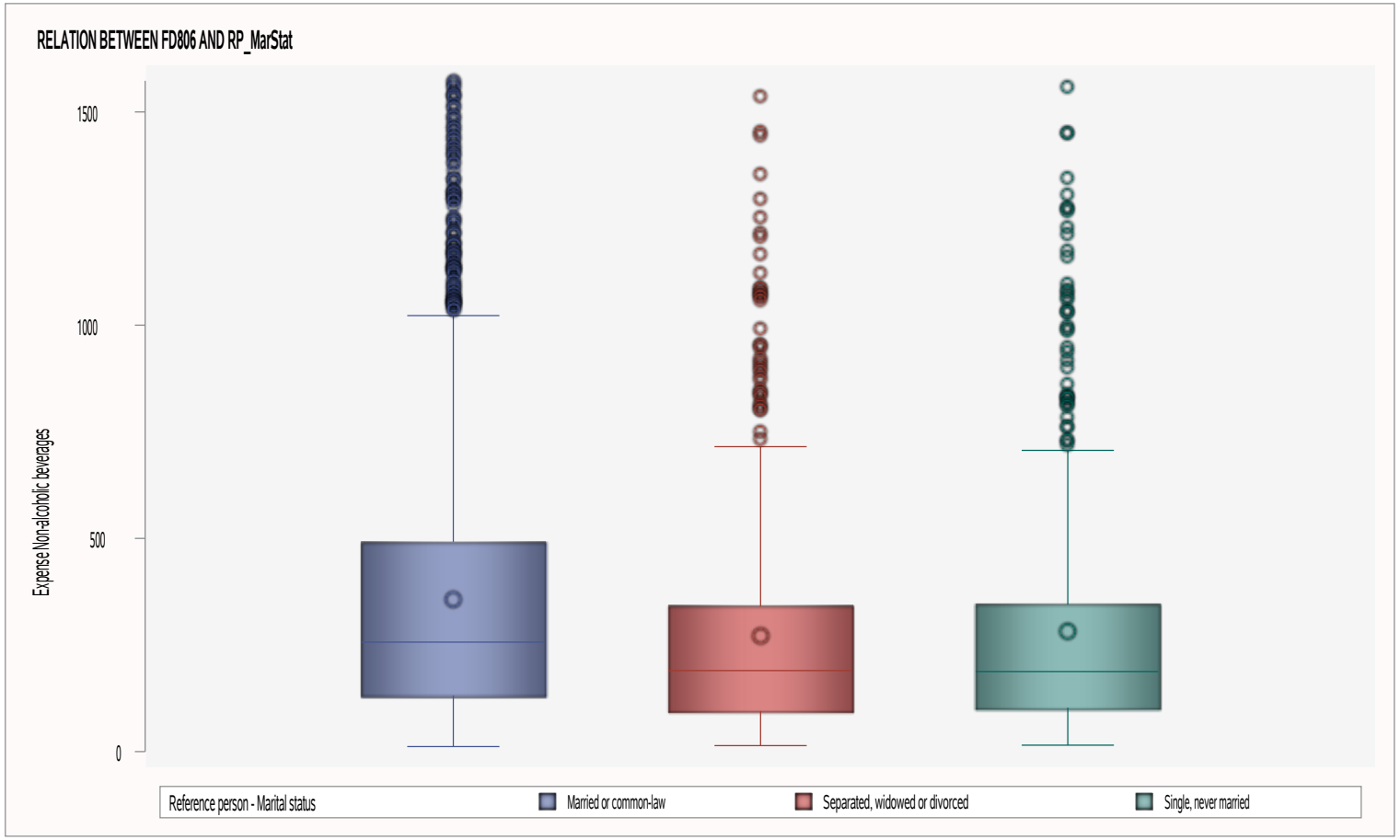
BIVARIATE ANALYSIS OF RP_MarStat AND FD806 FOR ANA.MODEL1
RELATION BETWEEN FD806 AND RP_MarStat

The MEANS Procedure

Analysis Variable : FD806 Expense Non-alcoholic beverages														
Reference person - Marital status	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
Married or common-law	5220895	5220895	0	11.83	127.36	256.75	355.37	491.92	1573.00	364.56	88.15	355.10	355.64	1.52
Separated, widowed or divorced	1548207	1548207	0	14.08	90.55	189.98	272.20	340.55	1536.76	250.00	95.19	271.79	272.61	1.84
Single, never married	1359774	1359774	0	15.08	98.80	187.72	279.65	344.76	1557.89	245.96	103.30	279.17	280.14	2.05



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

Reference person - Marital status=Married or common-law

Moments			
N	5220895	**Sum Weights**	5220895
Mean	355.371086	**Sum Observations**	1855355125
Std Deviation	313.259759	**Variance**	98131.6767
Skewness	1.52033375	**Kurtosis**	2.16622167

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

Freq: WeightD

Reference person - Marital status=Married or common-law

Moments			
Uncorrected SS	1.17167E12	**Corrected SS**	5.12335E11
Coeff Variation	88.1500414	**Std Error Mean**	0.13709832

Basic Statistical Measures			
Location		**Variability**	
Mean	355.3711	**Std Deviation**	313.25976
Median	256.7500	**Variance**	98132
Mode	657.3000	**Range**	1561
		Interquartile Range	364.56000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	2592.089	**Pr > |t|**	<.0001
Sign	**M**	2610448	**Pr >= |M|**	<.0001
Signed Rank	**S**	6.814E12	**Pr >= |S|**	<.0001

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.145877	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	40531.62	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	238096.5	**Pr > A-Sq**	<0.0050

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

Freq: WeightD

Reference person - Marital status=Married or common-law

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1573.00
99%	1464.24
95%	1001.63
90%	777.14
75% Q3	491.92
50% Median	256.75
25% Q1	127.36
10%	61.36
5%	51.10
1%	29.64
0% Min	11.83

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
11.83	3940	1427	1546.22	669	151
21.58	347	1436	1556.41	1360	1505
21.84	1372	274	1561.82	491	1355
21.84	3250	111	1569.37	4907	794
23.40	132	1069	1573.00	43	984

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

Freq: WeightD

Reference person - Marital status=Separated, widowed or divorced

Moments			
N	1548207	**Sum Weights**	1548207
Mean	272.200095	**Sum Observations**	421422092
Std Deviation	259.097423	**Variance**	67131.4747
Skewness	1.83781754	**Kurtosis**	3.27746988
Uncorrected SS	2.18644E11	**Corrected SS**	1.03933E11
Coeff Variation	95.1863824	**Std Error Mean**	0.20823253

Basic Statistical Measures			
Location		**Variability**	
Mean	272.2001	**Std Deviation**	259.09742
Median	189.9800	**Variance**	67131
Mode	487.1400	**Range**	1523
		Interquartile Range	250.00000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	1307.193	**Pr > |t|**	<.0001
Sign	**M**	774103.5	**Pr >= |M|**	<.0001
Signed Rank	**S**	5.992E11	**Pr >= |S|**	<.0001

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

Freq: WeightD

Reference person - Marital status=Separated, widowed or divorced

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.179555	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	16721.39	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	98921.46	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1536.76
99%	1217.58
95%	878.44
90%	644.27
75% Q3	340.55
50% Median	189.98
25% Q1	90.55
10%	51.88
5%	39.00
1%	20.54
0% Min	14.08

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

Freq: WeightD

Reference person - Marital status=Separated, widowed or divorced

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
14.08	15313	1979	1292.20	1554	1709
20.54	4062	1779	1354.61	1683	1961
21.58	840	1833	1444.43	1311	1901
24.96	1553	1904	1456.00	442	1578
25.74	1311	1732	1536.76	836	1688

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

Freq: WeightD

Reference person - Marital status=Single, never married

Moments			
N	1359774	Sum Weights	1359774
Mean	279.652937	Sum Observations	380264793
Std Deviation	288.884539	Variance	83454.277
Skewness	2.04765808	Kurtosis	3.93689601
Uncorrected SS	2.19821E11	Corrected SS	1.13479E11
Coeff Variation	103.301092	Std Error Mean	0.24773703

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

Freq: WeightD

Reference person - Marital status=Single, never married

Basic Statistical Measures			
Location		**Variability**	
Mean	279.6529	**Std Deviation**	288.88454
Median	187.7200	**Variance**	83454
Mode	386.9000	**Range**	1543
		Interquartile Range	245.96000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	1128.83	**Pr > |t|**	<.0001
Sign	**M**	679887	**Pr >= |M|**	<.0001
Signed Rank	**S**	4.622E11	**Pr >= |S|**	<.0001

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.187739	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	18325.68	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	107903.5	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1557.89
99%	1277.64
95%	1062.90
90%	606.21

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

Freq: WeightD

Reference person - Marital status=Single, never married

Quantiles (Definition 5)	
Level	**Quantile**
75% Q3	344.76
50% Median	187.72
25% Q1	98.80
10%	54.08
5%	30.16
1%	21.28
0% Min	15.08

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
15.08	1520	2043	1308.58	3082	2230
20.80	8060	2169	1346.28	5467	2108
21.28	7665	2288	1447.85	46	2223
22.36	381	2167	1449.41	505	1994
24.82	209	2036	1557.89	759	2291

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_MarStat	3	Married or common-law Separated, widowed or divorced Single, never married

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	2	11899823199	5949911599.5	66277.8	<.0001
Error	8.13E6	729747306642	89772.26076		
Corrected Total	8.13E6	741647129841			

R-Square	Coeff Var	Root MSE	FD806 Mean
0.016045	91.66492	299.6202	326.8646

Source	DF	Type I SS	Mean Square	F Value	Pr > F
RP_MarStat	2	11899823199	5949911599	66277.8	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
RP_MarStat	2	11899823199	5949911599	66277.8	<.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**
RP_MarStat	2	3.9955E9	1.9977E9	51021.0	<.0001
Error	8.13E6	3.183E11	39155.2		

Welch's ANOVA for FD806			
Source	**DF**	**F Value**	**Pr > F**
RP_MarStat	2.0000	72809.8	<.0001
Error	2969393		

The GLM Procedure

Level of RP_MarStat	**N**	**FD806**	
		Mean	**Std Dev**
Married or common-law	5220895	355.371086	313.259759
Separated, widowed or divorced	1548207	272.200095	259.097423
Single, never married	1359774	279.652937	288.884539

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

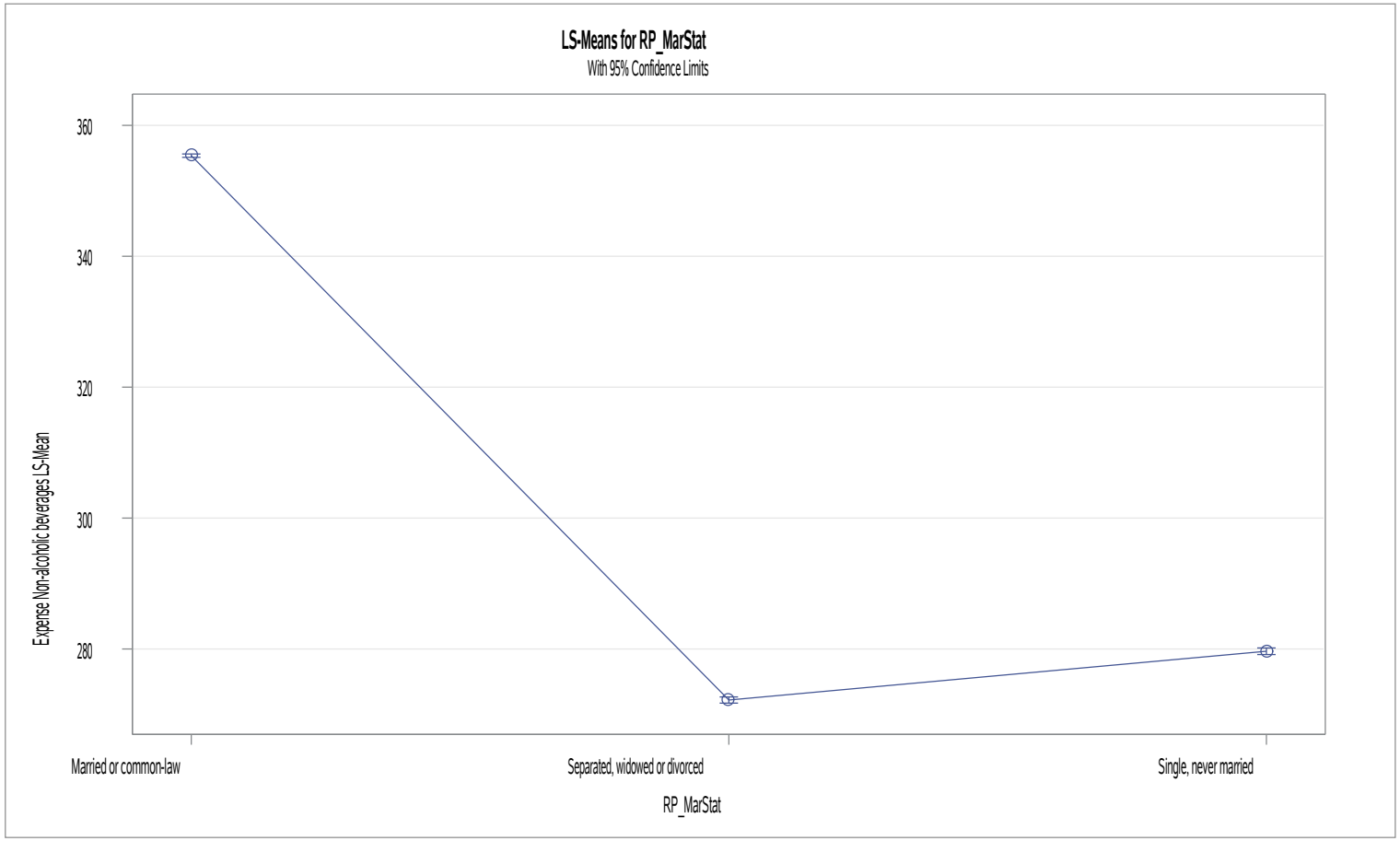
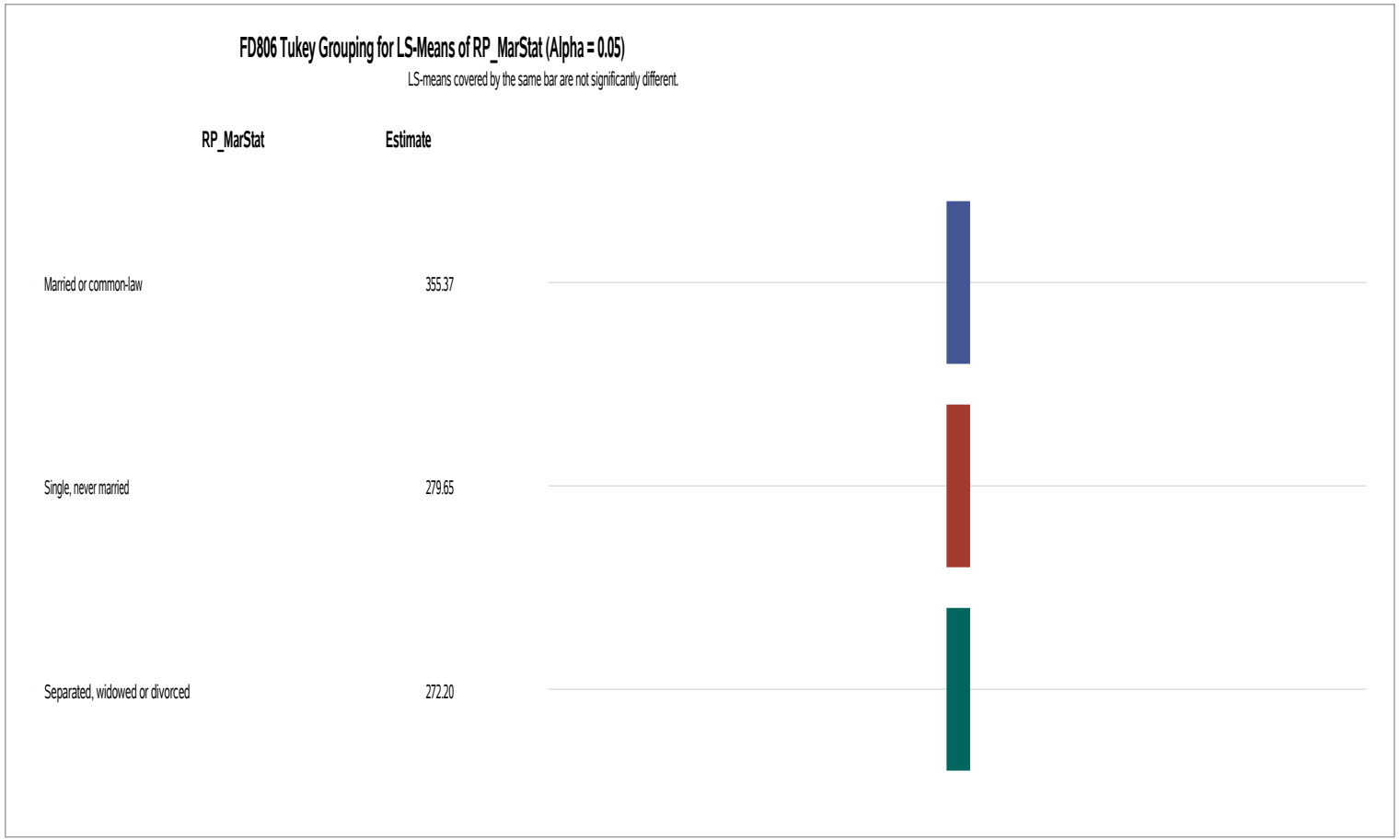
RP_MarStat	**FD806 LSMEAN**	**LSMEAN Number**
Married or common-law	355.371086	1
Separated, widowed or divorced	272.200095	2
Single, never married	279.652937	3

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

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

Dependent Variable: FD806

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



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