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

The MEANS Procedure

Analysis Variable : LOG_FD806														
Spouse - Sex	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
Female	2493849	2493849	0	2.47	4.86	5.49	5.49	6.25	7.36	1.40	17.02	5.49	5.49	-0.22
Male	2727046	2727046	0	3.07	4.84	5.56	5.48	6.18	7.36	1.34	16.80	5.48	5.48	-0.19
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 - Sex=Female

Moments			
N	2493849	**Sum Weights**	2493849
Mean	5.48775798	**Sum Observations**	13685639.7
Std Deviation	0.93420326	**Variance**	0.87273572
Skewness	-0.216224	**Kurtosis**	-0.6025601

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Sex=Female

Moments			
Uncorrected SS	77279948.9	**Corrected SS**	2176470.24
Coeff Variation	17.0234048	**Std Error Mean**	0.00059157

Basic Statistical Measures			
Location		**Variability**	
Mean	5.487758	**Std Deviation**	0.93420
Median	5.493761	**Variance**	0.87274
Mode	6.583603	**Range**	4.89010
		Interquartile Range	1.39589

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	9276.596	**Pr > |t|**	<.0001
Sign	**M**	1246925	**Pr >= |M|**	<.0001
Signed Rank	**S**	1.555E12	**Pr >= |S|**	<.0001

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.049375	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	1049.703	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	8131.745	**Pr > A-Sq**	<0.0050

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Sex=Female

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	7.36074
99%	7.23335
95%	6.90112
90%	6.64348
75% Q3	6.25416
50% Median	5.49376
25% Q1	4.85826
10%	4.18510
5%	3.94893
1%	3.31200
0% Min	2.47064

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
2.47064	3940	241	7.30971	58	426
3.21727	186	64	7.32095	5649	28
3.22764	463	605	7.33580	1566	144
3.23789	987	180	7.35014	1360	249
3.25810	2580	351	7.36074	43	416

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Sex=Male

Moments			
N	2727046	**Sum Weights**	2727046
Mean	5.48162384	**Sum Observations**	14948640.4
Std Deviation	0.92083002	**Variance**	0.84792793
Skewness	-0.1911797	**Kurtosis**	-0.6566523
Uncorrected SS	84255161.2	**Corrected SS**	2312337.63
Coeff Variation	16.7984898	**Std Error Mean**	0.00055761

Basic Statistical Measures			
Location		**Variability**	
Mean	5.481624	**Std Deviation**	0.92083
Median	5.560682	**Variance**	0.84793
Mode	6.488141	**Range**	4.28666
		Interquartile Range	1.33867

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	9830.509	**Pr > |t|**	<.0001
Sign	**M**	1363523	**Pr >= |M|**	<.0001
Signed Rank	**S**	1.859E12	**Pr >= |S|**	<.0001

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Sex=Male

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.04949	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	1427.212	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	9647.027	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	7.35843
99%	7.29261
95%	6.90938
90%	6.65639
75% Q3	6.18058
50% Median	5.56068
25% Q1	4.84190
10%	4.09351
5%	3.93104
1%	3.47259
0% Min	3.07177

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Sex=Male

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
3.07177	347	1236	7.33995	1345	791
3.08374	1372	1154	7.34003	776	1527
3.08374	3250	1030	7.34357	669	915
3.15274	132	1335	7.35361	491	1382
3.25810	975	793	7.35843	4907	1128

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Sex=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

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Sex=No spouse

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

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

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Sex=No spouse

Quantiles (Definition 5)	
Level	**Quantile**
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

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
2.64476	15313	1994	7.27783	46	1615
2.71337	1520	2209	7.27891	505	2118
3.02237	4062	1975	7.28345	442	1629
3.03495	8060	2018	7.33743	836	1656
3.05777	7665	2032	7.35109	759	1841

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_Sex	3	Female Male 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

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	152973.029	76486.514	88127.1	<.0001
Error	8.13E6	7055137.985	0.868		
Corrected Total	8.13E6	7208111.014			

R-Square	Coeff Var	Root MSE	LOG_FD806 Mean
0.021222	17.30926	0.931617	5.382190

Source	DF	Type I SS	Mean Square	F Value	Pr > F
SP_Sex	2	152973.0289	76486.5145	88127.1	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SP_Sex	2	152973.0289	76486.5145	88127.1	<.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_Sex	2	371.6	185.8	652.09	<.0001
Error	8.13E6	2316009	0.2849		

Welch's ANOVA for LOG_FD806			
Source	**DF**	**F Value**	**Pr > F**
SP_Sex	2.0000	87455.0	<.0001
Error	5372470		

The GLM Procedure

Level of SP_Sex	N	LOG_FD806	
		Mean	**Std Dev**
Female	2493849	5.48775798	0.93420326
Male	2727046	5.48162384	0.92083002
No spouse	2907981	5.19840979	0.93942162

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

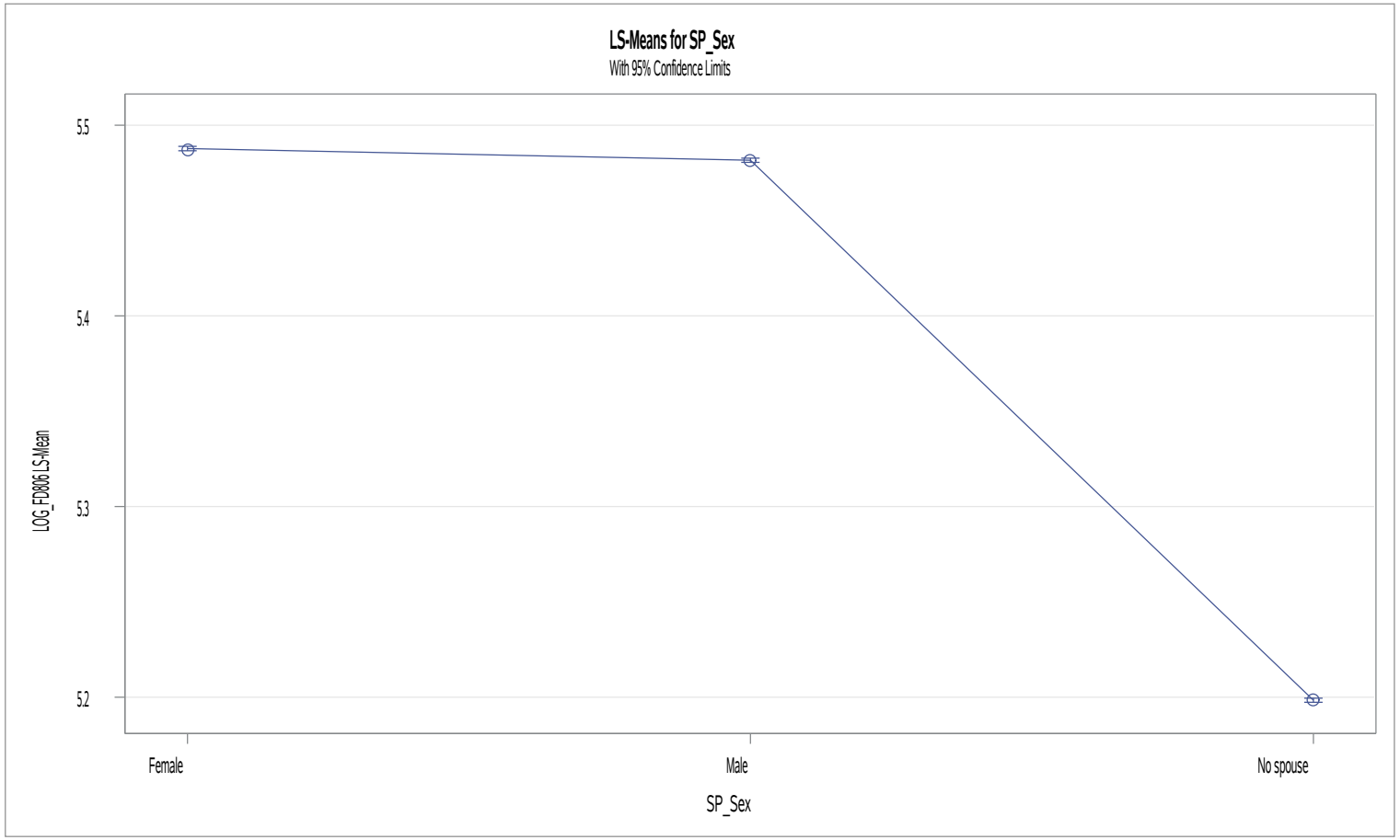
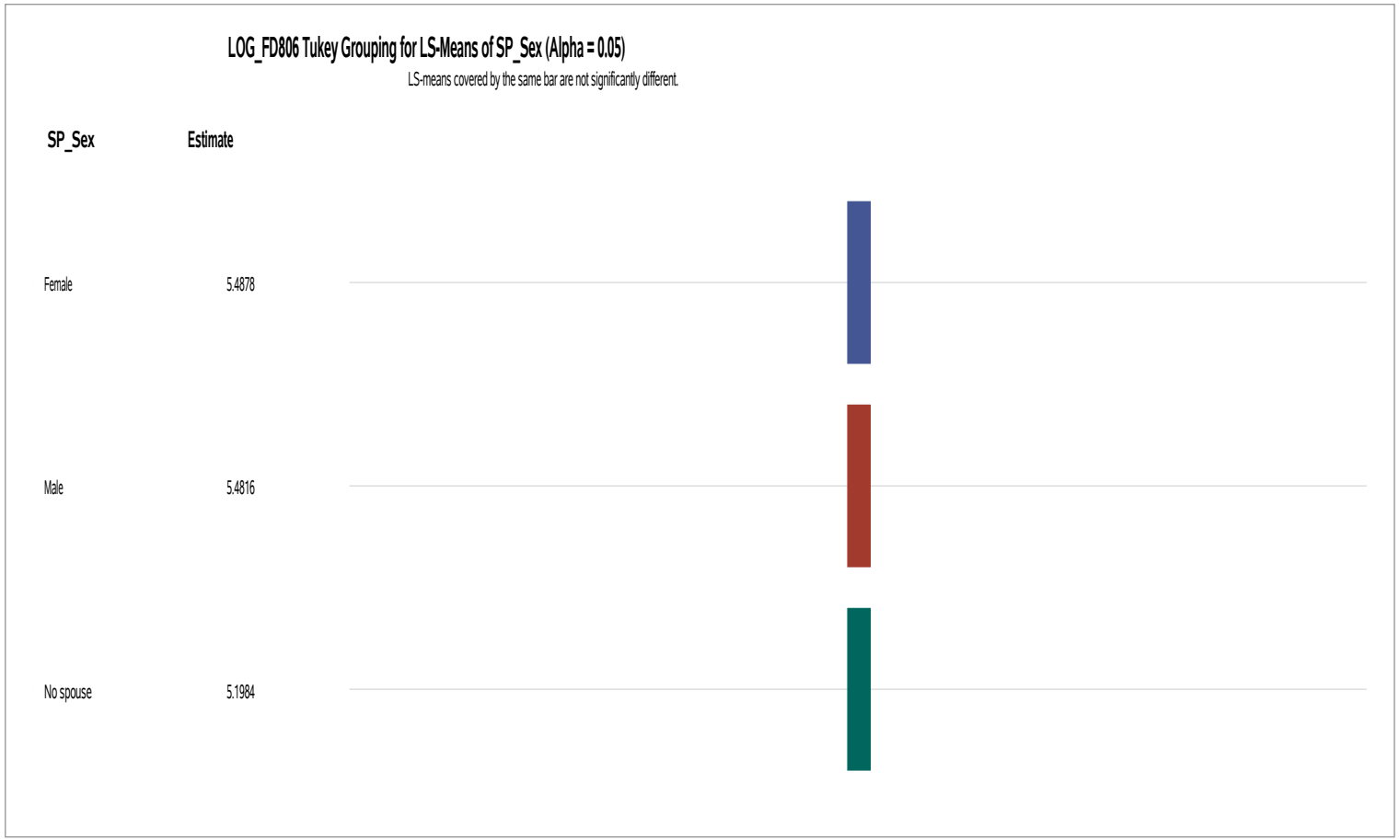
SP_Sex	LOG_FD806 LSMEAN	LSMEAN Number
Female	5.48775798	1
Male	5.48162384	2
No spouse	5.19840979	3

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

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

Dependent Variable: LOG_FD806

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



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