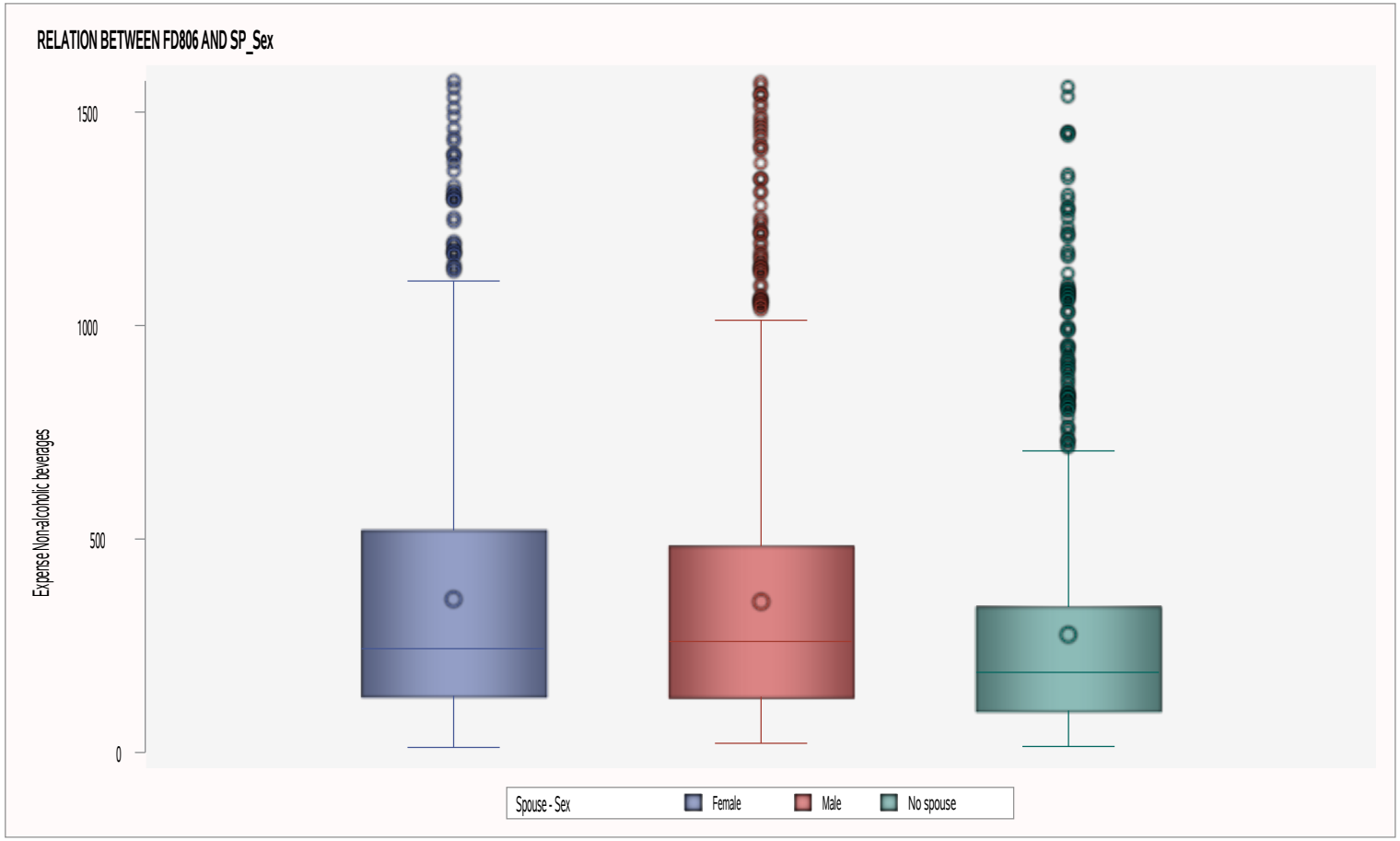
BIVARIATE ANALYSIS OF SP_Sex AND FD806 FOR ANA.MODEL1
RELATION BETWEEN FD806 AND SP_Sex

The MEANS Procedure

Analysis Variable : FD806 Expense Non-alcoholic beverages														
Spouse - Sex	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
Female	2493849	2493849	0	11.83	128.80	243.17	357.82	520.17	1573.00	391.37	87.38	357.43	358.21	1.38
Male	2727046	2727046	0	21.58	126.71	260.00	353.13	483.27	1569.37	356.56	88.86	352.76	353.51	1.65
No spouse	2907981	2907981	0	14.08	94.12	187.76	275.69	340.75	1557.89	246.63	99.19	275.37	276.00	1.97



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 - Sex=Female

Moments			
N	2493849	**Sum Weights**	2493849
Mean	357.817232	**Sum Observations**	892342146
Std Deviation	312.667427	**Variance**	97760.9198
Skewness	1.38193044	**Kurtosis**	1.49793282

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

Freq: WeightD

Spouse - Sex=Female

Moments			
Uncorrected SS	5.63096E11	**Corrected SS**	2.43801E11
Coeff Variation	87.3818807	**Std Error Mean**	0.19799196

Basic Statistical Measures			
Location		**Variability**	
Mean	357.8172	**Std Deviation**	312.66743
Median	243.1700	**Variance**	97761
Mode	723.1400	**Range**	1561
		Interquartile Range	391.37000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	1807.231	**Pr > |t|**	<.0001
Sign	**M**	1246925	**Pr >= |M|**	<.0001
Signed Rank	**S**	1.555E12	**Pr >= |S|**	<.0001

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.147967	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	20142.35	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	116122.4	**Pr > A-Sq**	<0.0050

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

Freq: WeightD

Spouse - Sex=Female

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1573.00
99%	1384.86
95%	993.39
90%	767.76
75% Q3	520.17
50% Median	243.17
25% Q1	128.80
10%	65.70
5%	51.88
1%	27.44
0% Min	11.83

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
11.83	3940	241	1494.74	58	426
24.96	186	64	1511.64	5649	28
25.22	463	605	1534.26	1566	144
25.48	987	180	1556.41	1360	249
26.00	2580	351	1573.00	43	416

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

Freq: WeightD

Spouse - Sex=Male

Moments			
N	2727046	**Sum Weights**	2727046
Mean	353.134116	**Sum Observations**	963012980
Std Deviation	313.783826	**Variance**	98460.2892
Skewness	1.64593682	**Kurtosis**	2.77700089
Uncorrected SS	6.08578E11	**Corrected SS**	2.68506E11
Coeff Variation	88.8568425	**Std Error Mean**	0.19001344

Basic Statistical Measures			
Location		**Variability**	
Mean	353.1341	**Std Deviation**	313.78383
Median	260.0000	**Variance**	98460
Mode	657.3000	**Range**	1548
		Interquartile Range	356.56000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	1858.469	**Pr > |t|**	<.0001
Sign	**M**	1363523	**Pr >= |M|**	<.0001
Signed Rank	**S**	1.859E12	**Pr >= |S|**	<.0001

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

Freq: WeightD

Spouse - Sex=Male

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.152391	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	20811.45	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	125168.6	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1569.37
99%	1469.40
95%	1001.63
90%	777.74
75% Q3	483.27
50% Median	260.00
25% Q1	126.71
10%	59.95
5%	50.96
1%	32.22
0% Min	21.58

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

Freq: WeightD

Spouse - Sex=Male

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
21.58	347	1236	1540.64	1345	791
21.84	1372	1154	1540.76	776	1527
21.84	3250	1030	1546.22	669	915
23.40	132	1335	1561.82	491	1382
26.00	975	793	1569.37	4907	1128

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

Freq: WeightD

Spouse - Sex=No spouse

Moments			
N	2907981	Sum Weights	2907981
Mean	275.685049	Sum Observations	801686886
Std Deviation	273.455354	Variance	74777.8306
Skewness	1.96559807	Kurtosis	3.75787583
Uncorrected SS	4.38466E11	Corrected SS	2.17452E11
Coeff Variation	99.1912164	Std Error Mean	0.16035801

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

Freq: WeightD

Spouse - Sex=No spouse

Basic Statistical Measures			
Location		**Variability**	
Mean	275.6850	**Std Deviation**	273.45535
Median	187.7600	**Variance**	74778
Mode	386.9000	**Range**	1544
		Interquartile Range	246.63000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	1719.185	**Pr > |t|**	<.0001
Sign	**M**	1453991	**Pr >= |M|**	<.0001
Signed Rank	**S**	2.114E12	**Pr >= |S|**	<.0001

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.173993	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	34913.71	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	204079.7	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1557.89
99%	1230.11
95%	915.33
90%	606.21

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

Freq: WeightD

Spouse - Sex=No spouse

Quantiles (Definition 5)	
Level	**Quantile**
75% Q3	340.75
50% Median	187.76
25% Q1	94.12
10%	52.40
5%	39.00
1%	21.28
0% Min	14.08

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
14.08	15313	1994	1447.85	46	1615
15.08	1520	2209	1449.41	505	2118
20.54	4062	1975	1456.00	442	1629
20.80	8060	2018	1536.76	836	1656
21.28	7665	2032	1557.89	759	1841

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_Sex	3	Female Male No spouse

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	11888180398	5944090198.9	66211.9	<.0001
Error	8.13E6	729758949444	89773.693038		
Corrected Total	8.13E6	741647129841			

R-Square	Coeff Var	Root MSE	FD806 Mean
0.016029	91.66565	299.6226	326.8646

Source	DF	Type I SS	Mean Square	F Value	Pr > F
SP_Sex	2	11888180398	5944090199	66211.9	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SP_Sex	2	11888180398	5944090199	66211.9	<.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_Sex	2	3.9126E9	1.9563E9	49966.3	<.0001
Error	8.13E6	3.183E11	39152.6		

Welch's ANOVA for FD806			
Source	**DF**	**F Value**	**Pr > F**
SP_Sex	2.0000	71492.3	<.0001
Error	5294240		

The GLM Procedure

Level of SP_Sex	**N**	**FD806**	
		Mean	**Std Dev**
Female	2493849	357.817232	312.667427
Male	2727046	353.134116	313.783826
No spouse	2907981	275.685049	273.455354

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

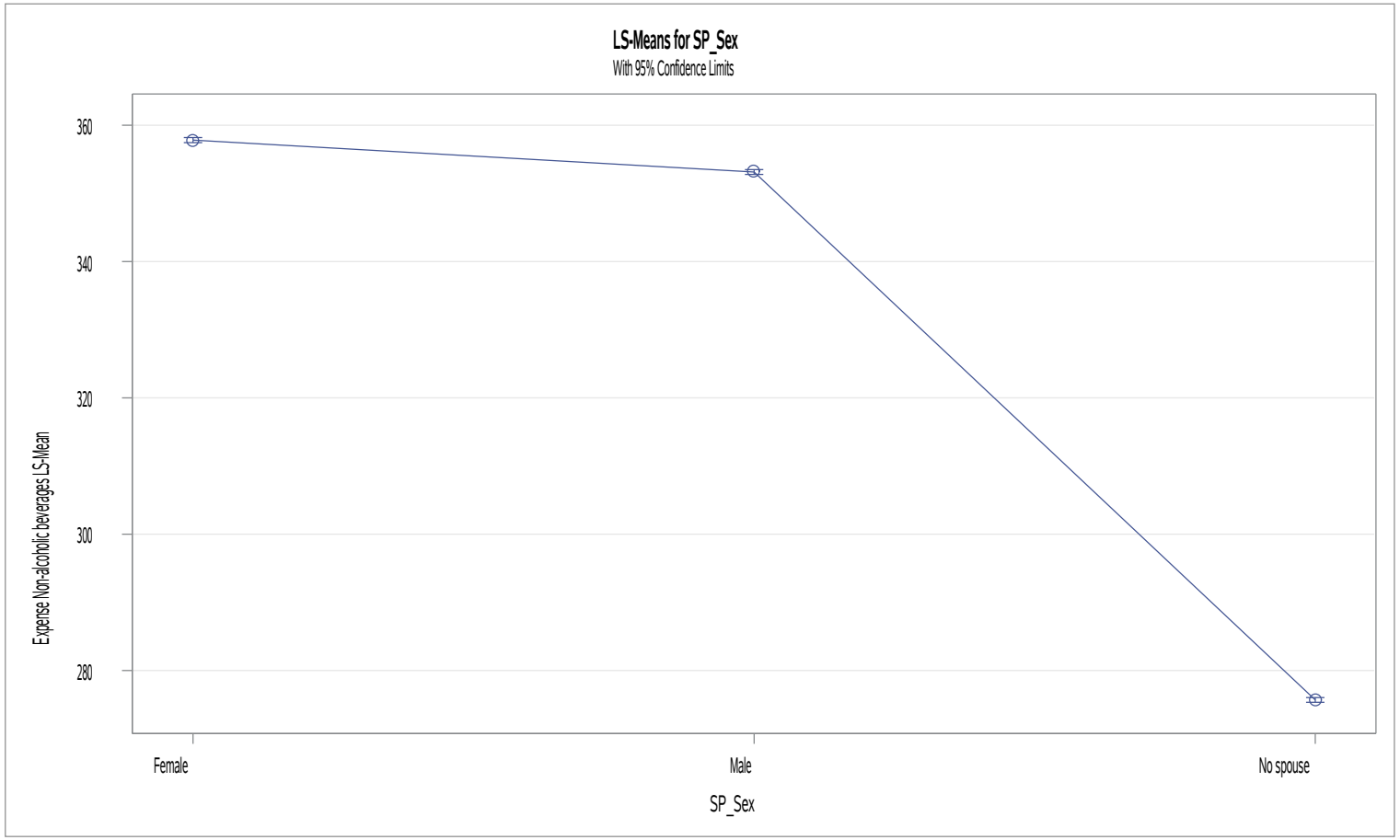
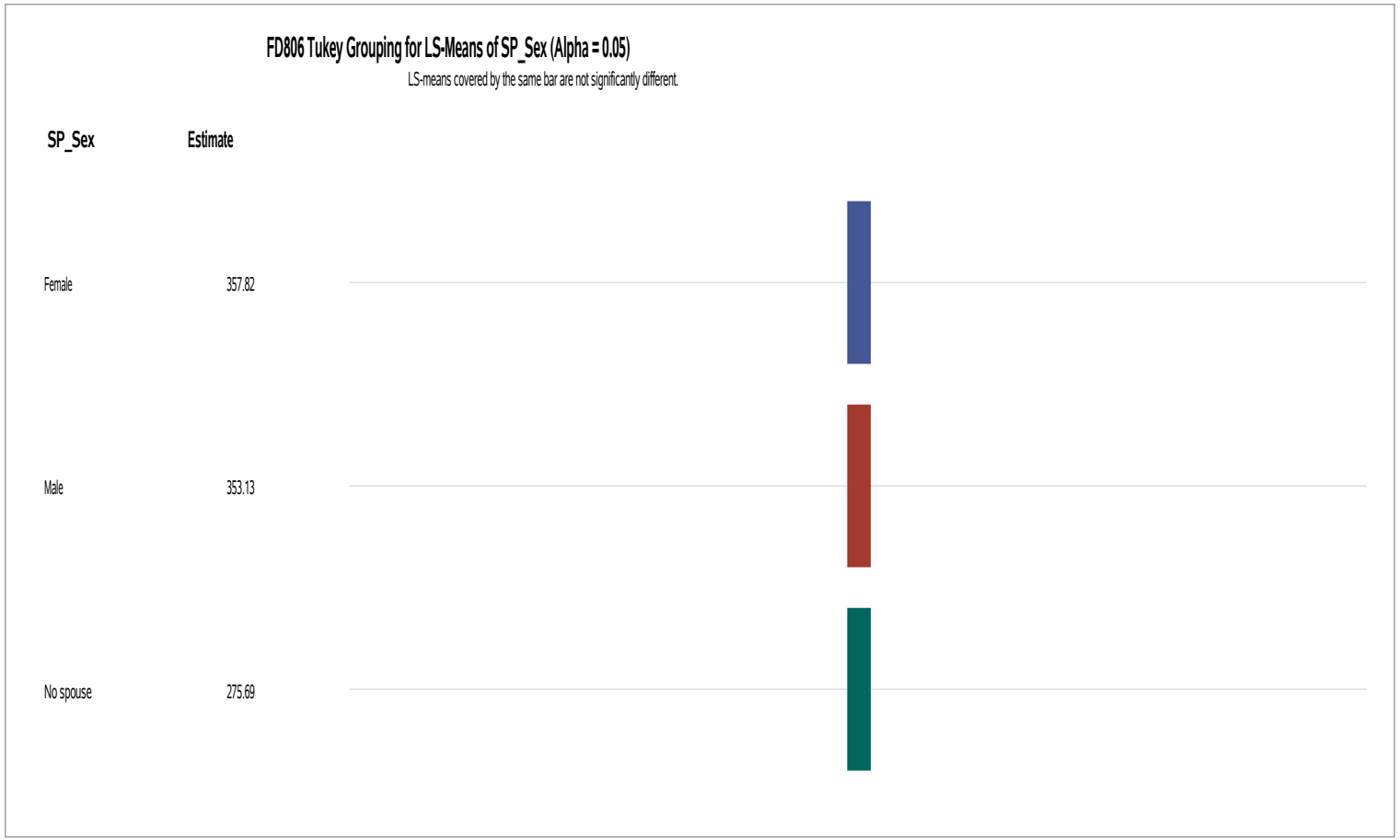
SP_Sex	**FD806 LSMEAN**	**LSMEAN Number**
Female	357.817232	1
Male	353.134116	2
No spouse	275.685049	3

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

Least Squares Means for effect SP_Sex
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