	888.680		2.170	.031	181.947	3674.464
	Beginning Salary	1.909	.047	.880	40.276	.000	1.816	2.003

a. Dependent Variable: Current Salary

Equation for the Slope:  $Y=a+bX$

$CURRENT\ SALARY = \$1928.206 + 1.909 * BEGINNING\ SALARY$

If current salary is \$0, then beginning salary would be \$1928.21. For every dollar in current salary is almost twice as much (1.909) as the beginning salary. In examining the slope, we can hypothesize \$20,000 current salary indicates a beginning salary of \$10,000.

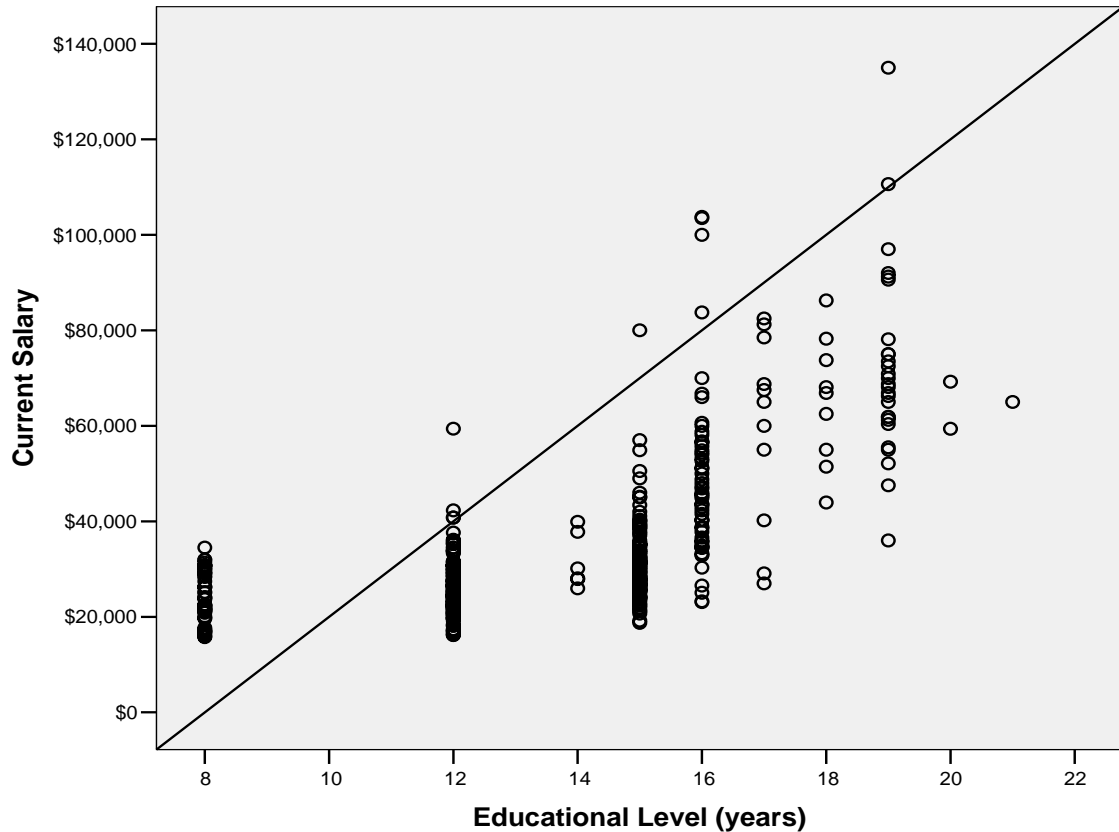


**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.097(a)	.009	.007	\$17,012.353

a Predictors: (Constant), Previous Experience (months)

From the regression analysis, we can see there is little correlation between salary and previous experience (in months). The scatterplot best fit line is barely a fit as the data points are all over the place and don't cluster around the line—under 100 months and \$40,000 salary. As the best fit line does not fit well, it's not useful to examine the slope.

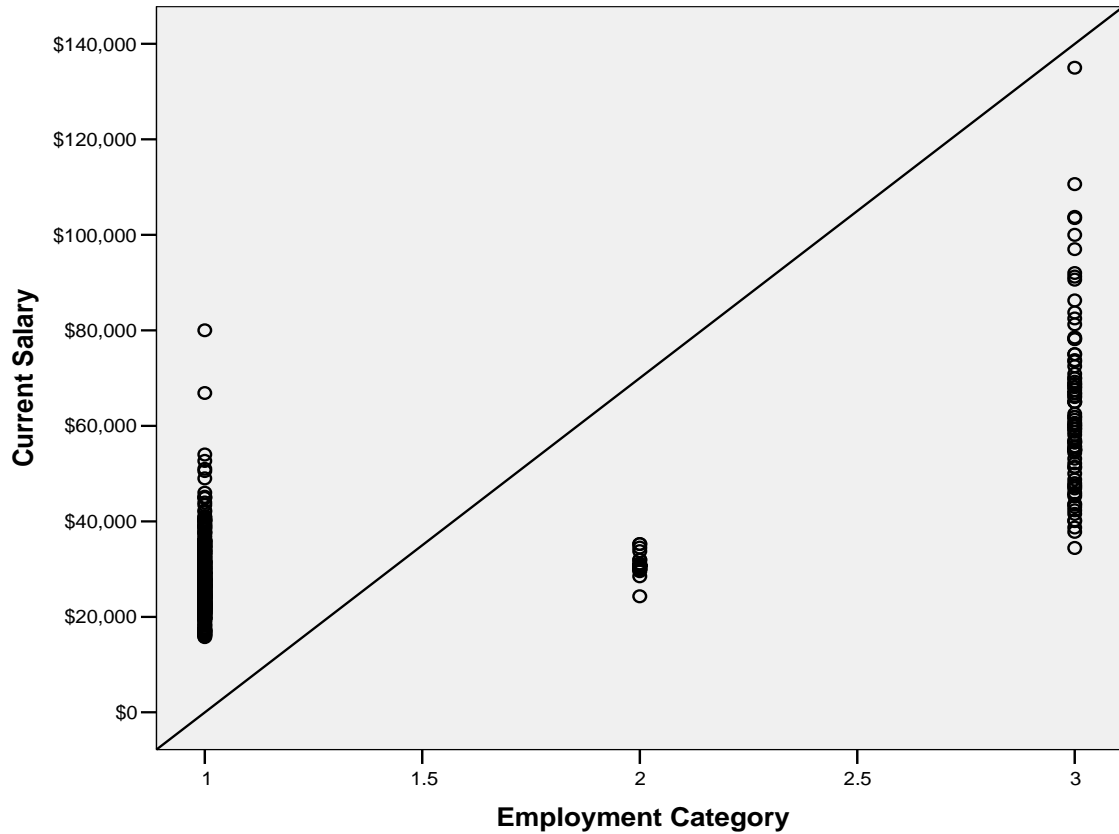


**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.661(a)	.436	.435	\$12,833.540

a Predictors: (Constant), Educational Level (years)

While the R in the model summary indicates a correlation of .661 between Current Salary and Educational level, the R squared offers a much lower correlation number of .436. This low correlation is confirmed by the scatterplot as most of the data does not cluster around the best fit line. This type of data is not well formatted for regression analysis as one of the variables (Education) is ordinal level data even though it lists here in years. As the best fit line does not fit well, it's not useful to examine the slope.

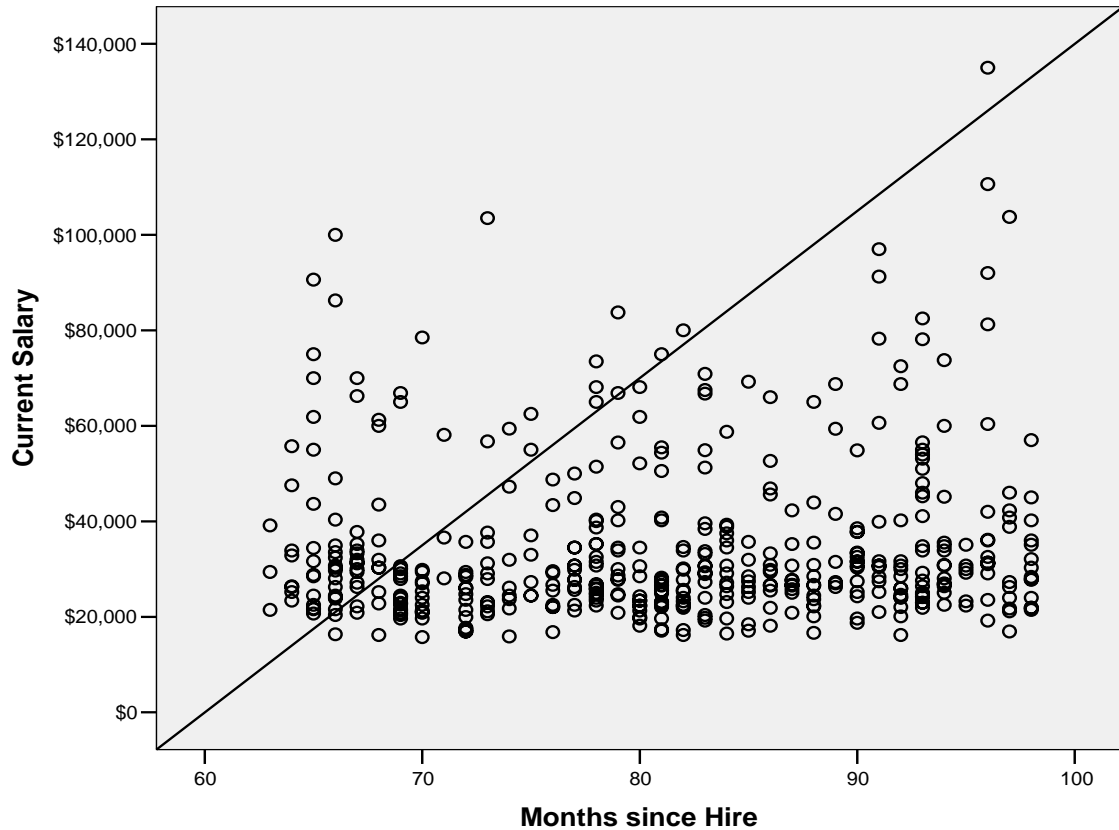


**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.780(a)	.609	.608	\$10,694.465

a Predictors: (Constant), Employment Category

The R number indicates a strong correlation between Current Salary and Employment Category and the R Square number confirms a slight correlation at .609. However, the scatterplot tells a different story as the best fit line does not come close to covering all the data. This is because Employment Category is ordinal level data which does not lend itself to regression analysis. As the best fit line does not fit well, it's not useful to examine the slope.

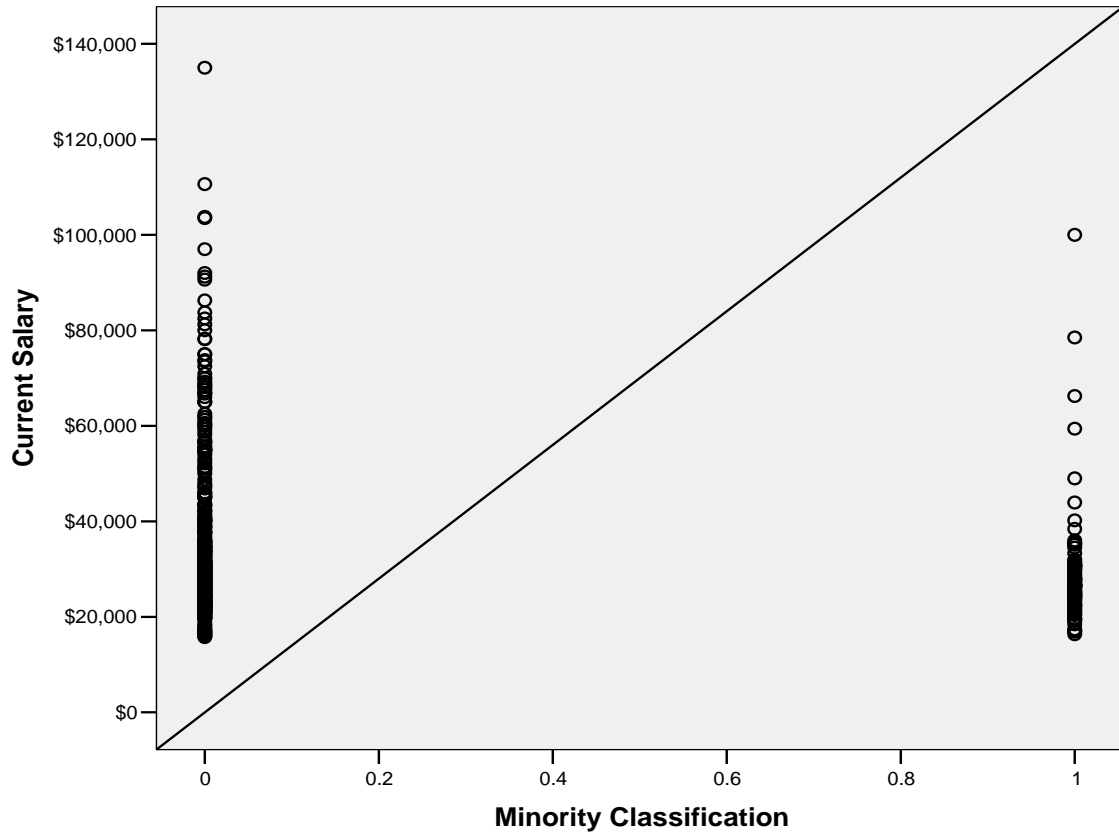


**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.084(a)	.007	.005	\$17,033.194

a Predictors: (Constant), Months since Hire

As both the model summary (R and R Square) and the scatterplot indicate, there is little correlation between current salary and months since hire. Despite what would seem like a reasonable hypothesis that a correlation would exist between the two variables, we can see from this regression analysis that it does not. Both variables are interval level data making it appropriate to use a regression. As the best fit line does not fit well, it's not useful to examine the slope.



**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.177(a)	.031	.029	\$16,822.807

a Predictors: (Constant), Minority Classification

As both the model summary (R and R Square) and the scatterplot indicate, there is little correlation between current salary and minority classification. As minority classification is nominal level data it is not appropriate to use a regression. As the best fit line does not fit well, it's not useful to examine the slope.

5.

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.917(a)	.841	.839	\$6,858.979

a Predictors: (Constant), Minority Classification, Months since Hire, Educational Level (years), Previous Experience (months), Employment Category, Beginning Salary

In examining the multiple regression of Current Salary against the independent variables listed above, we see that R of .917 and R Square of .841 indicate that the variables all together are highly correlated to current salary. These variables explain current salary by 84% and other variables account for 16% of the variance in current salary.

Coefficients<sup>a</sup>

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		
	B	Std. Error	Beta			Lower Bound	Upper Bound	
1	(Constant)	-14952.7	2995.233		-4.992	.000	-20838.492	-9066.887
	Educational Level (years)	541.455	152.799	.091	3.544	.000	241.196	841.713
	Employment Category	5834.279	625.873	.264	9.322	.000	4604.402	7064.155
	Beginning Salary	1.361	.069	.627	19.677	.000	1.225	1.497
	Months since Hire	155.941	31.509	.092	4.949	.000	94.024	217.858
	Previous Experience (months)	-19.139	3.288	-.117	-5.821	.000	-25.600	-12.678
	Minority Classification	-655.186	782.000	-.016	-.838	.403	-2191.859	881.488

a. Dependent Variable: Current Salary

Equation for the Slope:  $Y=a+bX$

CURRENT SALARY= $(-14952.7)$

$+(541.455*\text{EDUCATIONAL LEVEL})$

$+(5834.279*\text{EMPLOYMENT CATEGORY})$

$+(1.361*\text{BEGINNING SALARY})$

$+(155.941*\text{MONTHS SINCE HIRE})$

$+(-19.139*\text{PREVIOUS EXPERIENCE})$

$+(-655.186*\text{MINORITY CLASSIFICATION})$

If current salary is equal to \$0 then  $b= -\$14953$ .

For every dollar in current salary, then the associated variables increase or decrease by the numerical value assigned to b. This tells us that educational level, employment category, beginning salary, and months since hire will all increase current salary, while previous experience and minority classification will decrease current salary.