

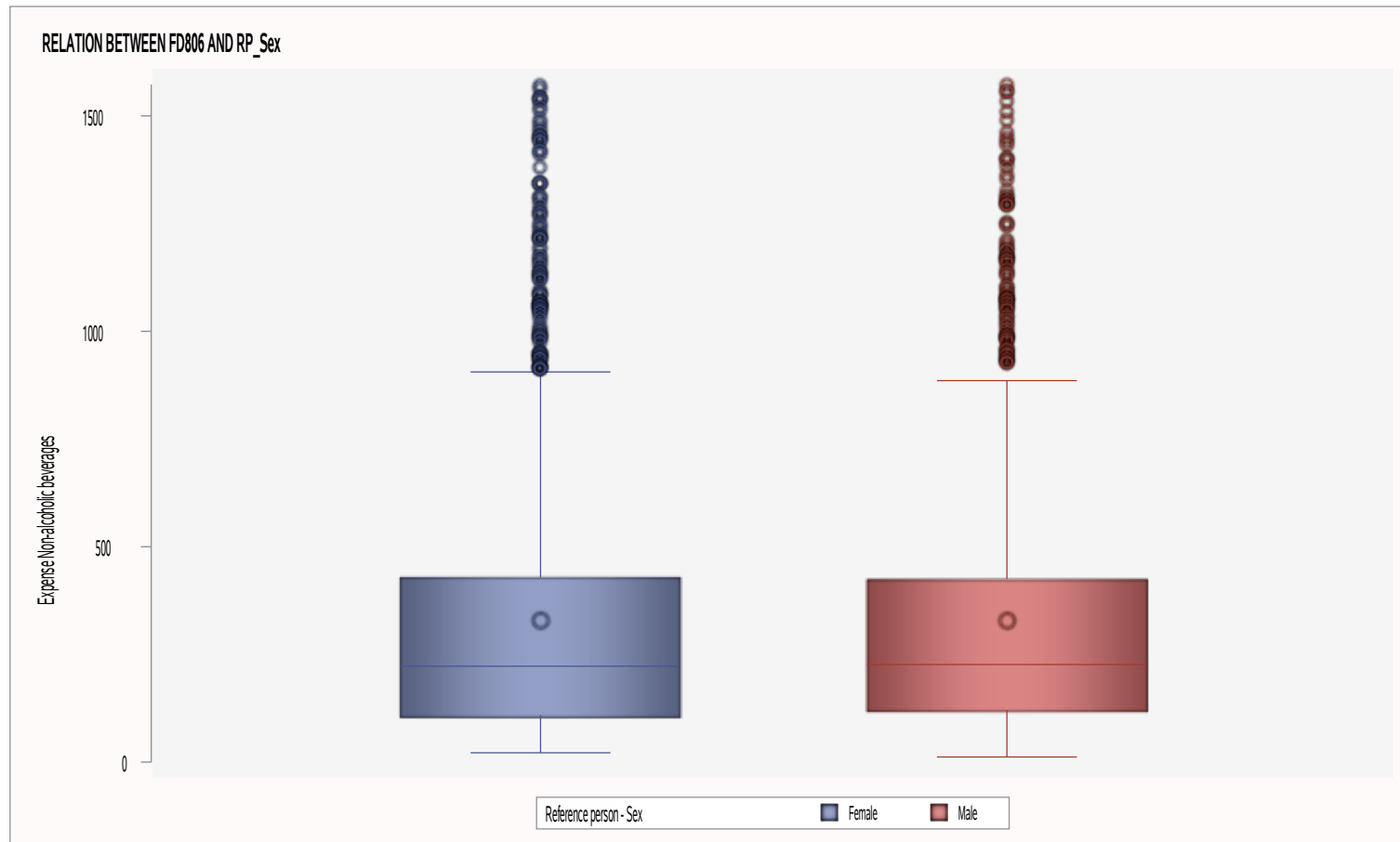
# BIVARIATE ANALYSIS OF RP\_Sex AND FD806 FOR ANA.MODEL1

## RELATION BETWEEN FD806 AND RP\_Sex

11:42 Saturday, November 20, 2021 1

### The MEANS Procedure

Analysis Variable : FD806 Expense Non-alcoholic beverages														
Reference person - 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	4262186	4262186	0	21.58	104.72	222.56	326.98	428.81	1569.37	324.09	94.95	326.68	327.27	1.72
Male	3866690	3866690	0	11.83	116.22	226.42	326.74	425.26	1573.00	309.04	89.52	326.45	327.03	1.57



Null hypothesis: There's no difference in means
Assumptions:
1.Sample distribution must be normal:
e.g:Shapiro (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
2.Groups are independent of one another.
3.There are no major outliers.
4.A check for unequal variances will help determine which version of an independent samples t-test is most appropriate:
(e.g:Levene's test, 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) t-test 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: FD806 (Expense Non-alcoholic beverages)**

**Freq: WeightD**

Reference person - Sex=Female

Moments			
<b>N</b>	4262186	<b>Sum Weights</b>	4262186
<b>Mean</b>	326.978148	<b>Sum Observations</b>	1393641686
<b>Std Deviation</b>	310.478872	<b>Variance</b>	96397.1302
<b>Skewness</b>	1.71690926	<b>Kurtosis</b>	2.83833535
<b>Uncorrected SS</b>	8.66553E11	<b>Corrected SS</b>	4.10862E11
<b>Coeff Variation</b>	94.9540127	<b>Std Error Mean</b>	0.15038892

Basic Statistical Measures			
Location		Variability	
<b>Mean</b>	326.9781	<b>Std Deviation</b>	310.47887
<b>Median</b>	222.5600	<b>Variance</b>	96397
<b>Mode</b>	657.3000	<b>Range</b>	1548
		<b>Interquartile Range</b>	324.09000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
<b>Student's t</b>	<b>t</b>	2174.217	<b>Pr &gt;  t </b>	<.0001
<b>Sign</b>	<b>M</b>	2131093	<b>Pr &gt;=  M </b>	<.0001
<b>Signed Rank</b>	<b>S</b>	4.542E12	<b>Pr &gt;=  S </b>	<.0001

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

**Freq: WeightD**

Reference person - Sex=Female

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.164481	Pr > D	<0.0100
Cramer-von Mises	W-Sq	41312.01	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	240753.3	Pr > A-Sq	<0.0050

Quantiles (Definition 5)	
Level	Quantile
100% Max	1569.37
99%	1464.24
95%	1001.63
90%	775.63
75% Q3	428.81
50% Median	222.56
25% Q1	104.72
10%	54.08
5%	46.66
1%	27.72
0% Min	21.58

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

**Freq: WeightD**

Reference person - Sex=Female

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
21.58	840	982	1540.64	1345	216
21.58	347	743	1540.76	776	632
21.84	1372	141	1546.22	669	79
21.84	3250	59	1561.82	491	704
22.36	381	1177	1569.37	4907	407

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

**Freq: WeightD**

Reference person - Sex=Male

Moments			
<b>N</b>	3866690	<b>Sum Weights</b>	3866690
<b>Mean</b>	326.739492	<b>Sum Observations</b>	1263400325
<b>Std Deviation</b>	292.484616	<b>Variance</b>	85547.2504
<b>Skewness</b>	1.57119709	<b>Kurtosis</b>	2.23729428
<b>Uncorrected SS</b>	7.43587E11	<b>Corrected SS</b>	3.30785E11
<b>Coeff Variation</b>	89.5161507	<b>Std Error Mean</b>	0.14874191

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

**Freq: WeightD**

Reference person - Sex=Male

Basic Statistical Measures			
Location		Variability	
Mean	326.7395	Std Deviation	292.48462
Median	226.4200	Variance	85547
Mode	386.9000	Range	1561
		Interquartile Range	309.04000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	2196.688	Pr >  t	<.0001
Sign	M	1933345	Pr >=  M	<.0001
Signed Rank	S	3.738E12	Pr >=  S	<.0001

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.156824	Pr > D	<0.0100
Cramer-von Mises	W-Sq	35050.33	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	201427.2	Pr > A-Sq	<0.0050

Quantiles (Definition 5)	
Level	Quantile
100% Max	1573.00
99%	1300.78
95%	985.14
90%	741.01

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

**Freq: WeightD**

Reference person - Sex=Male

Quantiles (Definition 5)	
Level	Quantile
75% Q3	425.26
50% Median	226.42
25% Q1	116.22
10%	65.70
5%	50.75
1%	21.28
0% Min	11.83

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
11.83	3940	1946	1511.64	5649	1395
14.08	15313	2157	1534.26	1566	1654
15.08	1520	2184	1556.41	1360	1997
20.54	4062	2091	1557.89	759	2302
20.80	8060	2250	1573.00	43	1742

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_Sex	2	Female Male

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

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	115475.02759	115475.02759	1.27	0.2606
Error	8.13E6	741647014366	91236.13115		
Corrected Total	8.13E6	741647129841			

R-Square	Coeff Var	Root MSE	FD806 Mean
0.000000	92.40926	302.0532	326.8646

Source	DF	Type I SS	Mean Square	F Value	Pr > F
RP_Sex	1	115475.0286	115475.0286	1.27	0.2606

Source	DF	Type III SS	Mean Square	F Value	Pr > F
RP_Sex	1	115475.0286	115475.0286	1.27	0.2606



### 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
RP_Sex	1	1.5525E8	1.5525E8	3886.75	<.0001
Error	8.13E6	3.247E11	39943.6		

Welch's ANOVA for FD806			
Source	DF	F Value	Pr > F
RP_Sex	1.0000	1.27	0.2592
Error	8117345		

### The GLM Procedure

Level of RP_Sex	N	FD806	
		Mean	Std Dev
Female	4262186	326.978148	310.478872
Male	3866690	326.739492	292.484616

### The TTEST Procedure

Variable: FD806 (Expense Non-alcoholic beverages)

Frequency: WeightD

RP_Sex	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Female		4262186	327.0	310.5	0.1504	21.5800	1569.4
Male		3866690	326.7	292.5	0.1487	11.8300	1573.0
Diff (1-2)	Pooled		0.2387	302.1	0.2121		
Diff (1-2)	Satterthwaite		0.2387		0.2115		

### The TTEST Procedure

Variable: FD806 (Expense Non-alcoholic beverages)

RP_Sex	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Female		327.0	326.7	327.3	310.5	310.3	310.7
Male		326.7	326.4	327.0	292.5	292.3	292.7
Diff (1-2)	Pooled	0.2387	-0.1771	0.6544	302.1	301.9	302.2
Diff (1-2)	Satterthwaite	0.2387	-0.1759	0.6532			

Method	Variances	DF	t Value	Pr >  t
Pooled	Equal	8.13E6	1.13	0.2606
Satterthwaite	Unequal	8.12E6	1.13	0.2592

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	4.26E6	3.87E6	1.13	<.0001

## The TTEST Procedure

Variable: FD806 (Expense Non-alcoholic beverages)

