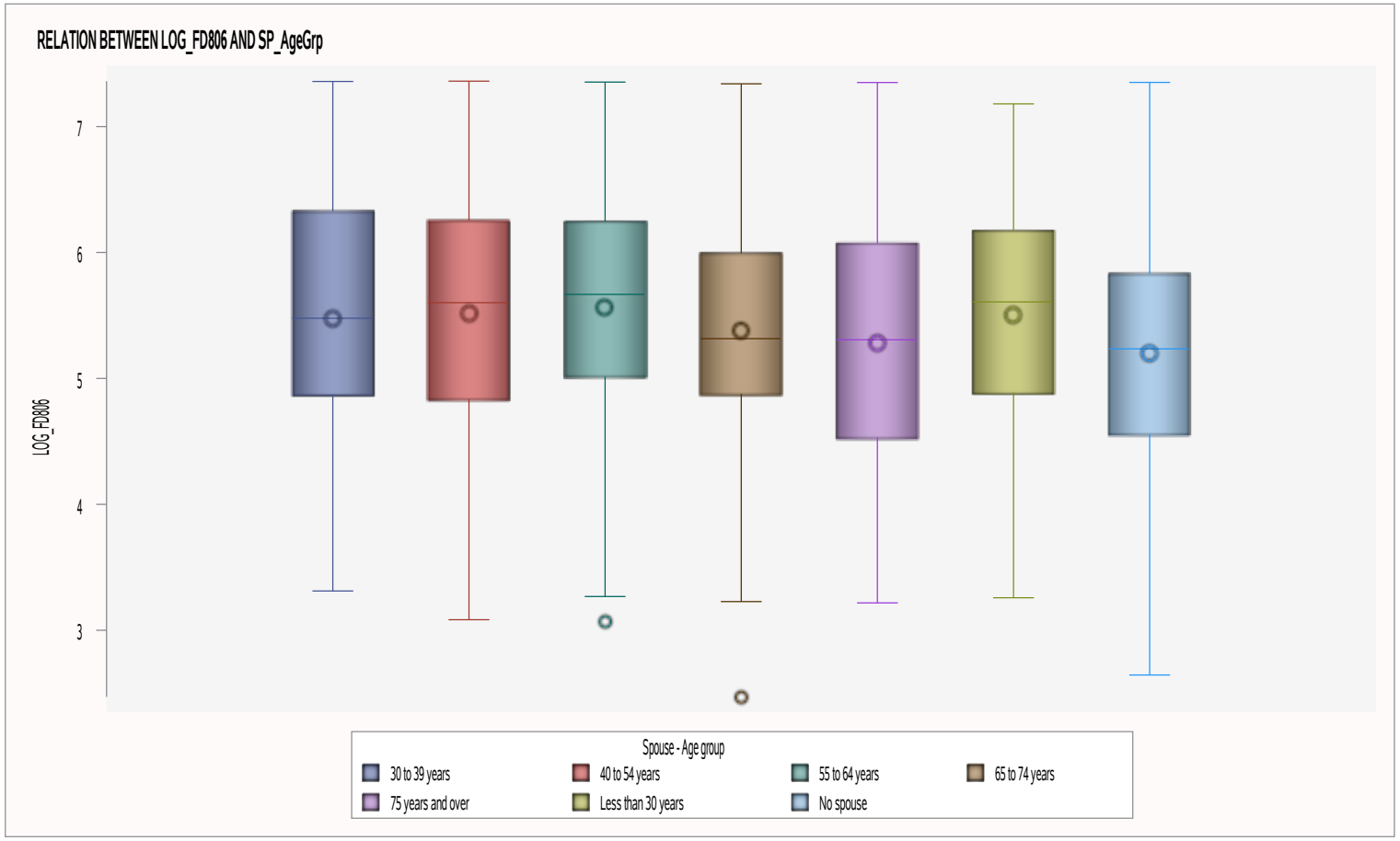
BIVARIATE ANALYSIS OF SP_AgeGrp AND LOG_FD806 FOR ANA.MODEL2
RELATION BETWEEN LOG_FD806 AND SP_AgeGrp

The MEANS Procedure

Analysis Variable : LOG_FD806														
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	3.31	4.86	5.48	5.47	6.33	7.36	1.47	17.93	5.47	5.48	-0.14
40 to 54 years	1667543	1667543	0	3.08	4.82	5.60	5.51	6.25	7.36	1.43	16.42	5.51	5.52	-0.19
55 to 64 years	1237655	1237655	0	3.07	5.00	5.67	5.56	6.25	7.35	1.25	16.36	5.56	5.56	-0.27
65 to 74 years	642970	642970	0	2.47	4.86	5.32	5.38	5.99	7.34	1.13	16.48	5.38	5.38	-0.13
75 years and over	329381	329381	0	3.22	4.52	5.31	5.28	6.07	7.35	1.56	18.03	5.28	5.28	-0.13
Less than 30 years	313417	313417	0	3.26	4.87	5.61	5.50	6.17	7.18	1.30	16.97	5.50	5.51	-0.44
No spouse	2907981	2907981	0	2.64	4.54	5.24	5.20	5.83	7.35	1.29	18.07	5.20	5.20	-0.06



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: LOG_FD806

Freq: WeightD

Spouse - Age group=30 to 39 years

Moments			
N	1029929	**Sum Weights**	1029929
Mean	5.47333516	**Sum Observations**	5637146.61
Std Deviation	0.98154514	**Variance**	0.96343087
Skewness	-0.1409459	**Kurtosis**	-0.7342995
Uncorrected SS	31846257.1	**Corrected SS**	992264.429
Coeff Variation	17.9332183	**Std Error Mean**	0.00096718

Basic Statistical Measures			
Location		**Variability**	
Mean	5.473335	**Std Deviation**	0.98155
Median	5.478178	**Variance**	0.96343
Mode	5.680343	**Range**	4.04643
		Interquartile Range	1.46998

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Age group=30 to 39 years

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	5659.074	**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**	0.062989	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	515.3351	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	4271.325	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	7.35843
99%	7.30953
95%	7.04349
90%	6.72002
75% Q3	6.32824
50% Median	5.47818
25% Q1	4.85826
10%	3.97669
5%	3.79661
1%	3.33506
0% Min	3.31200

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Age group=30 to 39 years

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
3.31200	9754	123	7.29261	12196	99
3.33506	5931	148	7.30953	8411	169
3.36246	3183	298	7.32818	59	256
3.46511	1141	79	7.33580	1566	230
3.57291	18	275	7.35843	4907	287

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Age group=40 to 54 years

Moments			
N	1667543	Sum Weights	1667543
Mean	5.51482803	Sum Observations	9196212.88
Std Deviation	0.90531471	Variance	0.81959472
Skewness	-0.1882319	Kurtosis	-0.7007128
Uncorrected SS	52082241.2	Corrected SS	1366708.62
Coeff Variation	16.4160098	Std Error Mean	0.00070107

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Age group=40 to 54 years

Basic Statistical Measures			
Location		**Variability**	
Mean	5.514828	**Std Deviation**	0.90531
Median	5.601824	**Variance**	0.81959
Mode	6.488141	**Range**	4.27700
		Interquartile Range	1.43078

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	7866.308	**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**	0.050564	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	1130.841	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	7464.001	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	7.36074
99%	7.17072
95%	6.89278
90%	6.63021

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Age group=40 to 54 years

Quantiles (Definition 5)	
Level	**Quantile**
75% Q3	6.25437
50% Median	5.60182
25% Q1	4.82358
10%	4.33990
5%	4.06010
1%	3.43108
0% Min	3.08374

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
3.08374	1372	563	7.27414	2361	606
3.08374	3250	545	7.28773	45	711
3.15274	132	341	7.34003	776	361
3.25810	1093	413	7.34357	669	306
3.30689	7416	499	7.36074	43	701

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Age group=55 to 64 years

Moments			
N	1237655	**Sum Weights**	1237655
Mean	5.55861908	**Sum Observations**	6879652.69
Std Deviation	0.90933558	**Variance**	0.82689119
Skewness	-0.2745805	**Kurtosis**	-0.714123
Uncorrected SS	39264773.9	**Corrected SS**	1023405.19
Coeff Variation	16.3590195	**Std Error Mean**	0.00081738

Basic Statistical Measures			
Location		**Variability**	
Mean	5.558619	**Std Deviation**	0.90934
Median	5.667499	**Variance**	0.82689
Mode	3.931041	**Range**	4.28184
		Interquartile Range	1.24774

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	6800.526	**Pr > |t|**	<.0001
Sign	**M**	618827.5	**Pr >= |M|**	<.0001
Signed Rank	**S**	3.829E11	**Pr >= |S|**	<.0001

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Age group=55 to 64 years

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.0712	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	1196.006	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	9385.734	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	7.35361
99%	7.30361
95%	6.92992
90%	6.63154
75% Q3	6.24630
50% Median	5.66750
25% Q1	4.99856
10%	4.08665
5%	3.93104
1%	3.85142
0% Min	3.07177

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Age group=55 to 64 years

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
3.07177	347	822	7.30361	5912	981
3.26805	1144	826	7.30971	58	1089
3.30689	400	758	7.32095	5649	1021
3.38608	846	985	7.32214	2134	769
3.47010	606	786	7.35361	491	814

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Age group=65 to 74 years

Moments			
N	642970	**Sum Weights**	642970
Mean	5.37779181	**Sum Observations**	3457758.8
Std Deviation	0.88643032	**Variance**	0.78575872
Skewness	-0.1254675	**Kurtosis**	-0.0499838
Uncorrected SS	19100325.5	**Corrected SS**	505218.495
Coeff Variation	16.4831654	**Std Error Mean**	0.00110548

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Age group=65 to 74 years

Basic Statistical Measures			
Location		**Variability**	
Mean	5.377792	**Std Deviation**	0.88643
Median	5.316059	**Variance**	0.78576
Mode	5.560682	**Range**	4.86931
		Interquartile Range	1.13166

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	4864.685	**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**	0.042582	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	200.3127	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	1359.2	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	7.33995
99%	7.28909
95%	6.83324
90%	6.68268

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Age group=65 to 74 years

Quantiles (Definition 5)	
Level	**Quantile**
75% Q3	5.99441
50% Median	5.31606
25% Q1	4.86275
10%	4.14599
5%	3.96878
1%	3.26805
0% Min	2.47064

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
2.47064	3940	1330	7.06875	502	1152
3.22764	463	1198	7.15446	52	1341
3.25810	975	1133	7.21716	680	1199
3.26805	1551	1290	7.28909	12702	1240
3.31637	442	1189	7.33995	1345	1129

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Age group=75 years and over

Moments			
N	329381	**Sum Weights**	329381
Mean	5.27988894	**Sum Observations**	1739095.1
Std Deviation	0.95193079	**Variance**	0.90617223
Skewness	-0.1292743	**Kurtosis**	-0.9024807
Uncorrected SS	9480703.98	**Corrected SS**	298475.01
Coeff Variation	18.0293715	**Std Error Mean**	0.00165866

Basic Statistical Measures			
Location		**Variability**	
Mean	5.279889	**Std Deviation**	0.95193
Median	5.307029	**Variance**	0.90617
Mode	4.725971	**Range**	4.13286
		Interquartile Range	1.55656

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	3183.235	**Pr > |t|**	<.0001
Sign	**M**	164690.5	**Pr >= |M|**	<.0001
Signed Rank	**S**	2.712E10	**Pr >= |S|**	<.0001

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Age group=75 years and over

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.079223	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	336.1111	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	2279.36	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	7.35014
99%	6.88457
95%	6.82113
90%	6.44626
75% Q3	6.07323
50% Median	5.30703
25% Q1	4.51667
10%	3.93378
5%	3.71625
1%	3.38912
0% Min	3.21727

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Age group=75 years and over

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
3.21727	186	1387	6.85654	1185	1396
3.23789	987	1389	6.88457	10418	1431
3.37143	1772	1439	6.95519	558	1445
3.38912	7954	1415	7.00661	800	1391
3.59868	628	1380	7.35014	1360	1398

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Age group=Less than 30 years

Moments			
N	313417	**Sum Weights**	313417
Mean	5.50197993	**Sum Observations**	1724414.04
Std Deviation	0.93349878	**Variance**	0.87141996
Skewness	-0.438389	**Kurtosis**	-0.343148
Uncorrected SS	9760808.41	**Corrected SS**	273116.96
Coeff Variation	16.9665973	**Std Error Mean**	0.00166745

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Age group=Less than 30 years

Basic Statistical Measures			
Location		**Variability**	
Mean	5.501980	**Std Deviation**	0.93350
Median	5.607565	**Variance**	0.87142
Mode	6.267238	**Range**	3.92252
		Interquartile Range	1.30247

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	3299.64	**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**	0.084817	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	513.1916	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	3191.598	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	7.18062
99%	7.08369
95%	7.03000
90%	6.76542

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Age group=Less than 30 years

Quantiles (Definition 5)	
Level	**Quantile**
75% Q3	6.17476
50% Median	5.60757
25% Q1	4.87229
10%	4.23208
5%	3.59457
1%	3.32215
0% Min	3.25810

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
3.25810	2580	1513	7.03000	3742	1530
3.32215	969	1465	7.05404	9366	1550
3.38912	1988	1480	7.08369	4418	1515
3.52812	3248	1522	7.13636	52	1545
3.59457	11993	1490	7.18062	1691	1521

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Age group=No spouse

Moments			
N	2907981	**Sum Weights**	2907981
Mean	5.19840979	**Sum Observations**	15116876.9
Std Deviation	0.93942162	**Variance**	0.88251299
Skewness	-0.0590322	**Kurtosis**	-0.3875657
Uncorrected SS	81150051.1	**Corrected SS**	2566330.12
Coeff Variation	18.0713268	**Std Error Mean**	0.00055089

Basic Statistical Measures			
Location		**Variability**	
Mean	5.198410	**Std Deviation**	0.93942
Median	5.235165	**Variance**	0.88251
Mode	5.958166	**Range**	4.70633
		Interquartile Range	1.28658

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	9436.387	**Pr > |t|**	<.0001
Sign	**M**	1453991	**Pr >= |M|**	<.0001
Signed Rank	**S**	2.114E12	**Pr >= |S|**	<.0001

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Age group=No spouse

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.034118	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	364.6552	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	2814.721	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	7.35109
99%	7.11486
95%	6.81928
90%	6.40723
75% Q3	5.83115
50% Median	5.23516
25% Q1	4.54457
10%	3.95891
5%	3.66356
1%	3.05777
0% Min	2.64476

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Age group=No spouse

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
2.64476	15313	1930	7.27783	46	2250
2.71337	1520	2145	7.27891	505	2054
3.02237	4062	1911	7.28345	442	1565
3.03495	8060	1954	7.33743	836	1592
3.05777	7665	1968	7.35109	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: LOG_FD806

Frequency: WeightD

The GLM Procedure

Dependent Variable: LOG_FD806

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	182592.191	30432.032	35211.3	<.0001
Error	8.13E6	7025518.822	0.864		
Corrected Total	8.13E6	7208111.014			

R-Square	Coeff Var	Root MSE	LOG_FD806 Mean
0.025331	17.27289	0.929660	5.382190

Source	DF	Type I SS	Mean Square	F Value	Pr > F
SP_AgeGrp	6	182592.1915	30432.0319	35211.3	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SP_AgeGrp	6	182592.1915	30432.0319	35211.3	<.0001

The GLM Procedure

Levene's Test for Homogeneity of LOG_FD806 Variance ANOVA of Absolute Deviations from Group Means					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
SP_AgeGrp	6	5076.0	846.0	2973.59	<.0001
Error	8.13E6	2312718	0.2845		

Welch's ANOVA for LOG_FD806			
Source	**DF**	**F Value**	**Pr > F**
SP_AgeGrp	6.0000	35242.5	<.0001
Error	1796535		

The GLM Procedure

Level of SP_AgeGrp	N	LOG_FD806	
		Mean	Std Dev
30 to 39 years	1029929	5.47333516	0.98154514
40 to 54 years	1667543	5.51482803	0.90531471
55 to 64 years	1237655	5.55861908	0.90933558
65 to 74 years	642970	5.37779181	0.88643032
75 years and over	329381	5.27988894	0.95193079
Less than 30 years	313417	5.50197993	0.93349878
No spouse	2907981	5.19840979	0.93942162

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

SP_AgeGrp	LOG_FD806 LSMEAN	LSMEAN Number
30 to 39 years	5.47333516	1
40 to 54 years	5.51482803	2
55 to 64 years	5.55861908	3
65 to 74 years	5.37779181	4
75 years and over	5.27988894	5
Less than 30 years	5.50197993	6
No spouse	5.19840979	7

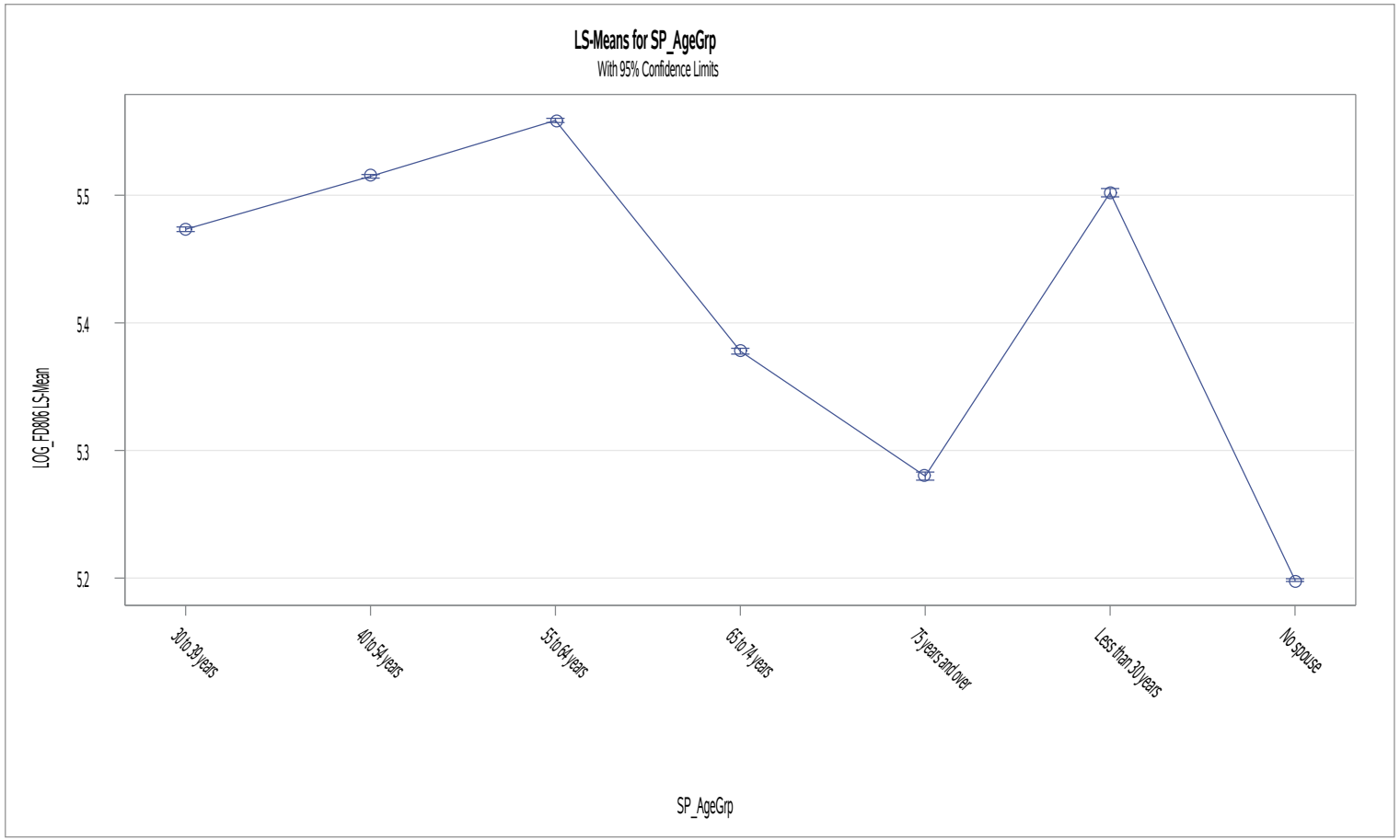
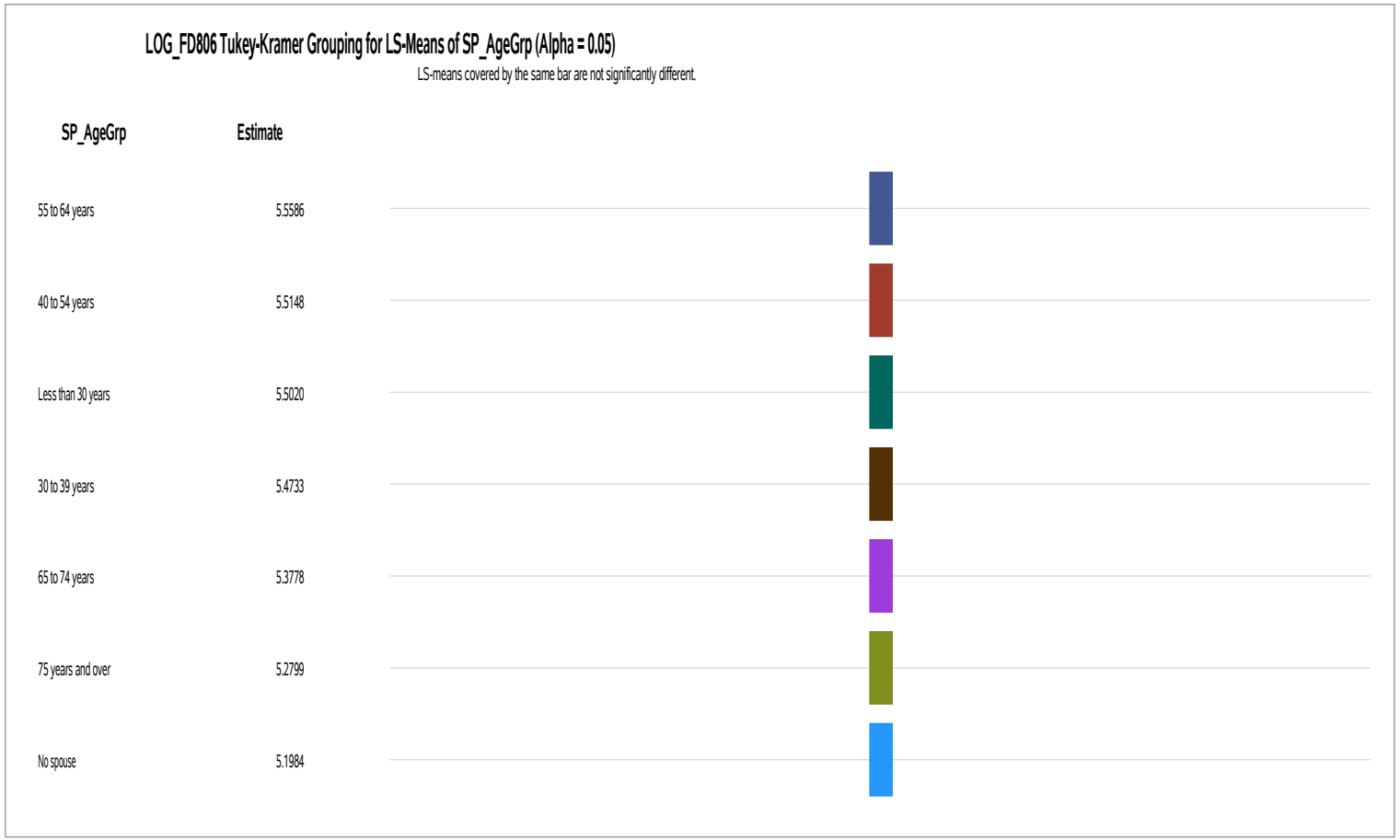
The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

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

Dependent Variable: LOG_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