OMIS 645

APPLIED BUSINESS ANALYTICS SAS

JOB POSTINGS AND FORTUNE 1000 COMPANIES

VENKATA ANIRUDH THOTA

KALYAN PRANEETH PERALA

SIVANI KALVALAPALLI

TANUSHREE CHUNGI

Introduction:

The theme of this project is job postings in indeed and fortune 1000 companies. The datasets that are used in the project are indeed job posting for data scientist, data analyst and data engineer positions which was then combined with the dataset that corresponds to fortune 1000 companies. The indeed job posting dataset is extracted from the data gov and data world and was further cleansed for our analysis. The second dataset is extracted from the fortune 1000 US companies is extracted from various sites including fortune 1000 information websites.

As a team, we agreed to run analysis on these datasets because these datasets are quite relevant to us and this course, because each of us aspire to land in a job in a fortune 1000 company or at the least in a data analyst related position. The datasets variables such as skills required, and salary sparked more interest. We were also interested in the assets the revenue, the number of employees and locations of these fortune 1000 US companies. In compliance, with the questions that we had as a team, we conducted various analysis and analyzed the results, we have interpreted the results for these questions.

The indeed job posting dataset has columns such as job title, queried salary, number of skills and company. The dataset also has information regarding the number of reviews and the number of stars for that posting, the location of the company and the industry to which the company belongs. Each row that is each level of analysis in this dataset is about a job posting.

The second dataset that is considered is the fortune 1000 companies' dataset. The dataset corresponds to the fortune 1000 US companies, here the level of analysis is a company. The columns in this dataset includes current rank and previous rank of the company, the revenue, profits, assets, market value, number of employees, CEO's name and title.

The dataset also contains information regarding the sector and industry these companies belong to and the city, state and location details including the latitude and longitude. The two datasets that is the fortune 1000 US dataset and the indeed dataset were combined based on the company name, that is the company name is used as the primary key. The analysis and the inferences in the report are extracted from these separate datasets as well as the combination of the two datasets.

Data Cleansing and Enhancement:

The datasets had many empty cells, and these were cleaned. To run the analysis, we also enhanced the dataset, in the fortune 1000 US companies' dataset we added the gender of the chief executive officer. In order to enhance the data and make it simpler to run the analysis we created segregated datasets, using the job type. The job type in the dataset is a categorical variable with three categories., we segregated the three categories made it into three variables with binary categories. In order to make the analysis more efficient, the indeed dataset has a column on salary that is in various ranges. We converted these values in ranges into six different levels and use that in salary levels excel file for analysis. The following are the salary levels: <80000 is considered as level 1, 80000-99999 is level 2, 100000-119999 is level 3, 120000-139999 is level 4, 140000-159999 is level 5 and the last one level 6 is >160000.

While combining the two datasets we used the company name as the primary key to join these two tables, we used the SQL queries to combine these two tables. Redundant information such as the latitude and longitude of the company location in the data sets was eliminated. Some columns such as the name of the CEO in the fortune1000 Us companies were deleted.We combined the two datasets by adding an additional column titled “Is Fortune”. The “Is Fortune” column is a binary categorical, if the value is 1, then the company is a part of the fortune 1000, if the value is 0, then the company is not a part of the fortune 1000.

QUESTIONS, RESULTS AND INTERPRETATIONS:

Question 1: Which job type has more posting in fortune 1000 companies?

Data Set: Indeed and Fortune 1000 Combined

Test: Chi-Square Test of Independence.

Row Variable: Is fortune

Column Variables: Job Type (Data Analyst, Data Engineer, Data Scientist)

$$X_{calc}^{2} = \Sigma^{r}_{f=1}\Sigma^{c}_{k=1}[\ (f_{f,k} - e_{f,k})^{2} / e_{f,k}\ ]$$

Statistic	DF	Value	Prob
Chi-Square	2	87.7453	<.0001
Likelihood Ratio Chi-Square	2	87.5891	<.0001
Mantel-Haenszel Chi-Square	1	79.2886	<.0001
Phi Coefficient		0.1239	
Contingency Coefficient		0.1230	
Cramer's V		0.1239	

Sample Size = 5715

Interpretation of results:

Setting up the hypothesis:

H_0: Variables are independent
H_1: Variables are dependent

Chi square value = 87.7453

The P value is <.0001 which is less than the level of the significance 0.05. We conclude that we reject the null hypothesis and conclude that the job postings are dependent on fortune 1000 companies.

The job postings of fortune 1000 companies depend on job type, and looking at the bar graph from the result in the appendix, we can state that data science role is most common job role for fortune 1000 and data analyst is most common role in non-fortune company.

Question 2: Which location has more postings of different job types?

Data Set: Indeed and Fortune 1000 Combined

Test: Chi-Square Test of Independence.

Row Variable: Job Type (Data Analyst, Data Engineer, Data Scientist)

Column Variables: Location

$$X_{calc}^{2} = \Sigma^{r}_{f=1}\Sigma^{c}_{k=1}[\ (f_{f,k} - e_{f,k})^{2} / e_{f,k}\ ]$$

Statistic	DF	Value	Prob
Chi-Square	100	218.4061	<.0001
Likelihood Ratio Chi-Square	100	223.5882	<.0001
Mantel-Haenszel Chi-Square	1	9.6629	0.0019
Phi Coefficient		0.1999	
Contingency Coefficient		0.1961	
Cramer's V		0.1414	

Interpretation of results:

Setting up the hypothesis:

H_0: Variables are independent
H_1: Variables are dependent

Chi square value = 218.4061

The P value is <.0001 which is less than the level of the significance 0.05. We conclude that we reject the null hypothesis and conclude that the variables are dependent on job type.

The job postings of different job types depend on location, and looking at the bar graph from the result in the appendix, we can state that data science, data engineer and data analyst job postings are more in CA, NY, VA, TX

Question 3: Is there any significant difference in salary level for each job type? Can we infer something information from that?

Test: Chi-Square Test of Independence

Dataset Used: Indeed

Row Variable: Job Type (Data Analyst, Data Engineer, Data Scientist)

Column Variables: Job levels (1,2,3,4,5,6) 6- being the highest

$$X_{calc}^{2} = \Sigma^{r}_{f=1}\Sigma^{c}_{k=1}[\ (f_{f,k} - e_{f,k})^{2} / e_{f,k}\ ]$$

Statistics for Table of Salary_level by Job_Type

Statistic	DF	Value	Prob
Chi-Square	10	2493.7994	<.0001
Likelihood Ratio Chi-Square	10	2621.4312	<.0001
Mantel-Haenszel Chi-Square	1	1464.3487	<.0001
Phi Coefficient		0.6606	
Contingency Coefficient		0.5512	
Cramer's V		0.4671	

Sample Size = 5715

Interpretation of results:

Setting up the hypothesis:

H_0: Variables are independent
H_1: Variables are dependent

Chi square value = 2493.79

The P value is <.0001 which is less than the level of the significance 0.05. We conclude that we reject the null hypothesis and conclude that the variables are dependent.

The job postings of different salary levels depend on job type, and looking at the bar graph from the result in the appendix, we can state that data scientist posting are more common on the higher level of salary while data analyst position are more common in the lower range of salary level.

Questions 4: Is there any significant different in number of skills required for each job type?

Test: Linear Regression

Dataset Used: Indeed

Dependent Variable: No_of_Skills

Classification Variables: Job Type (Data Analyst, Data Engineer, Data Scientist)

Least Squares Model (No Selection)

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	33598	16799	837.81	<.0001
Error	5712	114533	20.05135		
Corrected Total	5714	148132			

Root MSE	4.47787
Dependent Mean	7.80367
R-Square	0.2268
Adj R-Sq	0.2265
AIC	22855
AICC	22855
SBC	17158

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	8.493118	0.088797	95.65	<.0001
Job_Type data_analyst	1	-4.002321	0.138087	-28.98	<.0001
Job_Type data_engineer	1	2.346621	0.149751	15.67	<.0001
Job_Type data_scientist	0	0	.	.	.

12/3/2019 Results: Linear Regression

Dependent Variable: No_of_Skills No_of_Skills

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate
Intercept	Intercept	B	8.49312	0.08880	95.65	<.0001	0
Job_Type_data_analyst	Job_Type data_analyst	B	-4.00232	0.13809	-28.98	<.0001	-0.36477
Job_Type_data_engineer	Job_Type data_engineer	B	2.34662	0.14975	15.67	<.0001	0.19721
Job_Type_data_scientist	Job_Type data_scientist	0	0	.	.	.	.

Interpretation of results:

Setting up the hypothesis:

H_0: Job type does not predict number of skills
H_1: Job type predicts number of skills

F value = 837.8

The P value is <.0001 which is less than the level of the significance 0.05, we can conclude that the model is significant.

Looking at the R square value, we can say that 22.6% of variance in number of skills is predicted by the model.

From the ANNOVA table parameter estimates p-value which is also less than 0.05, we conclude that we reject the null hypothesis and conclude that job type predicts number of skills.

Also, we can say that data analyst job requires 4 (approx.) time less skills as compared to data scientist job positions, and data engineer job requires 2.3 time more skills as compared to data engineer.

Question 5: Does knowing SAS effect the chances of being qualified for data scientist job?

Analysis: Binary logistic regression

Data set used: Indeed and fortune combined

Y - Data scientist

X - SAS skill

Results:

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	7855.303	7749.100
SC	7861.954	7762.401
-2 Log L	7853.303	7745.100

R-Square	0.0188	Max-rescaled R-Square	0.0251

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	108.2038	1	<.0001
Score	108.7255	1	<.0001
Wald	105.8071	1	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
sas	1	105.8071	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	0.4028	0.0665	36.6631	<.0001
sas	0	1	-0.7480	0.0727	105.8071	<.0001
sas	1	0	0	.	.	.

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
sas 0 vs 1	0.473	0.410	0.546

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	19.5	Somers' D	0.103
Percent Discordant	9.2	Gamma	0.358
Percent Tied	71.2	Tau-a	0.051
Pairs	8066396	c	0.551

Interpretation of the result:

The Wald's Chi-Square value for the variable is 105.8071 and the p-value is <.0001

- The –2Log L values for only intercepts and intercepts with covariates is decreasing citing that the model is a good fit.
- From the p-value of Wald Chi-square we can conclude that the model is significant.
- The R square range for the model is 1.8 to 2.5%, that is the percent variance explained by the model.
- From the p-value in type 3 Analysis table, we can conclude that the skill "SAS" has a significant effect on the chances of being qualified for a data scientist job.

- From the Odds ratio estimates table, we can infer that knowing SAS increases your chances for Data scientist by 52.7%
- From the c-value, we can say 55.1% of rows in the current data are correctly predicted by the model

Question 6: Machine Learning influence on job posting for data scientist who already knows SAS.

Test: Moderation (binary logistic regression)

Data set used: Indeed and fortune combined

Y - Data scientist

X - SAS skill

M (moderator) - Machine Learning skill

- Results after adding Machine Learning as moderator

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	7855.303	5637.823
SC	7861.954	5664.426
-2 Log L	7853.303	5629.823

R-Square	0.3223	Max-rescaled R-Square	0.4315

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	2223.4803	3	<.0001
Score	2073.0466	3	<.0001
Wald	1707.2235	3	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
sas	1	80.8550	<.0001
machine_learning	1	790.1858	<.0001
sas*machine_learning	1	9.1725	0.0025

Analysis of Maximum Likelihood Estimates							
Parameter			DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept			1	1.8635	0.1436	168.3566	<.0001
sas	0		1	-0.5520	0.1543	12.7987	0.0003
sas	1		0	0	.	.	.
machine_learning	0		1	-2.3218	0.1693	188.0307	<.0001
machine_learning	1		0	0	.	.	.
sas*machine_learning	0	0	1	-0.5607	0.1851	9.1725	0.0025
sas*machine_learning	0	1	0	0	.	.	.
sas*machine_learning	1	0	0	0	.	.	.
sas*machine_learning	1	1	0	0	.	.	.

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	69.9	Somers' D	0.629
Percent Discordant	7.0	Gamma	0.819
Percent Tied	23.2	Tau-a	0.311
Pairs	8066396	c	0.814

$$Y = b0 + b1\ X + b2\ M + b3\ XM + e$$
$$Y = 1.8 - 0.5\ X - 2.3M - 0.5\ XM + e$$

Interpretation of results:

Hypothesis:

H_0 : There is no interaction effect

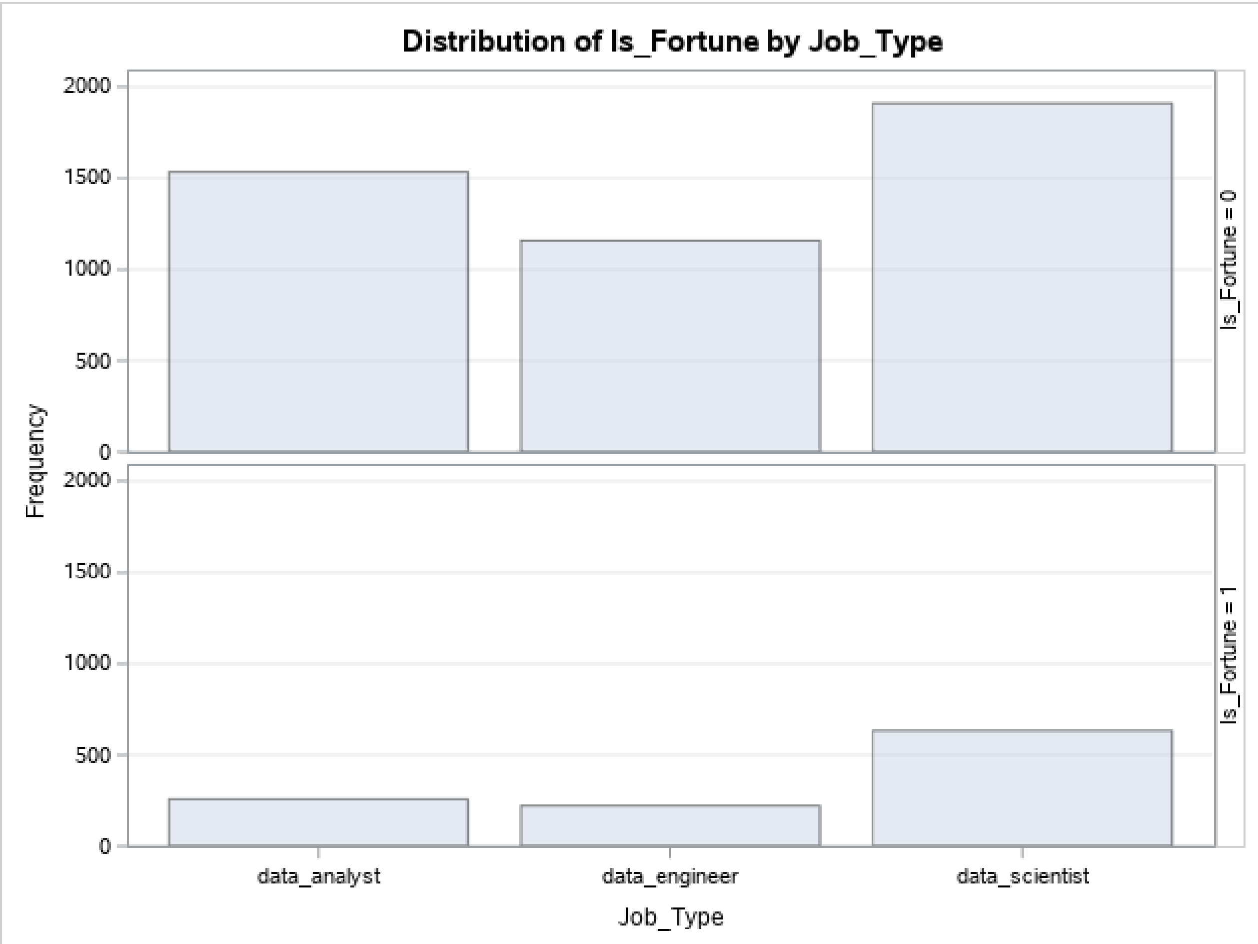
H_1: There is interaction effect

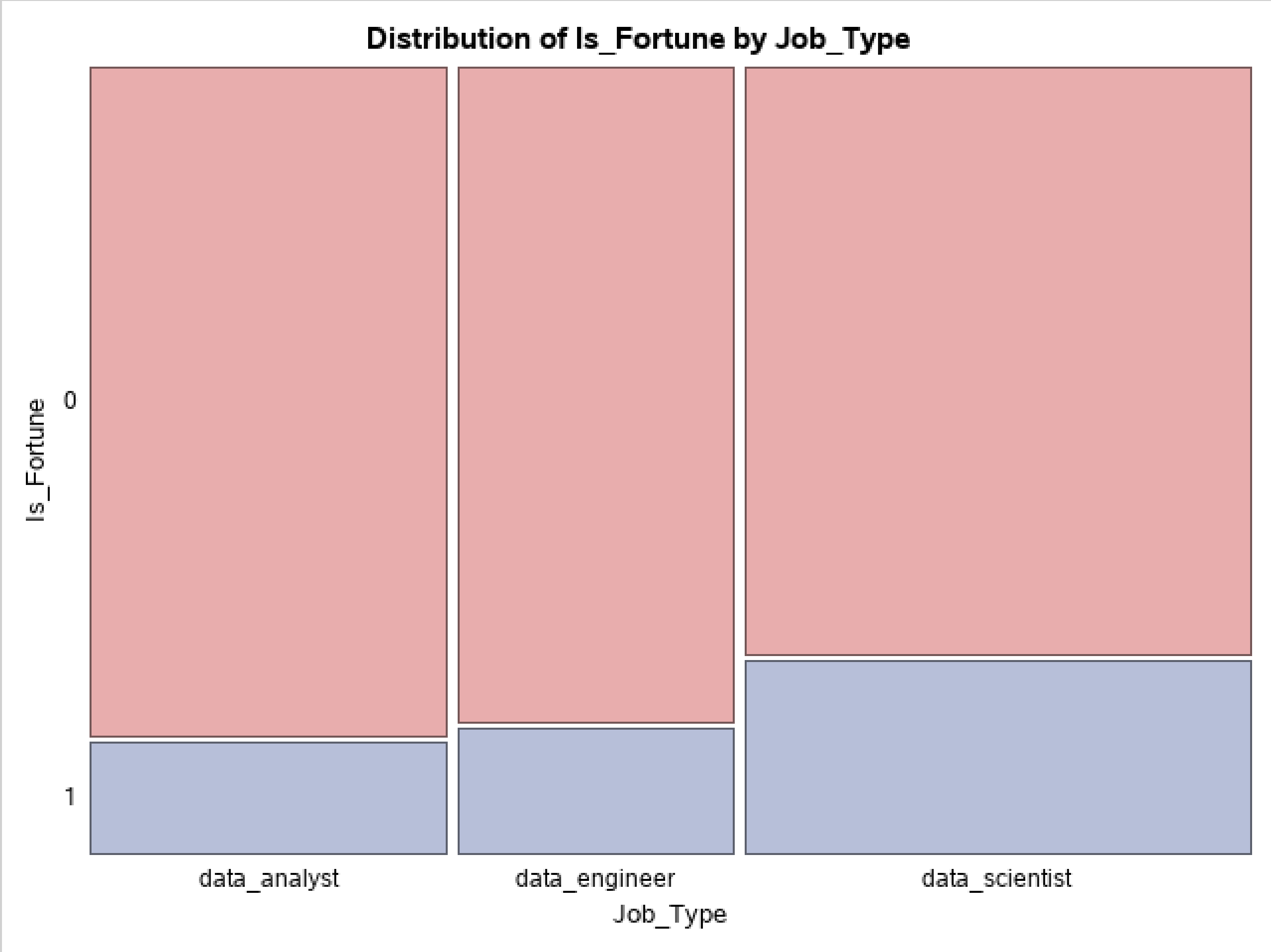
The p-value for the product of sas and machine_learning is 0.0025

- The –2Log L values for only intercepts and intercepts with covariates is decreasing citing that the model is a good fit.
- From the p-value of Wald Chi-square we can conclude that the model is significant.
- The R square range for the model has significantly increased to 32.2-43.1% after the moderator is added, that means that variance explained by this model has increased after adding machine learning as moderator
- From the p-value of Type 3 analysis table, we can reject the null hypothesis and conclude that there is an interaction effect between the variables. Thus, inferring that the skill “Machine Learning” increases chances of a job posting for data scientist with existing knowledge of SAS.
- From the c-value we can say that 81.4% of rows in current data are correctly predicted by the model

APPENDIX

Frequency	Table of Is_Fortune by Job_Type				
	Job_Type(Job_Type)				
Is_Fortune(Is_Fortune)	data_analyst	data_engineer	data_scientist	Total	
0	1536	1158	1911	4605	
1	257	221	632	1110	
Total	1793	1379	2543	5715	





Statistics for Table of Is_Fortune by Job_Type

Statistic	DF	Value	Prob
Chi-Square	2	87.7453	<.0001
Likelihood Ratio Chi-Square	2	87.5891	<.0001
Mantel-Haenszel Chi-Square	1	79.2886	<.0001
Phi Coefficient		0.1239	
Contingency Coefficient		0.1230	
Cramer's V		0.1239	

Sample Size = 5715

Frequency

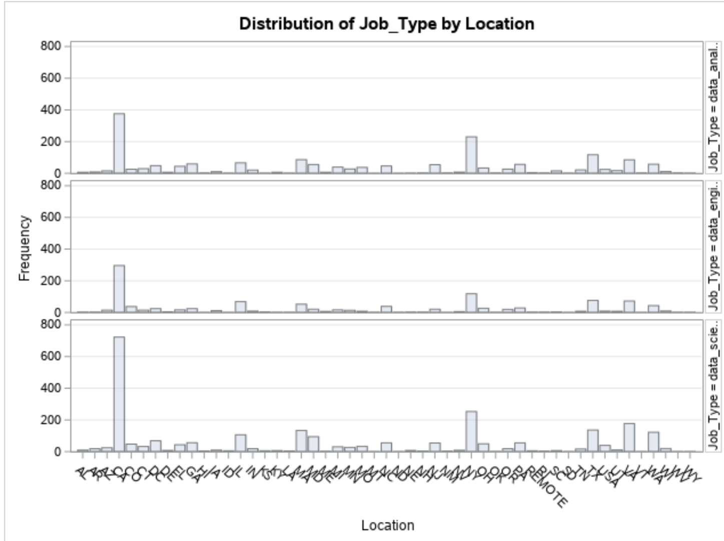
Table of Job_Type by Location

Location(Location)

Job_Type(Job_Type)	AL	AR	AZ	CA	CO	CT	DC	DE	FL	GA	HI	IA	ID	IL	IN	KS	KY	LA	MA	MD	ME	MI	MN	MO	MT	NC
data_analyst	6	8	15	376	25	29	48	6	44	59	2	10	1	66	20	1	6	1	86	55	6	39	27	37	0	46
data_engineer	2	2	14	296	37	14	24	4	16	24	0	11	0	68	8	3	1	1	52	20	6	16	13	7	1	38
data_scientist	9	18	24	723	47	32	68	7	43	56	4	9	5	106	18	5	6	4	133	94	2	30	26	33	0	55
Total	17	28	53	1395	109	75	140	17	103	139	6	30	6	240	46	9	13	6	271	169	14	85	66	77	1	139

Job_Type(Job_Type)	ND	NE	NH	NJ	NM	NV	NY	OH	OK	OR	PA	REMOTE	RI	SC	SD	TN	TX	USA	UT	VA	VT	WA	WI	WV	WY	Total
data_analyst	0	1	2	54	2	7	230	33	2	26	56	4	2	15	1	21	117	24	17	85	2	57	11	2	1	1791
data_engineer	1	2	2	20	0	5	118	26	1	19	28	2	2	3	0	7	76	8	7	72	0	43	9	1	1	1131
data_scientist	0	7	3	54	3	8	253	49	1	18	55	5	3	6	0	16	136	39	10	177	0	122	19	0	0	2541
Total	1	10	7	128	5	20	601	108	4	63	139	11	7	24	1	44	329	71	34	334	2	222	39	3	2	5463

Frequency Missing = 252



Statistics for Table of Job_Type by Location

Statistic	DF	Value	Prob
Chi-Square	100	218.4061	<.0001
Likelihood Ratio Chi-Square	100	223.5882	<.0001

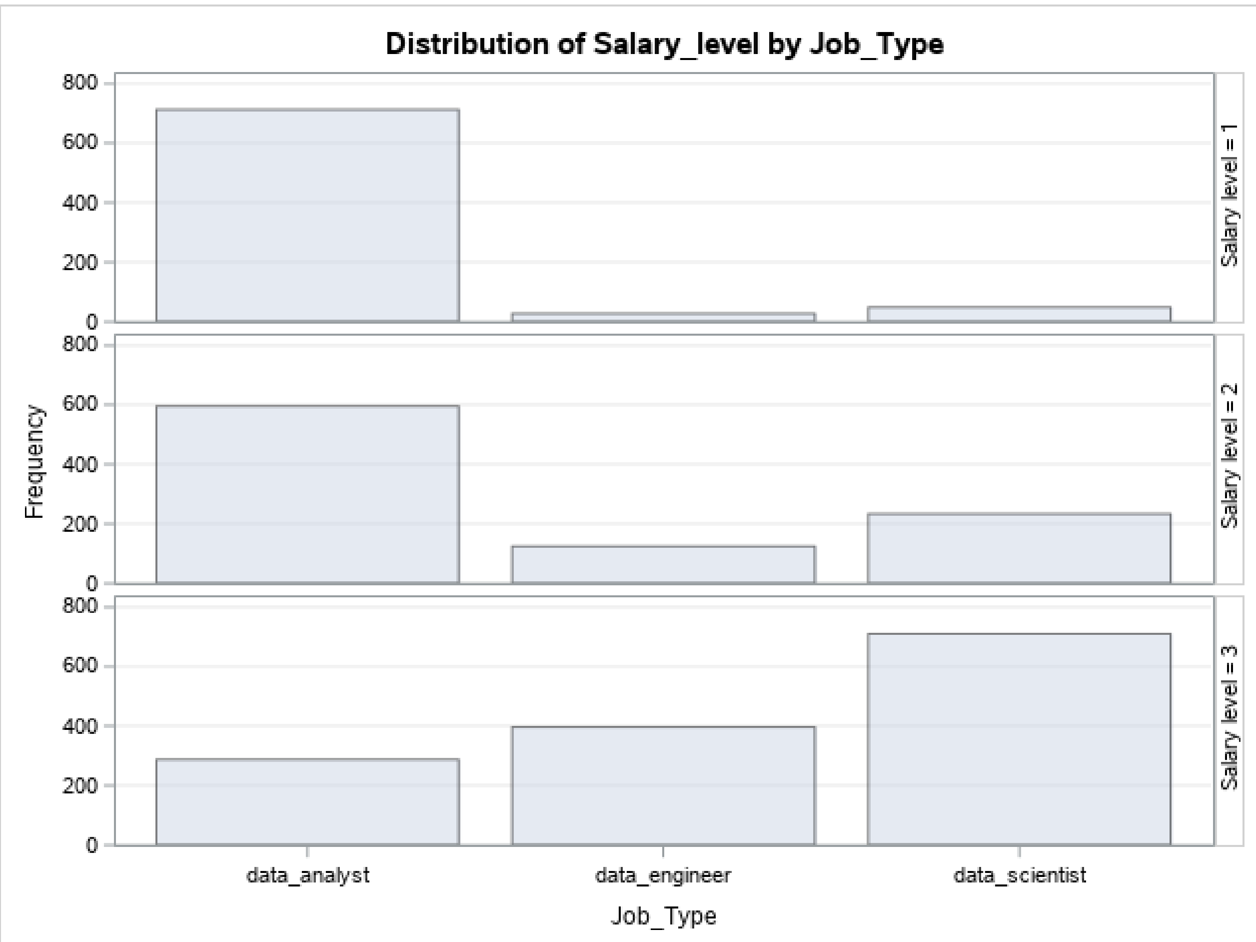
WARNING: 36% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

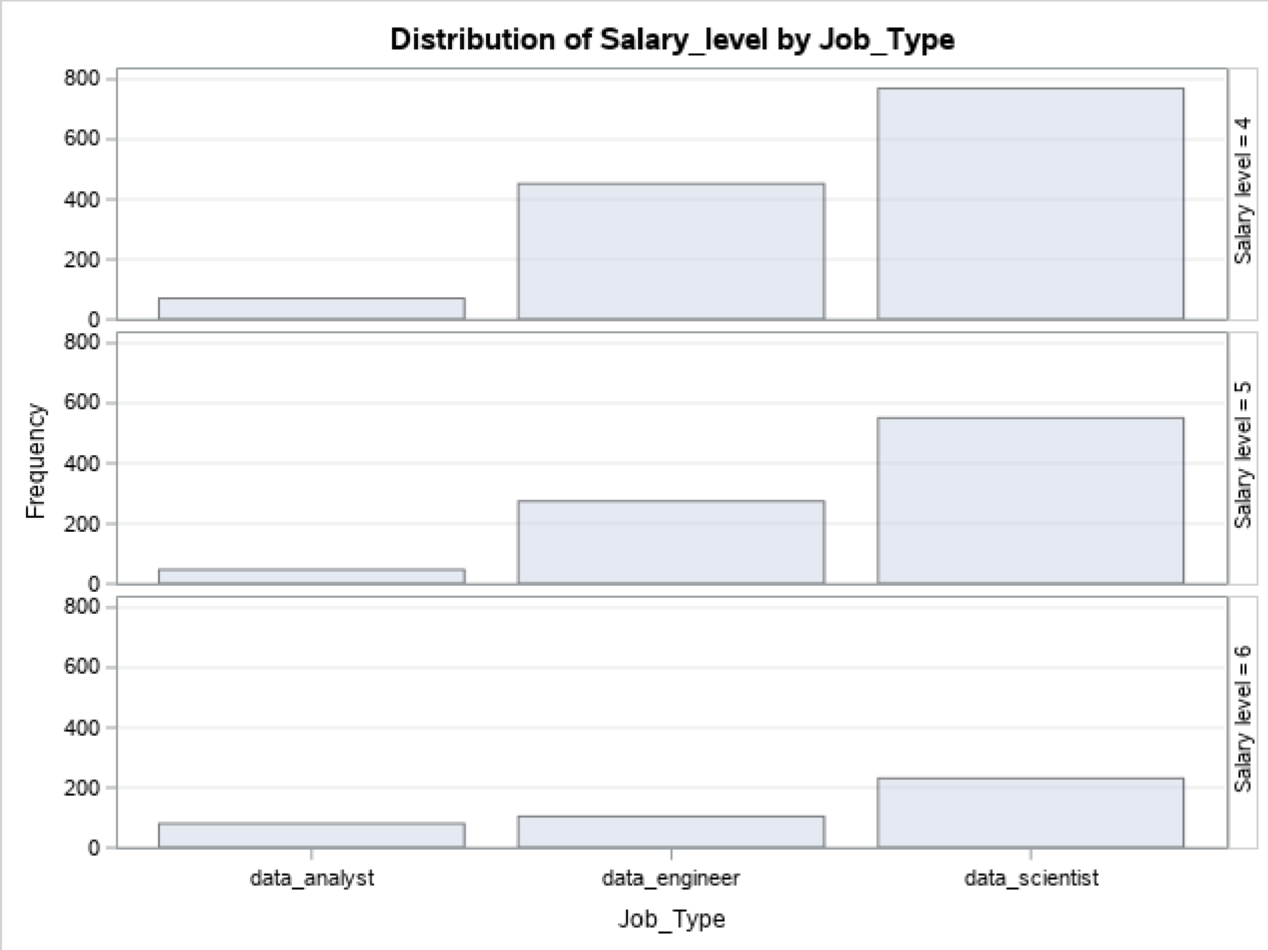
Statistic	DF	Value	Prob
Mantel-Haenszel Chi-Square	1	9.6629	0.0019
Phi Coefficient		0.1999	
Contingency Coefficient		0.1961	
Cramer's V		0.1414	
WARNING: 36% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

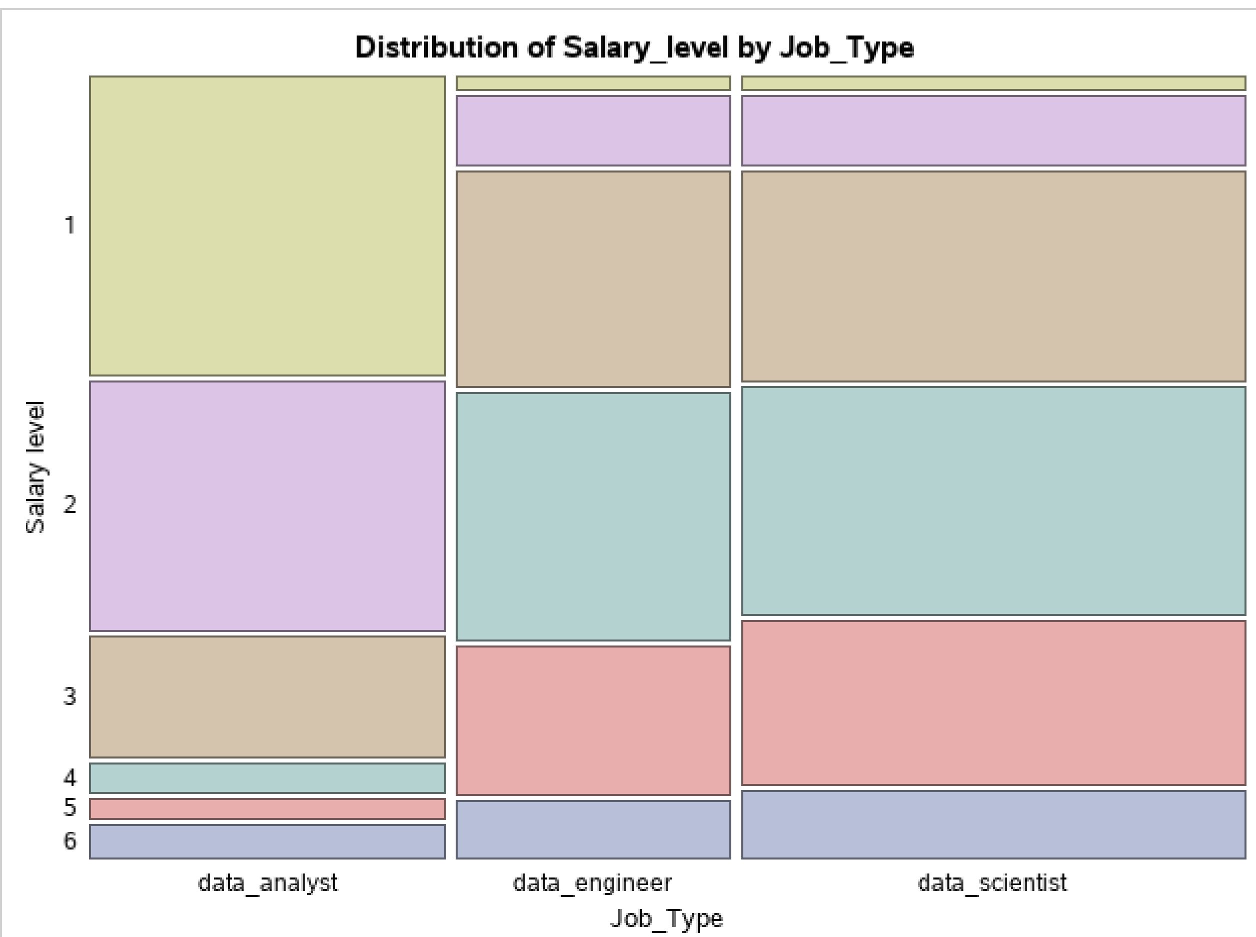
Sample Size = 5463
Frequency Missing = 252

Frequency Expected Deviation

Table of Salary_level by Job_Type				
	Job_Type(Job_Type)			
Salary_level(Salary level)	data_analyst	data_engineer	data_scientist	Total
1	713 247.22 465.78	27 190.14 -163.1	48 350.64 -302.6	788
2	595 298.99 296.01	125 229.95 -105	233 424.06 -191.1	953
3	288 437.35 -149.3	396 336.37 59.635	710 620.29 89.713	1394
4	70 405.35 -335.3	452 311.75 140.25	770 574.9 195.1	1292
5	47 273.89 -226.9	275 210.65 64.35	551 388.46 162.54	873
6	80 130.2 -50.2	104 100.14 3.8626	231 184.66 46.338	415
Total	1793	1379	2543	5715




Distribution of Salary_level by Job_Type
Frequency
800
600
400
200
0
Salary level = 4
Salary level = 5
Salary level = 6
data_analyst
data_engineer
data_scientist
Job_Type


Distribution of Salary_level by Job_Type
Salary level
1
2
3
4
5
6
data_analyst
data_engineer
data_scientist
Job_Type

Statistics for Table of Salary_level by Job_Type

Statistic	DF	Value	Prob
Chi-Square	10	2493.7994	<.0001
Likelihood Ratio Chi-Square	10	2621.4312	<.0001
Mantel-Haenszel Chi-Square	1	1464.3487	<.0001
Phi Coefficient		0.6606	
Contingency Coefficient		0.5512	
Cramer's V		0.4671	

Sample Size = 5715

Data Set	ANI.INDEED_ALL_PROJ
Dependent Variable	No_of_Skills
Selection Method	None

Number of Observations Read	5715
Number of Observations Used	5715

Class Level Information		
Class	**Levels**	**Values**
Job_Type	3	data_analyst data_engineer data_scientist

Dimensions	
Number of Effects	2
Number of Parameters	4

Least Squares Summary				
Step	**Effect Entered**	**Number Effects In**	**Number Parms In**	**SBC**
0	**Intercept**	1	1	18611.0193
1	**Job_Type**	2	3	17158.2161*
*** Optimal Value of Criterion**				

Least Squares Model (No Selection)

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	2	33598	16799	837.81	<.0001
Error	5712	114533	20.05135		
Corrected Total	5714	148132			

Root MSE	4.47787
Dependent Mean	7.80367
R-Square	0.2268
Adj R-Sq	0.2265
AIC	22855
AICC	22855
SBC	17158

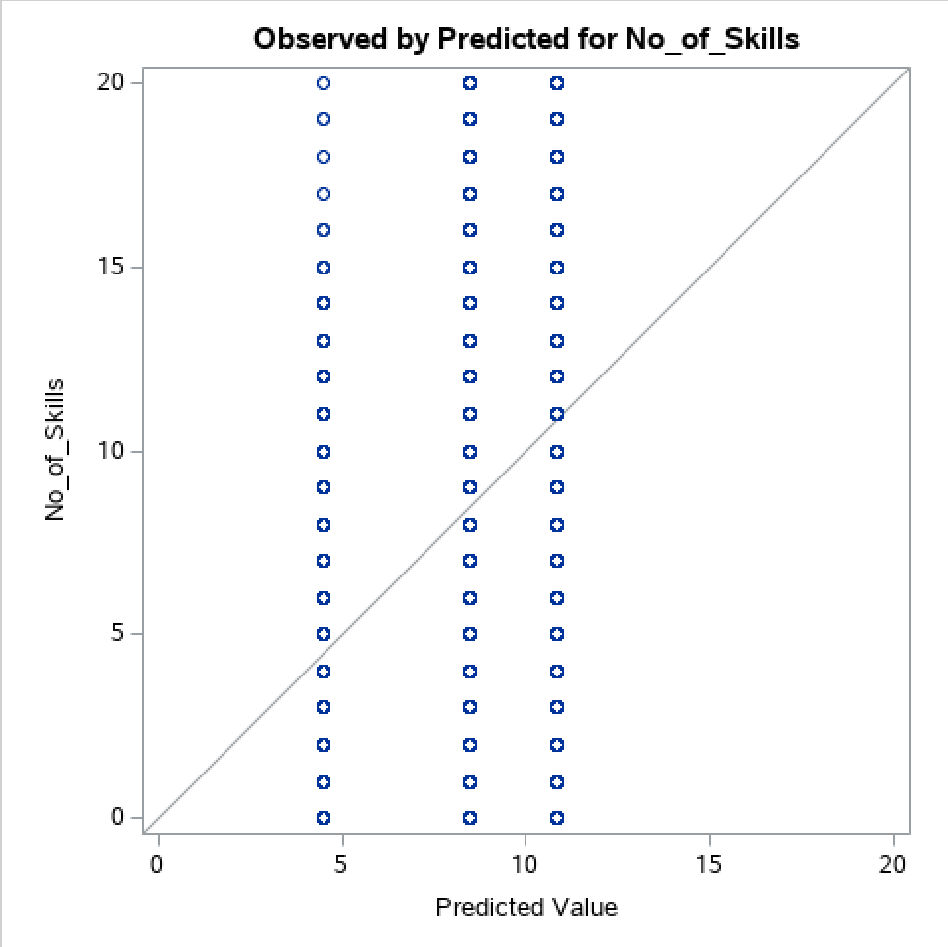
Parameter Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	8.493118	0.088797	95.65	<.0001
Job_Type data_analyst	1	-4.002321	0.138087	-28.98	<.0001
Job_Type data_engineer	1	2.346621	0.149751	15.67	<.0001
Job_Type data_scientist	0	0	.	.	.

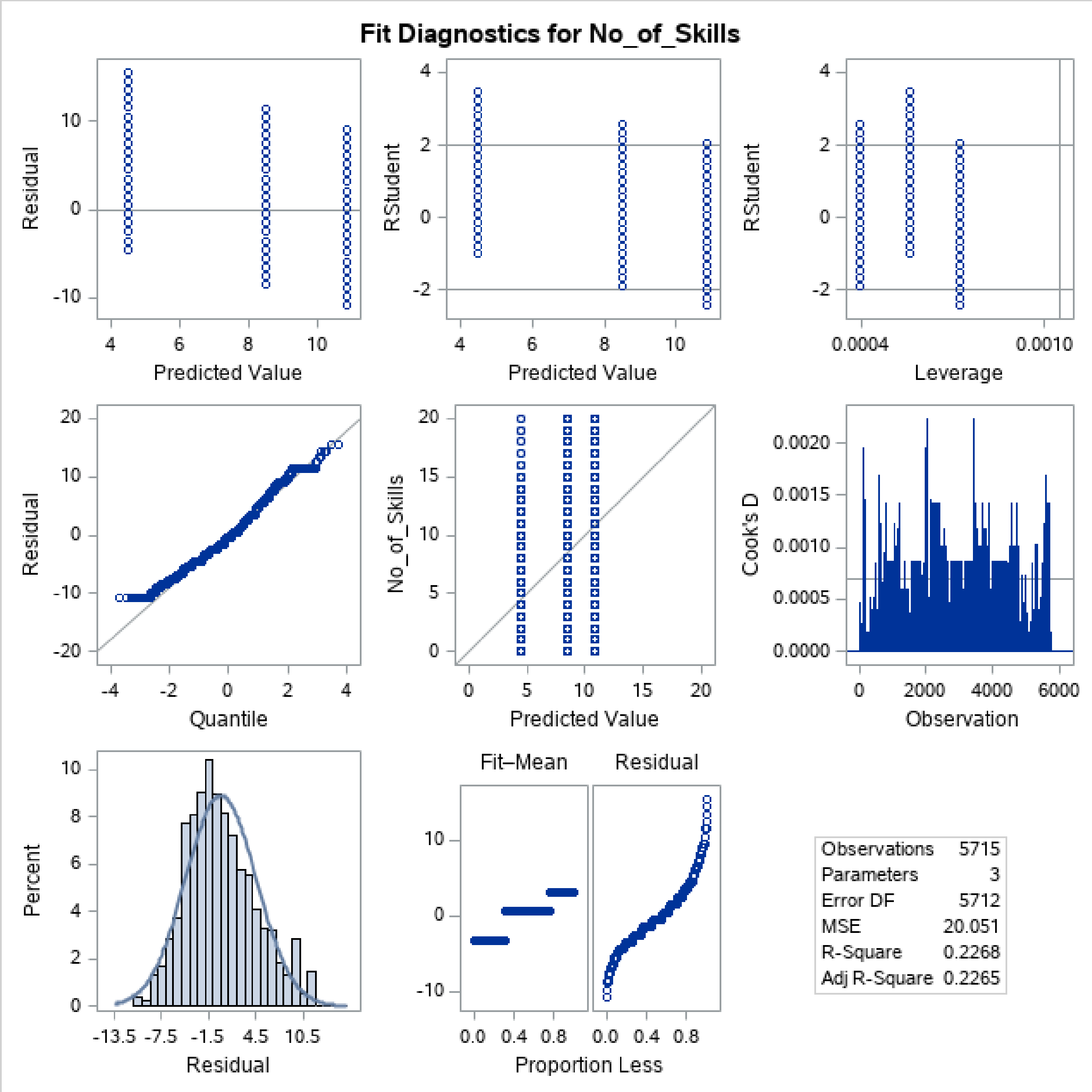
Model: MODEL1

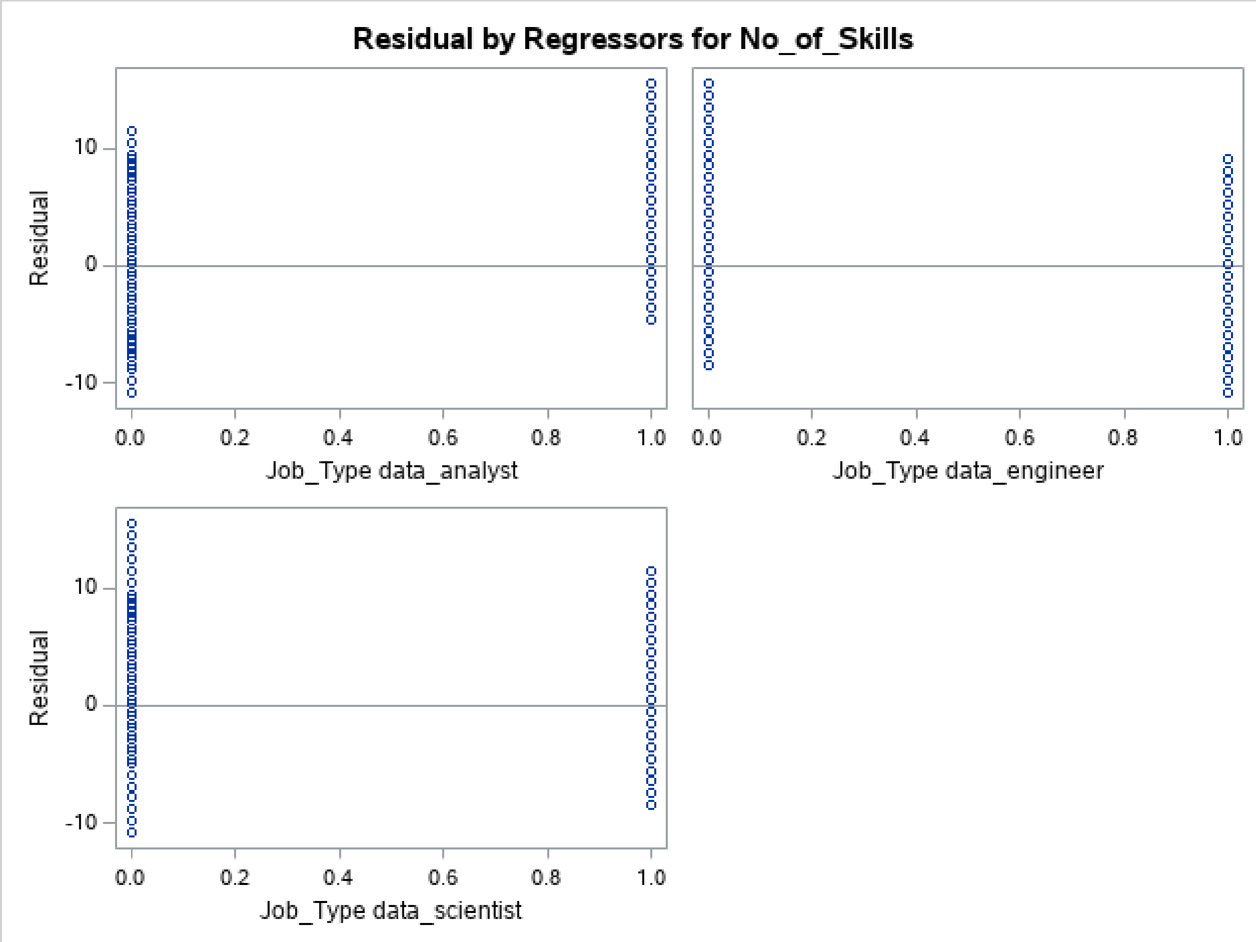
Dependent Variable: No_of_Skills No_of_Skills

Parameter Estimates							
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Standardized Estimate**
Intercept	Intercept	B	8.49312	0.08880	95.65	<.0001	0
Job_Type_data_analyst	Job_Type data_analyst	B	-4.00232	0.13809	-28.98	<.0001	-0.36477
Job_Type_data_engineer	Job_Type data_engineer	B	2.34662	0.14975	15.67	<.0001	0.19721
Job_Type_data_scientist	Job_Type data_scientist	0	0	.	.	.	.

Model: MODEL1
Dependent Variable: No_of_Skills No_of_Skills




Fit Diagnostics for No_of_Skills
Residual
Predicted Value
RStudent
Predicted Value
RStudent
Leverage
Residual
Quantile
No_of_Skills
Predicted Value
Cook's D
Observation
Percent
Residual
Fit–Mean
Residual
Proportion Less
Observations 5715
Parameters 3
Error DF 5712
MSE 20.051
R-Square 0.2268
Adj R-Square 0.2265


Residual by Regressors for No_of_Skills
Residual
10
0
-10
0.0
0.2
0.4
0.6
0.8
1.0
Job_Type data_analyst
Job_Type data_engineer
Job_Type data_scientist

Model Information		
Data Set	ANI.INDEED_JOB_SEG	
Response Variable	data_scientist	data_scientist
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	5715
Number of Observations Used	5715

Response Profile		
Ordered Value	**data_scientist**	**Total Frequency**
1	0	3172
2	1	2543

Probability modeled is data_scientist='1'.

Class Level Information			
Class	**Value**	**Design Variables**	
sas	**0**	1	0
	1	0	1

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	7855.303	7749.100
SC	7861.954	7762.401
-2 Log L	7853.303	7745.100

R-Square	0.0188	**Max-rescaled R-Square**	0.0251

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	108.2038	1	<.0001
Score	108.7255	1	<.0001
Wald	105.8071	1	<.0001

Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
sas	1	105.8071	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	0.4028	0.0665	36.6631	<.0001
sas	**0**	1	-0.7480	0.0727	105.8071	<.0001
sas	**1**	0	0	.	.	.

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
sas 0 vs 1	0.473	0.410	0.546

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	19.5	Somers' D	0.103
Percent Discordant	9.2	Gamma	0.358
Percent Tied	71.2	Tau-a	0.051
Pairs	8066396	c	0.551

Model Information		
Data Set	ANI.INDEED_JOB_SEG	
Response Variable	data_scientist	data_scientist
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	5715
Number of Observations Used	5715

Response Profile		
Ordered Value	**data_scientist**	**Total Frequency**
1	0	3172
2	1	2543

Probability modeled is data_scientist='1'.

Class Level Information			
Class	**Value**	**Design Variables**	
sas	**0**	1	0
	1	0	1
machine_learning	**0**	1	0
	1	0	1

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	7855.303	5637.823
SC	7861.954	5664.426
-2 Log L	7853.303	5629.823

R-Square	0.3223	**Max-rescaled R-Square**	0.4315

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	2223.4803	3	<.0001
Score	2073.0466	3	<.0001
Wald	1707.2235	3	<.0001

Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
sas	1	80.8550	<.0001
machine_learning	1	790.1858	<.0001
sas*machine_learning	1	9.1725	0.0025

Analysis of Maximum Likelihood Estimates							
Parameter			**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**

Analysis of Maximum Likelihood Estimates							
Parameter			DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept			1	1.8635	0.1436	168.3566	<.0001
sas	0		1	-0.5520	0.1543	12.7987	0.0003
sas	1		0	0	.	.	.
machine_learning	0		1	-2.3218	0.1693	188.0307	<.0001
machine_learning	1		0	0	.	.	.
sas*machine_learning	0	0	1	-0.5607	0.1851	9.1725	0.0025
sas*machine_learning	0	1	0	0	.	.	.
sas*machine_learning	1	0	0	0	.	.	.
sas*machine_learning	1	1	0	0	.	.	.

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	69.9	Somers' D	0.629
Percent Discordant	7.0	Gamma	0.819
Percent Tied	23.2	Tau-a	0.311
Pairs	8066396	c	0.814