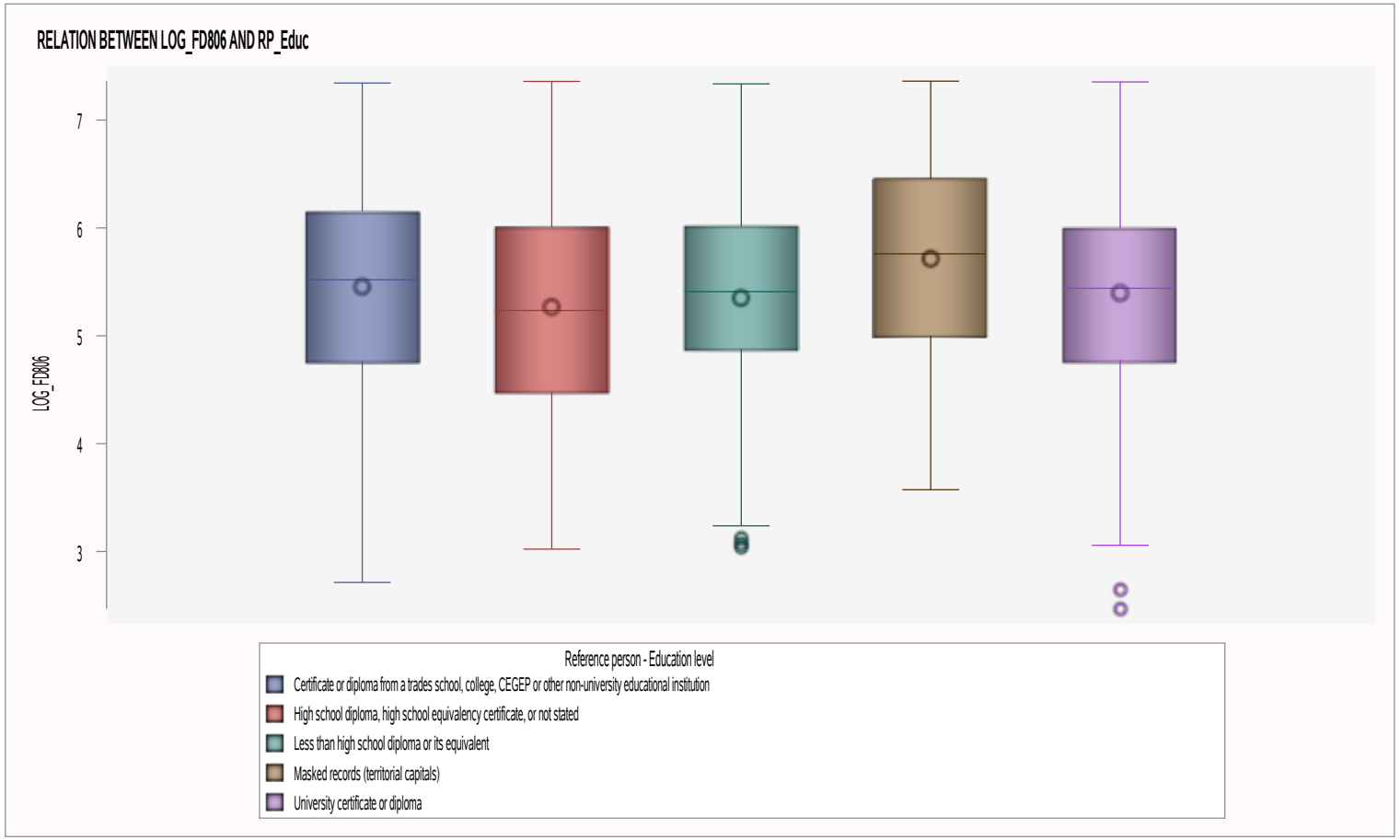
BIVARIATE ANALYSIS OF RP_Educ AND LOG_FD806 FOR ANA.MODEL2
RELATION BETWEEN LOG_FD806 AND RP_Educ

The MEANS Procedure

Analysis Variable : LOG_FD806														
Reference person - Education level	N Obs	N	N Miss	Minimum	Lower Quartile	Median	Mean	Upper Quartile	Maximum	Quartile Range	Coeff of Variation	Lower 95% CL for Mean	Upper 95% CL for Mean	Skewness
Certificate or diploma from a trades school, college, CEGEP or other non-university educational institution	2688986	2688986	0	2.71	4.76	5.52	5.45	6.15	7.34	1.40	17.51	5.45	5.45	-0.18
High school diploma, high school equivalency certificate, or not stated	1668821	1668821	0	3.02	4.47	5.23	5.27	6.00	7.36	1.54	18.14	5.27	5.27	0.08
Less than high school diploma or its equivalent	1043308	1043308	0	3.03	4.86	5.41	5.35	6.01	7.34	1.15	17.11	5.35	5.35	-0.33
Masked records (territorial capitals)	11711	11711	0	3.57	4.99	5.76	5.71	6.45	7.36	1.47	16.60	5.70	5.73	-0.28
University certificate or diploma	2716050	2716050	0	2.47	4.76	5.44	5.40	6.00	7.35	1.24	17.11	5.39	5.40	-0.22



One-way ANOVA Assumptions
In order to run a one-way ANOVA the following assumptions must be met:
1.The response of interest is continuous and normally distributed for each treatment group:
Normality test: PROC UNIVARIATE NORMAL and QQPlot for each group.
2.Treatment groups are independent of one another. Experimental units only receive one treatment,and they do not overlap.
3.There are no major outliers.
4.A check for unequal variances will help determine which version of a one-way ANOVA is most appropriate
(Levene’s test, Null hypothesis: variances are equal between groups):
A .If variances are equal, then the assumptions of a standard one-way ANOVA are met.
B. If variances are unequal, then a Welch’s one-way ANOVA is appropriate.

Normal Distribution?
Null hypothesis: sample has a normal distribution
CLT :
a.If it looks normal and each group have more than 30 observations
b.If moderately skewed, each group must have more than 100 observations
*rule of thumb: If skewness is between -1 and -0.5 or between 0.5 and 1, the distribution is moderately skewed.
*if the sample size is over 2000, the Kolmgorov test should be used. If the sample size is less than 2000, the Shapiro test is better.

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Education level=Certificate or diploma from a trades school, college, CEGEP or other non-university educational institution

Moments			
N	2688986	**Sum Weights**	2688986
Mean	5.44883434	**Sum Observations**	14651839.3
Std Deviation	0.95402066	**Variance**	0.91015542
Skewness	-0.1787741	**Kurtosis**	-0.6390754
Uncorrected SS	82282839.3	**Corrected SS**	2447394.27
Coeff Variation	17.508711	**Std Error Mean**	0.00058179

Basic Statistical Measures			
Location		**Variability**	
Mean	5.448834	**Std Deviation**	0.95402
Median	5.519940	**Variance**	0.91016
Mode	4.276944	**Range**	4.63020
		Interquartile Range	1.39797

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Education level=Certificate or diploma from a trades school, college, CEGEP or other non-university educational institution

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	9365.697	**Pr > |t|**	<.0001
Sign	**M**	1344493	**Pr >= |M|**	<.0001
Signed Rank	**S**	1.808E12	**Pr >= |S|**	<.0001

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.045039	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	1048.815	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	7193.385	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	7.34357
99%	7.23335
95%	7.03277
90%	6.65562
75% Q3	6.15346
50% Median	5.51994
25% Q1	4.75548
10%	4.20080
5%	3.86556
1%	3.30689
0% Min	2.71337

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Education level=Certificate or diploma from a trades school, college, CEGEP or other non-university educational institution

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
2.71337	1520	493	7.32095	5649	384
3.08374	1372	369	7.32214	2134	435
3.08374	3250	139	7.33743	836	687
3.21165	209	460	7.33995	1345	569
3.24649	1166	238	7.34357	669	170

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Education level=High school diploma, high school equivalency certificate, or not stated

Moments			
N	1668821	Sum Weights	1668821
Mean	5.27079456	Sum Observations	8796012.65
Std Deviation	0.95586754	Variance	0.91368275
Skewness	0.07583426	Kurtosis	-0.8531505
Uncorrected SS	47886747.7	Corrected SS	1524772.05
Coeff Variation	18.1351697	Std Error Mean	0.00073993

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Education level=High school diploma, high school equivalency certificate, or not stated

Basic Statistical Measures			
Location		**Variability**	
Mean	5.270795	**Std Deviation**	0.95587
Median	5.234951	**Variance**	0.91368
Mode	4.678792	**Range**	4.33606
		Interquartile Range	1.53639

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	7123.333	**Pr > |t|**	<.0001
Sign	**M**	834410.5	**Pr >= |M|**	<.0001
Signed Rank	**S**	6.962E11	**Pr >= |S|**	<.0001

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.059936	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	1063.758	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	7719.368	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	7.35843
99%	7.29261
95%	6.82113
90%	6.59266

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Education level=High school diploma, high school equivalency certificate, or not stated

Quantiles (Definition 5)	
Level	**Quantile**
75% Q3	6.00344
50% Median	5.23495
25% Q1	4.46706
10%	3.95124
5%	3.87141
1%	3.43817
0% Min	3.02237

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
3.02237	4062	1160	7.20104	1455	776
3.21727	186	1130	7.20510	5467	884
3.22764	463	1067	7.20541	1060	808
3.24805	1311	981	7.29261	12196	797
3.33220	6907	1140	7.35843	4907	1184

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Education level=Less than high school diploma or its equivalent

Moments			
N	1043308	**Sum Weights**	1043308
Mean	5.34949846	**Sum Observations**	5581174.54
Std Deviation	0.9155476	**Variance**	0.83822741
Skewness	-0.3257457	**Kurtosis**	-0.3181083
Uncorrected SS	30731013.2	**Corrected SS**	874528.524
Coeff Variation	17.1146437	**Std Error Mean**	0.00089634

Basic Statistical Measures			
Location		**Variability**	
Mean	5.349498	**Std Deviation**	0.91555
Median	5.409859	**Variance**	0.83823
Mode	6.188552	**Range**	4.30085
		Interquartile Range	1.15155

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	5968.132	**Pr > |t|**	<.0001
Sign	**M**	521654	**Pr >= |M|**	<.0001
Signed Rank	**S**	2.721E11	**Pr >= |S|**	<.0001

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Education level=Less than high school diploma or its equivalent

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.062322	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	575.4714	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	4043.612	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	7.33580
99%	7.07229
95%	6.83324
90%	6.46854
75% Q3	6.01430
50% Median	5.40986
25% Q1	4.86275
10%	4.05109
5%	3.66356
1%	3.25810
0% Min	3.03495

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Education level=Less than high school diploma or its equivalent

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
3.03495	8060	1253	7.07229	5611	1224
3.07177	840	1404	7.08978	1941	1241
3.10727	381	1243	7.10846	2000	1226
3.23789	987	1328	7.21716	680	1315
3.25810	1093	1227	7.33580	1566	1216

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Education level=Masked records (territorial capitals)

Moments			
N	11711	**Sum Weights**	11711
Mean	5.71392282	**Sum Observations**	66915.7502
Std Deviation	0.94859878	**Variance**	0.89983964
Skewness	-0.2810241	**Kurtosis**	-0.9449775
Uncorrected SS	392888.554	**Corrected SS**	10537.1222
Coeff Variation	16.6015329	**Std Error Mean**	0.00876568

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Education level=Masked records (territorial capitals)

Basic Statistical Measures			
Location		**Variability**	
Mean	5.713923	**Std Deviation**	0.94860
Median	5.759879	**Variance**	0.89984
Mode	4.574092	**Range**	3.78783
		Interquartile Range	1.46558

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	651.8517	**Pr > |t|**	<.0001
Sign	**M**	5855.5	**Pr >= |M|**	<.0001
Signed Rank	**S**	34289808	**Pr >= |S|**	<.0001

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.082104	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	16.42393	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	115.1312	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	7.36074
99%	7.30971
95%	7.08256
90%	6.93674

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Education level=Masked records (territorial capitals)

Quantiles (Definition 5)	
Level	**Quantile**
75% Q3	6.45063
50% Median	5.75988
25% Q1	4.98504
10%	4.25135
5%	4.13772
1%	3.78328
0% Min	3.57291

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
3.57291	18	1468	7.27783	46	1632
3.78328	115	1542	7.28773	45	1558
3.85757	157	1587	7.30971	58	1597
3.91562	58	1620	7.32818	59	1472
3.94119	33	1588	7.36074	43	1508

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Education level=University certificate or diploma

Moments			
N	2716050	**Sum Weights**	2716050
Mean	5.39578241	**Sum Observations**	14655214.8
Std Deviation	0.9232874	**Variance**	0.85245963
Skewness	-0.2173279	**Kurtosis**	-0.2870189
Uncorrected SS	81391672.4	**Corrected SS**	2315322.12
Coeff Variation	17.1112794	**Std Error Mean**	0.00056023

Basic Statistical Measures			
Location		**Variability**	
Mean	5.395782	**Std Deviation**	0.92329
Median	5.440771	**Variance**	0.85246
Mode	6.488141	**Range**	4.88297
		Interquartile Range	1.24077

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	9631.333	**Pr > |t|**	<.0001
Sign	**M**	1358025	**Pr >= |M|**	<.0001
Signed Rank	**S**	1.844E12	**Pr >= |S|**	<.0001

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Education level=University certificate or diploma

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.047387	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	679.7678	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	4947.206	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	7.35361
99%	7.26582
95%	6.89946
90%	6.62514
75% Q3	5.99849
50% Median	5.44077
25% Q1	4.75772
10%	4.07923
5%	3.93613
1%	3.07177
0% Min	2.47064

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Education level=University certificate or diploma

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
2.47064	3940	2141	7.30953	8411	1639
2.64476	15313	2265	7.34003	776	1838
3.05777	7665	2127	7.35014	1360	2242
3.07177	347	2168	7.35109	759	2170
3.15274	132	1679	7.35361	491	2024

Null hypothesis: equal variances
a.If variances are equal, then a pooled t-test is appropriate
b.If variances are unequal, then a Satterthwaite (also known as Welch's) test is appropriate

The GLM Procedure

Class Level Information		
Class	**Levels**	**Values**
RP_Educ	5	Certificate or diploma from a trades school, college, CEGEP or other non-university educational institution High school diploma, high school equivalency certificate, or not stated Less than high school diploma or its equivalent Masked records (territorial capitals) University certificate or diploma

Number of Observations Read	2327
Number of Observations Used	2327
Sum of Frequencies Read	8128876
Sum of Frequencies Used	8128876

The GLM Procedure

Dependent Variable: LOG_FD806

Frequency: WeightD

The GLM Procedure

Dependent Variable: LOG_FD806

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	35556.930	8889.233	10074.4	<.0001
Error	8.13E6	7172554.083	0.882		
Corrected Total	8.13E6	7208111.014			

R-Square	Coeff Var	Root MSE	LOG_FD806 Mean
0.004933	17.45271	0.939338	5.382190

Source	DF	Type I SS	Mean Square	F Value	Pr > F
RP_Educ	4	35556.93035	8889.23259	10074.4	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
RP_Educ	4	35556.93035	8889.23259	10074.4	<.0001

The GLM Procedure

Levene's Test for Homogeneity of LOG_FD806 Variance ANOVA of Absolute Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
RP_Educ	4	6042.3	1510.6	5215.73	<.0001
Error	8.13E6	2354283	0.2896		

Welch's ANOVA for LOG_FD806			
Source	DF	F Value	Pr > F
RP_Educ	4.0000	9784.21	<.0001
Error	92390.9		

The GLM Procedure

Level of RP_Educ	N	LOG_FD806 Mean	LOG_FD806 Std Dev
Certificate or diploma from a trades school, college, CEGEP or other non-university educational institution	2688986	5.44883434	0.95402066
High school diploma, high school equivalency certificate, or not stated	1668821	5.27079456	0.95586754
Less than high school diploma or its equivalent	1043308	5.34949846	0.91554760
Masked records (territorial capitals)	11711	5.71392282	0.94859878
University certificate or diploma	2716050	5.39578241	0.92328740

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

RP_Educ	LOG_FD806 LSMEAN	LSMEAN Number
Certificate or diploma from a trades school, college, CEGEP or other non-university educational institution	5.44883434	1
High school diploma, high school equivalency certificate, or not stated	5.27079456	2
Less than high school diploma or its equivalent	5.34949846	3
Masked records (territorial capitals)	5.71392282	4
University certificate or diploma	5.39578241	5

Least Squares Means for effect RP_Educ
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: LOG_FD806

i/j	1	2	3	4	5
1		<.0001	<.0001	<.0001	<.0001
2	<.0001		<.0001	<.0001	<.0001
3	<.0001	<.0001		<.0001	<.0001
4	<.0001	<.0001	<.0001		<.0001
5	<.0001	<.0001	<.0001	<.0001	

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer