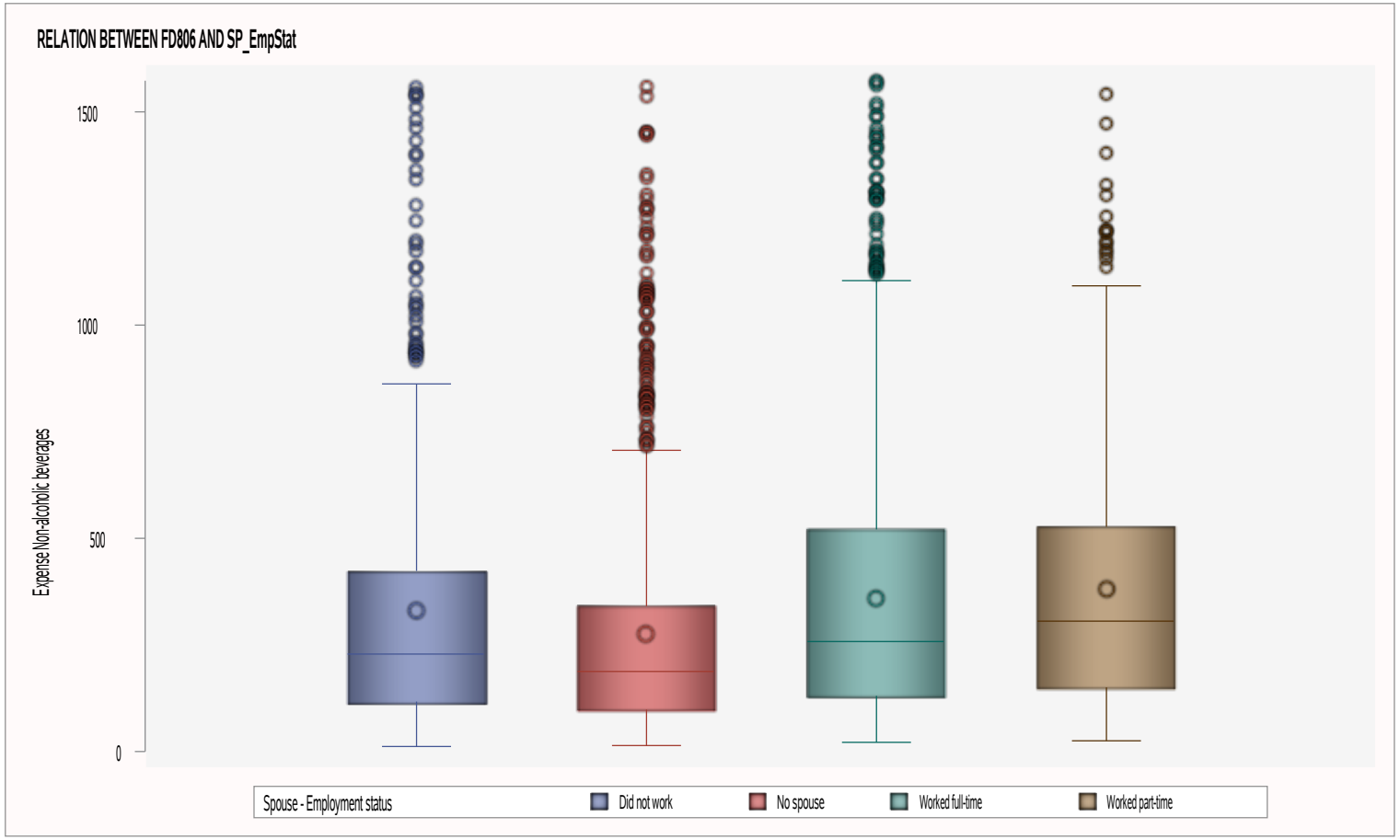
BIVARIATE ANALYSIS OF SP_EmpStat AND FD806 FOR ANA.MODEL1
RELATION BETWEEN FD806 AND SP_EmpStat

The MEANS Procedure

Analysis Variable : FD806 Expense Non-alcoholic beverages														
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	11.83	112.84	228.52	330.34	423.54	1556.41	310.70	95.90	329.83	330.86	1.80
No spouse	2907981	2907981	0	14.08	94.12	187.76	275.69	340.75	1557.89	246.63	99.19	275.37	276.00	1.97
Worked full-time	2599458	2599458	0	21.58	126.24	257.92	357.87	520.84	1573.00	394.60	87.19	357.49	358.25	1.46
Worked part-time	1168123	1168123	0	25.22	146.12	305.75	380.96	527.02	1540.76	380.90	81.18	380.40	381.52	1.33



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: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

Spouse - Employment status=Did not work

Moments			
N	1453314	**Sum Weights**	1453314
Mean	330.3412	**Sum Observations**	480089491
Std Deviation	316.792144	**Variance**	100357.262
Skewness	1.80452787	**Kurtosis**	3.1846025

The UNIVARIATE Procedure
Variable: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

Spouse - Employment status=Did not work

Moments			
Uncorrected SS	3.04444E11	**Corrected SS**	1.45851E11
Coeff Variation	95.8984661	**Std Error Mean**	0.26278143

Basic Statistical Measures			
Location		**Variability**	
Mean	330.3412	**Std Deviation**	316.79214
Median	228.5200	**Variance**	100357
Mode	50.9600	**Range**	1545
		Interquartile Range	310.70000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	1257.095	**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**	0.168675	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	15103.96	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	88670.19	**Pr > A-Sq**	<0.0050

The UNIVARIATE Procedure
Variable: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

Spouse - Employment status=Did not work

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1556.41
99%	1485.65
95%	985.14
90%	779.48
75% Q3	423.54
50% Median	228.52
25% Q1	112.84
10%	56.56
5%	46.54
1%	29.64
0% Min	11.83

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
11.83	3940	415	1511.64	5649	45
24.96	186	112	1534.26	1566	239
25.48	987	309	1540.64	1345	57
26.00	975	59	1546.22	669	361
26.26	1551	297	1556.41	1360	428

The UNIVARIATE Procedure
Variable: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

Spouse - Employment status=No spouse

Moments			
N	2907981	**Sum Weights**	2907981
Mean	275.685049	**Sum Observations**	801686886
Std Deviation	273.455354	**Variance**	74777.8306
Skewness	1.96559807	**Kurtosis**	3.75787583
Uncorrected SS	4.38466E11	**Corrected SS**	2.17452E11
Coeff Variation	99.1912164	**Std Error Mean**	0.16035801

Basic Statistical Measures			
Location		**Variability**	
Mean	275.6850	**Std Deviation**	273.45535
Median	187.7600	**Variance**	74778
Mode	386.9000	**Range**	1544
		Interquartile Range	246.63000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	1719.185	**Pr > |t|**	<.0001
Sign	**M**	1453991	**Pr >= |M|**	<.0001
Signed Rank	**S**	2.114E12	**Pr >= |S|**	<.0001

The UNIVARIATE Procedure
Variable: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

Spouse - Employment status=No spouse

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.173993	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	34913.71	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	204079.7	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1557.89
99%	1230.11
95%	915.33
90%	606.21
75% Q3	340.75
50% Median	187.76
25% Q1	94.12
10%	52.40
5%	39.00
1%	21.28
0% Min	14.08

The UNIVARIATE Procedure
Variable: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

Spouse - Employment status=No spouse

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
14.08	15313	894	1447.85	46	515
15.08	1520	1109	1449.41	505	1018
20.54	4062	875	1456.00	442	529
20.80	8060	918	1536.76	836	556
21.28	7665	932	1557.89	759	741

The UNIVARIATE Procedure
Variable: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

Spouse - Employment status=Worked full-time

Moments			
N	2599458	Sum Weights	2599458
Mean	357.866075	Sum Observations	930257832
Std Deviation	312.023086	Variance	97358.4062
Skewness	1.46032209	Kurtosis	1.9696261
Uncorrected SS	5.85987E11	Corrected SS	2.53079E11
Coeff Variation	87.1899036	Std Error Mean	0.19352868

The UNIVARIATE Procedure
Variable: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

Spouse - Employment status=Worked full-time

Basic Statistical Measures			
Location		**Variability**	
Mean	357.8661	**Std Deviation**	312.02309
Median	257.9200	**Variance**	97358
Mode	657.3000	**Range**	1551
		Interquartile Range	394.60000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	1849.163	**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**	0.154426	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	19337.43	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	113291.6	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1573.00
99%	1410.32
95%	991.74
90%	770.38

The UNIVARIATE Procedure
Variable: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

Spouse - Employment status=Worked full-time

Quantiles (Definition 5)	
Level	**Quantile**
75% Q3	520.84
50% Median	257.92
25% Q1	126.24
10%	61.28
5%	52.00
1%	27.44
0% Min	21.58

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
21.58	347	1647	1513.44	2134	1398
21.84	1372	1533	1522.61	59	1663
21.84	3250	1349	1561.82	491	1909
23.40	132	1806	1569.37	4907	1496
26.00	2580	1465	1573.00	43	1702

The UNIVARIATE Procedure
Variable: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

Spouse - Employment status=Worked part-time

Moments			
N	1168123	**Sum Weights**	1168123
Mean	380.959713	**Sum Observations**	445007803
Std Deviation	309.25005	**Variance**	95635.5936
Skewness	1.3255941	**Kurtosis**	1.5187142
Uncorrected SS	2.81244E11	**Corrected SS**	1.11714E11
Coeff Variation	81.1765758	**Std Error Mean**	0.28613138

Basic Statistical Measures			
Location		**Variability**	
Mean	380.9597	**Std Deviation**	309.25005
Median	305.7500	**Variance**	95636
Mode	315.4500	**Range**	1516
		Interquartile Range	380.90000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	1331.415	**Pr > |t|**	<.0001
Sign	**M**	584061.5	**Pr >= |M|**	<.0001
Signed Rank	**S**	3.411E11	**Pr >= |S|**	<.0001

The UNIVARIATE Procedure
Variable: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

Spouse - Employment status=Worked part-time

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.143278	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	7057.565	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	43108.33	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1540.76
99%	1469.40
95%	1090.18
90%	777.14
75% Q3	527.02
50% Median	305.75
25% Q1	146.12
10%	84.28
5%	55.40
1%	34.06
0% Min	25.22

The UNIVARIATE Procedure
Variable: FD806 (Expense Non-alcoholic beverages)

Freq: WeightD

Spouse - Employment status=Worked part-time

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
25.22	463	2057	1301.74	1072	2127
26.26	1144	2128	1327.30	3791	2154
27.56	442	2054	1400.18	966	2007
29.64	1988	2164	1469.40	12196	2084
30.91	3237	2274	1540.76	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: FD806 Expense Non-alcoholic beverages

Frequency: WeightD

The GLM Procedure

Dependent Variable: FD806 Expense Non-alcoholic beverages

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	13551148732	4517049577.2	50430.9	<.0001
Error	8.13E6	728095981110	89569.128547		
Corrected Total	8.13E6	741647129841			

R-Square	Coeff Var	Root MSE	FD806 Mean
0.018272	91.56115	299.2810	326.8646

Source	DF	Type I SS	Mean Square	F Value	Pr > F
SP_EmpStat	3	13551148732	4517049577	50430.9	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SP_EmpStat	3	13551148732	4517049577	50430.9	<.0001

The GLM Procedure

Levene's Test for Homogeneity of FD806 Variance ANOVA of Absolute Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
SP_EmpStat	3	3.7654E9	1.2551E9	32010.2	<.0001
Error	8.13E6	3.187E11	39210.7		

Welch's ANOVA for FD806			
Source	DF	F Value	Pr > F
SP_EmpStat	3.0000	53743.9	<.0001
Error	3465115		

The GLM Procedure

Level of SP_EmpStat	N	FD806	
		Mean	Std Dev
Did not work	1453314	330.341200	316.792144
No spouse	2907981	275.685049	273.455354
Worked full-time	2599458	357.866075	312.023086
Worked part-time	1168123	380.959713	309.250050

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

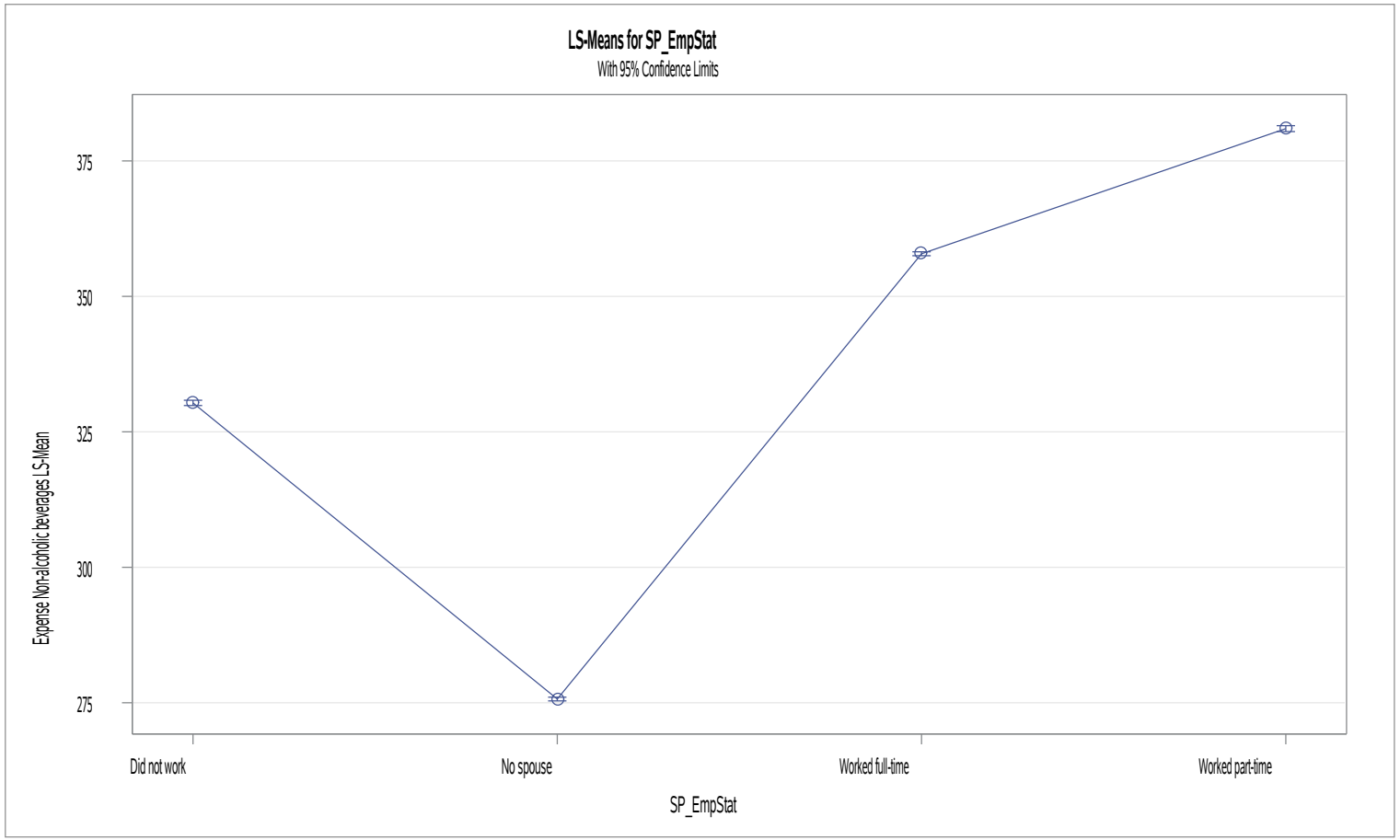
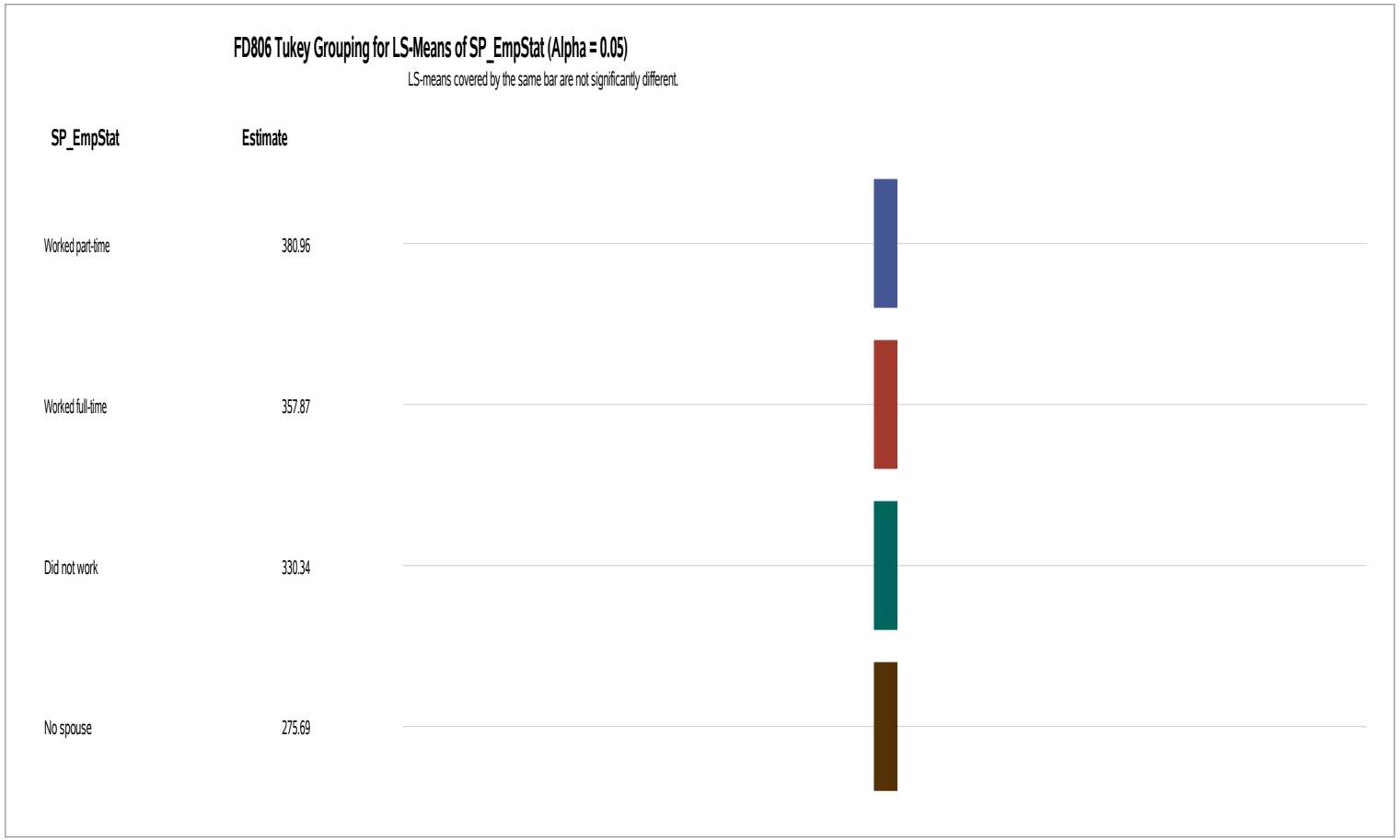
SP_EmpStat	FD806 LSMEAN	LSMEAN Number
Did not work	330.341200	1
No spouse	275.685049	2
Worked full-time	357.866075	3
Worked part-time	380.959713	4

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

Dependent Variable: 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