Project: Attrition Analysis – Data Science with SAS

Imported Dataset, Descriptive Analysis, Logistic Regression

CODE:

```
%web_drop_table(WORK.IMPORT1);


PROC IMPORT OUT= IMPORT1 (keep= Employee_ID Sex_Indicator Relocation_indicator Retain_indicator
Marital_Status)
 DATAFILE="/folders/myfolders/Project 03_Attrition Analysis_Datasets.xlsx"
 DBMS=XLSX replace;
 GETNAMES=YES;
RUN;
PROC CONTENTS DATA=WORK.IMPORT1; RUN;



proc means data=IMPORT1;
run;
proc freq data=IMPORT1;
run;
proc univariate data=IMPORT1;
run;
proc corr data=IMPORT1;
 var Retain_Indicator;
run;
```

CONTENTS Procedure

The Contents Procedure

WORK.IMPORT1

Attributes

Data Set Name	WORK.IMPORT1	Observations	50
Member Type	DATA	Variables	5
Engine	V9	Indexes	0
Created	2018-03-29 19:39:18	Observation Length	40
Last Modified	2018-03-29 19:39:18	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Information

Engine/Host Dependent Information	
Data Set Page Size	65536
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	1632
Obs in First Data Page	50
Number of Data Set Repairs	0
Filename	/tmp/SAS_workB42A000016DA_localhost.localdomain/SAS_workAD0C000016DA_localhost.localdomain/import1.sas7bdat
Release Created	9.0401M5
Host Created	Linux
Inode Number	670890
Access Permission	rw-rw-r--
Owner Name	sasdemo
File Size	128KB
File Size (bytes)	131072

Variables

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Label

1	Employee_ID	Num	8	BEST.	Employee_ID
5	Marital_Status	Num	8	BEST.	Marital_Status
4	Relocation_Indicator	Num	8	BEST.	Relocation_Indicator
2	Retain_Indicator	Num	8	BEST.	Retain_Indicator
3	Sex_Indicator	Num	8	BEST.	Sex_Indicator

The MEANS Procedure

The Means Procedure

Summary statistics

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
Employee_ID	**Employee_ID**	50	25.5000000	14.5773797	1.0000000	50.0000000
Retain_Indicator	**Retain_Indicator**	50	0.5600000	0.5014265	0	1.0000000
Sex_Indicator	**Sex_Indicator**	50	0.5600000	0.5014265	0	1.0000000
Relocation_Indicator	**Relocation_Indicator**	50	0.4800000	0.5046720	0	1.0000000
Marital_Status	**Marital_Status**	50	0.4200000	0.4985694	0	1.0000000

The FREQ Procedure

The Freq Procedure

Table Employee_ID

One-Way Frequencies

Employee_ID				
Employee_ID	**Frequency**	**Percent**	**Cumulative Frequency**	**Cumulative Percent**
1	1	2.00	1	2.00
2	1	2.00	2	4.00
3	1	2.00	3	6.00
4	1	2.00	4	8.00
5	1	2.00	5	10.00
6	1	2.00	6	12.00

7	1	2.00	7	14.00
8	1	2.00	8	16.00
9	1	2.00	9	18.00
10	1	2.00	10	20.00
11	1	2.00	11	22.00
12	1	2.00	12	24.00
13	1	2.00	13	26.00
14	1	2.00	14	28.00
15	1	2.00	15	30.00
16	1	2.00	16	32.00
17	1	2.00	17	34.00
18	1	2.00	18	36.00
19	1	2.00	19	38.00
20	1	2.00	20	40.00
21	1	2.00	21	42.00
22	1	2.00	22	44.00
23	1	2.00	23	46.00
24	1	2.00	24	48.00
25	1	2.00	25	50.00
26	1	2.00	26	52.00
27	1	2.00	27	54.00
28	1	2.00	28	56.00
29	1	2.00	29	58.00
30	1	2.00	30	60.00
31	1	2.00	31	62.00
32	1	2.00	32	64.00
33	1	2.00	33	66.00
34	1	2.00	34	68.00
35	1	2.00	35	70.00
36	1	2.00	36	72.00
37	1	2.00	37	74.00
38	1	2.00	38	76.00
39	1	2.00	39	78.00
40	1	2.00	40	80.00
41	1	2.00	41	82.00
42	1	2.00	42	84.00
43	1	2.00	43	86.00
44	1	2.00	44	88.00
45	1	2.00	45	90.00
46	1	2.00	46	92.00
47	1	2.00	47	94.00
48	1	2.00	48	96.00
49	1	2.00	49	98.00
50	1	2.00	50	100.00

Table Retain_Indicator

One-Way Frequencies

Retain_Indicator				
Retain_Indicator	**Frequency**	**Percent**	**Cumulative Frequency**	**Cumulative Percent**
0	22	44.00	22	44.00
1	28	56.00	50	100.00

Table Sex_Indicator

One-Way Frequencies

Sex_Indicator				
Sex_Indicator	**Frequency**	**Percent**	**Cumulative Frequency**	**Cumulative Percent**
0	22	44.00	22	44.00
1	28	56.00	50	100.00

Table Relocation_Indicator

One-Way Frequencies

Relocation_Indicator				
Relocation_Indicator	**Frequency**	**Percent**	**Cumulative Frequency**	**Cumulative Percent**
0	26	52.00	26	52.00
1	24	48.00	50	100.00

Table Marital_Status

One-Way Frequencies

Marital_Status				
Marital_Status	**Frequency**	**Percent**	**Cumulative Frequency**	**Cumulative Percent**
0	29	58.00	29	58.00
1	21	42.00	50	100.00

The UNIVARIATE Procedure

Variable: Employee_ID (Employee_ID)

The Univariate Procedure

Employee_ID

Moments

Moments			
N	50	**Sum Weights**	50
Mean	25.5	**Sum Observations**	1275
Std Deviation	14.5773797	**Variance**	212.5
Skewness	0	**Kurtosis**	-1.2
Uncorrected SS	42925	**Corrected SS**	10412.5
Coeff Variation	57.166195	**Std Error Mean**	2.06155281

Basic Measures of Location and Variability

Basic Statistical Measures			
Location		**Variability**	
Mean	25.50000	**Std Deviation**	14.57738
Median	25.50000	**Variance**	212.50000
Mode	.	**Range**	49.00000
		Interquartile Range	25.00000

Tests For Location

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	12.36932	**Pr > |t|**	<.0001
Sign	**M**	25	**Pr >= |M|**	<.0001
Signed Rank	**S**	637.5	**Pr >= |S|**	<.0001

Quantiles

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	50.0
99%	50.0
95%	48.0
90%	45.5
75% Q3	38.0
50% Median	25.5
25% Q1	13.0
10%	5.5
5%	3.0
1%	1.0

0% Min	1.0

Extreme Observations

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
1	1	46	46
2	2	47	47
3	3	48	48
4	4	49	49
5	5	50	50

The UNIVARIATE Procedure

Variable: Retain_Indicator (Retain_Indicator)

Retain_Indicator

Moments

Moments			
N	50	Sum Weights	50
Mean	0.56	Sum Observations	28
Std Deviation	0.50142654	Variance	0.25142857
Skewness	-0.2492888	Kurtosis	-2.0203699
Uncorrected SS	28	Corrected SS	12.32
Coeff Variation	89.5404529	Std Error Mean	0.07091242

Basic Measures of Location and Variability

Basic Statistical Measures			
Location		Variability	
Mean	0.560000	Std Deviation	0.50143
Median	1.000000	Variance	0.25143
Mode	1.000000	Range	1.00000
		Interquartile Range	1.00000

Tests For Location

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	7.897065	Pr > |t|	<.0001
Sign	M	14	Pr >= |M|	<.0001

Signed Rank	S	203	Pr >= |S|	<.0001

Quantiles

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1
99%	1
95%	1
90%	1
75% Q3	1
50% Median	1
25% Q1	0
10%	0
5%	0
1%	0
0% Min	0

Extreme Observations

Extreme Observations			
Lowest		**Highest**	
Value	**Obs**	**Value**	**Obs**
0	48	1	44
0	46	1	45
0	43	1	47
0	36	1	49
0	33	1	50

The UNIVARIATE Procedure

Variable: Sex_Indicator (Sex_Indicator)

Sex_Indicator

Moments

Moments			
N	50	**Sum Weights**	50
Mean	0.56	**Sum Observations**	28
Std Deviation	0.50142654	**Variance**	0.25142857
Skewness	-0.2492888	**Kurtosis**	-2.0203699
Uncorrected SS	28	**Corrected SS**	12.32
Coeff Variation	89.5404529	**Std Error Mean**	0.07091242

Basic Measures of Location and Variability

Basic Statistical Measures			
Location		**Variability**	
Mean	0.560000	**Std Deviation**	0.50143
Median	1.000000	**Variance**	0.25143
Mode	1.000000	**Range**	1.00000
		Interquartile Range	1.00000

Tests For Location

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	7.897065	**Pr > |t|**	<.0001
Sign	**M**	14	**Pr >= |M|**	<.0001
Signed Rank	**S**	203	**Pr >= |S|**	<.0001

Quantiles

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1
99%	1
95%	1
90%	1
75% Q3	1
50% Median	1
25% Q1	0
10%	0
5%	0
1%	0
0% Min	0

Extreme Observations

Extreme Observations			
Lowest		**Highest**	
Value	**Obs**	**Value**	**Obs**
0	50	1	42
0	49	1	43
0	48	1	44
0	46	1	45
0	39	1	47

The UNIVARIATE Procedure

Variable: Relocation_Indicator (Relocation_Indicator)

Relocation_Indicator

Moments

Moments			
N	50	**Sum Weights**	50
Mean	0.48	**Sum Observations**	24
Std Deviation	0.50467205	**Variance**	0.25469388
Skewness	0.08256187	**Kurtosis**	-2.0780057
Uncorrected SS	24	**Corrected SS**	12.48
Coeff Variation	105.14001	**Std Error Mean**	0.07137141

Basic Measures of Location and Variability

Basic Statistical Measures			
Location		**Variability**	
Mean	0.480000	**Std Deviation**	0.50467
Median	0.000000	**Variance**	0.25469
Mode	0.000000	**Range**	1.00000
		Interquartile Range	1.00000

Tests For Location

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	6.725382	**Pr > |t|**	<.0001
Sign	**M**	12	**Pr >= |M|**	<.0001
Signed Rank	**S**	150	**Pr >= |S|**	<.0001

Quantiles

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1
99%	1
95%	1
90%	1
75% Q3	1
50% Median	0
25% Q1	0
10%	0
5%	0

1%	0
0% Min	0

Extreme Observations

Extreme Observations			
Lowest		**Highest**	
Value	**Obs**	**Value**	**Obs**
0	49	1	40
0	48	1	44
0	46	1	45
0	43	1	47
0	42	1	50

The UNIVARIATE Procedure

Variable: Marital_Status (Marital_Status)

Marital_Status

Moments

Moments			
N	50	**Sum Weights**	50
Mean	0.42	**Sum Observations**	21
Std Deviation	0.49856938	**Variance**	0.24857143
Skewness	0.33428982	**Kurtosis**	-1.9686965
Uncorrected SS	21	**Corrected SS**	12.18
Coeff Variation	118.706996	**Std Error Mean**	0.07050836

Basic Measures of Location and Variability

Basic Statistical Measures			
Location		**Variability**	
Mean	0.420000	**Std Deviation**	0.49857
Median	0.000000	**Variance**	0.24857
Mode	0.000000	**Range**	1.00000
		Interquartile Range	1.00000

Tests For Location

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	5.956741	**Pr > |t|**	<.0001

Sign	**M**	10.5	**Pr >= |M|**	<.0001
Signed Rank	**S**	115.5	**Pr >= |S|**	<.0001

Quantiles

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1
99%	1
95%	1
90%	1
75% Q3	1
50% Median	0
25% Q1	0
10%	0
5%	0
1%	0
0% Min	0

Extreme Observations

Extreme Observations			
Lowest		**Highest**	
Value	**Obs**	**Value**	**Obs**
0	50	1	37
0	49	1	38
0	46	1	44
0	45	1	47
0	43	1	48

The CORR Procedure

The Corr Procedure

Variables Information

1 Variables:	**Retain_Indicator**

Simple Statistics

Simple Statistics							
Variable	**N**	**Mean**	**Std Dev**	**Sum**	**Minimum**	**Maximum**	**Label**
Retain_Indicator	50	0.56000	0.50143	28.00000	0	1.00000	Retain_Indicator

Pearson Correlations

Pearson Correlation Coefficients, N = 50 Prob > |r| under H0: Rho=0	
	Retain_Indicator
Retain_Indicator **Retain_Indicator**	1.00000

CODE for ANOVA:

```
%web_drop_table(WORK.IMPORT1);


PROC IMPORT OUT= IMPORT1 (keep= Employee_ID Sex_Indicator Relocation_indicator Retain_indicator
Marital_Status)
 DATAFILE="/folders/myfolders/Project 03_Attrition Analysis_Datasets.xlsx"
 DBMS=XLSX replace;
 GETNAMES=YES;
RUN;
proc anova data=WORK.IMPORT1;
 title 'ANOVA';
 class Retain_Indicator Sex_Indicator Relocation_Indicator Marital_Status;
 Model Retain_Indicator = Sex_Indicator Relocation_Indicator Marital_Status;
run;
```

ANOVA

The ANOVA Procedure

Class Level Information		
Class	Levels	Values
Retain_Indicator	2	0 1
Sex_Indicator	2	0 1
Relocation_Indicator	2	0 1
Marital_Status	2	0 1
Number of Observations Read		50
Number of Observations Used		50

ANOVA

The ANOVA Procedure

Dependent Variable: Retain_Indicator Retain_Indicator

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	0.59074713	0.19691571	0.77	0.5155
Error	46	11.72925287	0.25498376		
Corrected Total	49	12.32000000			

R-Square	Coeff Var	Root MSE	Retain_Indicator Mean
0.047950	90.17128	0.504959	0.560000

Source	DF	Anova SS	Mean Square	F Value	Pr > F

Sex_Indicator	1	0.14142857	0.14142857	0.55	0.4602
Relocation_Indicator	1	0.19500000	0.19500000	0.76	0.3864
Marital_Status	1	0.25431856	0.25431856	1.00	0.3232

CODE for LOGISTICS REGRESSION:

```
%web_drop_table(WORK.IMPORT);


FILENAME REFFILE '/folders/myfolders/sasuser.v94/Attrition.xlsx';

PROC IMPORT DATAFILE=REFFILE
 DBMS=XLSX
 OUT=WORK.Attrition;
 GETNAMES=YES;
RUN;


proc logistic data=Attrition;
 Model Retain_Indicator = Sex_Indicator Marital_Status Relocation_Indicator;
run;
```

The LOGISTIC Procedure

The Logistic Procedure

Model Information

Model Information		
Data Set	WORK.ATTRITION	
Response Variable	Retain_Indicator	Retain_Indicator
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Observations Summary

Number of Observations Read	**50**
Number of Observations Used	50

Response Profile

Response Profile		
Ordered Value	**Retain_Indicator**	**Total Frequency**
1	0	22
2	1	28

Probability modeled is Retain_Indicator='0'.

Convergence Status

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

Fit Statistics

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	70.593	74.180
SC	72.505	81.829
-2 Log L	68.593	66.180

Global Tests

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	2.4125	3	0.4913
Score	2.3714	3	0.4990
Wald	2.2750	3	0.5173

Parameter Estimates

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-0.0861	0.5560	0.0240	0.8769
Sex_Indicator	1	-0.2796	0.5967	0.2196	0.6394
Marital_Status	1	0.6589	0.6089	1.1709	0.2792
Relocation_Indicator	1	-0.5960	0.6038	0.9744	0.3236

Odds Ratios

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
Sex_Indicator	0.756	0.235	2.435
Marital_Status	1.933	0.586	6.375
Relocation_Indicator	0.551	0.169	1.799

Association Statistics

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	56.0	**Somers' D**	0.250
Percent Discordant	31.0	**Gamma**	0.287
Percent Tied	13.0	**Tau-a**	0.126
Pairs	616	**c**	0.625

Frequency

Frequency of churn: 28/50 = ~56%. The frequency is 28.

Maximum and Minimum

Max: 0.8124

Min: 0.1599

New Dataset

Analysis of Maximum Likelihood Estimates						
Parameter	**EMPLOYEE_ID**	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	**25**	1	-0.5292	0.5014	1.1141	0.2912
Intercept	**28**	1	-0.2809	0.4975	0.3188	0.5723
Intercept	**29**	1	-0.1935	0.4967	0.1517	0.6969
Intercept	**30**	1	-0.1045	0.4963	0.0444	0.8332
Intercept	**31**	1	-0.0162	0.4962	0.0011	0.9739
Intercept	**32**	1	0.0718	0.4965	0.0209	0.8851
Intercept	**34**	1	0.2518	0.4984	0.2552	0.6135
Intercept	**35**	1	0.3447	0.5001	0.4751	0.4907
Intercept	**37**	1	0.5426	0.5055	1.1522	0.2831
Intercept	**38**	1	0.6502	0.5094	1.6291	0.2018
Sex_Indicator		1	0.5321	0.5090	1.0930	0.2958
Marital_Status		1	0.7278	0.5179	1.9748	0.1599

Relocation_ Indicator		1	-0.1192	0.5025	0.0563	0.8124

HIghlighted are about Maximum value.

Max: 0.8124

Min: 0.1599