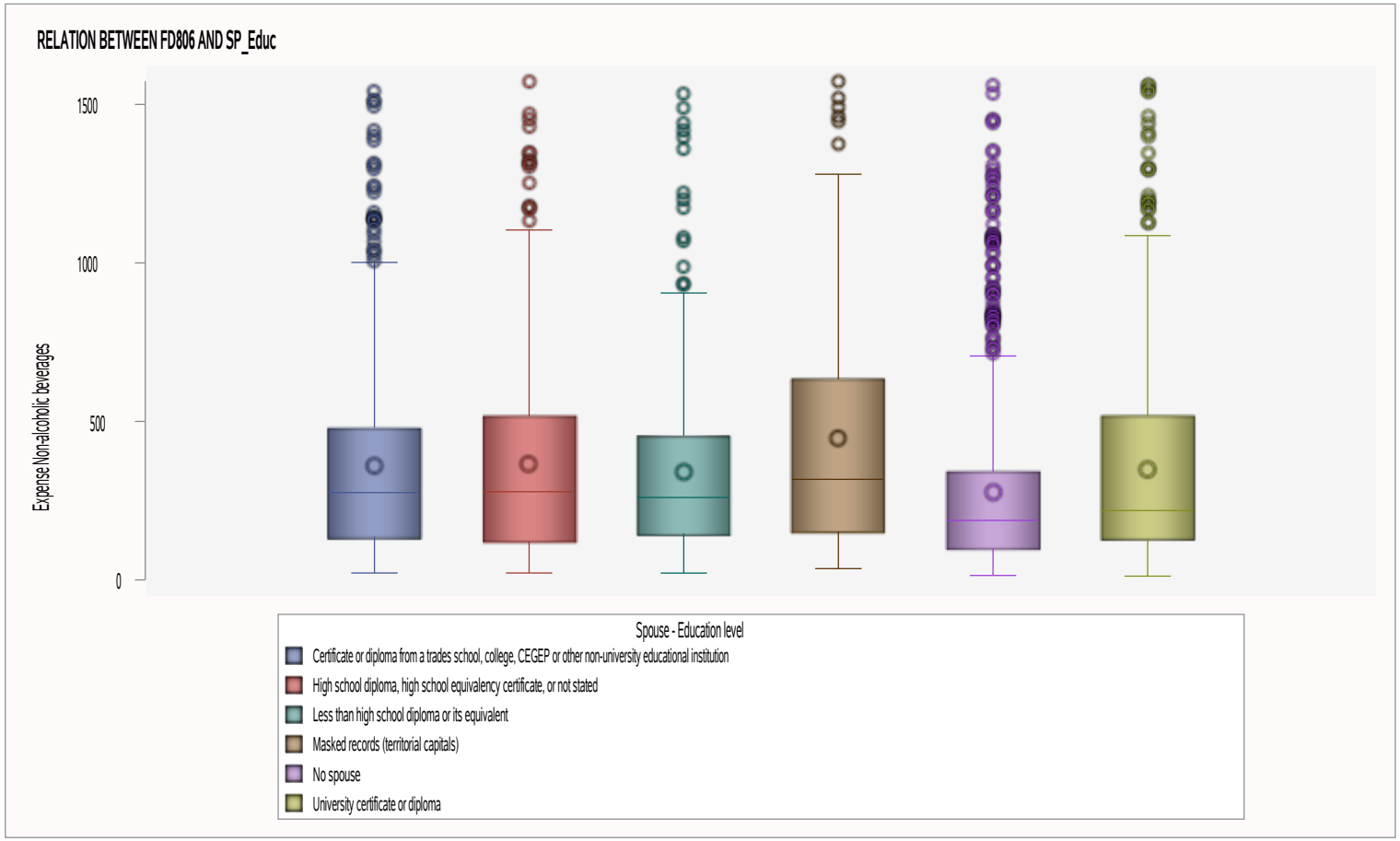
BIVARIATE ANALYSIS OF SP_Educ AND FD806 FOR ANA.MODEL1
RELATION BETWEEN FD806 AND SP_Educ

The MEANS Procedure

Analysis Variable : FD806 Expense Non-alcoholic beverages														
Spouse - 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	1636840	1636840	0	21.84	131.39	275.47	359.03	480.47	1540.64	349.08	85.47	358.56	359.50	1.52
High school diploma, high school equivalency certificate, or not stated	1358512	1358512	0	21.84	116.48	278.20	366.33	516.10	1569.37	399.62	87.55	365.79	366.87	1.37
Less than high school diploma or its equivalent	546094	546094	0	21.58	141.22	260.00	339.50	455.00	1534.26	313.78	85.05	338.73	340.27	1.68
Masked records (territorial capitals)	11711	11711	0	35.62	146.21	317.31	447.33	633.10	1573.00	486.89	82.03	440.68	453.98	1.02
No spouse	2903242	2903242	0	14.08	94.12	187.72	275.52	340.75	1557.89	246.63	99.21	275.21	275.84	1.97
University certificate or diploma	1672477	1672477	0	11.83	124.41	218.72	347.48	516.62	1561.82	392.21	92.10	347.00	347.97	1.60



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: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

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

Moments			
N	1636840	**Sum Weights**	1636840
Mean	359.031927	**Sum Observations**	587677820
Std Deviation	306.85644	**Variance**	94160.8747
Skewness	1.52284442	**Kurtosis**	2.30066606
Uncorrected SS	3.65121E11	**Corrected SS**	1.54126E11
Coeff Variation	85.4677305	**Std Error Mean**	0.23984581

Basic Statistical Measures			
Location		**Variability**	
Mean	359.0319	**Std Deviation**	306.85644
Median	275.4700	**Variance**	94161
Mode	50.9600	**Range**	1519
		Interquartile Range	349.08000

The UNIVARIATE Procedure
Variable: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

Spouse - 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**	1496.928	**Pr > |t|**	<.0001
Sign	**M**	818420	**Pr >= |M|**	<.0001
Signed Rank	**S**	6.698E11	**Pr >= |S|**	<.0001

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.143985	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	11545.29	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	68254.96	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1540.64
99%	1494.48
95%	1001.63
90%	777.74
75% Q3	480.47
50% Median	275.47
25% Q1	131.39
10%	60.77
5%	50.96
1%	27.44
0% Min	21.84

The UNIVARIATE Procedure
Variable: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

Spouse - 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
21.84	3250	188	1421.16	1473	4
24.96	186	439	1494.48	8411	61
25.22	463	392	1511.64	5649	331
26.00	2580	472	1513.44	2134	253
26.00	975	366	1540.64	1345	364

The UNIVARIATE Procedure
Variable: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

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

Moments			
N	1358512	Sum Weights	1358512
Mean	366.33257	Sum Observations	497667193
Std Deviation	320.731999	Variance	102869.015
Skewness	1.37263355	Kurtosis	1.56999037
Uncorrected SS	3.2206E11	Corrected SS	1.39749E11
Coeff Variation	87.5521385	Std Error Mean	0.275176

The UNIVARIATE Procedure
Variable: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

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

Basic Statistical Measures			
Location		**Variability**	
Mean	366.3326	**Std Deviation**	320.73200
Median	278.2000	**Variance**	102869
Mode	293.0500	**Range**	1548
		Interquartile Range	399.62000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	1331.266	**Pr > |t|**	<.0001
Sign	**M**	679256	**Pr >= |M|**	<.0001
Signed Rank	**S**	4.614E11	**Pr >= |S|**	<.0001

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.145881	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	9864.007	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	59132.74	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1569.37
99%	1469.40
95%	985.14
90%	873.08

The UNIVARIATE Procedure
Variable: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

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

Quantiles (Definition 5)	
Level	**Quantile**
75% Q3	516.10
50% Median	278.20
25% Q1	116.48
10%	59.95
5%	52.00
1%	34.06
0% Min	21.84

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
21.84	1372	590	1346.69	1060	592
26.26	1144	647	1430.56	1379	656
27.30	7416	577	1454.18	1128	484
27.72	969	811	1469.40	12196	500
29.64	1988	815	1569.37	4907	533

The UNIVARIATE Procedure
Variable: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

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

Moments			
N	546094	**Sum Weights**	546094
Mean	339.49993	**Sum Observations**	185398875
Std Deviation	288.742559	**Variance**	83372.2651
Skewness	1.67706691	**Kurtosis**	3.52417518
Uncorrected SS	1.08472E11	**Corrected SS**	4.5529E10
Coeff Variation	85.049372	**Std Error Mean**	0.39073031

Basic Statistical Measures			
Location		**Variability**	
Mean	339.4999	**Std Deviation**	288.74256
Median	260.0000	**Variance**	83372
Mode	260.0000	**Range**	1513
		Interquartile Range	313.78000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	868.8856	**Pr > |t|**	<.0001
Sign	**M**	273047	**Pr >= |M|**	<.0001
Signed Rank	**S**	7.455E10	**Pr >= |S|**	<.0001

The UNIVARIATE Procedure
Variable: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

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

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.151661	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	3115.277	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	19171.28	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1534.26
99%	1485.65
95%	905.07
90%	729.72
75% Q3	455.00
50% Median	260.00
25% Q1	141.22
10%	54.87
5%	37.95
1%	29.64
0% Min	21.58

The UNIVARIATE Procedure
Variable: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

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

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
21.58	347	887	1397.07	495	890
25.48	987	981	1420.62	1377	869
26.26	1551	967	1441.41	1081	858
29.64	7954	989	1485.65	5912	909
32.14	606	884	1534.26	1566	850

The UNIVARIATE Procedure
Variable: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

Spouse - Education level=Masked records (territorial capitals)

Moments			
N	11711	**Sum Weights**	11711
Mean	447.32884	**Sum Observations**	5238668.05
Std Deviation	366.931505	**Variance**	134638.729
Skewness	1.0227851	**Kurtosis**	0.2342045
Uncorrected SS	3920026824	**Corrected SS**	1576619520
Coeff Variation	82.0272407	**Std Error Mean**	3.39068942

The UNIVARIATE Procedure
Variable: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

Spouse - Education level=Masked records (territorial capitals)

Basic Statistical Measures			
Location		**Variability**	
Mean	447.3288	**Std Deviation**	366.93150
Median	317.3100	**Variance**	134639
Mode	96.9400	**Range**	1537
		Interquartile Range	486.89000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	131.9286	**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**	0.155138	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	69.88919	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	422.3186	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1573.00
99%	1494.74
95%	1191.01
90%	1029.41

The UNIVARIATE Procedure
Variable: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

Spouse - Education level=Masked records (territorial capitals)

Quantiles (Definition 5)	
Level	**Quantile**
75% Q3	633.10
50% Median	317.31
25% Q1	146.21
10%	70.20
5%	62.66
1%	43.96
0% Min	35.62

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
35.62	18	1022	1447.85	46	1177
43.96	115	1144	1462.25	45	1066
47.35	157	1150	1494.74	58	1091
50.18	58	1176	1522.61	59	1008
51.48	33	1151	1573.00	43	1057

The UNIVARIATE Procedure
Variable: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

Spouse - Education level=No spouse

Moments			
N	2903242	**Sum Weights**	2903242
Mean	275.521431	**Sum Observations**	799905392
Std Deviation	273.358491	**Variance**	74724.8647
Skewness	1.96763749	**Kurtosis**	3.76805449
Uncorrected SS	4.37335E11	**Corrected SS**	2.16944E11
Coeff Variation	99.2149648	**Std Error Mean**	0.16043198

Basic Statistical Measures			
Location		**Variability**	
Mean	275.5214	**Std Deviation**	273.35849
Median	187.7200	**Variance**	74725
Mode	386.9000	**Range**	1544
		Interquartile Range	246.63000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	1717.372	**Pr > |t|**	<.0001
Sign	**M**	1451621	**Pr >= |M|**	<.0001
Signed Rank	**S**	2.107E12	**Pr >= |S|**	<.0001

The UNIVARIATE Procedure
Variable: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

Spouse - Education level=No spouse

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.174038	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	34880.43	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	203889.5	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1557.89
99%	1230.11
95%	915.33
90%	606.21
75% Q3	340.75
50% Median	187.72
25% Q1	94.12
10%	52.40
5%	39.00
1%	21.28
0% Min	14.08

The UNIVARIATE Procedure
Variable: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

Spouse - Education level=No spouse

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
14.08	15313	1556	1444.43	1311	1266
15.08	1520	1771	1449.41	505	1680
20.54	4062	1537	1456.00	442	1191
20.80	8060	1580	1536.76	836	1218
21.28	7665	1594	1557.89	759	1403

The UNIVARIATE Procedure
Variable: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

Spouse - Education level=University certificate or diploma

Moments			
N	1672477	**Sum Weights**	1672477
Mean	347.48105	**Sum Observations**	581154064
Std Deviation	320.027662	**Variance**	102417.705
Skewness	1.59612068	**Kurtosis**	2.2269118
Uncorrected SS	3.73231E11	**Corrected SS**	1.71291E11
Coeff Variation	92.0993137	**Std Error Mean**	0.24746139

The UNIVARIATE Procedure
Variable: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

Spouse - Education level=University certificate or diploma

Basic Statistical Measures			
Location		**Variability**	
Mean	347.4811	**Std Deviation**	320.02766
Median	218.7200	**Variance**	102418
Mode	657.3000	**Range**	1550
		Interquartile Range	392.21000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	1404.183	**Pr > |t|**	<.0001
Sign	**M**	836238.5	**Pr >= |M|**	<.0001
Signed Rank	**S**	6.993E11	**Pr >= |S|**	<.0001

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.182151	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	17158.06	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	97746.84	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1561.82
99%	1442.51
95%	1176.76
90%	758.65

The UNIVARIATE Procedure
Variable: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

Spouse - Education level=University certificate or diploma

Quantiles (Definition 5)	
Level	**Quantile**
75% Q3	516.62
50% Median	218.72
25% Q1	124.41
10%	68.86
5%	52.26
1%	30.52
0% Min	11.83

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
11.83	3940	2267	1464.24	12702	2244
23.40	132	2002	1540.76	776	2013
28.08	5931	1948	1546.22	669	2000
28.46	523	2268	1556.41	1360	2280
28.60	1062	2014	1561.82	491	2153

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**
SP_Educ	6	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) No spouse 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: FD806 Expense Non-alcoholic beverages

Frequency: WeightD

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	12431174942	2486234988.4	27715.1	<.0001
Error	8.13E6	729215954899	89706.927888		
Corrected Total	8.13E6	741647129841			

R-Square	Coeff Var	Root MSE	FD806 Mean
0.016762	91.63156	299.5111	326.8646

Source	DF	Type I SS	Mean Square	F Value	Pr > F
SP_Educ	5	12431174942	2486234988	27715.1	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SP_Educ	5	12431174942	2486234988	27715.1	<.0001

The GLM Procedure

Levene's Test for Homogeneity of FD806 Variance ANOVA of Absolute Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
SP_Educ	5	4.4463E9	8.8926E8	22741.3	<.0001
Error	8.13E6	3.179E11	39103.4		

Welch's ANOVA for FD806			
Source	DF	F Value	Pr > F
SP_Educ	5.0000	29536.8	<.0001
Error	132153		

The GLM Procedure

Level of SP_Educ	N	FD806	
		Mean	Std Dev
Certificate or diploma from a trades school, college, CEGEP or other non-university educational institution	1636840	359.031927	306.856440
High school diploma, high school equivalency certificate, or not stated	1358512	366.332570	320.731999
Less than high school diploma or its equivalent	546094	339.499930	288.742559
Masked records (territorial capitals)	11711	447.328840	366.931505
No spouse	2903242	275.521431	273.358491
University certificate or diploma	1672477	347.481050	320.027662

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

SP_Educ	FD806 LSMEAN	LSMEAN Number
Certificate or diploma from a trades school, college, CEGEP or other non-university educational institution	359.031927	1
High school diploma, high school equivalency certificate, or not stated	366.332570	2
Less than high school diploma or its equivalent	339.499930	3
Masked records (territorial capitals)	447.328840	4
No spouse	275.521431	5
University certificate or diploma	347.481050	6

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

Dependent Variable: FD806

i/j	1	2	3	4	5	6
1		<.0001	<.0001	<.0001	<.0001	<.0001
2	<.0001		<.0001	<.0001	<.0001	<.0001
3	<.0001	<.0001		<.0001	<.0001	<.0001
4	<.0001	<.0001	<.0001		<.0001	<.0001
5	<.0001	<.0001	<.0001	<.0001		<.0001
6	<.0001	<.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