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

The MEANS Procedure

Analysis Variable : FD806 Expense Non-alcoholic beverages														
Region	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
Atlantic provinces	629768	629768	0	21.58	144.56	310.70	404.59	573.30	1561.82	428.74	84.86	403.75	405.44	1.36
Central	4987062	4987062	0	14.08	101.14	195.16	294.35	382.58	1469.40	281.44	95.09	294.10	294.59	1.81
Prairie	1530428	1530428	0	11.83	131.33	258.76	387.31	575.82	1534.26	444.49	86.91	386.77	387.84	1.33
Territorial capitals	11711	11711	0	35.62	146.21	317.31	447.33	633.10	1573.00	486.89	82.03	440.68	453.98	1.02
West	969907	969907	0	24.96	143.48	258.15	346.77	446.82	1569.37	303.34	86.11	346.17	347.36	1.68

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

Region=Atlantic provinces

Moments			
N	629768	**Sum Weights**	629768
Mean	404.594424	**Sum Observations**	254800621
Std Deviation	343.356582	**Variance**	117893.743
Skewness	1.35825064	**Kurtosis**	1.46298435

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

Freq: WeightD

Region=Atlantic provinces

Moments			
Uncorrected SS	1.77336E11	**Corrected SS**	7.42456E10
Coeff Variation	84.8643882	**Std Error Mean**	0.43266831

Basic Statistical Measures			
Location		**Variability**	
Mean	404.5944	**Std Deviation**	343.35658
Median	310.7000	**Variance**	117894
Mode	76.1800	**Range**	1540
		Interquartile Range	428.74000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	935.1145	**Pr > |t|**	<.0001
Sign	**M**	314884	**Pr >= |M|**	<.0001
Signed Rank	**S**	9.915E10	**Pr >= |S|**	<.0001

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.133816	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	4094.321	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	25014.54	**Pr > A-Sq**	<0.0050

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

Freq: WeightD

Region=Atlantic provinces

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1561.82
99%	1513.44
95%	1133.16
90%	844.40
75% Q3	573.30
50% Median	310.70
25% Q1	144.56
10%	75.69
5%	49.00
1%	26.00
0% Min	21.58

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
21.58	840	331	1540.76	776	205
21.58	347	241	1546.22	669	35
23.40	132	176	1556.41	1360	634
24.82	209	688	1557.89	759	717
24.96	186	527	1561.82	491	226

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

Freq: WeightD

Region=Central

Moments			
N	4987062	**Sum Weights**	4987062
Mean	294.346271	**Sum Observations**	1467923104
Std Deviation	279.893261	**Variance**	78340.2375
Skewness	1.80974594	**Kurtosis**	3.35664873
Uncorrected SS	8.22765E11	**Corrected SS**	3.90688E11
Coeff Variation	95.0897933	**Std Error Mean**	0.12533433

Basic Statistical Measures			
Location		**Variability**	
Mean	294.3463	**Std Deviation**	279.89326
Median	195.1600	**Variance**	78340
Mode	72.0200	**Range**	1455
		Interquartile Range	281.44000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	2348.489	**Pr > |t|**	<.0001
Sign	**M**	2493531	**Pr >= |M|**	<.0001
Signed Rank	**S**	6.218E12	**Pr >= |S|**	<.0001

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

Freq: WeightD

Region=Central

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.162274	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	51027.26	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	293746.8	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1469.40
99%	1309.62
95%	905.69
90%	675.29
75% Q3	382.58
50% Median	195.16
25% Q1	101.14
10%	54.08
5%	48.08
1%	26.52
0% Min	14.08

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

Freq: WeightD

Region=Central

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
14.08	15313	1259	1384.86	5785	1060
20.54	4062	1240	1403.07	6213	1146
20.80	8060	1283	1410.32	2957	888
21.28	7665	1297	1464.24	12702	903
26.52	16121	1013	1469.40	12196	792

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

Freq: WeightD

Region=Prairie

Moments			
N	1530428	Sum Weights	1530428
Mean	387.308019	Sum Observations	592747037
Std Deviation	336.62841	Variance	113318.687
Skewness	1.32875966	Kurtosis	1.23576465
Uncorrected SS	4.03002E11	Corrected SS	1.73426E11
Coeff Variation	86.9149084	Std Error Mean	0.27210988

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

Freq: WeightD

Region=Prairie

Basic Statistical Measures			
Location		**Variability**	
Mean	387.3080	**Std Deviation**	336.62841
Median	258.7600	**Variance**	113319
Mode	626.0800	**Range**	1522
		Interquartile Range	444.49000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	1423.352	**Pr > |t|**	<.0001
Sign	**M**	765214	**Pr >= |M|**	<.0001
Signed Rank	**S**	5.856E11	**Pr >= |S|**	<.0001

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.155942	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	12359.2	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	71097.72	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1534.26
99%	1494.48
95%	1169.08
90%	839.54

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

Freq: WeightD

Region=Prairie

Quantiles (Definition 5)	
Level	**Quantile**
75% Q3	575.82
50% Median	258.76
25% Q1	131.33
10%	75.92
5%	55.28
1%	29.55
0% Min	11.83

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
11.83	3940	1853	1449.41	505	1585
15.08	1520	1916	1485.65	5912	1353
21.84	1372	1347	1494.48	8411	1436
21.84	3250	1321	1511.64	5649	1694
22.36	381	1617	1534.26	1566	1781

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

Freq: WeightD

Region=Territorial capitals

Moments			
N	11711	**Sum Weights**	11711
Mean	447.32884	**Sum Observations**	5238668.05
Std Deviation	366.931505	**Variance**	134638.729
Skewness	1.0227851	**Kurtosis**	0.2342045
Uncorrected SS	3920026824	**Corrected SS**	1576619520
Coeff Variation	82.0272407	**Std Error Mean**	3.39068942

Basic Statistical Measures			
Location		**Variability**	
Mean	447.3288	**Std Deviation**	366.93150
Median	317.3100	**Variance**	134639
Mode	96.9400	**Range**	1537
		Interquartile Range	486.89000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	131.9286	**Pr > |t|**	<.0001
Sign	**M**	5855.5	**Pr >= |M|**	<.0001
Signed Rank	**S**	34289808	**Pr >= |S|**	<.0001

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

Freq: WeightD

Region=Territorial capitals

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.155138	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	69.88919	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	422.3186	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1573.00
99%	1494.74
95%	1191.01
90%	1029.41
75% Q3	633.10
50% Median	317.31
25% Q1	146.21
10%	70.20
5%	62.66
1%	43.96
0% Min	35.62

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

Freq: WeightD

Region=Territorial capitals

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
35.62	18	2074	1447.85	46	2131
43.96	115	2058	1462.25	45	2096
47.35	157	2064	1494.74	58	2104
50.18	58	2130	1522.61	59	1964
51.48	33	2065	1573.00	43	2084

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

Freq: WeightD

Region=West

Moments			
N	969907	**Sum Weights**	969907
Mean	346.767865	**Sum Observations**	336332580
Std Deviation	298.615022	**Variance**	89170.9313
Skewness	1.68376037	**Kurtosis**	2.83724693
Uncorrected SS	2.03117E11	**Corrected SS**	8.64874E10
Coeff Variation	86.1138103	**Std Error Mean**	0.30321215

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

Freq: WeightD

Region=West

Basic Statistical Measures			
Location		**Variability**	
Mean	346.7679	**Std Deviation**	298.61502
Median	258.1500	**Variance**	89171
Mode	516.1000	**Range**	1544
		Interquartile Range	303.34000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	1143.648	**Pr > |t|**	<.0001
Sign	**M**	484953.5	**Pr >= |M|**	<.0001
Signed Rank	**S**	2.352E11	**Pr >= |S|**	<.0001

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.141356	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	7488.665	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	46314.23	**Pr > A-Sq**	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1569.37
99%	1346.28
95%	1145.04
90%	741.98

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

Freq: WeightD

Region=West

Quantiles (Definition 5)	
Level	**Quantile**
75% Q3	446.82
50% Median	258.15
25% Q1	143.48
10%	63.31
5%	48.36
1%	28.86
0% Min	24.96

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
24.96	1553	2301	1217.58	9539	2207
28.00	6907	2300	1327.30	3791	2255
28.86	3183	2272	1346.28	5467	2235
31.20	4303	2210	1354.61	1683	2307
37.95	3779	2260	1569.37	4907	2170

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**
Region	5	Atlantic provinces Central Prairie Territorial capitals West

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	4	15223979649	3805994912.1	42590.1	<.0001
Error	8.13E6	726423150193	89363.350728		
Corrected Total	8.13E6	741647129841			

R-Square	Coeff Var	Root MSE	FD806 Mean
0.020527	91.45592	298.9370	326.8646

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Region	4	15223979649	3805994912	42590.1	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Region	4	15223979649	3805994912	42590.1	<.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
Region	4	5.5724E9	1.3931E9	35980.8	<.0001
Error	8.13E6	3.147E11	38717.7		

Welch's ANOVA for FD806			
Source	DF	F Value	Pr > F
Region	4.0000	37624.5	<.0001
Error	90957.2		

The GLM Procedure

Level of Region	N	FD806	
		Mean	Std Dev
Atlantic provinces	629768	404.594424	343.356582
Central	4987062	294.346271	279.893261
Prairie	1530428	387.308019	336.628410
Territorial capitals	11711	447.328840	366.931505
West	969907	346.767865	298.615022

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

Region	FD806 LSMEAN	LSMEAN Number
Atlantic provinces	404.594424	1
Central	294.346271	2
Prairie	387.308019	3

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

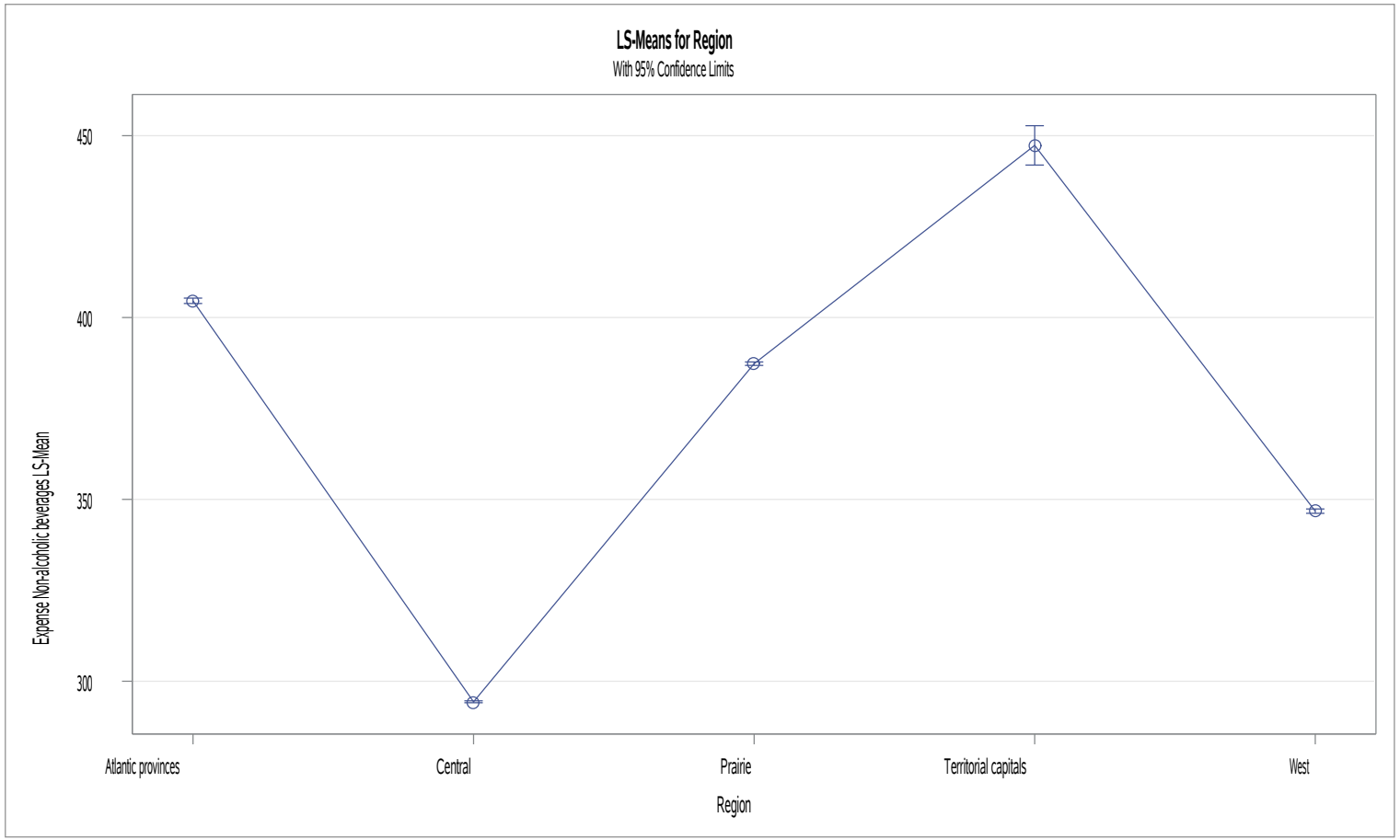
Region	FD806 LSMEAN	LSMEAN Number
Territorial capitals	447.328840	4
West	346.767865	5

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

Dependent Variable: FD806

i/j	1	2	3	4	5
1		<.0001	<.0001	<.0001	<.0001
2	<.0001		<.0001	<.0001	<.0001
3	<.0001	<.0001		<.0001	<.0001
4	<.0001	<.0001	<.0001		<.0001
5	<.0001	<.0001	<.0001	<.0001	

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



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