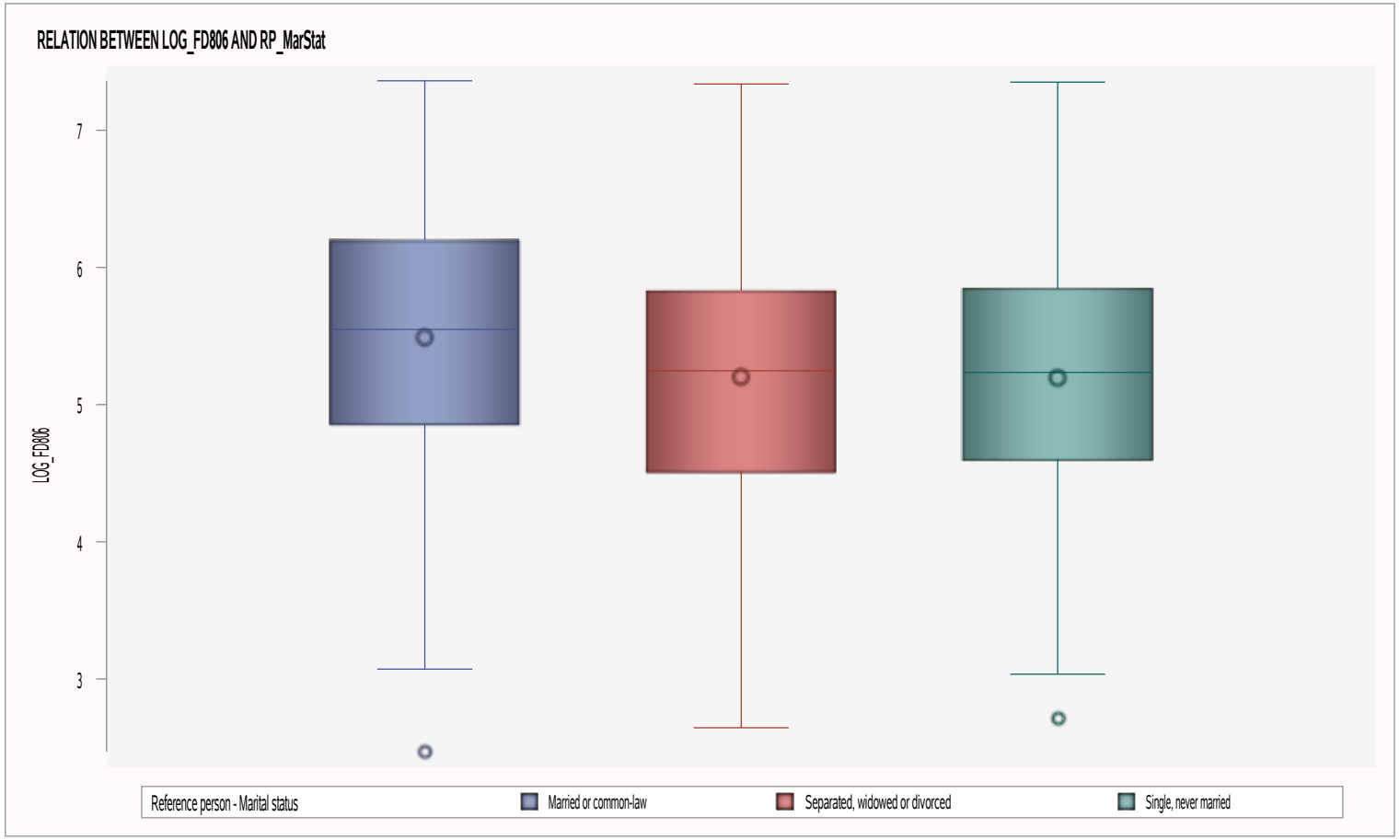
BIVARIATE ANALYSIS OF RP_MarStat AND LOG_FD806 FOR ANA.MODEL2
RELATION BETWEEN LOG_FD806 AND RP_MarStat

The MEANS Procedure

Analysis Variable : LOG_FD806														
Reference person - Marital status	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
Married or common-law	5220895	5220895	0	2.47	4.85	5.55	5.48	6.20	7.36	1.35	16.91	5.48	5.49	-0.20
Separated, widowed or divorced	1548207	1548207	0	2.64	4.51	5.25	5.20	5.83	7.34	1.32	17.94	5.20	5.20	-0.13
Single, never married	1359774	1359774	0	2.71	4.59	5.23	5.20	5.84	7.35	1.25	18.22	5.19	5.20	0.02



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

Reference person - Marital status=Married or common-law

Moments			
N	5220895	**Sum Weights**	5220895
Mean	5.48455392	**Sum Observations**	28634280.1
Std Deviation	0.92724701	**Variance**	0.85978702
Skewness	-0.2032826	**Kurtosis**	-0.6297533

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Marital status=Married or common-law

Moments			
Uncorrected SS	161535110	**Corrected SS**	4488856.88
Coeff Variation	16.9065164	**Std Error Mean**	0.00040581

Basic Statistical Measures			
Location		**Variability**	
Mean	5.484554	**Std Deviation**	0.92725
Median	5.548103	**Variance**	0.85979
Mode	6.488141	**Range**	4.89010
		Interquartile Range	1.35130

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	13515.07	**Pr > |t|**	<.0001
Sign	**M**	2610448	**Pr >= |M|**	<.0001
Signed Rank	**S**	6.814E12	**Pr >= |S|**	<.0001

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.03549	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	2087.027	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	15396.55	**Pr > A-Sq**	<0.0050

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Marital status=Married or common-law

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	7.36074
99%	7.28909
95%	6.90938
90%	6.65562
75% Q3	6.19832
50% Median	5.54810
25% Q1	4.84702
10%	4.11676
5%	3.93378
1%	3.38912
0% Min	2.47064

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
2.47064	3940	1429	7.34357	669	135
3.07177	347	1453	7.35014	1360	1507
3.08374	1372	272	7.35361	491	1356
3.08374	3250	109	7.35843	4907	792
3.15274	132	1093	7.36074	43	979

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Marital status=Separated, widowed or divorced

Moments			
N	1548207	**Sum Weights**	1548207
Mean	5.20017174	**Sum Observations**	8050942.3
Std Deviation	0.93282761	**Variance**	0.87016735
Skewness	-0.1333123	**Kurtosis**	-0.3866875
Uncorrected SS	43213481	**Corrected SS**	1347198.32
Coeff Variation	17.9384001	**Std Error Mean**	0.0007497

Basic Statistical Measures			
Location		**Variability**	
Mean	5.200172	**Std Deviation**	0.93283
Median	5.246919	**Variance**	0.87017
Mode	6.188552	**Range**	4.69268
		Interquartile Range	1.32466

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	6936.347	**Pr > |t|**	<.0001
Sign	**M**	774103.5	**Pr >= |M|**	<.0001
Signed Rank	**S**	5.992E11	**Pr >= |S|**	<.0001

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Marital status=Separated, widowed or divorced

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.041738	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	520.2227	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	3469.689	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	7.33743
99%	7.10462
95%	6.77815
90%	6.46812
75% Q3	5.83056
50% Median	5.24692
25% Q1	4.50590
10%	3.94893
5%	3.66356
1%	3.02237
0% Min	2.64476

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Marital status=Separated, widowed or divorced

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
2.64476	15313	1979	7.16410	1554	1709
3.02237	4062	1780	7.21127	1683	1961
3.07177	840	1832	7.27547	1311	1901
3.21727	1553	1905	7.28345	442	1575
3.24805	1311	1732	7.33743	836	1687

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Marital status=Single, never married

Moments			
N	1359774	Sum Weights	1359774
Mean	5.19640367	Sum Observations	7065934.61
Std Deviation	0.94686986	Variance	0.89656253
Skewness	0.02203074	Kurtosis	-0.3884498
Uncorrected SS	37936570.1	Corrected SS	1219121.52
Coeff Variation	18.2216379	Std Error Mean	0.000812

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Marital status=Single, never married

Basic Statistical Measures			
Location		**Variability**	
Mean	5.196404	**Std Deviation**	0.94687
Median	5.234951	**Variance**	0.89656
Mode	5.958166	**Range**	4.63772
		Interquartile Range	1.24975

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	6399.499	**Pr > |t|**	<.0001
Sign	**M**	679887	**Pr >= |M|**	<.0001
Signed Rank	**S**	4.622E11	**Pr >= |S|**	<.0001

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.040272	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	318.9522	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	2449.614	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	7.35109
99%	7.15277
95%	6.96876
90%	6.40723

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Marital status=Single, never married

Quantiles (Definition 5)	
Level	**Quantile**
75% Q3	5.84285
50% Median	5.23495
25% Q1	4.59310
10%	3.99046
5%	3.40652
1%	3.05777
0% Min	2.71337

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
2.71337	1520	2043	7.17670	3082	2230
3.03495	8060	2169	7.20510	5467	2108
3.05777	7665	2289	7.27783	46	2223
3.10727	381	2167	7.27891	505	1994
3.21165	209	2037	7.35109	759	2291

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_MarStat	3	Married or common-law Separated, widowed or divorced Single, never married

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	2	152934.293	76467.147	88104.3	<.0001
Error	8.13E6	7055176.720	0.868		
Corrected Total	8.13E6	7208111.014			

R-Square	Coeff Var	Root MSE	LOG_FD806 Mean
0.021217	17.30931	0.931620	5.382190

Source	DF	Type I SS	Mean Square	F Value	Pr > F
RP_MarStat	2	152934.2933	76467.1466	88104.3	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
RP_MarStat	2	152934.2933	76467.1466	88104.3	<.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**
RP_MarStat	2	224.2	112.1	393.28	<.0001
Error	8.13E6	2317449	0.2851		

Welch's ANOVA for LOG_FD806			
Source	**DF**	**F Value**	**Pr > F**
RP_MarStat	2.0000	87448.3	<.0001
Error	2765839		

The GLM Procedure

Level of RP_MarStat	N	LOG_FD806	
		Mean	**Std Dev**
Married or common-law	5220895	5.48455392	0.92724701
Separated, widowed or divorced	1548207	5.20017174	0.93282761
Single, never married	1359774	5.19640367	0.94686986

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

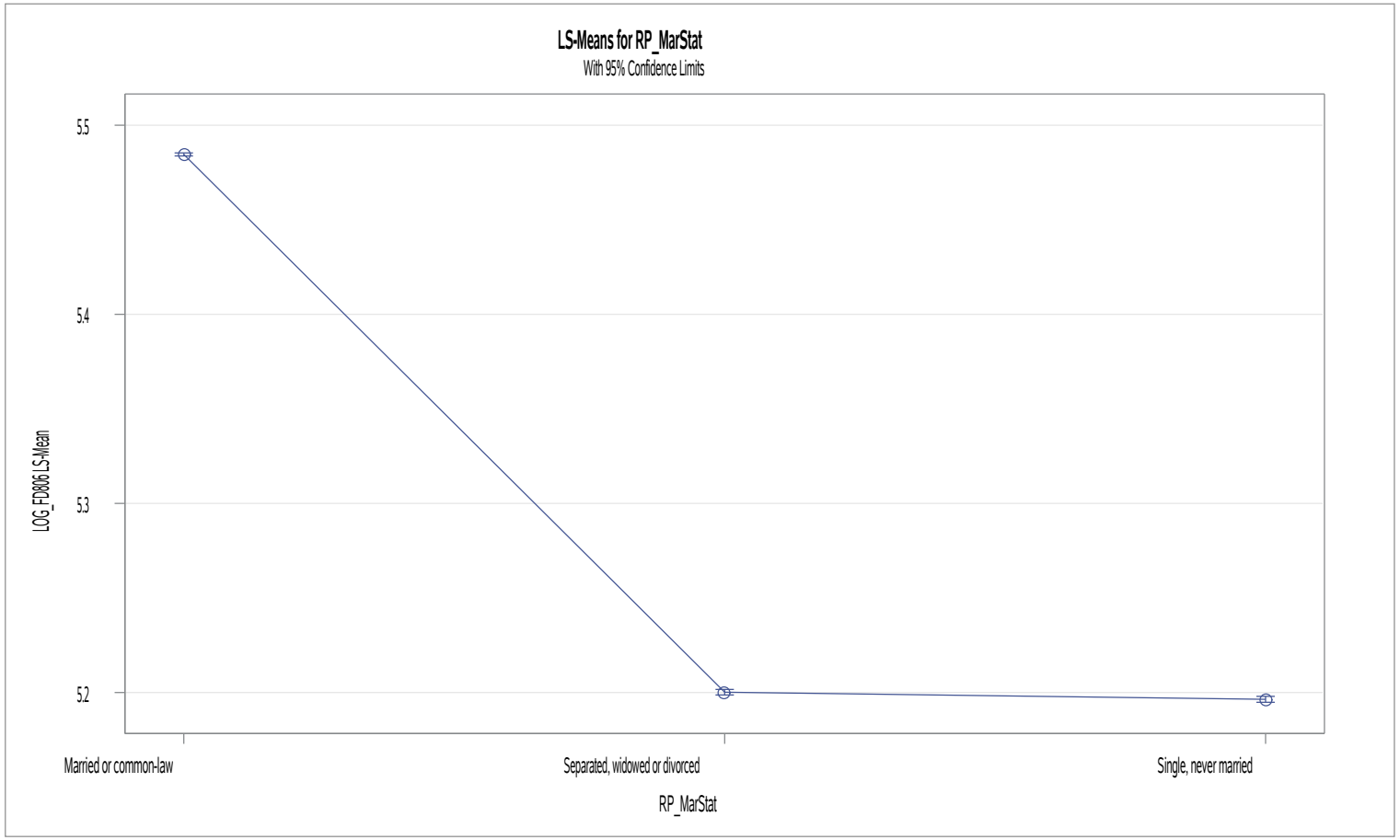
RP_MarStat	LOG_FD806 LSMEAN	LSMEAN Number
Married or common-law	5.48455392	1
Separated, widowed or divorced	5.20017174	2
Single, never married	5.19640367	3

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

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

Dependent Variable: LOG_FD806

i/j	1	2	3
1		<.0001	<.0001
2	<.0001		0.0017
3	<.0001	0.0017	



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