

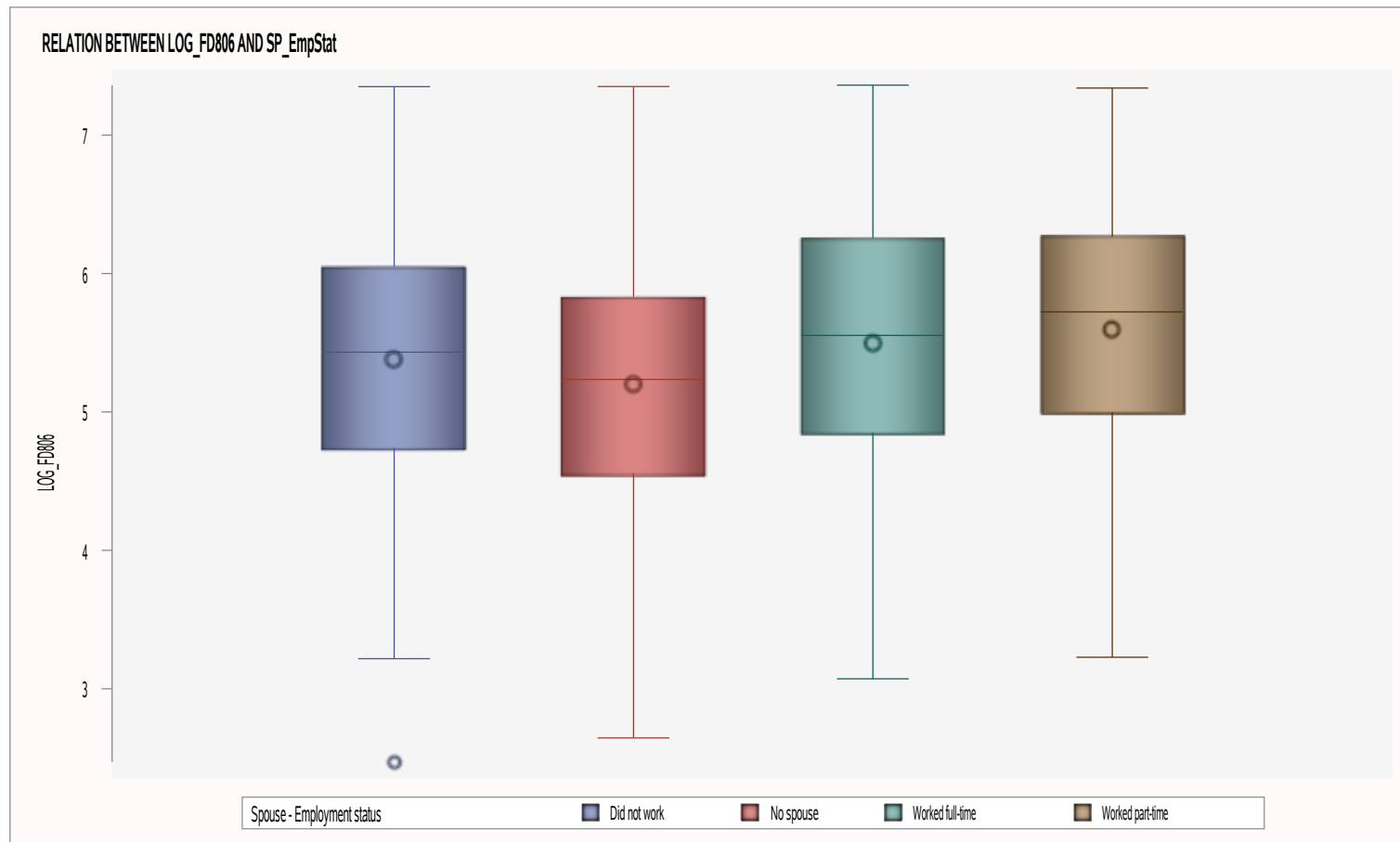
BIVARIATE ANALYSIS OF SP_EmpStat AND LOG_FD806 FOR ANA.MODEL2

RELATION BETWEEN LOG_FD806 AND SP_EmpStat

14:29 Sunday, November 21, 2021 1

The MEANS Procedure

Analysis Variable : LOG_FD806														
Spouse - Employment 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
Did not work	1453314	1453314	0	2.47	4.73	5.43	5.38	6.05	7.35	1.32	17.58	5.38	5.38	-0.09
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
Worked full-time	2599458	2599458	0	3.07	4.84	5.55	5.49	6.26	7.36	1.42	16.96	5.49	5.49	-0.22
Worked part-time	1168123	1168123	0	3.23	4.98	5.72	5.60	6.27	7.34	1.28	15.72	5.60	5.60	-0.28



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

Spouse - Employment status=Did not work

Moments			
N	1453314	Sum Weights	1453314
Mean	5.38144915	Sum Observations	7820935.39
Std Deviation	0.94589277	Variance	0.89471314
Skewness	-0.0869046	Kurtosis	-0.5717134

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Employment status=Did not work

Moments			
Uncorrected SS	43388264.3	Corrected SS	1300298.23
Coeff Variation	17.5769155	Std Error Mean	0.00078463

Basic Statistical Measures			
Location		Variability	
Mean	5.381449	Std Deviation	0.94589
Median	5.431624	Variance	0.89471
Mode	3.931041	Range	4.87950
		Interquartile Range	1.32268

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	6858.625	Pr > t 	<.0001
Sign	M	726657	Pr >= M 	<.0001
Signed Rank	S	5.28E11	Pr >= S 	<.0001

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.04721	Pr > D	<0.0100
Cramer-von Mises	W-Sq	445.1515	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	3585.523	Pr > A-Sq	<0.0050

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Employment status=Did not work

Quantiles (Definition 5)	
Level	Quantile
100% Max	7.35014
99%	7.30361
95%	6.89278
90%	6.65863
75% Q3	6.04865
50% Median	5.43162
25% Q1	4.72597
10%	4.03530
5%	3.84031
1%	3.38912
0% Min	2.47064

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
2.47064	3940	415	7.32095	5649	45
3.21727	186	112	7.33580	1566	239
3.23789	987	309	7.33995	1345	57
3.25810	975	59	7.34357	669	361
3.26805	1551	297	7.35014	1360	428

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Employment status=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 - Employment status=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 - Employment status=No spouse

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
2.64476	15313	894	7.27783	46	515
2.71337	1520	1109	7.27891	505	1018
3.02237	4062	875	7.28345	442	529
3.03495	8060	918	7.33743	836	556
3.05777	7665	932	7.35109	759	741

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Employment status=Worked full-time

Moments			
N	2599458	Sum Weights	2599458
Mean	5.49141261	Sum Observations	14274696.4
Std Deviation	0.93114944	Variance	0.86703928
Skewness	-0.2216507	Kurtosis	-0.6721872
Uncorrected SS	80642079.2	Corrected SS	2253831.32
Coeff Variation	16.9564647	Std Error Mean	0.00057753

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Employment status=Worked full-time

Basic Statistical Measures			
Location		Variability	
Mean	5.491413	Std Deviation	0.93115
Median	5.552649	Variance	0.86704
Mode	6.488141	Range	4.28897
		Interquartile Range	1.41726

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	9508.37	Pr > t 	<.0001
Sign	M	1299729	Pr >= M 	<.0001
Signed Rank	S	1.689E12	Pr >= S 	<.0001

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.052267	Pr > D	<0.0100
Cramer-von Mises	W-Sq	1378.334	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	10073.71	Pr > A-Sq	<0.0050

Quantiles (Definition 5)	
Level	Quantile
100% Max	7.36074
99%	7.25157
95%	6.89946
90%	6.64688

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Employment status=Worked full-time

Quantiles (Definition 5)	
Level	Quantile
75% Q3	6.25544
50% Median	5.55265
25% Q1	4.83818
10%	4.11545
5%	3.95124
1%	3.31200
0% Min	3.07177

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
3.07177	347	1647	7.32214	2134	1398
3.08374	1372	1533	7.32818	59	1663
3.08374	3250	1349	7.35361	491	1909
3.15274	132	1806	7.35843	4907	1496
3.25810	2580	1465	7.36074	43	1702

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Employment status=Worked part-time

Moments			
N	1168123	Sum Weights	1168123
Mean	5.59756833	Sum Observations	6538648.3
Std Deviation	0.87982595	Variance	0.7740937
Skewness	-0.2826993	Kurtosis	-0.5841081
Uncorrected SS	37504766.5	Corrected SS	904235.886
Coeff Variation	15.7180029	Std Error Mean	0.00081405

Basic Statistical Measures			
Location		Variability	
Mean	5.597568	Std Deviation	0.87983
Median	5.722768	Variance	0.77409
Mode	5.754000	Range	4.11239
		Interquartile Range	1.28281

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	6876.175	Pr > t 	<.0001
Sign	M	584061.5	Pr >= M 	<.0001
Signed Rank	S	3.411E11	Pr >= S 	<.0001

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Employment status=Worked part-time

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.062088	Pr > D	<0.0100
Cramer-von Mises	W-Sq	753.4548	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	5086.833	Pr > A-Sq	<0.0050

Quantiles (Definition 5)	
Level	Quantile
100% Max	7.34003
99%	7.29261
95%	6.99410
90%	6.65562
75% Q3	6.26724
50% Median	5.72277
25% Q1	4.98443
10%	4.43414
5%	4.01458
1%	3.52812
0% Min	3.22764

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Spouse - Employment status=Worked part-time

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
3.22764	463	2057	7.17146	1072	2127
3.26805	1144	2128	7.19090	3791	2154
3.31637	442	2054	7.24436	966	2007
3.38912	1988	2164	7.29261	12196	2084
3.43108	3237	2274	7.34003	776	2255

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_EmpStat	4	Did not work No spouse Worked full-time Worked part-time

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	3	183415.454	61138.485	70748.5	<.0001
Error	8.13E6	7024695.560	0.864		
Corrected Total	8.13E6	7208111.014			

R-Square	Coeff Var	Root MSE	LOG_FD806 Mean
0.025446	17.27188	0.929605	5.382190

Source	DF	Type I SS	Mean Square	F Value	Pr > F
SP_EmpStat	3	183415.4536	61138.4845	70748.5	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SP_EmpStat	3	183415.4536	61138.4845	70748.5	<.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_EmpStat	3	1662.8	554.3	1954.08	<.0001
Error	8.13E6	2305716	0.2836		

Welch's ANOVA for LOG_FD806			
Source	DF	F Value	Pr > F
SP_EmpStat	3.0000	72279.5	<.0001
Error	3572185		

The GLM Procedure

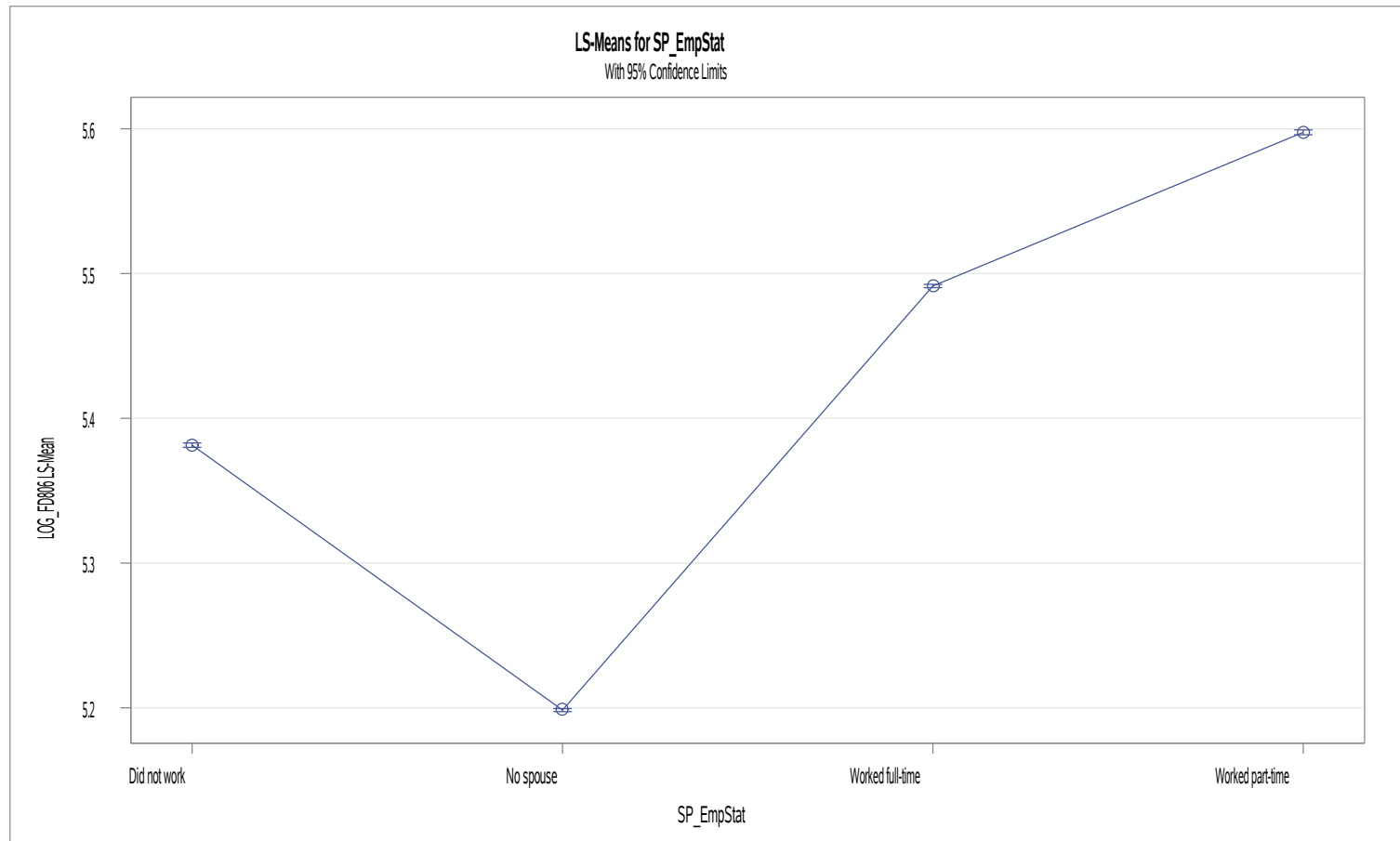
Level of SP_EmpStat	N	LOG_FD806	
		Mean	Std Dev
Did not work	1453314	5.38144915	0.94589277
No spouse	2907981	5.19840979	0.93942162
Worked full-time	2599458	5.49141261	0.93114944
Worked part-time	1168123	5.59756833	0.87982595

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

SP_EmpStat	LOG_FD806 LSMEAN	LSMEAN Number
Did not work	5.38144915	1
No spouse	5.19840979	2
Worked full-time	5.49141261	3
Worked part-time	5.59756833	4

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

