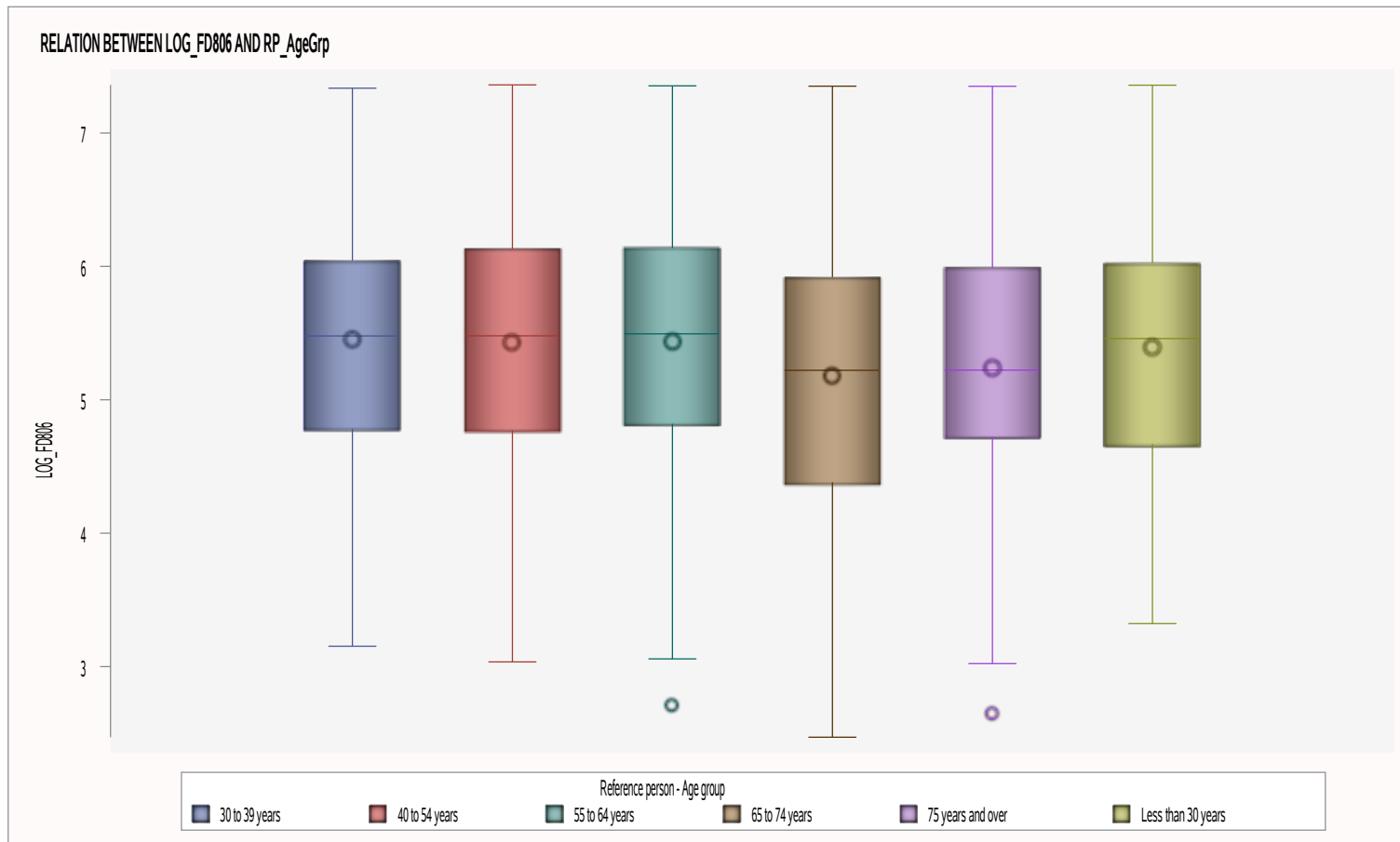


**BIVARIATE ANALYSIS OF RP_AgeGrp AND LOG_FD806 FOR ANA.MODEL2
RELATION BETWEEN LOG_FD806 AND RP_AgeGrp**

11:42 Saturday, November 20, 2021 1

The MEANS Procedure

Analysis Variable : LOG_FD806														
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	3.15	4.77	5.48	5.46	6.04	7.34	1.28	16.28	5.45	5.46	-0.03
40 to 54 years	2595446	2595446	0	3.03	4.76	5.48	5.43	6.13	7.36	1.37	17.40	5.43	5.43	-0.23
55 to 64 years	1693126	1693126	0	2.71	4.81	5.49	5.44	6.14	7.35	1.33	17.29	5.43	5.44	-0.14
65 to 74 years	1124539	1124539	0	2.47	4.37	5.22	5.18	5.92	7.35	1.55	18.77	5.18	5.18	0.09
75 years and over	670281	670281	0	2.64	4.71	5.22	5.24	5.99	7.35	1.28	18.79	5.24	5.24	-0.37
Less than 30 years	645612	645612	0	3.32	4.65	5.46	5.39	6.03	7.36	1.37	16.30	5.39	5.39	-0.16



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 Kolmogorov 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

Reference person - Age group=30 to 39 years

Moments			
N	1399872	Sum Weights	1399872
Mean	5.45501049	Sum Observations	7636316.45
Std Deviation	0.88833792	Variance	0.78914425
Skewness	-0.0277237	Kurtosis	-0.5363713
Uncorrected SS	42760886.5	Corrected SS	1104700.16
Coeff Variation	16.2848067	Std Error Mean	0.00075082

Basic Statistical Measures			
Location		Variability	
Mean	5.455010	Std Deviation	0.88834
Median	5.478595	Variance	0.78914
Mode	5.958166	Range	4.18307
		Interquartile Range	1.27562

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Age group=30 to 39 years

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	7265.434	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	0.032795	Pr > D	<0.0100
Cramer-von Mises	W-Sq	279.4447	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	2269.057	Pr > A-Sq	<0.0050

Quantiles (Definition 5)	
Level	Quantile
100% Max	7.33580
99%	7.29261
95%	6.93874
90%	6.65562
75% Q3	6.04444
50% Median	5.47860
25% Q1	4.76882
10%	4.27694
5%	3.96119
1%	3.36246
0% Min	3.15274

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Age group=30 to 39 years

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
3.15274	132	42	7.28220	1128	99
3.25810	2580	207	7.29261	12196	220
3.31200	9754	279	7.30953	8411	2
3.35061	1018	134	7.32818	59	341
3.36246	3183	146	7.33580	1566	196

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Age group=40 to 54 years

Moments			
N	2595446	Sum Weights	2595446
Mean	5.43239591	Sum Observations	14099490.2
Std Deviation	0.94537734	Variance	0.89373831
Skewness	-0.226726	Kurtosis	-0.6273494
Uncorrected SS	78913661.8	Corrected SS	2319648.62
Coeff Variation	17.4025854	Std Error Mean	0.00058681

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Age group=40 to 54 years

Basic Statistical Measures			
Location		Variability	
Mean	5.432396	Std Deviation	0.94538
Median	5.479472	Variance	0.89374
Mode	6.488141	Range	4.32579
		Interquartile Range	1.37222

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	9257.468	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	0.041734	Pr > D	<0.0100
Cramer-von Mises	W-Sq	1057.332	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	7707.702	Pr > A-Sq	<0.0050

Quantiles (Definition 5)	
Level	Quantile
100% Max	7.36074
99%	7.17072
95%	6.91766
90%	6.60970

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Age group=40 to 54 years

Quantiles (Definition 5)	
Level	Quantile
75% Q3	6.12994
50% Median	5.47947
25% Q1	4.75772
10%	4.08665
5%	3.90076
1%	3.30689
0% Min	3.03495

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
3.03495	8060	916	7.27547	1311	621
3.08374	1372	990	7.28345	442	611
3.08374	3250	415	7.34003	776	492
3.10727	381	762	7.34357	669	530
3.21727	1553	687	7.36074	43	1014

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Age group=55 to 64 years

Moments			
N	1693126	Sum Weights	1693126
Mean	5.43501741	Sum Observations	9202169.29
Std Deviation	0.93966324	Variance	0.88296701
Skewness	-0.1408492	Kurtosis	-0.6203097
Uncorrected SS	51508923.8	Corrected SS	1494973.51
Coeff Variation	17.2890567	Std Error Mean	0.00072215

Basic Statistical Measures			
Location		Variability	
Mean	5.435017	Std Deviation	0.93966
Median	5.494295	Variance	0.88297
Mode	7.114859	Range	4.64024
		Interquartile Range	1.33164

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	7526.158	Pr > t 	<.0001
Sign	M	846563	Pr >= M 	<.0001
Signed Rank	S	7.167E11	Pr >= S 	<.0001

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Age group=55 to 64 years

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.045528	Pr > D	<0.0100
Cramer-von Mises	W-Sq	572.5071	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	4947.638	Pr > A-Sq	<0.0050

Quantiles (Definition 5)	
Level	Quantile
100% Max	7.35361
99%	7.24213
95%	6.97337
90%	6.65639
75% Q3	6.13729
50% Median	5.49430
25% Q1	4.80566
10%	4.09201
5%	3.95124
1%	3.27790
0% Min	2.71337

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Age group=55 to 64 years

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
2.71337	1520	1355	7.28773	45	1554
3.05777	7665	1482	7.30361	5912	1079
3.21165	209	1257	7.32095	5649	1085
3.24805	1311	1237	7.32214	2134	1172
3.26805	1144	1130	7.35361	491	1123

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Age group=65 to 74 years

Moments			
N	1124539	Sum Weights	1124539
Mean	5.17856193	Sum Observations	5823494.85
Std Deviation	0.97185179	Variance	0.9444959
Skewness	0.0934121	Kurtosis	-0.7052274
Uncorrected SS	31219450.3	Corrected SS	1062121.53
Coeff Variation	18.7668276	Std Error Mean	0.00091646

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Age group=65 to 74 years

Basic Statistical Measures			
Location		Variability	
Mean	5.178562	Std Deviation	0.97185
Median	5.221004	Variance	0.94450
Mode	5.560682	Range	4.88045
		Interquartile Range	1.55215

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	5650.624	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	0.060224	Pr > D	<0.0100
Cramer-von Mises	W-Sq	743.6633	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	5697.853	Pr > A-Sq	<0.0050

Quantiles (Definition 5)	
Level	Quantile
100% Max	7.35109
99%	7.28909
95%	6.69241
90%	6.45104

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Age group=65 to 74 years

Quantiles (Definition 5)	
Level	Quantile
75% Q3	5.91919
50% Median	5.22100
25% Q1	4.36704
10%	3.93378
5%	3.84031
1%	3.38912
0% Min	2.47064

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
2.47064	3940	1592	7.21716	680	1612
3.07177	347	1633	7.28909	12702	1805
3.22764	463	1665	7.30971	58	1945
3.23789	987	1640	7.33995	1345	1615
3.25810	975	1657	7.35109	759	1722

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Age group=75 years and over

Moments			
N	670281	Sum Weights	670281
Mean	5.23742365	Sum Observations	3510545.56
Std Deviation	0.98415502	Variance	0.9685611
Skewness	-0.3698554	Kurtosis	-0.1377097
Uncorrected SS	19035421.5	Corrected SS	649207.131
Coeff Variation	18.7908232	Std Error Mean	0.00120208

Basic Statistical Measures			
Location		Variability	
Mean	5.237424	Std Deviation	0.98416
Median	5.222408	Variance	0.96856
Mode	6.188552	Range	4.70538
		Interquartile Range	1.28041

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	4356.951	Pr > t 	<.0001
Sign	M	335140.5	Pr >= M 	<.0001
Signed Rank	S	1.123E11	Pr >= S 	<.0001

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Age group=75 years and over

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.060496	Pr > D	<0.0100
Cramer-von Mises	W-Sq	411.4269	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	3004.562	Pr > A-Sq	<0.0050

Quantiles (Definition 5)	
Level	Quantile
100% Max	7.35014
99%	6.99866
95%	6.82113
90%	6.49147
75% Q3	5.98859
50% Median	5.22241
25% Q1	4.70818
10%	3.87867
5%	3.58574
1%	2.64476
0% Min	2.64476

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Age group=75 years and over

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
2.64476	15313	2099	6.95519	558	2138
3.02237	4062	2127	6.99866	5455	2129
3.07177	840	2035	7.00661	800	1992
3.21727	186	1980	7.33743	836	2014
3.33220	6907	2056	7.35014	1360	1986

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Age group=Less than 30 years

Moments			
N	645612	Sum Weights	645612
Mean	5.38890329	Sum Observations	3479140.63
Std Deviation	0.87832954	Variance	0.77146278
Skewness	-0.1637828	Kurtosis	-0.4783547
Uncorrected SS	19246817.3	Corrected SS	498064.856
Coeff Variation	16.2988551	Std Error Mean	0.00109313

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Age group=Less than 30 years

Basic Statistical Measures			
Location		Variability	
Mean	5.388903	Std Deviation	0.87833
Median	5.458777	Variance	0.77146
Mode	4.654532	Range	4.03628
		Interquartile Range	1.37126

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	4929.793	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	0.082811	Pr > D	<0.0100
Cramer-von Mises	W-Sq	602.5488	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	3129.265	Pr > A-Sq	<0.0050

Quantiles (Definition 5)	
Level	Quantile
100% Max	7.35843
99%	7.09923
95%	6.75944
90%	6.48150

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Age group=Less than 30 years

Quantiles (Definition 5)	
Level	Quantile
75% Q3	6.02579
50% Median	5.45878
25% Q1	4.65453
10%	4.23642
5%	3.78918
1%	3.40652
0% Min	3.32215

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
3.32215	969	2189	7.09923	7410	2288
3.38912	1988	2194	7.13636	52	2319
3.40652	12248	2290	7.20471	1074	2199
3.58269	838	2306	7.27783	46	2323
3.59457	11993	2258	7.35843	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: LOG_FD806

Frequency: WeightD

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	79395.211	15879.042	18106.9	<.0001
Error	8.13E6	7128715.803	0.877		
Corrected Total	8.13E6	7208111.014			

R-Square	Coeff Var	Root MSE	LOG_FD806 Mean
0.011015	17.39929	0.936463	5.382190

Source	DF	Type I SS	Mean Square	F Value	Pr > F
RP_AgeGrp	5	79395.21071	15879.04214	18106.9	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
RP_AgeGrp	5	79395.21071	15879.04214	18106.9	<.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
RP_AgeGrp	5	6310.5	1262.1	4419.43	<.0001
Error	8.13E6	2321432	0.2856		

Welch's ANOVA for LOG_FD806			
Source	DF	F Value	Pr > F
RP_AgeGrp	5.0000	17164.4	<.0001
Error	2737876		

The GLM Procedure

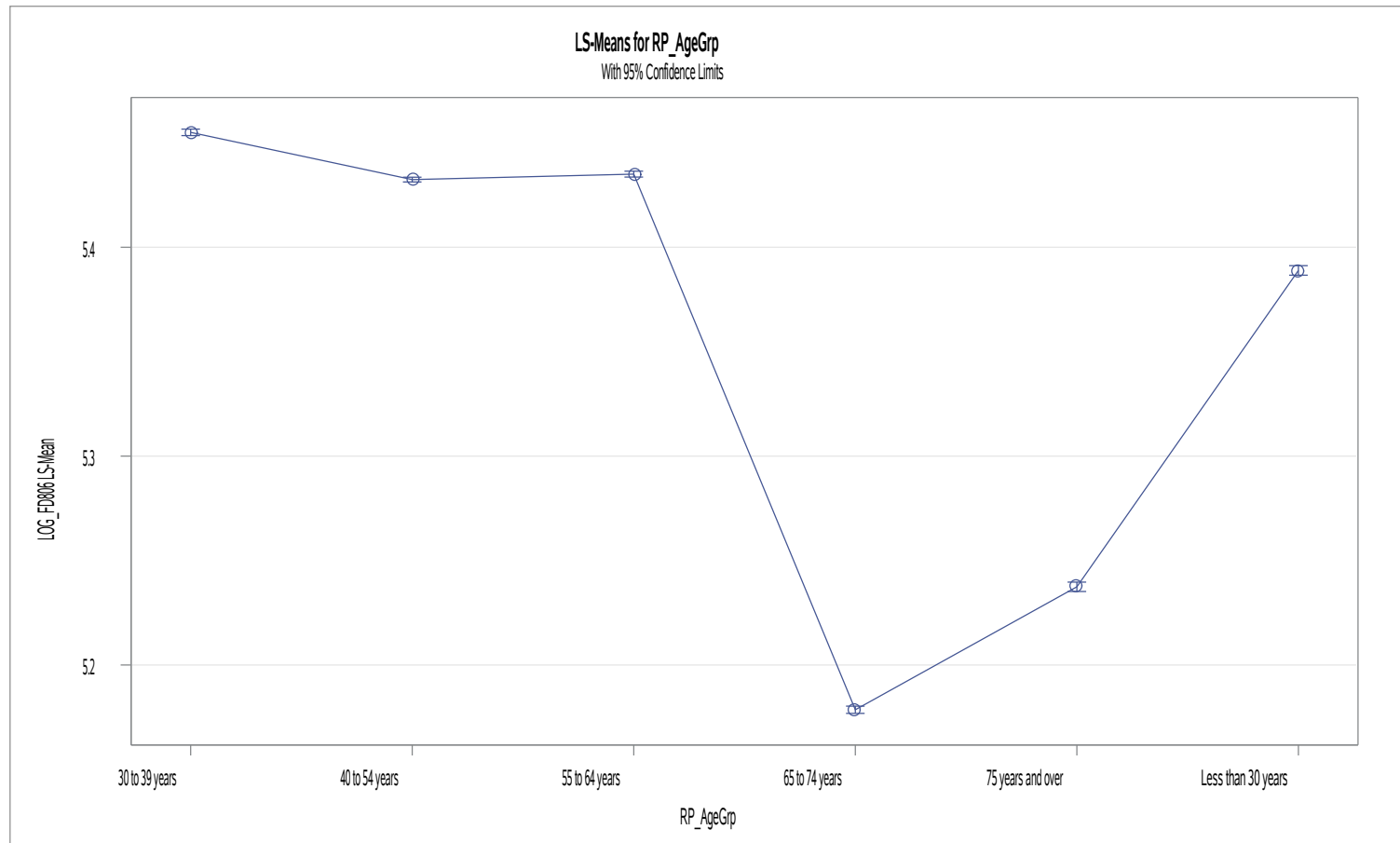
Level of RP_AgeGrp	N	LOG_FD806	
		Mean	Std Dev
30 to 39 years	1399872	5.45501049	0.88833792
40 to 54 years	2595446	5.43239591	0.94537734
55 to 64 years	1693126	5.43501741	0.93966324
65 to 74 years	1124539	5.17856193	0.97185179
75 years and over	670281	5.23742365	0.98415502
Less than 30 years	645612	5.38890329	0.87832954

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

RP_AgeGrp	LOG_FD806 LSMEAN	LSMEAN Number
30 to 39 years	5.45501049	1
40 to 54 years	5.43239591	2
55 to 64 years	5.43501741	3
65 to 74 years	5.17856193	4
75 years and over	5.23742365	5
Less than 30 years	5.38890329	6

Least Squares Means for effect RP_AgeGrp Pr > t for H0: LSMean(i)=LSMean(j) Dependent Variable: LOG_FD806						
i/j	1	2	3	4	5	6
1		<.0001	<.0001	<.0001	<.0001	<.0001
2	<.0001		0.0523	<.0001	<.0001	<.0001
3	<.0001	0.0523		<.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

LOG_FD806 Tukey Grouping for LS-Means of RP_AgeGrp (Alpha = 0.05)

LS-means covered by the same bar are not significantly different.

