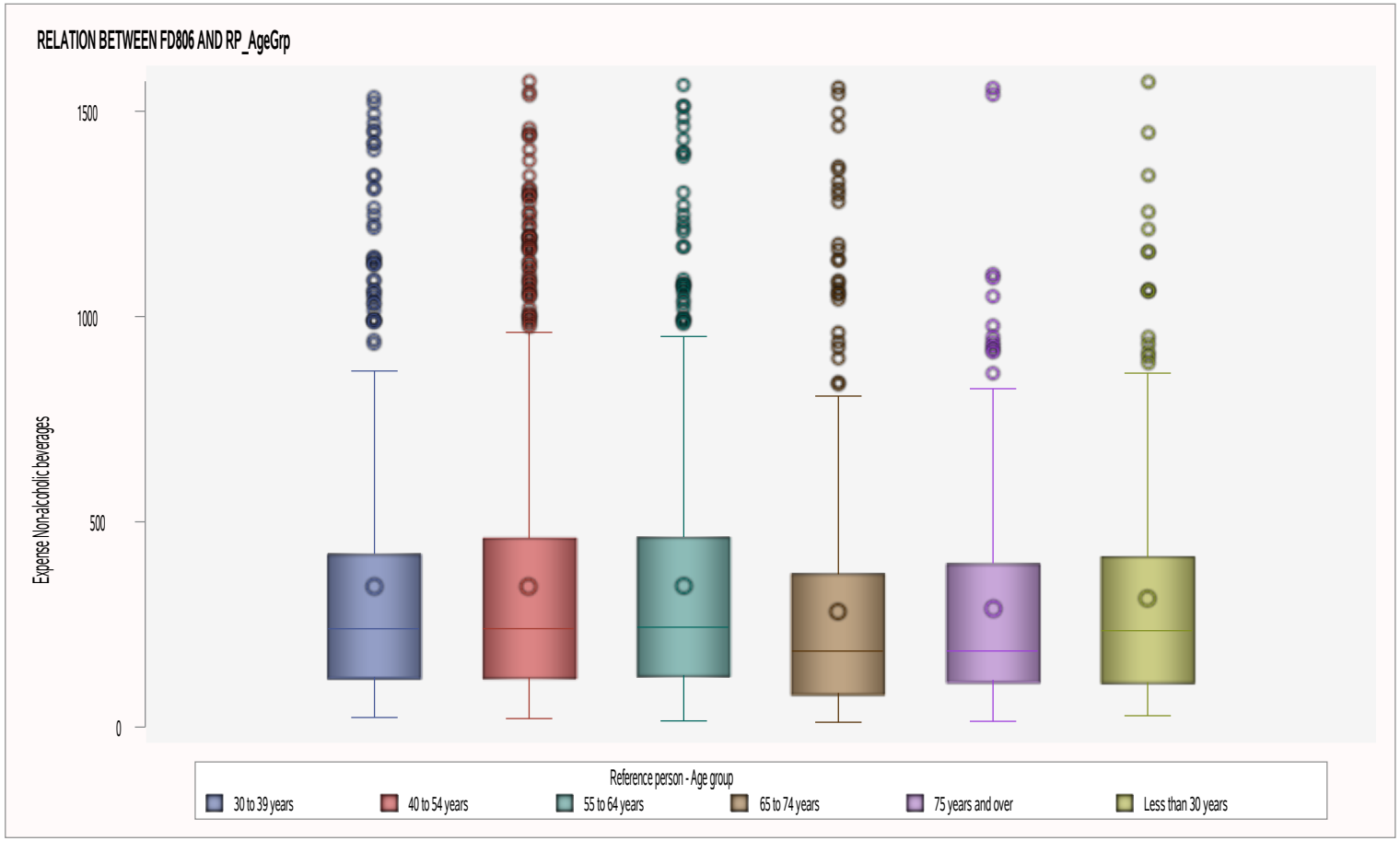
BIVARIATE ANALYSIS OF RP_AgeGrp AND FD806 FOR ANA.MODEL1
RELATION BETWEEN FD806 AND RP_AgeGrp

The MEANS Procedure

Analysis Variable : FD806 Expense Non-alcoholic beverages														
Reference person - 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	1399872	1399872	0	23.40	117.78	239.51	341.41	421.76	1534.26	303.98	92.24	340.89	341.93	1.76
40 to 54 years	2595446	2595446	0	20.80	116.48	239.72	341.19	459.41	1573.00	342.93	88.48	340.82	341.56	1.44
55 to 64 years	1693126	1693126	0	15.08	122.20	243.30	343.97	462.80	1561.82	340.60	91.36	343.49	344.44	1.55
65 to 74 years	1124539	1124539	0	11.83	78.81	185.12	281.24	372.11	1557.89	293.30	104.79	280.70	281.79	2.10
75 years and over	670281	670281	0	14.08	110.85	185.38	287.33	398.85	1556.41	288.00	91.83	286.70	287.97	1.58
Less than 30 years	645612	645612	0	27.72	105.06	234.81	313.39	413.97	1569.37	308.91	88.71	312.71	314.07	1.90



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 - Age group=30 to 39 years

Moments			
N	1399872	**Sum Weights**	1399872
Mean	341.412313	**Sum Observations**	477933537
Std Deviation	314.923089	**Variance**	99176.5519
Skewness	1.76437761	**Kurtosis**	2.951746
Uncorrected SS	3.02007E11	**Corrected SS**	1.38834E11
Coeff Variation	92.2412804	**Std Error Mean**	0.26617076

Basic Statistical Measures			
Location		**Variability**	
Mean	341.4123	**Std Deviation**	314.92309
Median	239.5100	**Variance**	99177
Mode	386.9000	**Range**	1511
		Interquartile Range	303.98000

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

Freq: WeightD

Reference person - Age group=30 to 39 years

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	1282.682	**Pr > |t|**	<.0001
Sign	**M**	699936	**Pr >= |M|**	<.0001
Signed Rank	**S**	4.899E11	**Pr >= |S|**	<.0001

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.176638	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	15212.95	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	86233.74	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1534.26
99%	1469.40
95%	1031.47
90%	777.14
75% Q3	421.76
50% Median	239.51
25% Q1	117.78
10%	72.02
5%	52.52
1%	28.86
0% Min	23.40

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

Freq: WeightD

Reference person - Age group=30 to 39 years

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
23.40	132	42	1454.18	1128	99
26.00	2580	207	1469.40	12196	220
27.44	9754	279	1494.48	8411	2
28.52	1018	134	1522.61	59	341
28.86	3183	146	1534.26	1566	196

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

Freq: WeightD

Reference person - Age group=40 to 54 years

Moments			
N	2595446	Sum Weights	2595446
Mean	341.190122	Sum Observations	885540536
Std Deviation	301.901198	Variance	91144.3333
Skewness	1.43681183	Kurtosis	1.66956865
Uncorrected SS	5.38698E11	Corrected SS	2.3656E11
Coeff Variation	88.4847417	Std Error Mean	0.18739536

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

Freq: WeightD

Reference person - Age group=40 to 54 years

Basic Statistical Measures			
Location		**Variability**	
Mean	341.1901	**Std Deviation**	301.90120
Median	239.7200	**Variance**	91144
Mode	657.3000	**Range**	1552
		Interquartile Range	342.93000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	1820.697	**Pr > |t|**	<.0001
Sign	**M**	1297723	**Pr >= |M|**	<.0001
Signed Rank	**S**	1.684E12	**Pr >= |S|**	<.0001

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.145525	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	20699.38	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	121804.7	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1573.00
99%	1300.78
95%	1009.95
90%	742.26

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

Freq: WeightD

Reference person - Age group=40 to 54 years

Quantiles (Definition 5)	
Level	**Quantile**
75% Q3	459.41
50% Median	239.72
25% Q1	116.48
10%	59.54
5%	49.44
1%	27.30
0% Min	20.80

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
20.80	8060	916	1444.43	1311	621
21.84	1372	990	1456.00	442	611
21.84	3250	415	1540.76	776	492
22.36	381	762	1546.22	669	530
24.96	1553	687	1573.00	43	1014

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

Freq: WeightD

Reference person - Age group=55 to 64 years

Moments			
N	1693126	**Sum Weights**	1693126
Mean	343.965174	**Sum Observations**	582376379
Std Deviation	314.235729	**Variance**	98744.0934
Skewness	1.54590469	**Kurtosis**	2.02279568
Uncorrected SS	3.67503E11	**Corrected SS**	1.67186E11
Coeff Variation	91.3568445	**Std Error Mean**	0.24149655

Basic Statistical Measures			
Location		**Variability**	
Mean	343.965	**Std Deviation**	314.23573
Median	243.300	**Variance**	98744
Mode	1230.110	**Range**	1547
		Interquartile Range	340.60000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	1424.307	**Pr > |t|**	<.0001
Sign	**M**	846563	**Pr >= |M|**	<.0001
Signed Rank	**S**	7.167E11	**Pr >= |S|**	<.0001

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

Freq: WeightD

Reference person - Age group=55 to 64 years

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.172793	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	16268.12	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	92730.63	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1561.82
99%	1397.07
95%	1067.82
90%	777.74
75% Q3	462.80
50% Median	243.30
25% Q1	122.20
10%	59.86
5%	52.00
1%	26.52
0% Min	15.08

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

Freq: WeightD

Reference person - Age group=55 to 64 years

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
15.08	1520	1355	1462.25	45	1554
21.28	7665	1482	1485.65	5912	1079
24.82	209	1257	1511.64	5649	1085
25.74	1311	1237	1513.44	2134	1172
26.26	1144	1130	1561.82	491	1123

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

Freq: WeightD

Reference person - Age group=65 to 74 years

Moments			
N	1124539	**Sum Weights**	1124539
Mean	281.24273	**Sum Observations**	316268418
Std Deviation	294.706163	**Variance**	86851.7223
Skewness	2.10159601	**Kurtosis**	4.74181518
Uncorrected SS	1.86616E11	**Corrected SS**	9.76681E10
Coeff Variation	104.787122	**Std Error Mean**	0.27790858

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

Freq: WeightD

Reference person - Age group=65 to 74 years

Basic Statistical Measures			
Location		**Variability**	
Mean	281.2427	**Std Deviation**	294.70616
Median	185.1200	**Variance**	86852
Mode	260.0000	**Range**	1546
		Interquartile Range	293.30000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	1011.997	**Pr > |t|**	<.0001
Sign	**M**	562269.5	**Pr >= |M|**	<.0001
Signed Rank	**S**	3.161E11	**Pr >= |S|**	<.0001

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.188683	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	13919.64	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	79676.56	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1557.89
99%	1464.24
95%	806.26
90%	633.36

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

Freq: WeightD

Reference person - Age group=65 to 74 years

Quantiles (Definition 5)	
Level	**Quantile**
75% Q3	372.11
50% Median	185.12
25% Q1	78.81
10%	51.10
5%	46.54
1%	29.64
0% Min	11.83

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
11.83	3940	1592	1362.61	680	1612
21.58	347	1633	1464.24	12702	1805
25.22	463	1665	1494.74	58	1945
25.48	987	1640	1540.64	1345	1615
26.00	975	1657	1557.89	759	1722

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

Freq: WeightD

Reference person - Age group=75 years and over

Moments			
N	670281	**Sum Weights**	670281
Mean	287.333545	**Sum Observations**	192594216
Std Deviation	263.856346	**Variance**	69620.1716
Skewness	1.58228006	**Kurtosis**	2.37703292
Uncorrected SS	1.02004E11	**Corrected SS**	4.6665E10
Coeff Variation	91.8292873	**Std Error Mean**	0.32228426

Basic Statistical Measures			
Location		**Variability**	
Mean	287.3335	**Std Deviation**	263.85635
Median	185.3800	**Variance**	69620
Mode	487.1400	**Range**	1542
		Interquartile Range	288.00000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	891.5532	**Pr > |t|**	<.0001
Sign	**M**	335140.5	**Pr >= |M|**	<.0001
Signed Rank	**S**	1.123E11	**Pr >= |S|**	<.0001

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

Freq: WeightD

Reference person - Age group=75 years and over

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.190185	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	6346.603	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	37824.56	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1556.41
99%	1095.16
95%	917.02
90%	659.49
75% Q3	398.85
50% Median	185.38
25% Q1	110.85
10%	48.36
5%	36.08
1%	14.08
0% Min	14.08

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

Freq: WeightD

Reference person - Age group=75 years and over

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
14.08	15313	2099	1048.58	558	2138
20.54	4062	2127	1095.16	5455	2129
21.58	840	2035	1103.91	800	1992
24.96	186	1980	1536.76	836	2014
28.00	6907	2056	1556.41	1360	1986

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

Freq: WeightD

Reference person - Age group=Less than 30 years

Moments			
N	645612	Sum Weights	645612
Mean	313.390897	Sum Observations	202328924
Std Deviation	278.024624	Variance	77297.6915
Skewness	1.89878152	Kurtosis	4.23407486
Uncorrected SS	1.13312E11	Corrected SS	4.99042E10
Coeff Variation	88.7149647	Std Error Mean	0.34601702

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

Freq: WeightD

Reference person - Age group=Less than 30 years

Basic Statistical Measures			
Location		**Variability**	
Mean	313.3909	**Std Deviation**	278.02462
Median	234.8100	**Variance**	77298
Mode	105.0600	**Range**	1542
		Interquartile Range	308.91000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	905.7095	**Pr > |t|**	<.0001
Sign	**M**	322806	**Pr >= |M|**	<.0001
Signed Rank	**S**	1.042E11	**Pr >= |S|**	<.0001

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.18209	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	5269.803	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	32176.04	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1569.37
99%	1211.03
95%	862.16
90%	652.95

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

Freq: WeightD

Reference person - Age group=Less than 30 years

Quantiles (Definition 5)	
Level	**Quantile**
75% Q3	413.97
50% Median	234.81
25% Q1	105.06
10%	69.16
5%	44.22
1%	30.16
0% Min	27.72

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
27.72	969	2189	1211.03	7410	2288
29.64	1988	2194	1256.84	52	2319
30.16	12248	2290	1345.76	1074	2199
35.97	838	2306	1447.85	46	2323
36.40	11993	2258	1569.37	4907	2230

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_AgeGrp	6	30 to 39 years 40 to 54 years 55 to 64 years 65 to 74 years 75 years and over Less than 30 years

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	4829243074.5	965848614.9	10655.6	<.0001
Error	8.13E6	736817886767	90642.10484		
Corrected Total	8.13E6	741647129841			

R-Square	Coeff Var	Root MSE	FD806 Mean
0.006512	92.10794	301.0683	326.8646

Source	DF	Type I SS	Mean Square	F Value	Pr > F
RP_AgeGrp	5	4829243074	965848615	10655.6	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
RP_AgeGrp	5	4829243074	965848615	10655.6	<.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_AgeGrp	5	1.5922E9	3.1843E8	8054.15	<.0001
Error	8.13E6	3.214E11	39536.7		

Welch's ANOVA for FD806			
Source	**DF**	**F Value**	**Pr > F**
RP_AgeGrp	5.0000	11401.8	<.0001
Error	2777219		

The GLM Procedure

Level of RP_AgeGrp	**N**	**FD806**	
		Mean	**Std Dev**
30 to 39 years	1399872	341.412313	314.923089
40 to 54 years	2595446	341.190122	301.901198
55 to 64 years	1693126	343.965174	314.235729
65 to 74 years	1124539	281.242730	294.706163
75 years and over	670281	287.333545	263.856346
Less than 30 years	645612	313.390897	278.024624

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

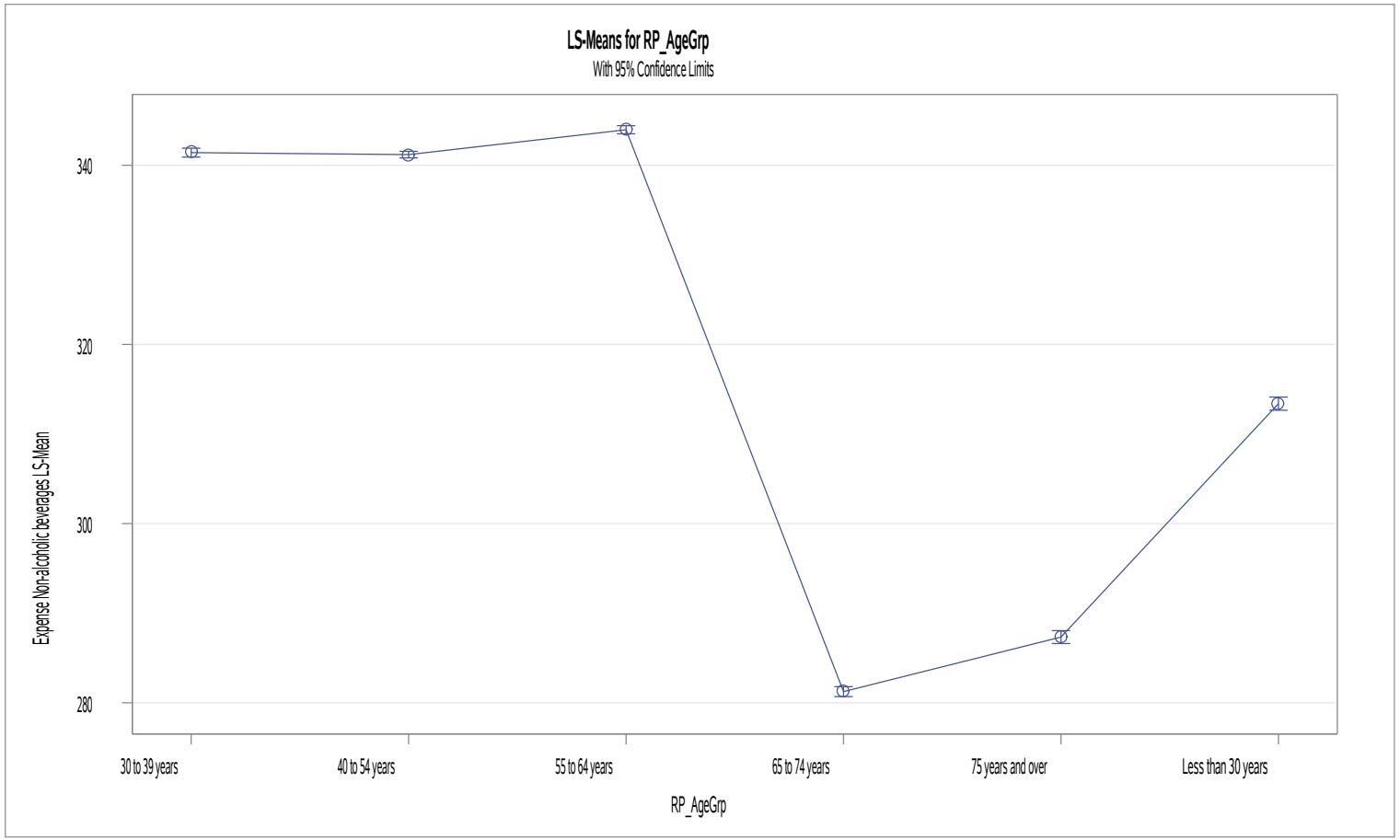
RP_AgeGrp	FD806 LSMEAN	LSMEAN Number
30 to 39 years	341.412313	1
40 to 54 years	341.190122	2
55 to 64 years	343.965174	3
65 to 74 years	281.242730	4
75 years and over	287.333545	5
Less than 30 years	313.390897	6

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

Dependent Variable: FD806

i/j	1	2	3	4	5	6
1		0.9816	<.0001	<.0001	<.0001	<.0001
2	0.9816		<.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



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