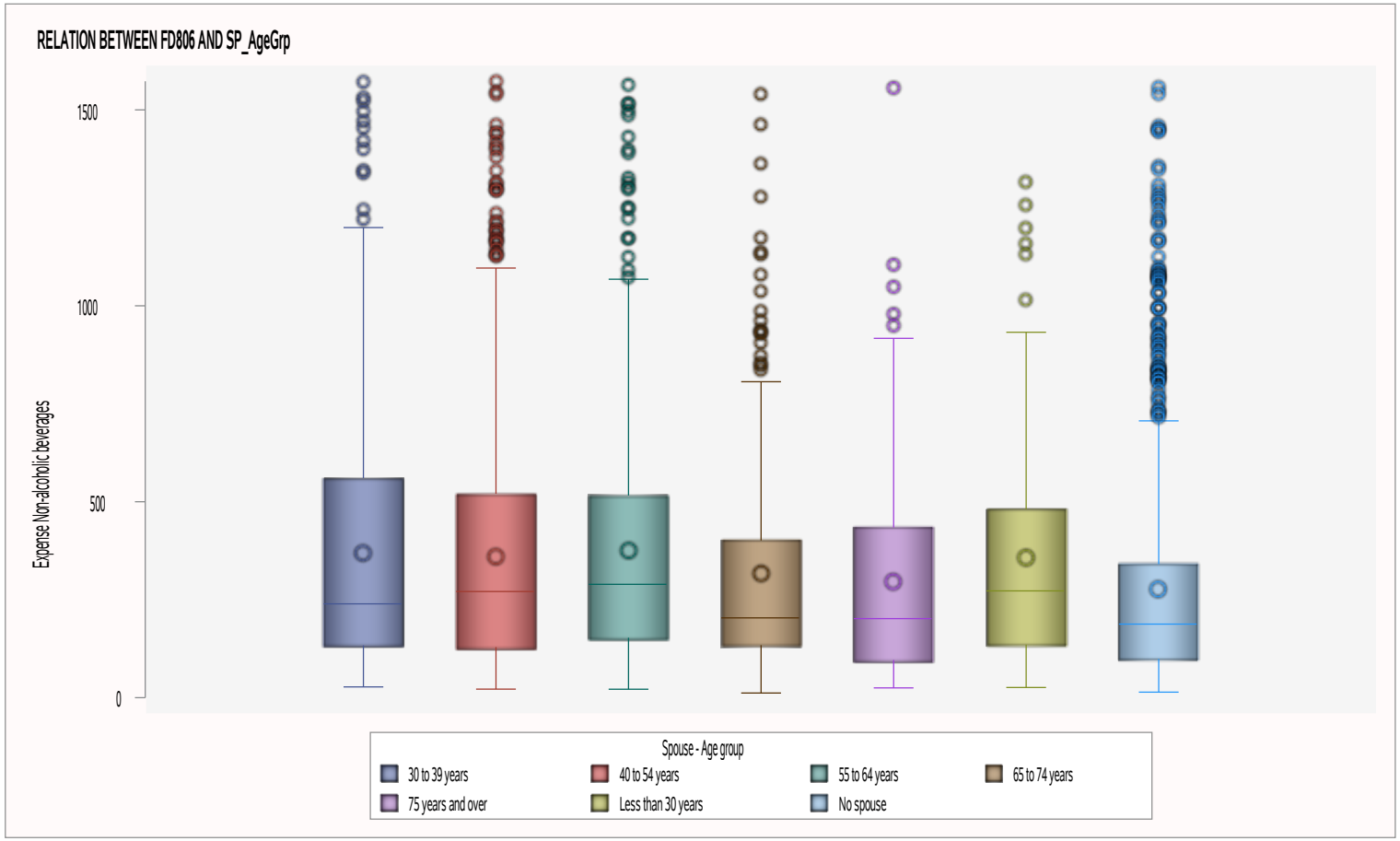
BIVARIATE ANALYSIS OF SP_AgeGrp AND FD806 FOR ANA.MODEL1
RELATION BETWEEN FD806 AND SP_AgeGrp

The MEANS Procedure

Analysis Variable : FD806 Expense Non-alcoholic beverages														
Spouse - Age group	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
30 to 39 years	1029929	1029929	0	27.44	128.80	239.41	369.03	560.17	1569.37	431.37	93.86	368.36	369.69	1.54
40 to 54 years	1667543	1667543	0	21.84	124.41	270.92	360.00	520.28	1573.00	395.87	85.01	359.53	360.46	1.36
55 to 64 years	1237655	1237655	0	21.58	148.20	289.31	374.40	516.10	1561.82	367.90	84.20	373.84	374.95	1.48
65 to 74 years	642970	642970	0	11.83	129.38	203.58	314.79	401.18	1540.64	271.80	93.21	314.07	315.51	1.94
75 years and over	329381	329381	0	24.96	91.53	201.75	296.11	434.08	1556.41	342.55	89.48	295.20	297.01	1.46
Less than 30 years	313417	313417	0	26.00	130.62	272.48	356.29	480.47	1313.72	349.85	84.23	355.24	357.34	1.44
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 - Age group=30 to 39 years

Moments			
N	1029929	**Sum Weights**	1029929
Mean	369.025814	**Sum Observations**	380070387
Std Deviation	346.381291	**Variance**	119979.999
Skewness	1.54041756	**Kurtosis**	2.02236752
Uncorrected SS	2.63827E11	**Corrected SS**	1.23571E11
Coeff Variation	93.8637024	**Std Error Mean**	0.34131139

Basic Statistical Measures			
Location		**Variability**	
Mean	369.0258	**Std Deviation**	346.38129
Median	239.4100	**Variance**	119980
Mode	293.0500	**Range**	1542
		Interquartile Range	431.37000

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

Freq: WeightD

Spouse - Age group=30 to 39 years

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	1081.2	**Pr > |t|**	<.0001
Sign	**M**	514964.5	**Pr >= |M|**	<.0001
Signed Rank	**S**	2.652E11	**Pr >= |S|**	<.0001

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.189568	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	9989.625	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	57031.7	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1569.37
99%	1494.48
95%	1145.38
90%	828.83
75% Q3	560.17
50% Median	239.41
25% Q1	128.80
10%	53.34
5%	44.55
1%	28.08
0% Min	27.44

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

Freq: WeightD

Spouse - Age group=30 to 39 years

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
27.44	9754	128	1469.40	12196	98
28.08	5931	147	1494.48	8411	169
28.86	3183	298	1522.61	59	258
31.98	1141	79	1534.26	1566	230
35.62	18	272	1569.37	4907	288

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

Freq: WeightD

Spouse - Age group=40 to 54 years

Moments			
N	1667543	Sum Weights	1667543
Mean	359.997011	Sum Observations	600310496
Std Deviation	306.024853	Variance	93651.2105
Skewness	1.3576604	Kurtosis	1.48273825
Uncorrected SS	3.72277E11	Corrected SS	1.56167E11
Coeff Variation	85.0076093	Std Error Mean	0.23698354

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

Freq: WeightD

Spouse - Age group=40 to 54 years

Basic Statistical Measures			
Location		**Variability**	
Mean	359.9970	**Std Deviation**	306.02485
Median	270.9200	**Variance**	93651
Mode	657.3000	**Range**	1551
		Interquartile Range	395.87000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	1519.08	**Pr > |t|**	<.0001
Sign	**M**	833771.5	**Pr >= |M|**	<.0001
Signed Rank	**S**	6.952E11	**Pr >= |S|**	<.0001

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.149476	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	12052.25	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	71536.48	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1573.00
99%	1300.78
95%	985.14
90%	757.64

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

Freq: WeightD

Spouse - Age group=40 to 54 years

Quantiles (Definition 5)	
Level	**Quantile**
75% Q3	520.28
50% Median	270.92
25% Q1	124.41
10%	76.70
5%	57.98
1%	30.91
0% Min	21.84

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
21.84	1372	564	1442.51	2361	604
21.84	3250	547	1462.25	45	710
23.40	132	341	1540.76	776	356
26.00	1093	413	1546.22	669	319
27.30	7416	497	1573.00	43	701

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

Freq: WeightD

Spouse - Age group=55 to 64 years

Moments			
N	1237655	**Sum Weights**	1237655
Mean	374.396247	**Sum Observations**	463373388
Std Deviation	315.246419	**Variance**	99380.3049
Skewness	1.47657637	**Kurtosis**	2.13984315
Uncorrected SS	2.96484E11	**Corrected SS**	1.22998E11
Coeff Variation	84.2012765	**Std Error Mean**	0.28336771

Basic Statistical Measures			
Location		**Variability**	
Mean	374.3962	**Std Deviation**	315.24642
Median	289.3100	**Variance**	99380
Mode	50.9600	**Range**	1540
		Interquartile Range	367.90000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	1321.238	**Pr > |t|**	<.0001
Sign	**M**	618827.5	**Pr >= |M|**	<.0001
Signed Rank	**S**	3.829E11	**Pr >= |S|**	<.0001

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

Freq: WeightD

Spouse - Age group=55 to 64 years

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.14239	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	7627.705	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	47238.31	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1561.82
99%	1485.65
95%	1022.41
90%	758.65
75% Q3	516.10
50% Median	289.31
25% Q1	148.20
10%	59.54
5%	50.96
1%	47.06
0% Min	21.58

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

Freq: WeightD

Spouse - Age group=55 to 64 years

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
21.58	347	821	1485.65	5912	981
26.26	1144	826	1494.74	58	1089
27.30	400	757	1511.64	5649	1022
29.55	846	985	1513.44	2134	770
32.14	606	789	1561.82	491	815

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

Freq: WeightD

Spouse - Age group=65 to 74 years

Moments			
N	642970	**Sum Weights**	642970
Mean	314.790434	**Sum Observations**	202400805
Std Deviation	293.420112	**Variance**	86095.3624
Skewness	1.93519119	**Kurtosis**	4.02318842
Uncorrected SS	1.1907E11	**Corrected SS**	5.53566E10
Coeff Variation	93.2112546	**Std Error Mean**	0.36592706

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

Freq: WeightD

Spouse - Age group=65 to 74 years

Basic Statistical Measures			
Location		**Variability**	
Mean	314.7904	**Std Deviation**	293.42011
Median	203.5800	**Variance**	86095
Mode	260.0000	**Range**	1529
		Interquartile Range	271.80000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	860.2546	**Pr > |t|**	<.0001
Sign	**M**	321485	**Pr >= |M|**	<.0001
Signed Rank	**S**	1.034E11	**Pr >= |S|**	<.0001

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.197389	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	7435.149	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	41715.1	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1540.64
99%	1464.24
95%	928.19
90%	798.46

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

Freq: WeightD

Spouse - Age group=65 to 74 years

Quantiles (Definition 5)	
Level	**Quantile**
75% Q3	401.18
50% Median	203.58
25% Q1	129.38
10%	63.18
5%	52.92
1%	26.26
0% Min	11.83

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
11.83	3940	1328	1174.68	502	1152
25.22	463	1196	1279.80	52	1342
26.00	975	1138	1362.61	680	1201
26.26	1551	1290	1464.24	12702	1240
27.56	442	1178	1540.64	1345	1129

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

Freq: WeightD

Spouse - Age group=75 years and over

Moments			
N	329381	**Sum Weights**	329381
Mean	296.10823	**Sum Observations**	97532424.8
Std Deviation	264.955341	**Variance**	70201.3327
Skewness	1.46043543	**Kurtosis**	2.18287343
Uncorrected SS	5.20031E10	**Corrected SS**	2.31229E10
Coeff Variation	89.4792222	**Std Error Mean**	0.46166123

Basic Statistical Measures			
Location		**Variability**	
Mean	296.1082	**Std Deviation**	264.95534
Median	201.7500	**Variance**	70201
Mode	112.8400	**Range**	1531
		Interquartile Range	342.55000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	641.3972	**Pr > |t|**	<.0001
Sign	**M**	164690.5	**Pr >= |M|**	<.0001
Signed Rank	**S**	2.712E10	**Pr >= |S|**	<.0001

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

Freq: WeightD

Spouse - Age group=75 years and over

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.180882	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	2598.705	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	15516.08	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1556.41
99%	977.08
95%	917.02
90%	630.34
75% Q3	434.08
50% Median	201.75
25% Q1	91.53
10%	51.10
5%	41.11
1%	29.64
0% Min	24.96

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

Freq: WeightD

Spouse - Age group=75 years and over

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
24.96	186	1387	950.07	1185	1398
25.48	987	1388	977.08	10418	1432
29.12	1772	1437	1048.58	558	1445
29.64	7954	1415	1103.91	800	1392
36.55	628	1380	1556.41	1360	1396

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

Freq: WeightD

Spouse - Age group=Less than 30 years

Moments			
N	313417	Sum Weights	313417
Mean	356.290901	Sum Observations	111667625
Std Deviation	300.106069	Variance	90063.6529
Skewness	1.43731221	Kurtosis	1.54615028
Uncorrected SS	6.80135E10	Corrected SS	2.82274E10
Coeff Variation	84.2306298	Std Error Mean	0.53606013

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

Freq: WeightD

Spouse - Age group=Less than 30 years

Basic Statistical Measures			
Location		**Variability**	
Mean	356.2909	**Std Deviation**	300.10607
Median	272.4800	**Variance**	90064
Mode	527.0200	**Range**	1288
		Interquartile Range	349.85000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	664.6473	**Pr > |t|**	<.0001
Sign	**M**	156708.5	**Pr >= |M|**	<.0001
Signed Rank	**S**	2.456E10	**Pr >= |S|**	<.0001

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.148445	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	2152.469	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	14896.33	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1313.72
99%	1192.36
95%	1130.03
90%	867.33

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

Freq: WeightD

Spouse - Age group=Less than 30 years

Quantiles (Definition 5)	
Level	**Quantile**
75% Q3	480.47
50% Median	272.48
25% Q1	130.62
10%	68.86
5%	36.40
1%	27.72
0% Min	26.00

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
26.00	2580	1513	1130.03	3742	1531
27.72	969	1465	1157.52	9366	1549
29.64	1988	1479	1192.36	4418	1515
34.06	3248	1522	1256.84	52	1545
36.40	11993	1490	1313.72	1691	1521

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

Freq: WeightD

Spouse - Age group=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

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

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

Freq: WeightD

Spouse - Age group=No spouse

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
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

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

Freq: WeightD

Spouse - Age group=No spouse

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
14.08	15313	1930	1447.85	46	2250
15.08	1520	2145	1449.41	505	2054
20.54	4062	1911	1456.00	442	1565
20.80	8060	1954	1536.76	836	1592
21.28	7665	1968	1557.89	759	1777

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_AgeGrp	7	30 to 39 years 40 to 54 years 55 to 64 years 65 to 74 years 75 years and over Less than 30 years 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

The GLM Procedure

Dependent Variable: FD806 Expense Non-alcoholic beverages

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	14751221204	2458536867.4	27493.8	<.0001
Error	8.13E6	726895908637	89421.530675		
Corrected Total	8.13E6	741647129841			

R-Square	Coeff Var	Root MSE	FD806 Mean
0.019890	91.48568	299.0343	326.8646

Source	DF	Type I SS	Mean Square	F Value	Pr > F
SP_AgeGrp	6	14751221204	2458536867	27493.8	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SP_AgeGrp	6	14751221204	2458536867	27493.8	<.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
SP_AgeGrp	6	5.2508E9	8.7513E8	22470.4	<.0001
Error	8.13E6	3.166E11	38945.8		

Welch's ANOVA for FD806			
Source	DF	F Value	Pr > F
SP_AgeGrp	6.0000	28293.2	<.0001
Error	1791803		

The GLM Procedure

Level of SP_AgeGrp	N	FD806	
		Mean	Std Dev
30 to 39 years	1029929	369.025814	346.381291
40 to 54 years	1667543	359.997011	306.024853
55 to 64 years	1237655	374.396247	315.246419
65 to 74 years	642970	314.790434	293.420112
75 years and over	329381	296.108230	264.955341
Less than 30 years	313417	356.290901	300.106069
No spouse	2907981	275.685049	273.455354

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

SP_AgeGrp	FD806 LSMEAN	LSMEAN Number
30 to 39 years	369.025814	1
40 to 54 years	359.997011	2
55 to 64 years	374.396247	3
65 to 74 years	314.790434	4
75 years and over	296.108230	5
Less than 30 years	356.290901	6
No spouse	275.685049	7

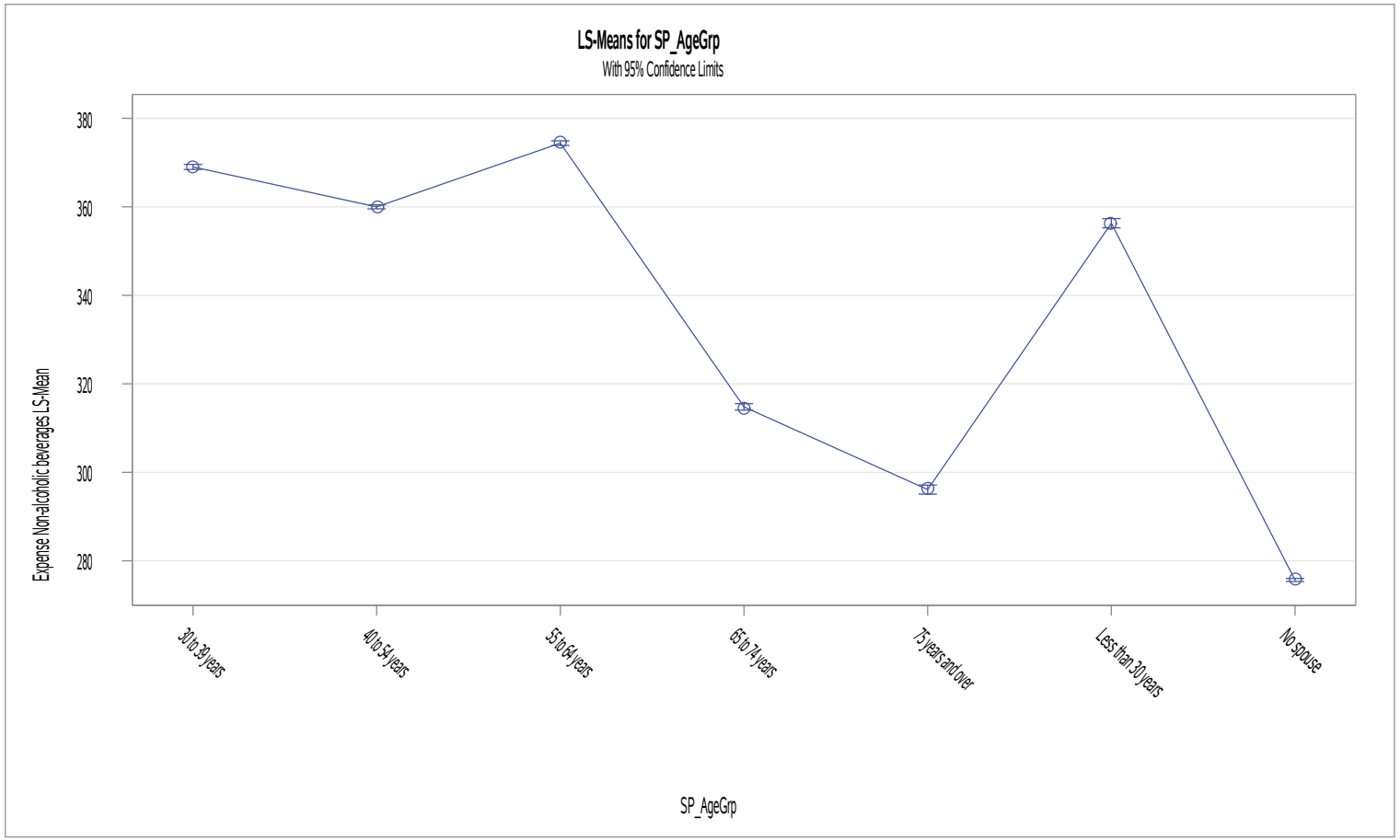
The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

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

Dependent Variable: FD806

i/j	1	2	3	4	5	6	7
1		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
2	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001
3	<.0001	<.0001		<.0001	<.0001	<.0001	<.0001
4	<.0001	<.0001	<.0001		<.0001	<.0001	<.0001
5	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001
6	<.0001	<.0001	<.0001	<.0001	<.0001		<.0001
7	<.0001	<.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