Food Expense

What impacts Store Expense Behaviour?

Metro College of Technology
November 2021
Ana Clara Tupinambá Freitas, oriented by Professor Arkar Min

Outline

- Introduction
- Executive Summary
- Business Questions
- Methodology
  - Phase 1:Identification of target variable
    - Data profiling
    - Data preparation
    - Visuals
  - Phase2: Model
    - EDA /Data preparation
    - Hypothesis
    - Modeling
- Results
- Conclusion & Recommendations

*Appendix

- Tabular summaries, Extra Graphs and SAS Script

Introduction

- The **Survey of Household Spending (SHS)** collects detailed spending information (User Guide from the SHS PUMF,2017 from Statistics Canada)

- Shape: 4012 x 437 weight more than 14 millions datapoints
  - 45 categorical (CaseId)
  - 394 numerical (WeightD – Frequency)

- Study data refers to **Food Expenses** based on a daily expenditure that spans over a **2-week period** […]that was annualized. […]Data on **income** […]come mainly from […]the **Canada Revenue Agency** from the preceding year. (User Guide from the SHS PUMF,2017)

- **SAS** Enterprise Guide, Excel and Adobe Acrobat reader was used in this study.

https://www150.statcan.gc.ca/n1/pub/62m0004x/62m0004x2017001-eng.htm

Executive Summary

To know how much a household spends on food can help stores decide which products maintain in shelfs and give insights to marketing.

This study has **2 phases**:

1. Identification of **product** with **biggest expense nationally**
2. Identification of **what impacts the expense** of the identified product.

To **maintain revenue**, the recommendation is that **special attention** should be applied for **stock keeping and marketing** for **areas with characteristics that leads to higher spending**.

Investigate and address areas with less spending to increase sales.

Business Questions

- Which product expense is the biggest regarding store purchase?(target)
- Is the expense diverse regarding geography (Region and Province)?
- Is the expense diverse regarding Household Type?
- Does higher Household Income mean higher expense?
- What impacts the expense of the previously identified product?

Which product expense is the biggest regarding store purchase?

Original Dataset Profiling

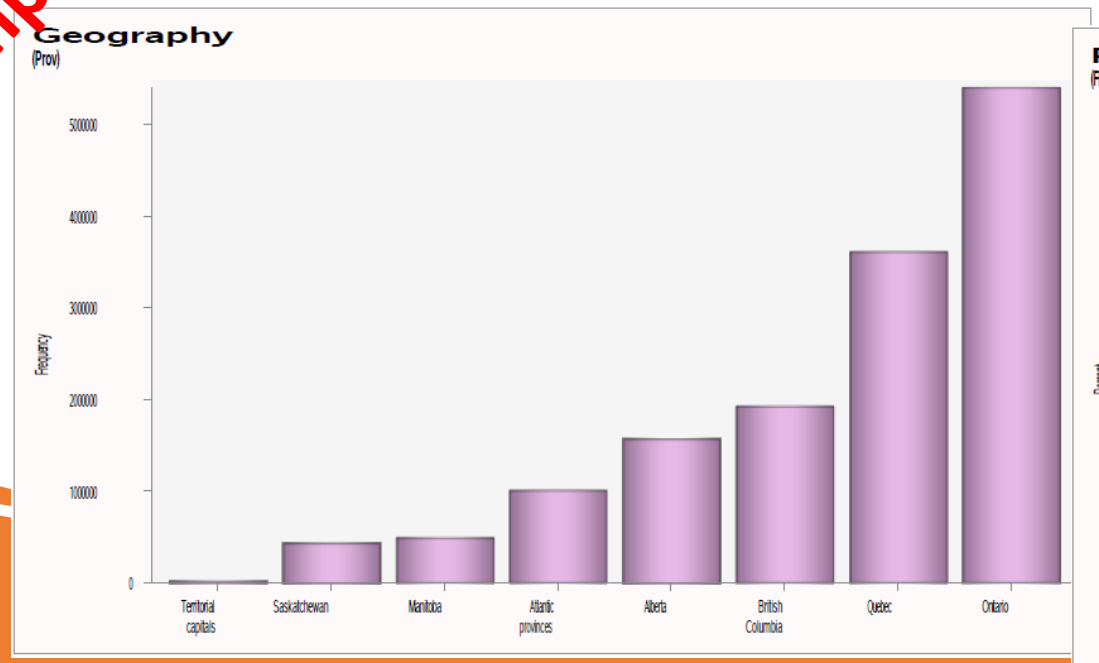
The CONTENTS Procedure

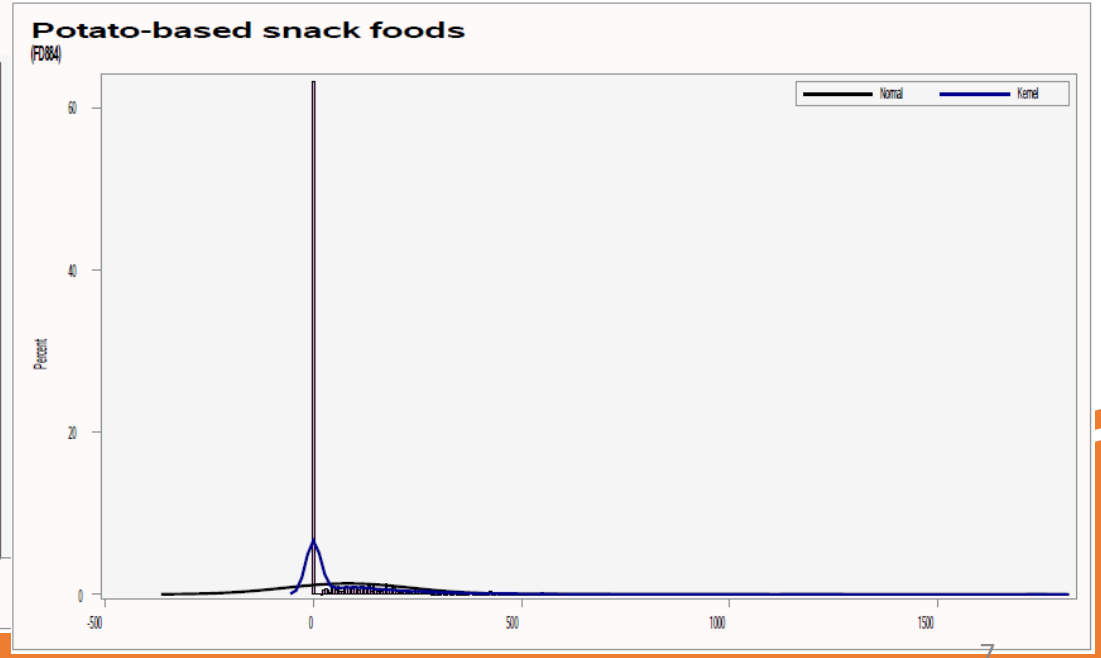
Data Set Name	ANA.DEM1	Observations	4012
Member Type	DATA	Variables	437
Engine	V9	Indexes	0
Created	11/18/2021 10:49:29	Observation Length	3864
Last Modified	11/18/2021 10:49:29	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	WINDOWS_64		
Encoding	wlatin1 Western (Windows)		

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat	Label
23	CC001	Num	8	BEST12.	BEST32.	Child care
24	CC001_C	Num	8	BEST12.	BEST32.	Child care - Interview
25	CC001_D	Num	8	BEST12.	BEST32.	Child care - Diary
26	CF001	Num	8	BEST12.	BEST32.	Women's and girls' wear (4 years and over)
27	CI001	Num	8	BEST12.	BEST32.	Children's wear (under 4 years)
⋮						
433	TVCon_SatDish	Char	3			Type of television services - Satellite dish
393	TX001	Num	8	BEST12.	BEST32.	Income taxes
415	Tenure	Char	22			Dwelling tenure
435	VehicleYN	Char	3			Owned, leased or operated a vehicle
2	WeightD	Num	8	BEST12.	BEST32.	
417	YearBuilt	Char	37			Period of construction of the dwelling

Examples





Data Preparation

The CONTENTS Procedure

Data Set Name	WORK.FOOD1	Observations	4012
Member Type	DATA	Variables	153
		Indexes	0
		Observation Length	1240
		Deleted Observations	0
		Compressed	NO
		Sorted	NO

Obs	CaseID	WeightD	FD003	FD100	FD1001	FD1002	...	FD881	FD882	FD883	FD884	FD885	FD889	Prov
1	10	1174.4303	8532.42	591.24	0	0		0	0	0	0	0	0	Atlantic provinces
2	40	107.7481	5420.36	90.74	0	0		0	0	107.9	107.9	0	0	Territorial capitals
3	60	7339.229	13105.37	330.84	0	0		324.55	0	0	0	0	0	British Columbia
4	130	274.6872	18158.2	2539.54	111.44	0		0	0	322.55	322.55	0	0	Atlantic provinces
5	140	10845.9064	2412.54	130.26	0	0		0	0	0	0	0	0	Quebec
6	150	1390.9806	2259.29	272.4	0	0		0	0	0	0	0	0	Saskatchewan
7	160	931.971	7464.18	1148.79	217.31	0		0	0	521.97	380.74	141.23	0	Atlantic provinces
8	190	3130.3169	8853.52	129.22	0	0		0	0	0	0	0	0	British Columbia
9	240	1160.7444	9922.64	1372.54	0	238.94		0	0	0	0	0	0	Atlantic provinces
10	270	61.2528	13687.79	392.83	0	0		0	0	0	0	0	0	Territorial capitals

Product Category

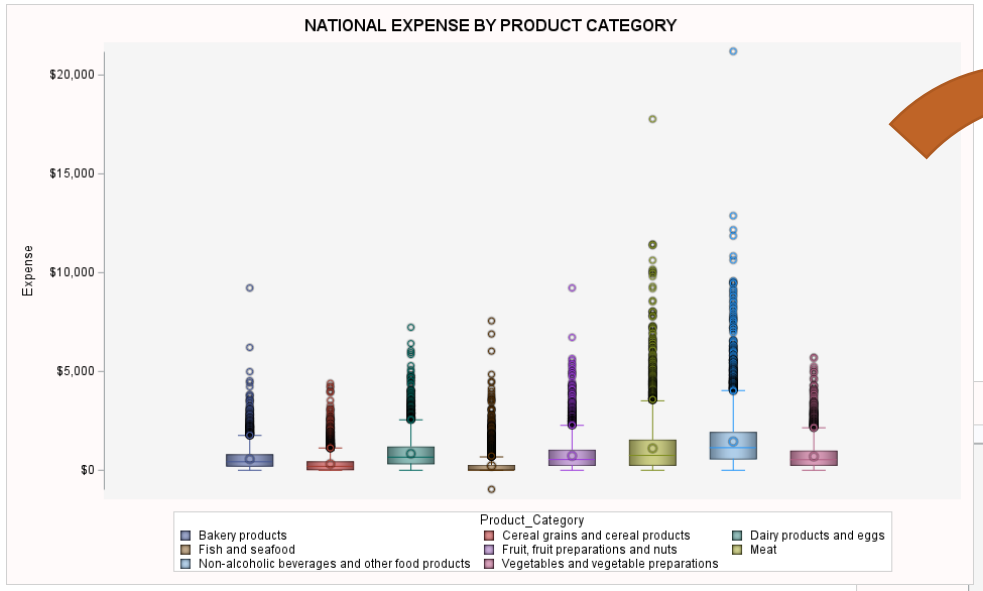
Product SubCategory

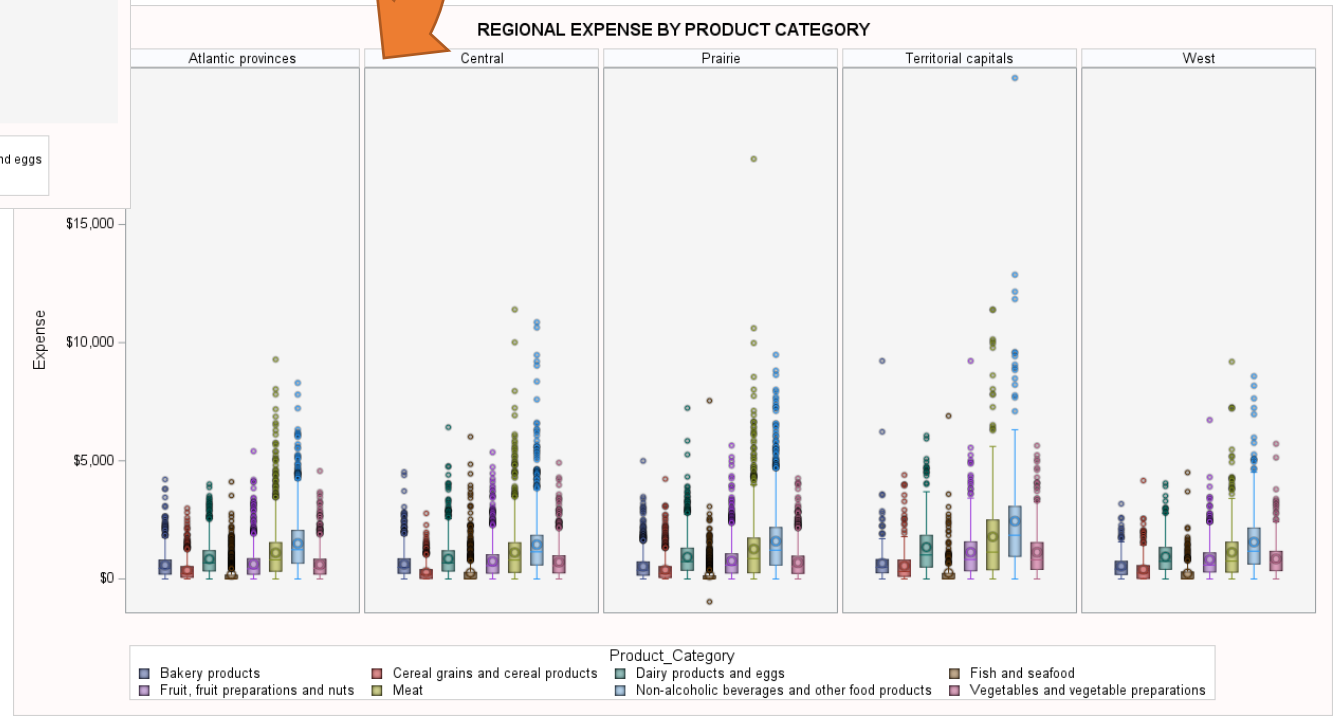
Product

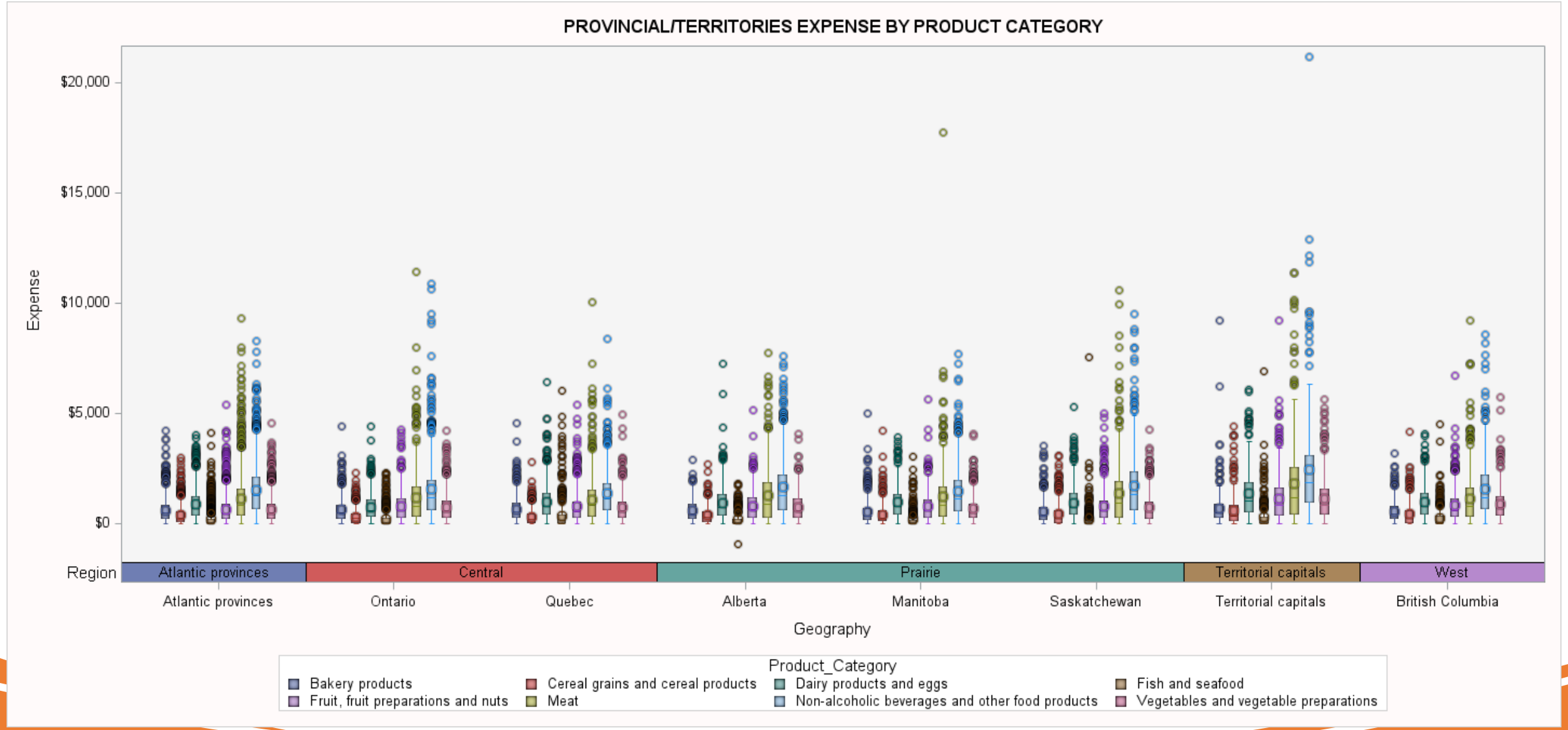
e CONTENTS Procedure

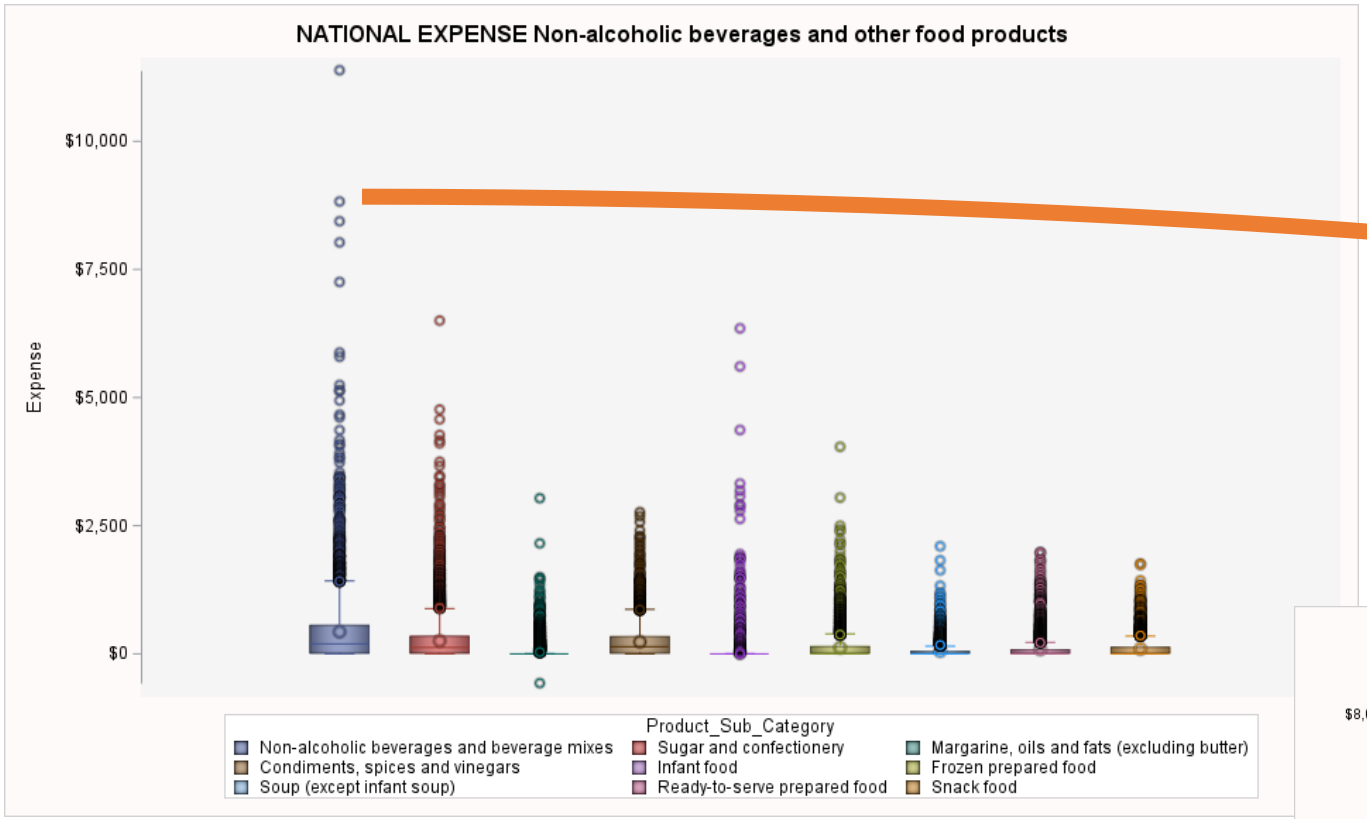
Data Set Name	ANA.FOOD	Observations	601800
Member Type	DATA	Variables	9

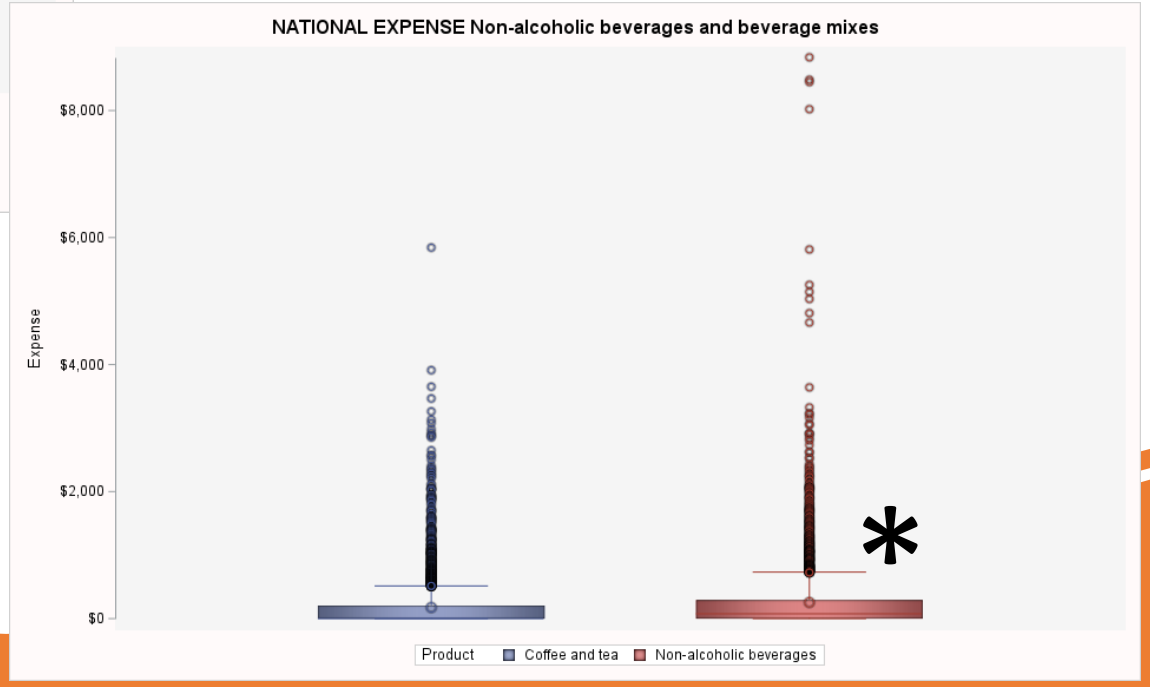
Obs	CaseID	WeightD	Region	Prov	FD	Product_Category	Product_Sub_Category	Product	Expense
1	10	1174.4303	Atlantic provinces	Atlantic provinces	FD003				$8,532.42
2	10	1174.4303	Atlantic provinces	Atlantic provinces	FD1001		Frozen side dishes and other frozen prepared food		$0.00
3	10	1174.4303	Atlantic provinces	Atlantic provinces	FD1002			Other ready-to-serve prepared food	$0.00
4	10	1174.4303	Atlantic provinces	Atlantic provinces	FD1003			Cod, flounder, sole and haddock (fresh or frozen, uncooked)	$0.00
5	10	1174.4303	Atlantic provinces	Atlantic provinces	FD1004			Other oils and fats	$0.00
6	10	1174.4303	Atlantic provinces	Atlantic provinces	FD101			Bread and unsweetened rolls and buns	$350.74
7	10	1174.4303	Atlantic provinces	Atlantic provinces	FD102			Bread	$90.74
8	10	1174.4303	Atlantic provinces	Atlantic provinces	FD103			Unsweetened rolls and buns	$260.00
9	10	1174.4303	Atlantic provinces	Atlantic provinces	FD104		Cookies and crackers		$0.00
10	10	1174.4303	Atlantic provinces	Atlantic provinces	FD105			Cookies and sweet biscuits	$0.00


NATIONAL EXPENSE BY PRODUCT CATEGORY
$20,000
$15,000
$10,000
$5,000
$0
Expense
Product_Category
Bakery products
Cereal grains and cereal products
Dairy products and eggs
Fish and seafood
Fruit, fruit preparations and nuts
Meat
Non-alcoholic beverages and other food products
Vegetables and vegetable preparations

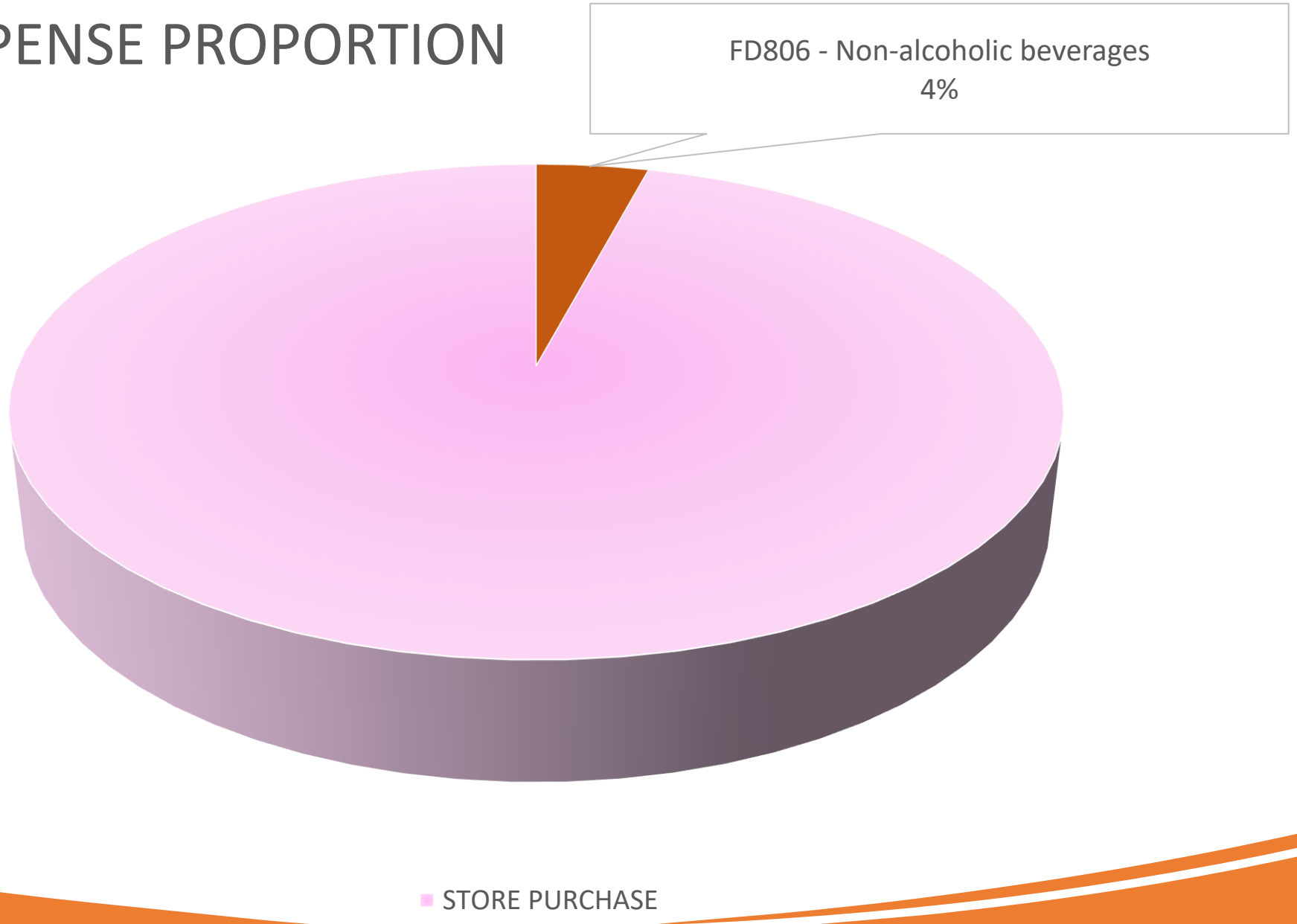

REGIONAL EXPENSE BY PRODUCT CATEGORY
Atlantic provinces
Central
Prairie
Territorial capitals
West
$15,000
$10,000
$5,000
$0
Expense
Product_Category
Bakery products
Cereal grains and cereal products
Dairy products and eggs
Fish and seafood
Fruit, fruit preparations and nuts
Meat
Non-alcoholic beverages and other food products
Vegetables and vegetable preparations


PROVINCIAL/TERRITORIES EXPENSE BY PRODUCT CATEGORY
$20,000
$15,000
$10,000
$5,000
$0
Expense
Region
Atlantic provinces
Central
Prairie
Territorial capitals
West
Atlantic provinces
Ontario
Quebec
Alberta
Manitoba
Saskatchewan
Territorial capitals
British Columbia
Geography
Product_Category
Bakery products
Cereal grains and cereal products
Dairy products and eggs
Fish and seafood
Fruit, fruit preparations and nuts
Meat
Non-alcoholic beverages and other food products
Vegetables and vegetable preparations

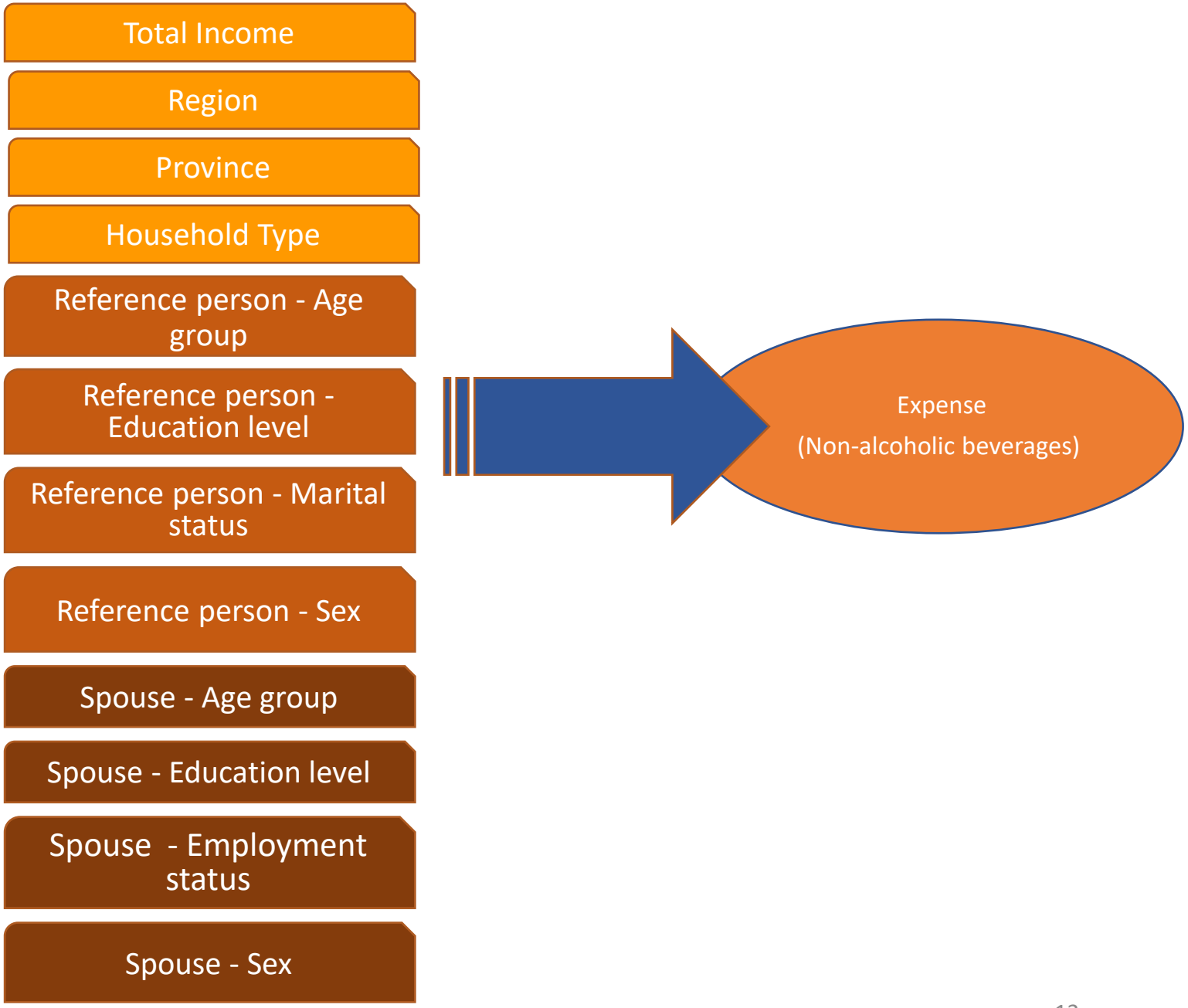

NATIONAL EXPENSE Non-alcoholic beverages and other food products
$10,000
$7,500
$5,000
$2,500
$0
Expense
Product_Sub_Category
Non-alcoholic beverages and beverage mixes
Condiments, spices and vinegars
Soup (except infant soup)
Sugar and confectionery
Infant food
Ready-to-serve prepared food
Margarine, oils and fats (excluding butter)
Frozen prepared food
Snack food


NATIONAL EXPENSE Non-alcoholic beverages and beverage mixes
$8,000
$6,000
$4,000
$2,000
$0
Expense
*
Product
Coffee and tea
Non-alcoholic beverages

NATIONAL EXPENSE PROPORTION



Study Framework



Exploratory Data Analysis

Data Preparation

- Sub setting
- Exclusion of outliers
- Exclusion zero expense

Data Set Name	ANA.MODEL	Observations	2421
Member Type	DATA	Variables	13
Engine	V9	Indexes	0
Created	11/19/2021 08:59:23	Observation Length	408
Last Modified	11/19/2021 08:59:23	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	WINDOWS_64		
Encoding	wlatin1 Western (Windows)		

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat	Label
3	FD806	Num	8	BEST12.	BEST32.	Non-alcoholic beverages
5	HHType6	Char	49			Household type
2	HH_TotInc	Num	8	BEST12.	BEST32.	Household - Total income
4	Prov	Char	20			Geography
6	RP_AgeGrp	Char	18			Reference person - Age group
9	RP_Educ	Char	107			Reference person - Education level
8	RP_MarStat	Char	30			Reference person - Marital status
7	RP_Sex	Char	6			Reference person - Sex
10	SP_AgeGrp	Char	18			Spouse - Age group
12	SP_Educ	Char	107			Spouse - Education level
13	SP_EmpStat	Char	16			Spouse - Employment status
11	SP_Sex	Char	9			Spouse - Sex
1	WeightD	Num	8	BEST12.	BEST32.	

Hypotheses

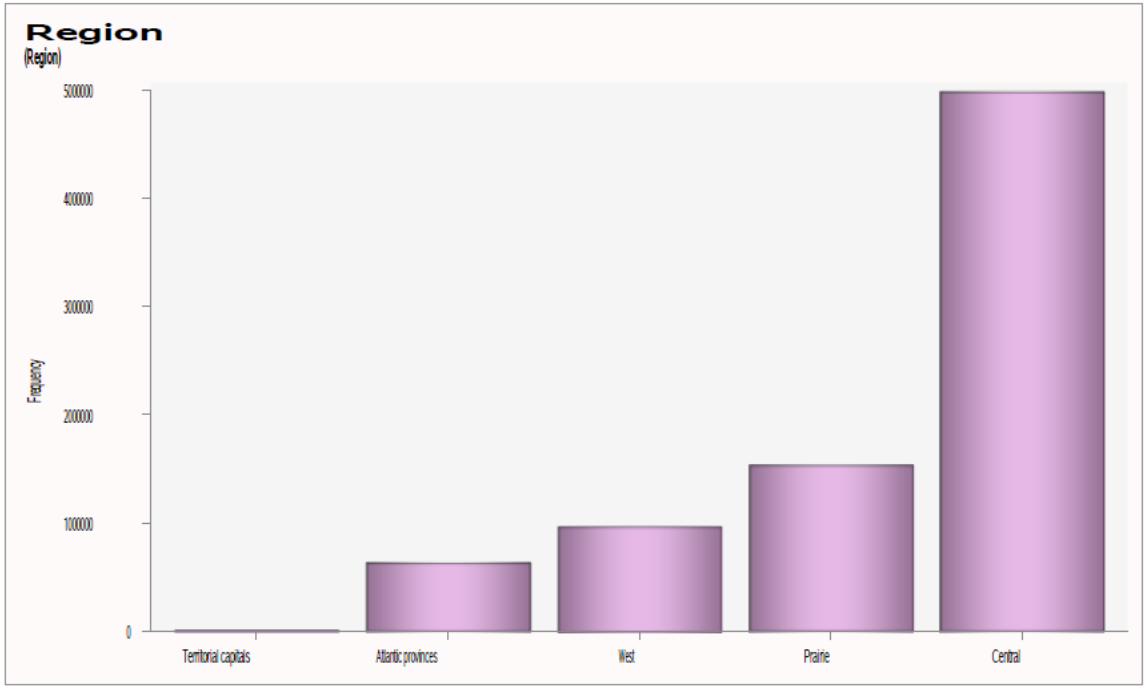
Different expense groups :

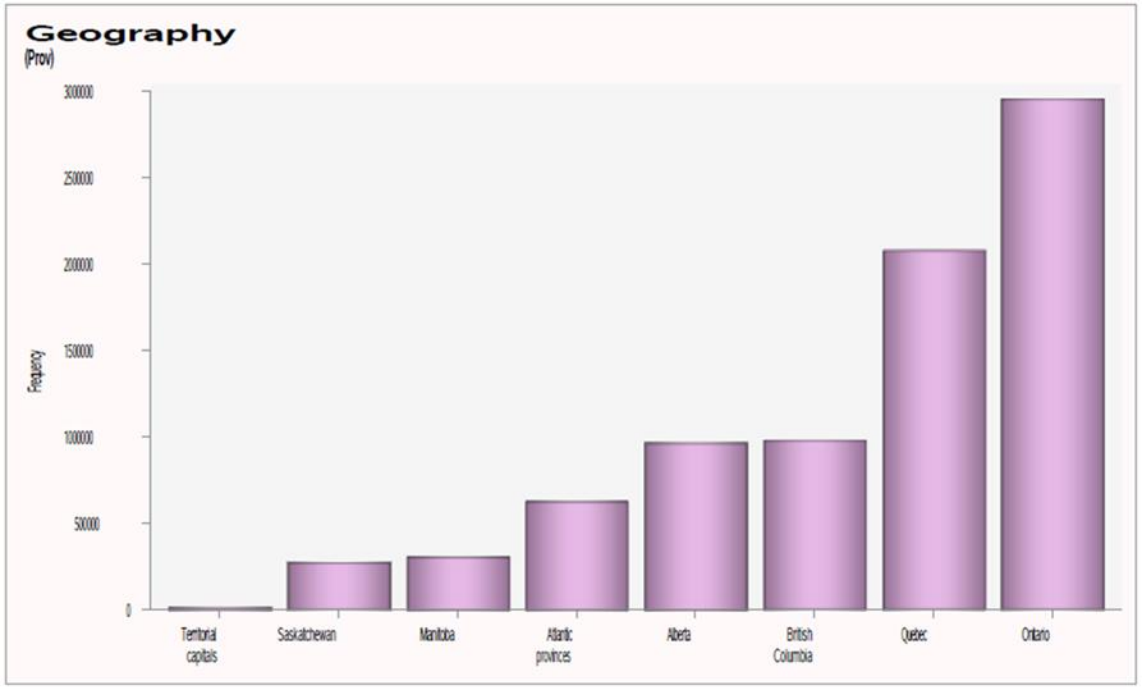
- Household dimension:
  - Region
  - Geography(Province)
  - Household Type
- Reference Person dimension:
  - Age Group
  - Marital Status
  - Education Level
  - Sex
- Reference Person dimension:
  - Age Group
  - Employment Status
  - Education Level
  - Sex

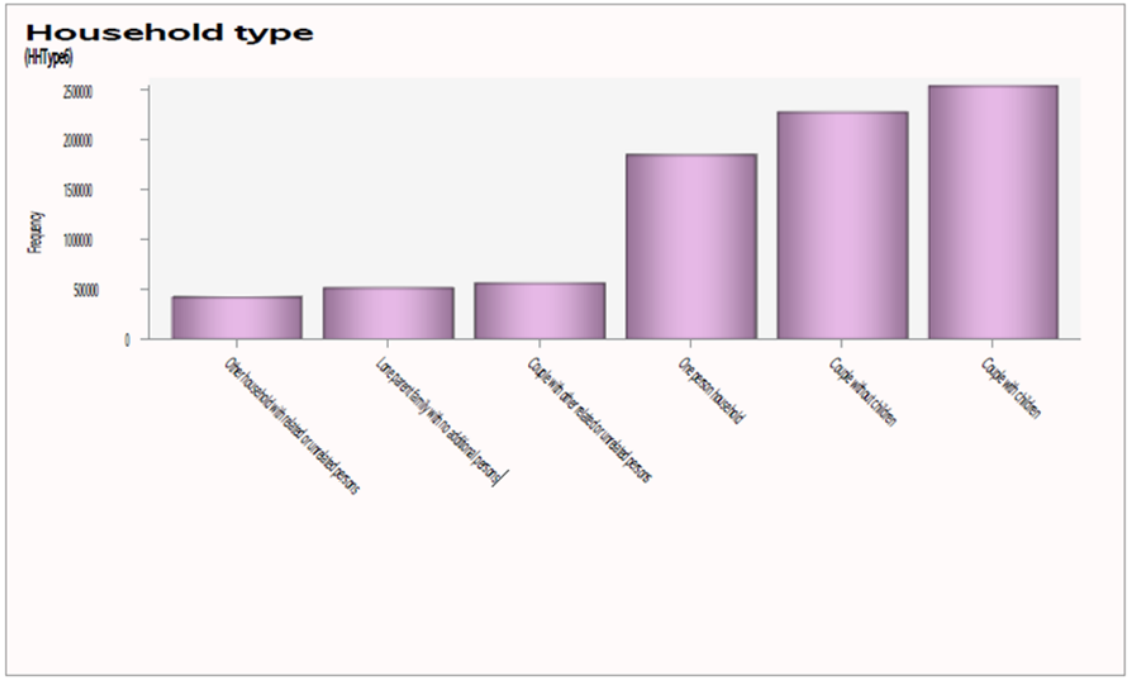
- Household Income is associated with higher Spending

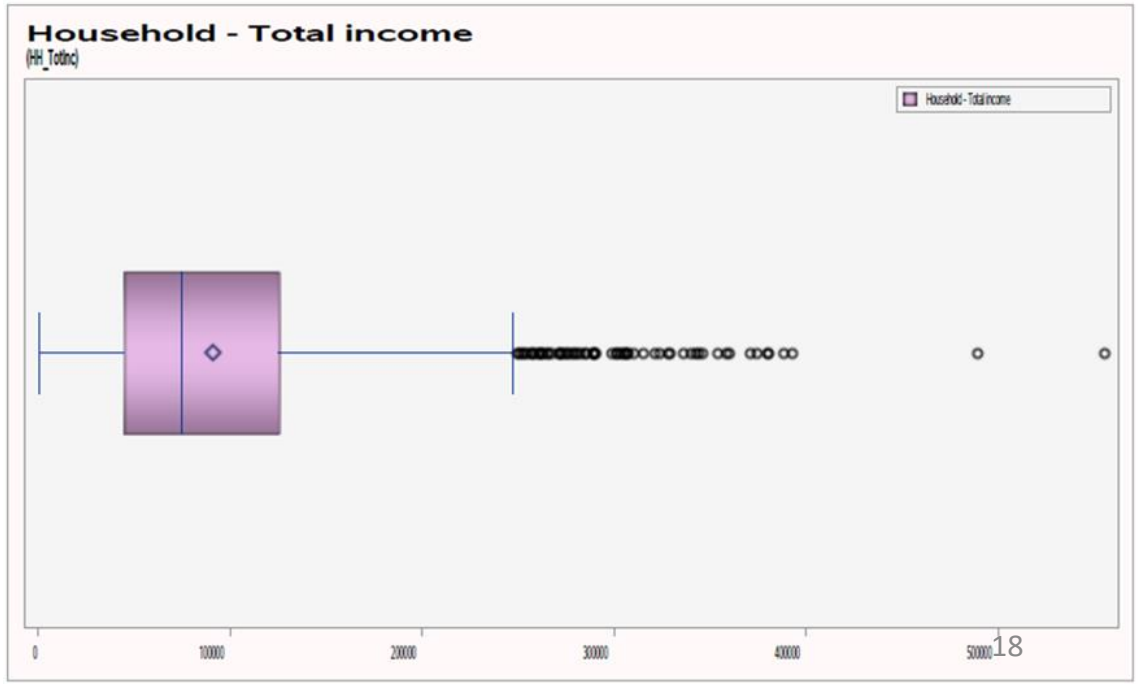
Univariate Analysis

Household

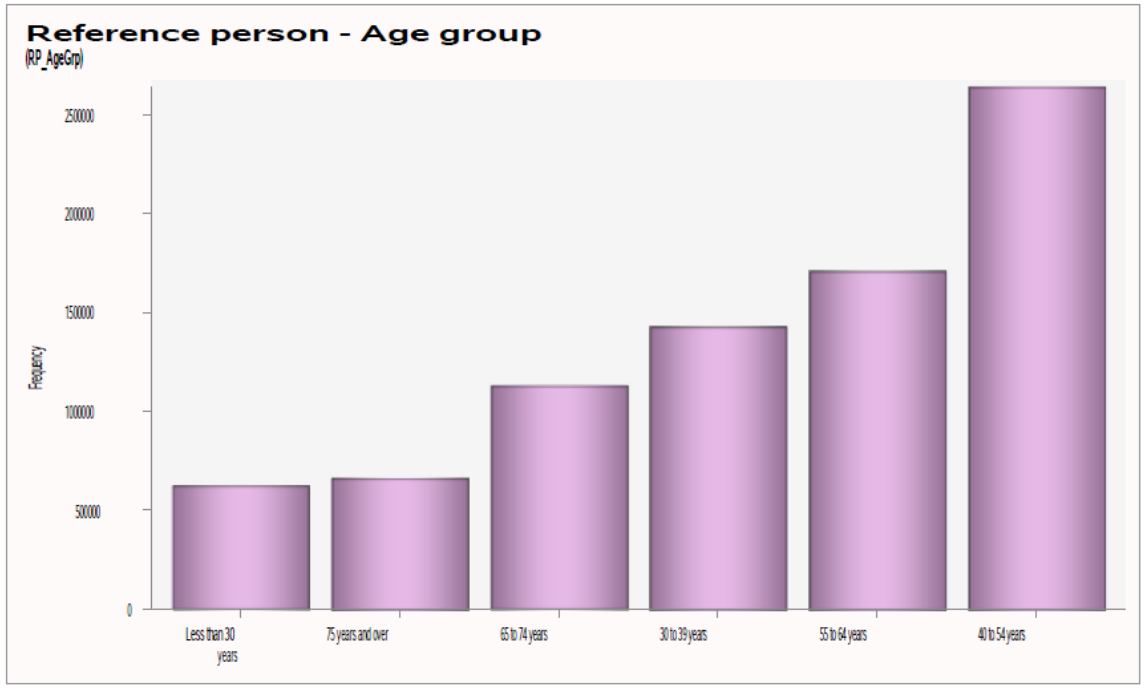

Region
(Region)
Frequency
5000000
4000000
3000000
2000000
1000000
0
Territorial capitals
Atlantic provinces
West
Prairie
Central

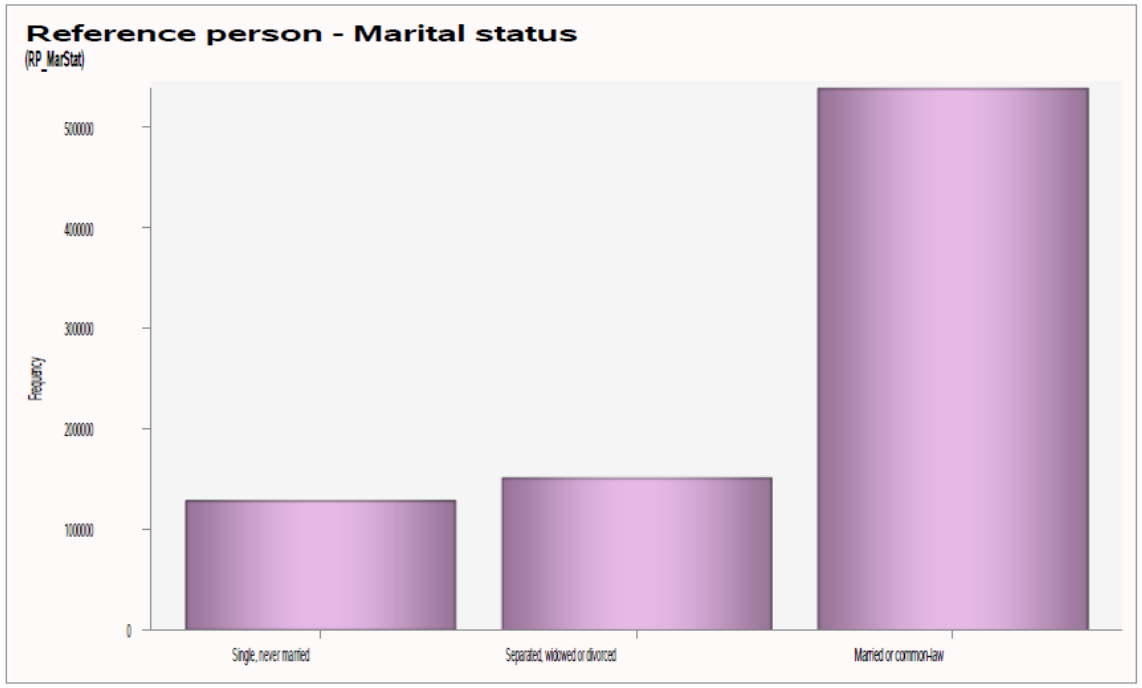

Geography
(Prov)
Frequency
3000000
2500000
2000000
1500000
1000000
500000
0
Territorial capitals
Saskatchewan
Manitoba
Atlantic provinces
Alberta
British Columbia
Quebec
Ontario

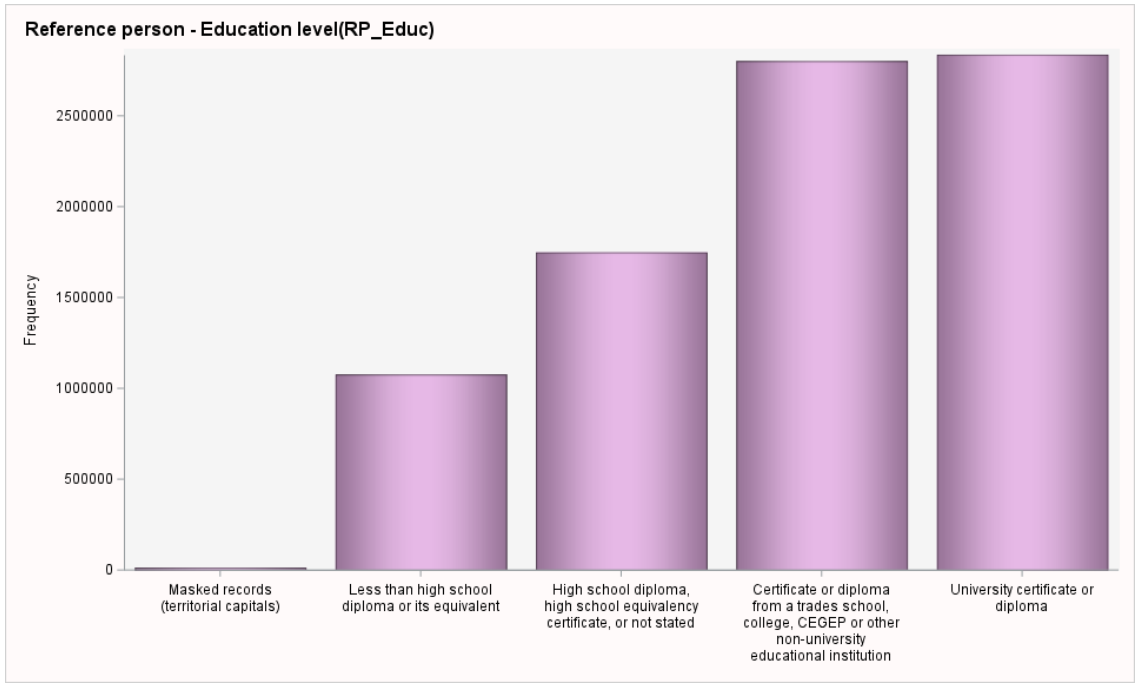

Household type
(HHType6)
Frequency
2500000
2000000
1500000
1000000
500000
0
One household with related or unrelated persons
Lone-parent family with no additional persons
Couple with other related or unrelated persons
One person household
Couple without children
Couple with children

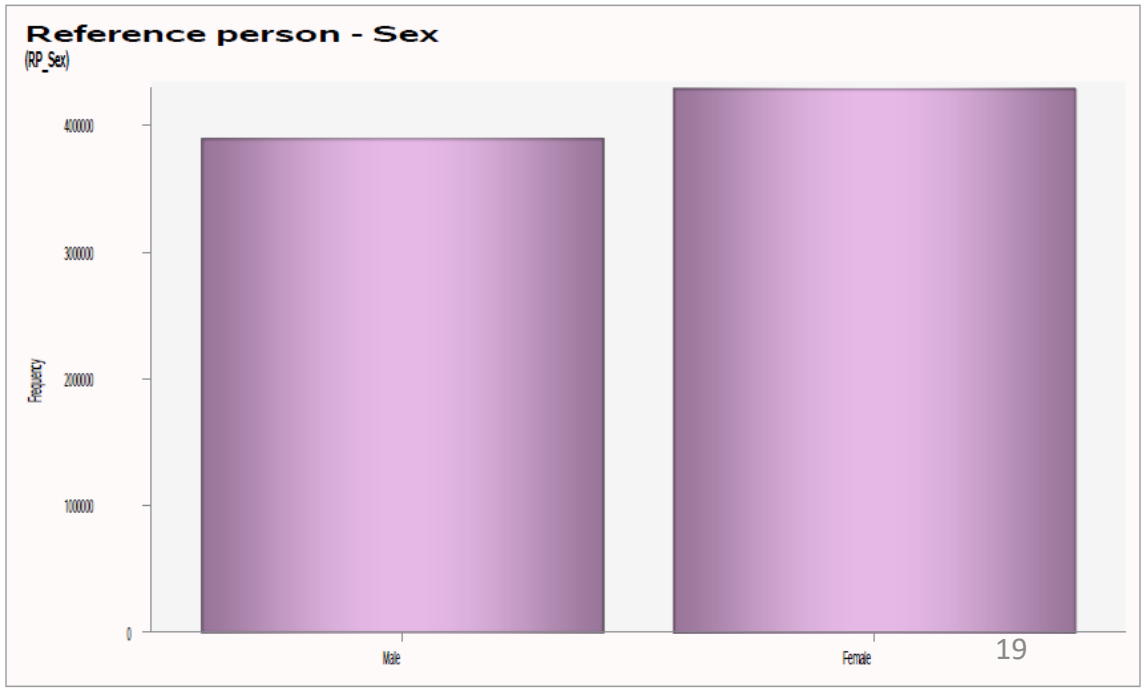

Household - Total income
(HH_Totinc)
Household - Total income
0
100000
200000
300000
400000
500000

Reference Person

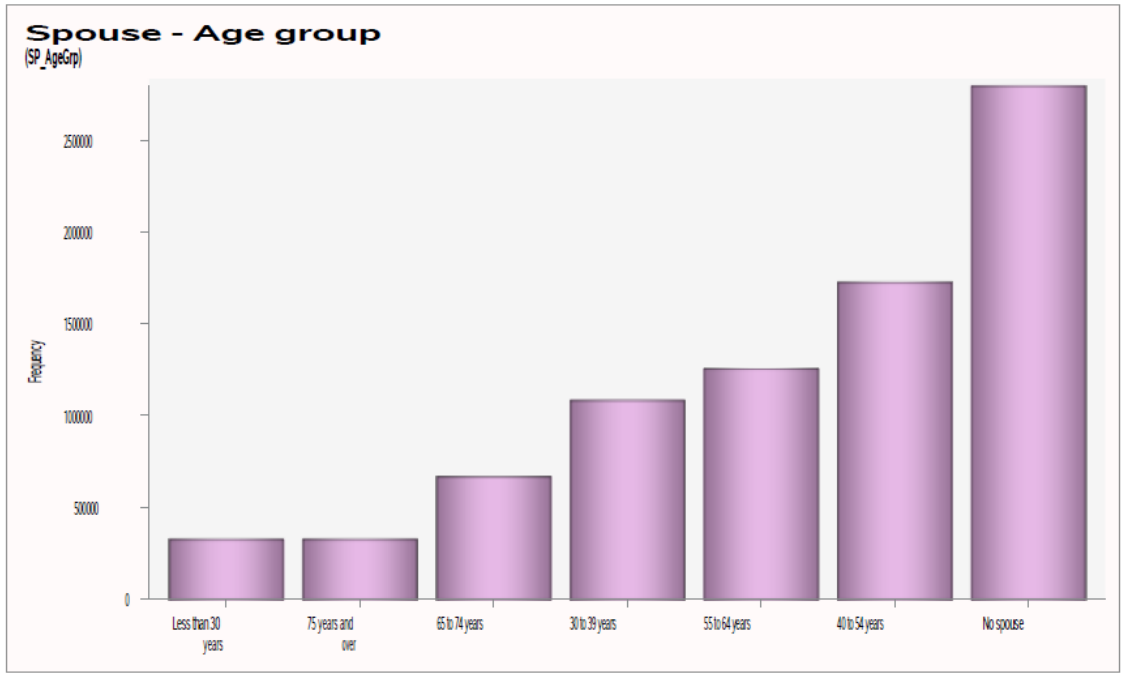


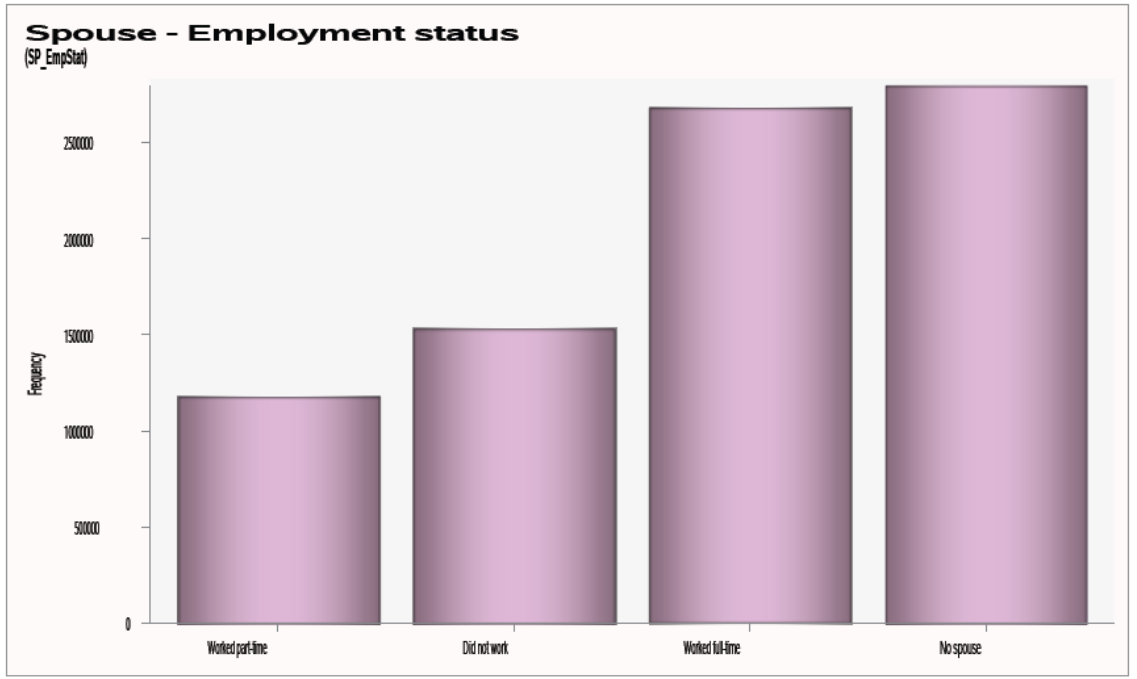


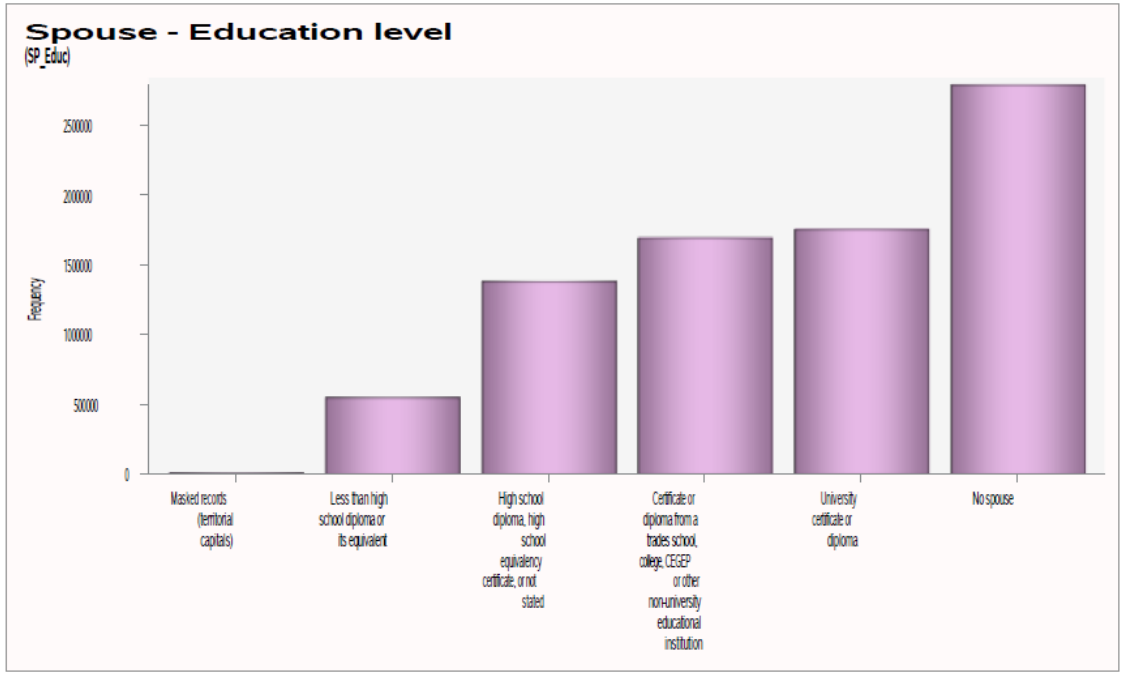


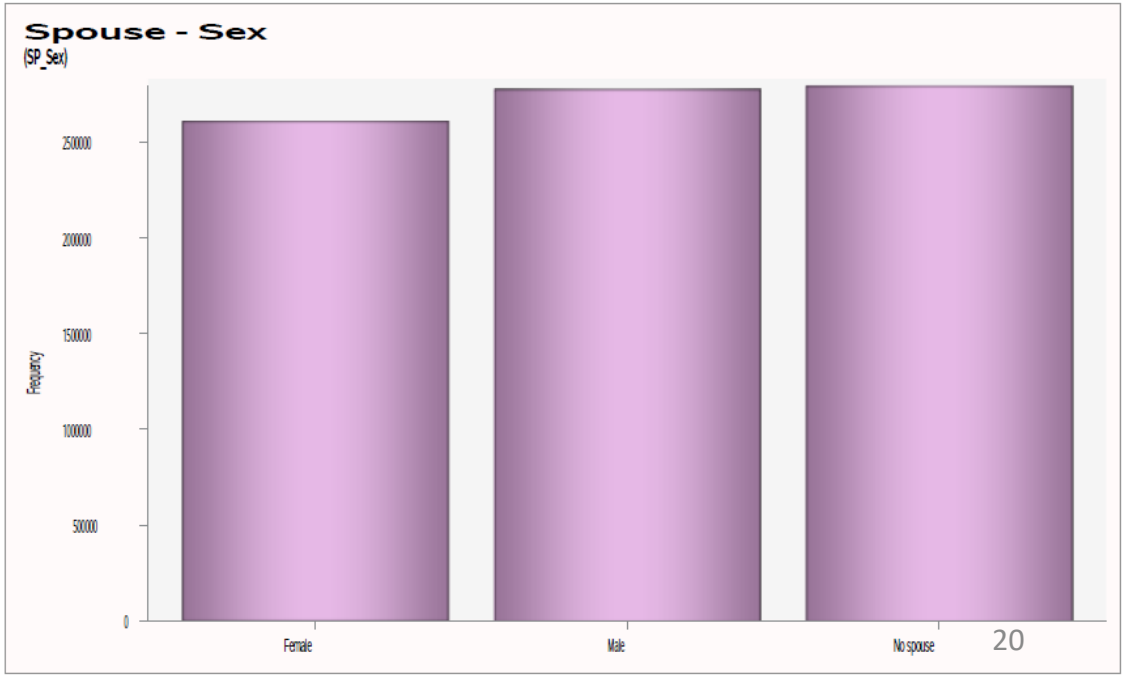


Spouse


Spouse - Age group
(SP_AgeGrp)
Frequency
2500000
2000000
1500000
1000000
500000
0
Less than 30 years
75 years and over
65 to 74 years
30 to 39 years
55 to 64 years
40 to 54 years
No spouse


Spouse - Employment status
(SP_EmpStat)
Frequency
2500000
2000000
1500000
1000000
500000
0
Worked part-time
Did not work
Worked full-time
No spouse


Spouse - Education level
(SP_Educ)
Frequency
2500000
2000000
1500000
1000000
500000
0
Masked records (territorial capitals)
Less than high school diploma or its equivalent
High school diploma, high school equivalency certificate, or not stated
Certificate or diploma from a trades school, college, CEGEP or other non-university educational institution
University certificate or diploma
No spouse

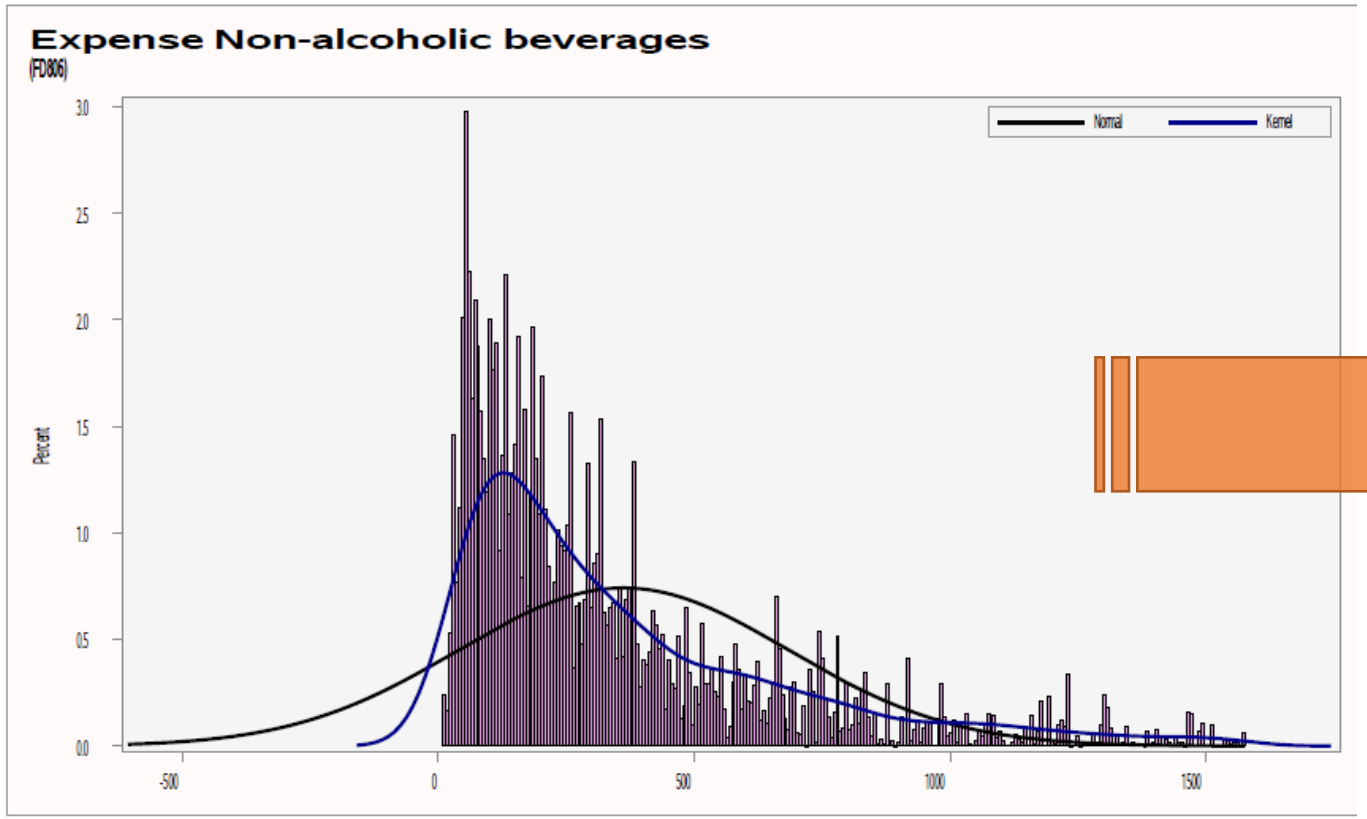

Spouse - Sex
(SP_Sex)
Frequency
2500000
2000000
1500000
1000000
500000
0
Female
Male
No spouse

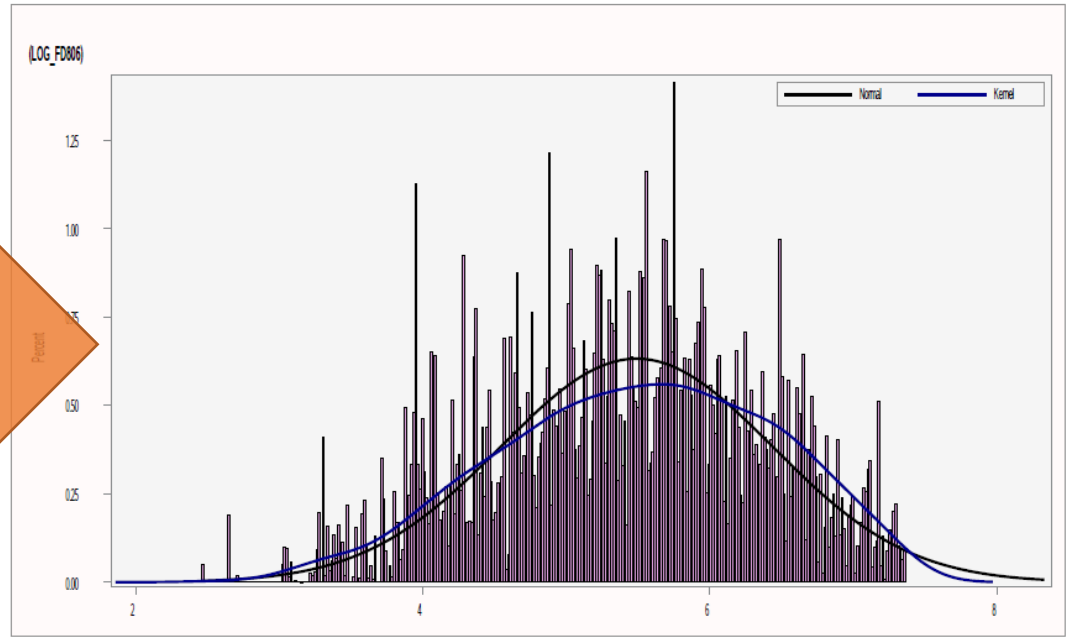
Expense

NUMERICAL UNIVARIATE ANALYSIS FOR ANA.MODEL1

The MEANS Procedure

Analysis Variable : FD806 Expense Non-alcoholic beverages										
N	N Miss	Mean	Median	Mode	Minimum	Maximum	Range	Quartile Range	Lower 95% CL for Mean	Upper 95% CL for Mean
8128876	0	326.86	224.12	72.02	11.83	1573.00	1561.17	316.56	326.66	327.07

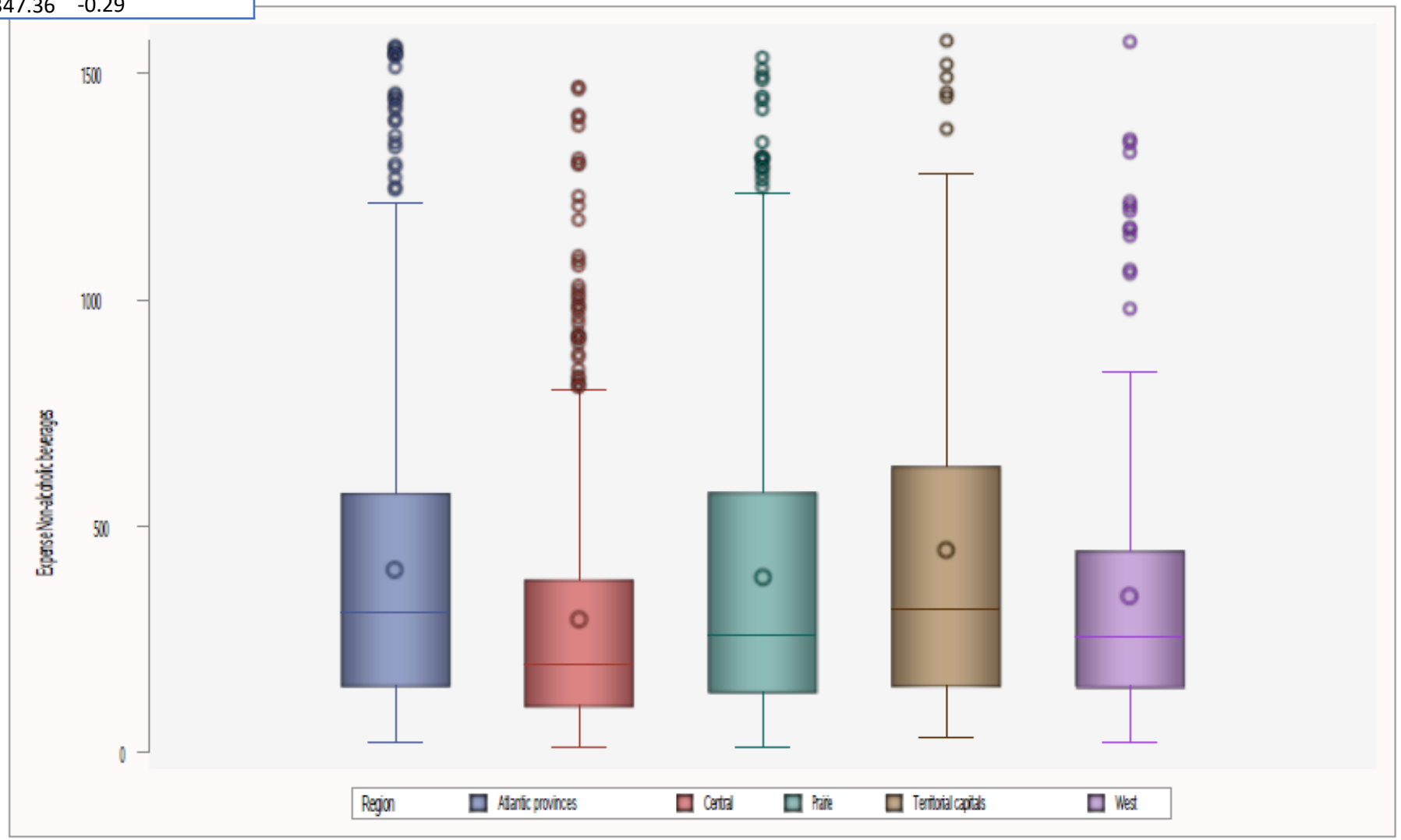




Bivariate Analysis

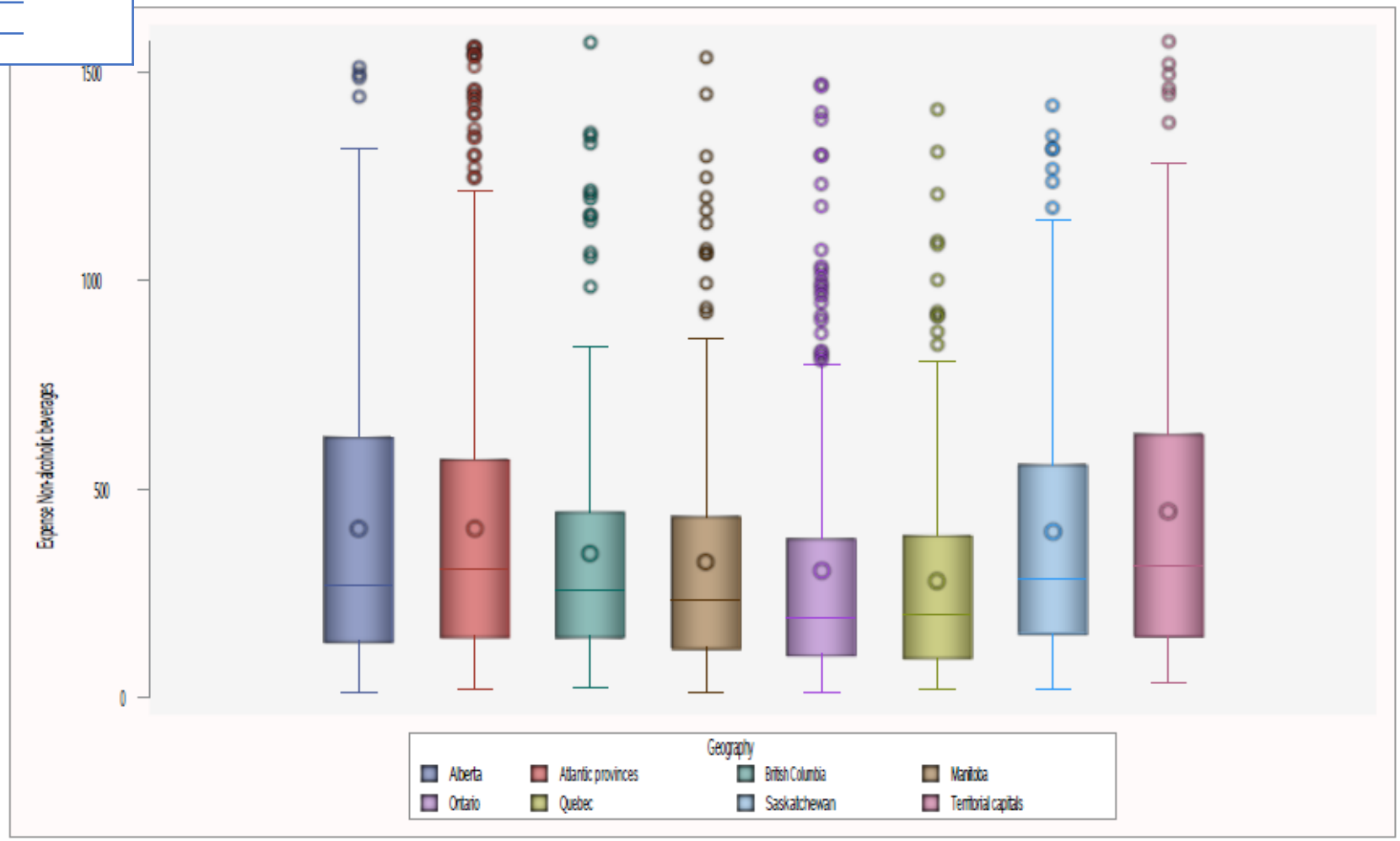
Level	N	Mean	95% CL (Lower)	95% CL (Upper)	Skewness	P-value
Atlantic provinces	629768	404.59	403.75	405.44	-0.41	<.0001
Central	4987062	294.35	294.1	294.59	-0.07	
Prairie	1530428	387.31	386.77	387.84	-0.28	
Territorial capitals	11711	447.33	440.68	453.98	-0.28	
West	969907	346.77	346.17	347.36	-0.29	

Region



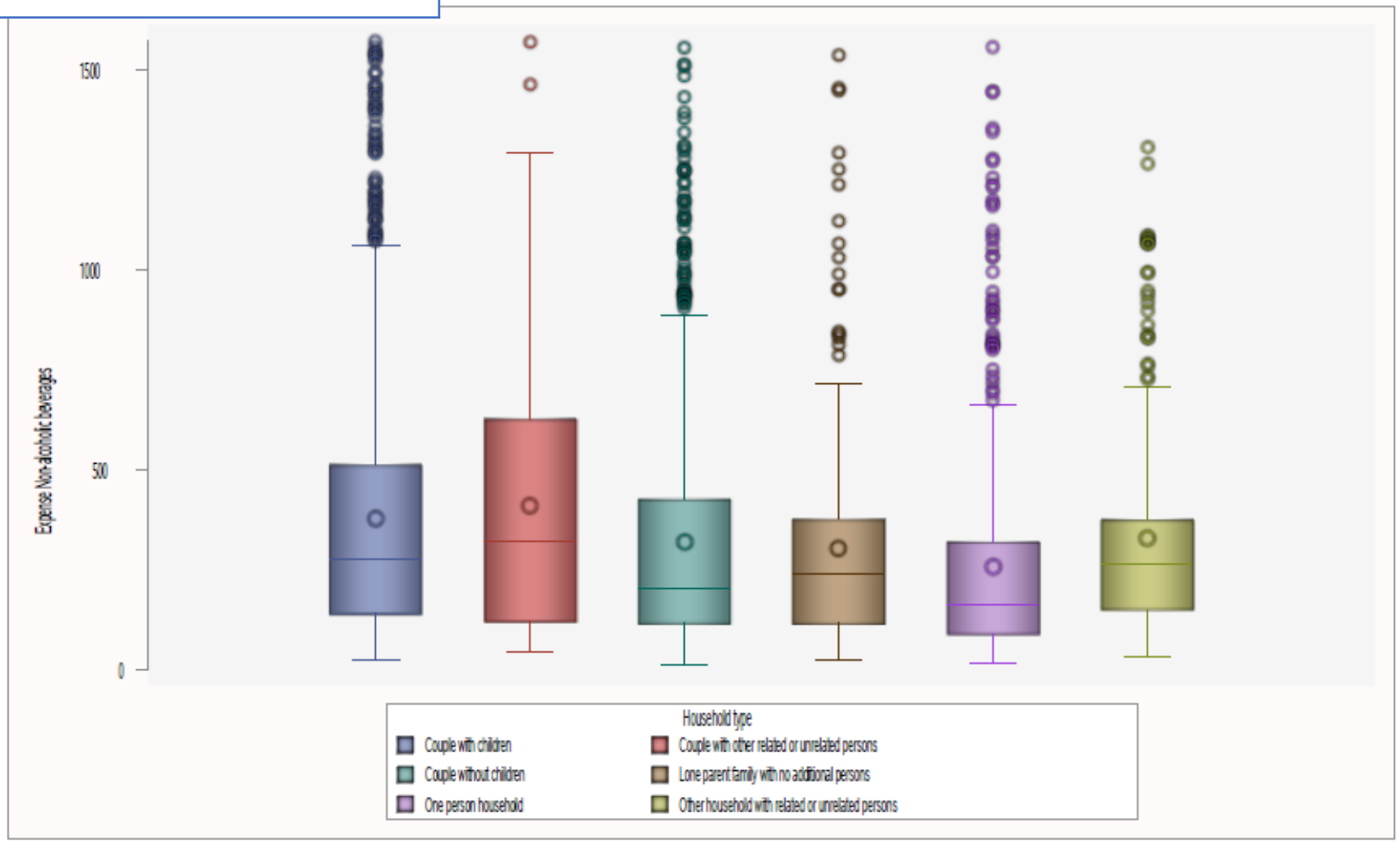
Level	N	Mean	95% CL (Lower)	95% CL (Upper)	Skewness	P-value
Alberta	958697	404.36	403.65	405.06	-0.29	
Atlantic provinces	629768	404.59	403.75	405.44	-0.41	
British Columbia	969907	346.77	346.17	347.36	-0.29	
Manitoba	299857	324.21	323.17	325.24	-0.23	<.0001
Ontario	2882375	305.37	305.03	305.72	0.02	
Quebec	2104687	279.25	278.91	279.58	-0.20	
Saskatchewan	271874	396.78	395.59	397.97	-0.29	
Territorial capitals	11711	447.33	440.68	453.98	-0.28	

Geography



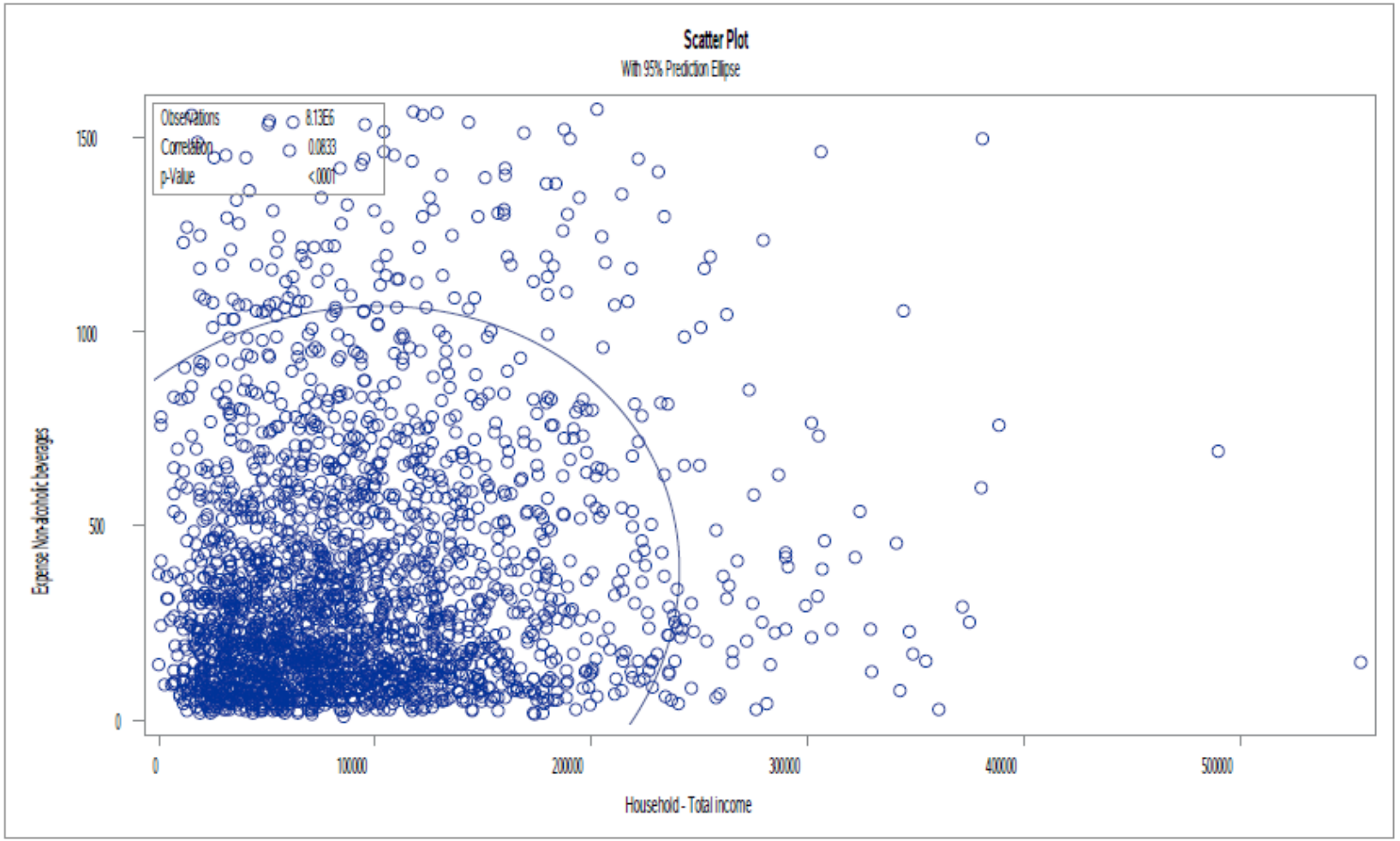
Level	N	Mean	95% CL (Lower)	95% CL (Upper)	Skewness	P-value
Couple with children	2457237	377.09	376.69	377.5	-0.29	
Couple with other related or unrelated persons	557180	409.09	408.21	409.96	-0.38	
Couple without children	2206478	317.62	317.23	318	-0.08	
Lone parent family with no additional persons	513225	302.25	301.51	302.98	-0.24	<.0001
One person household	1964999	257.47	257.08	257.85	0.08	
Other household with related or unrelated persons	429757	327.27	326.48	328.06	-0.35	

Household type



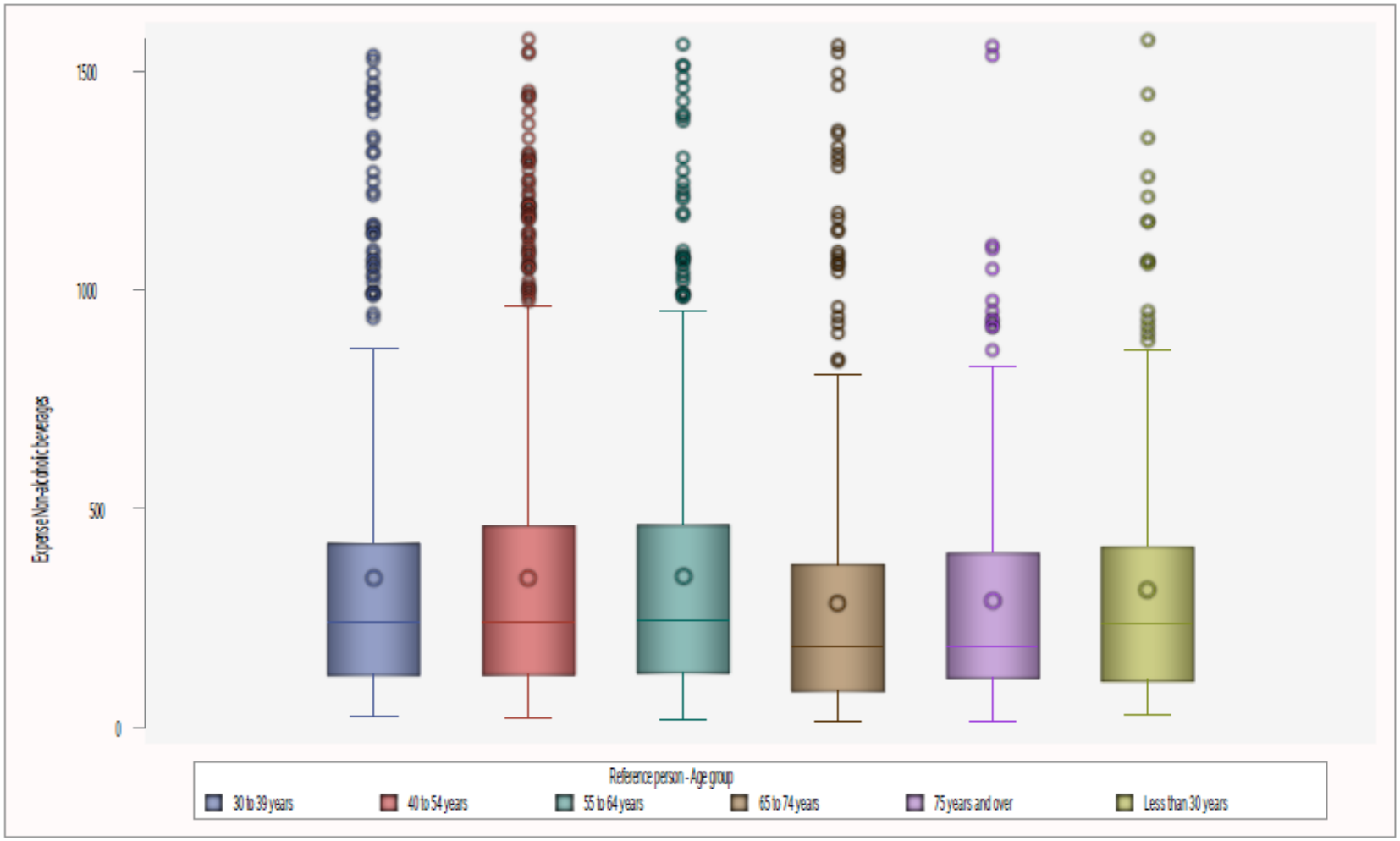
	Pearson	Spearman
Factor	PERC_EXPENSE	PERC_EXPENSE
HH_TotInc Household - Total income	0.08326 <.0001	0.09047 <.0001

Household - Total income



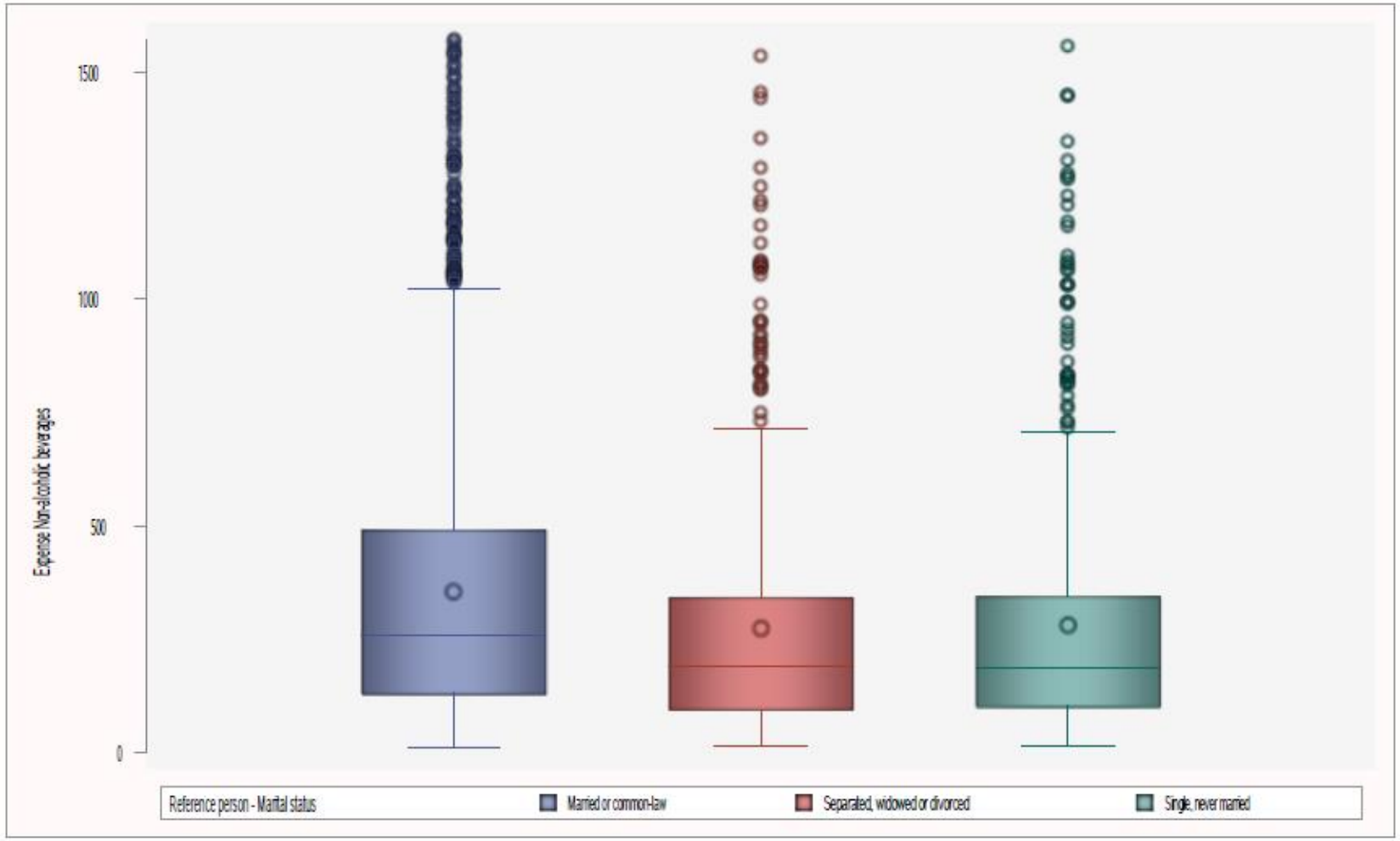
Level	N	Mean	95% CL (Lower)	95% CL (Upper)	Skewness	P-value
30 to 39 years	1399872	341.41	340.89	341.93	-0.03	
40 to 54 years	2595446	341.19	340.82	341.56	-0.23	
55 to 64 years	1693126	343.97	343.49	344.44	-0.14	<.0001
65 to 74 years	1124539	281.24	280.7	281.79	0.09	
75 years and over	670281	287.33	286.7	287.97	-0.37	
Less than 30 years	645612	313.39	312.71	314.07	-0.16	

Reference person – Age group



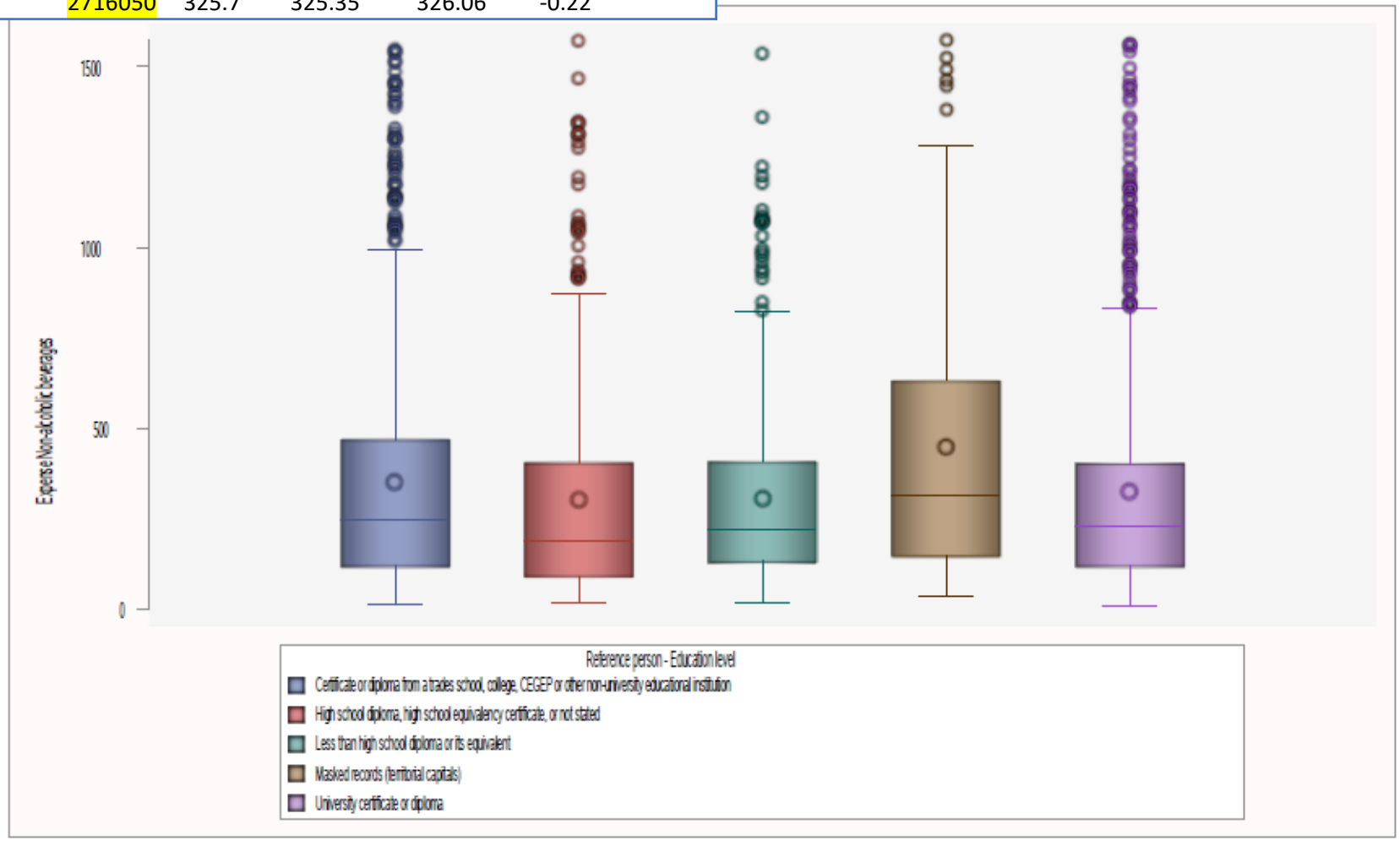
Level	N	Mean	95% CL (Lower)	95% CL (Upper)	Skewness	P-value
Married or common-law	5220895	355.37	355.1	355.64	-0.20	
Separated, widowed or	1548207	272.2	271.79	272.61	-0.13	<.0001
Single, never married	1359774	279.65	279.17	280.14	0.02	

Reference person – Marital status



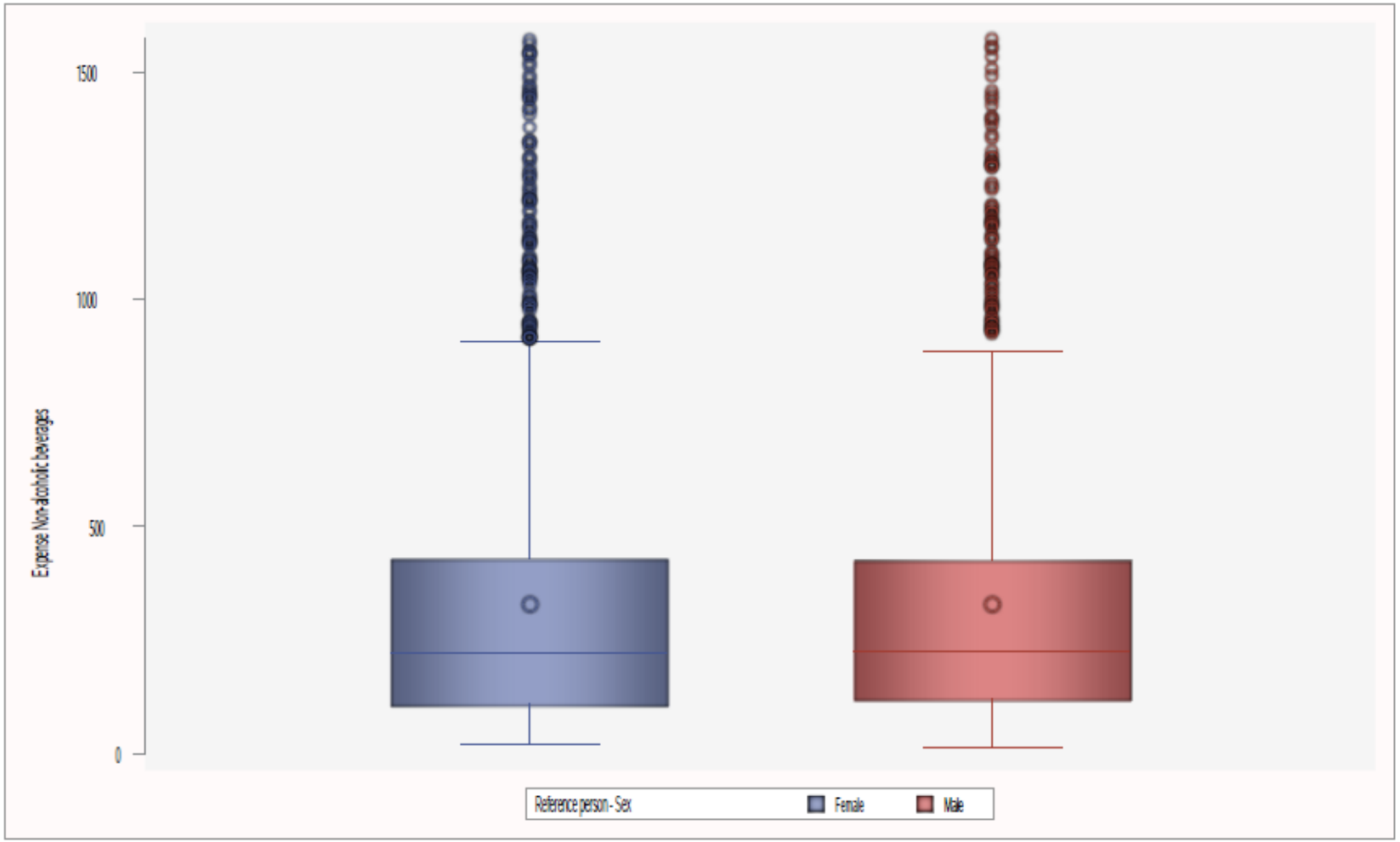
Level	N	Mean	95% CL (Lower)	95% CL (Upper)	Skewness	P-value
Certificate or diploma from a trades school, college, CEGEP or other non-university educational institution	2688986	351.14	350.76	351.53	-0.18	<.0001
High school diploma, high school equivalency certificate, or not stated	1668821	302.08	301.62	302.53	0.08	
Less than high school diploma or its equivalent	1043308	305.61	305.1	306.12	-0.33	
Masked records (Territorial capitals)	11711	447.33	440.68	453.98	-0.28	
University certificate or diploma	2716050	325.7	325.35	326.06	-0.22	

Reference person - Education level



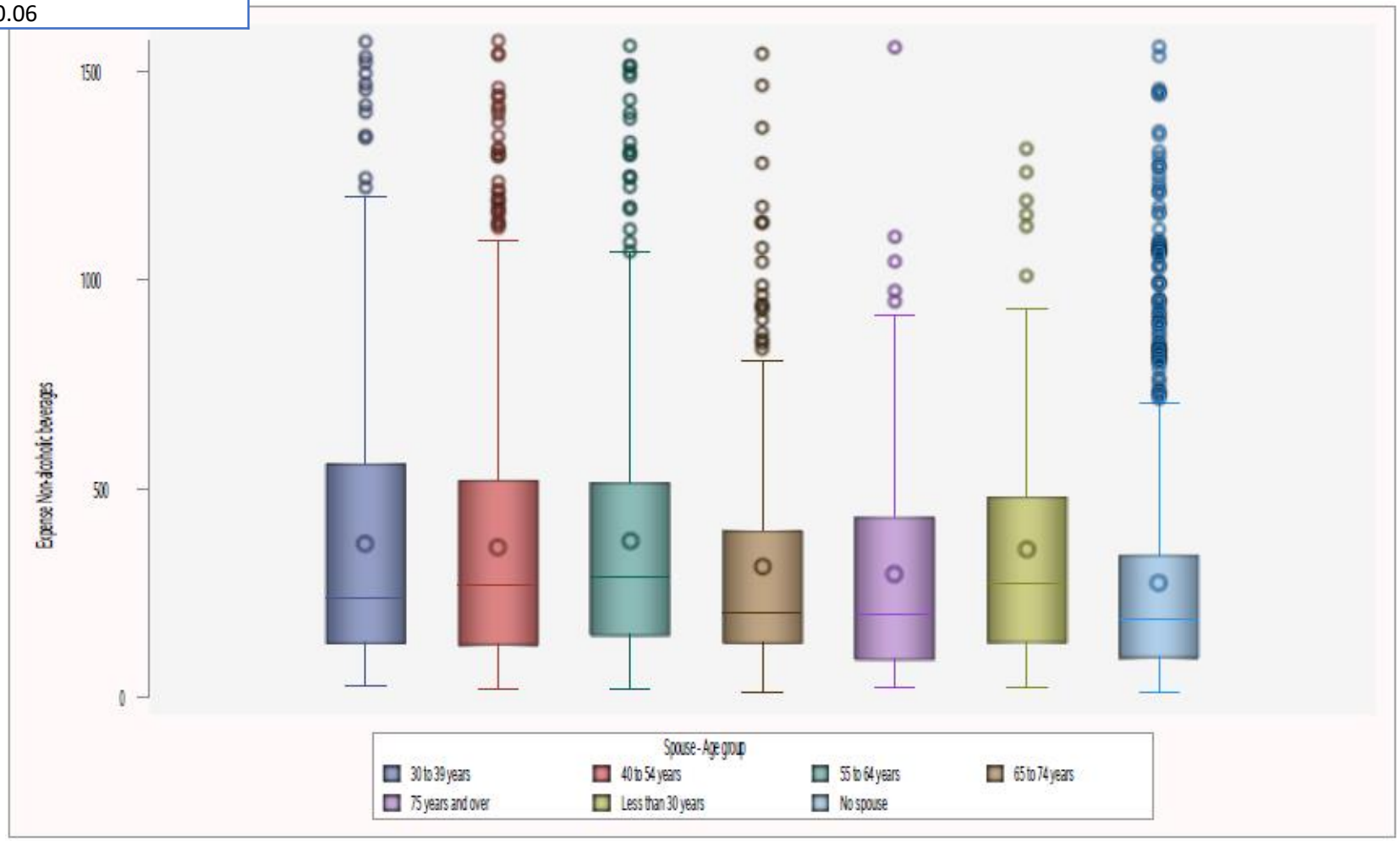
Level	N	Mean	95% CL (Lower)	95% CL (Upper)	Skewness	P-value
Female	4262186	326.98	326.68	327.27	-0.08	<.0001
Male	3866690	326.74	326.45	327.03	-0.23	

Reference person - Sex



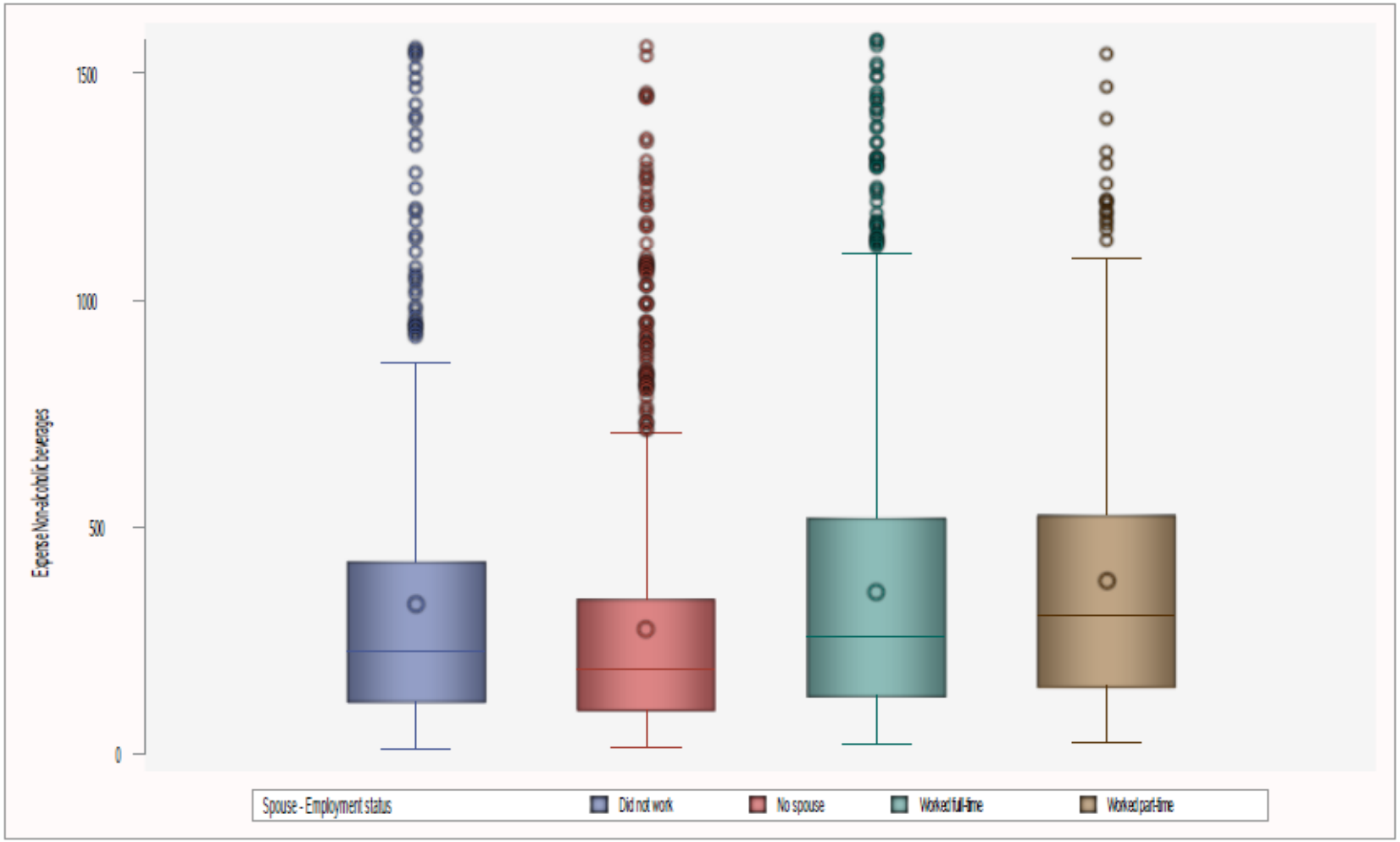
Level	N	Mean	95% CL (Lower)	95% CL (Upper)	Skewness	P-value
30 to 39 years	1029929	369.03	368.36	369.69	-0.14	
40 to 54 years	1667543	360	359.53	360.46	-0.19	
55 to 64 years	1237655	374.4	373.84	374.95	-0.27	
65 to 74 years	642970	314.79	314.07	315.51	-0.13	<.0001
75 years and	329381	296.11	295.2	297.01	-0.13	
Less than 30	313417	356.29	355.24	357.34	-0.44	
No spouse	2907981	275.69	275.37	276	-0.06	

Spouse - Age group



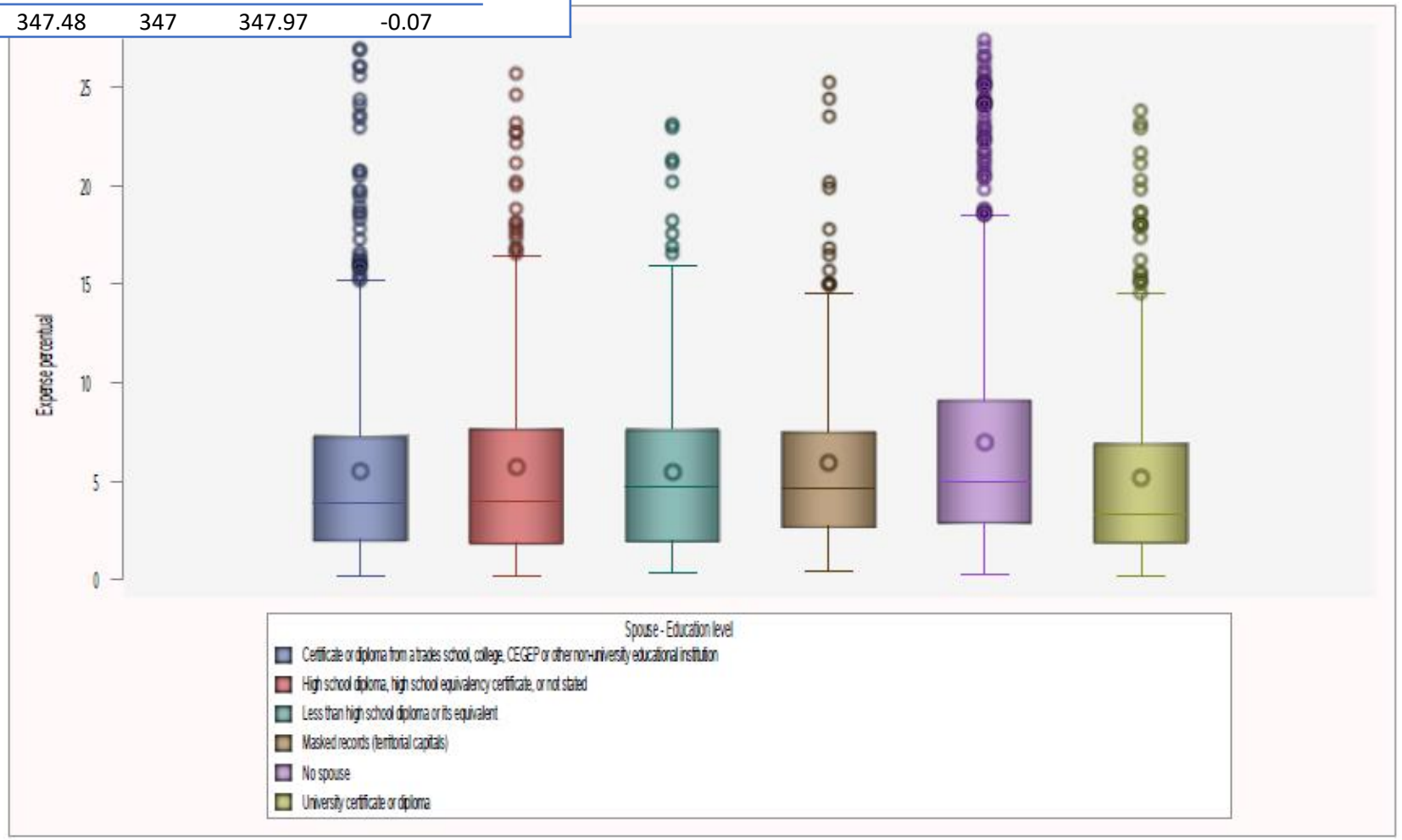
Level	N	Mean	95% CL (Lower)	95% CL (Upper)	Skewness	P-value
Did not work	1453314	330.34	329.83	330.86	-0.09	<.0001
No spouse	2907981	275.69	275.37	276	-0.06	
Worked full-time	2599458	357.87	357.49	358.25	-0.22	
Worked part-time	1168123	380.96	380.4	381.52	-0.28	

Spouse - Employment status



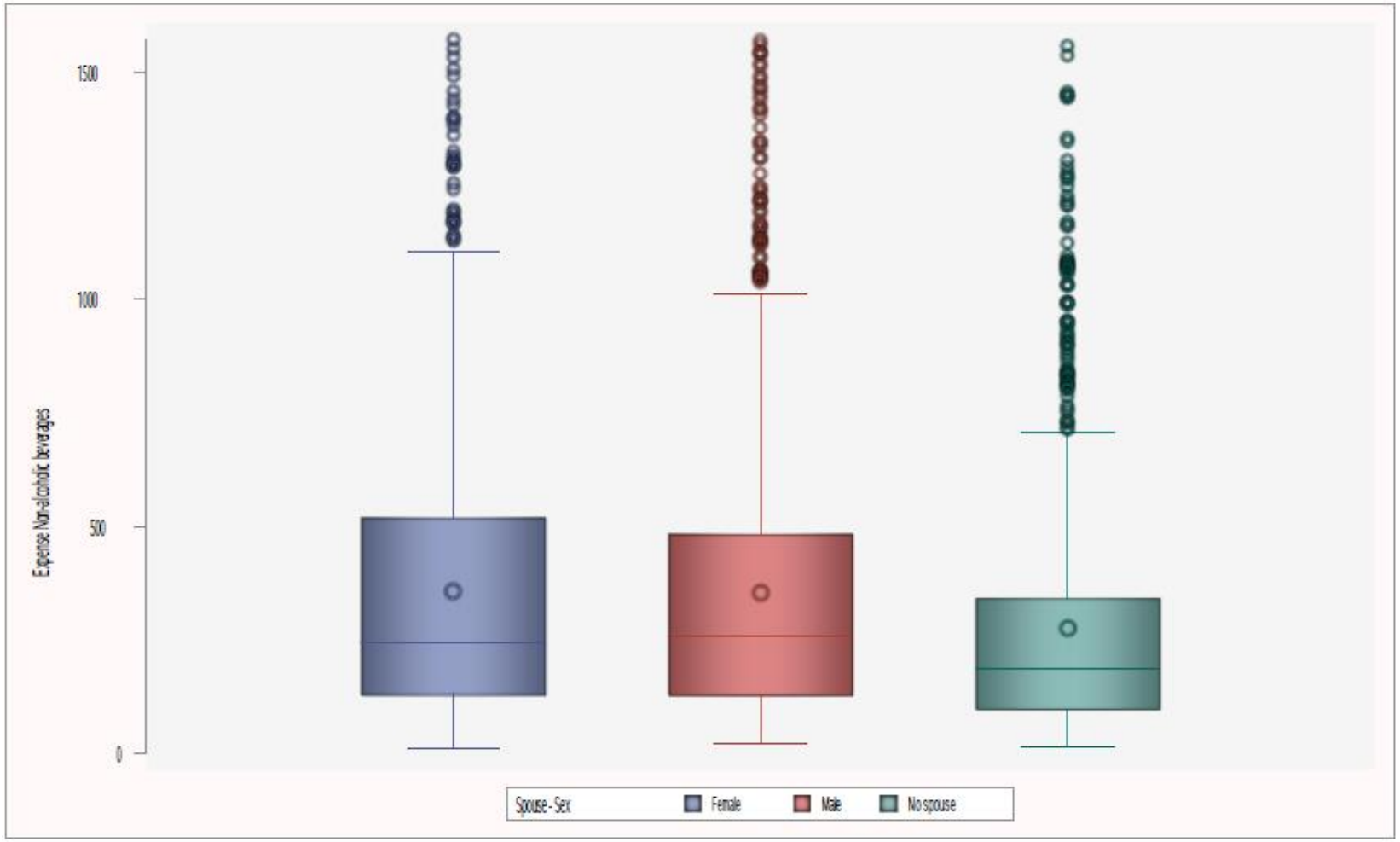
Level	N	Mean	95% CL (Lower)	95% CL (Upper)	Skewness	P-value
Certificate or diploma from a trades school, college, CEGEP or other non-university educational institution	1636840	359.03	358.56	359.5	-0.29	
High school diploma, high school equivalency certificate,or not stated	1358512	366.33	365.79	366.87	-0.16	
Less than high school diploma or its equivalent	546094	339.5	338.73	340.27	-0.46	<.0001
Masked records (territorial capitals)	11711	447.33	440.68	453.98	-0.28	
No spouse	2903242	275.52	275.21	275.84	-0.06	
University certificate or diploma	1672477	347.48	347	347.97	-0.07	

Spouse - Education level



Level	N	Mean	95% CL (Lower)	95% CL (Upper)	Skewness	P-value
Female	2493849	357.82	357.43	358.21	-0.22	
Male	2727046	353.13	352.76	353.51	-0.19	<.0001
No spouse	2907981	275.69	275.37	276	-0.06	

Spouse - Sex



Model

CORRELATION OF WeightD AND LOG_FD806 FOR ANA.MODEL2

The GLMSELECT Procedure
Selected Model

The selected model is the model at the last step (Step 10).

Effects:	Intercept Prov HHType6 RP_AgeGrp RP_Sex RP_MarStat RP_Educ SP_AgeGrp SP_Sex SP_Educ SP_EmpStat

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	34	537931	15821	19281.4
Error	8.13E6	6670180	0.82056	
Corrected Total	8.13E6	7208111		

Root MSE	0.90585
Dependent Mean	5.38219
R-Square	0.0746
Adj R-Sq	0.0746
AIC	6521253
AICC	6521253
SBC	-1607138

The GLMSELECT Procedure

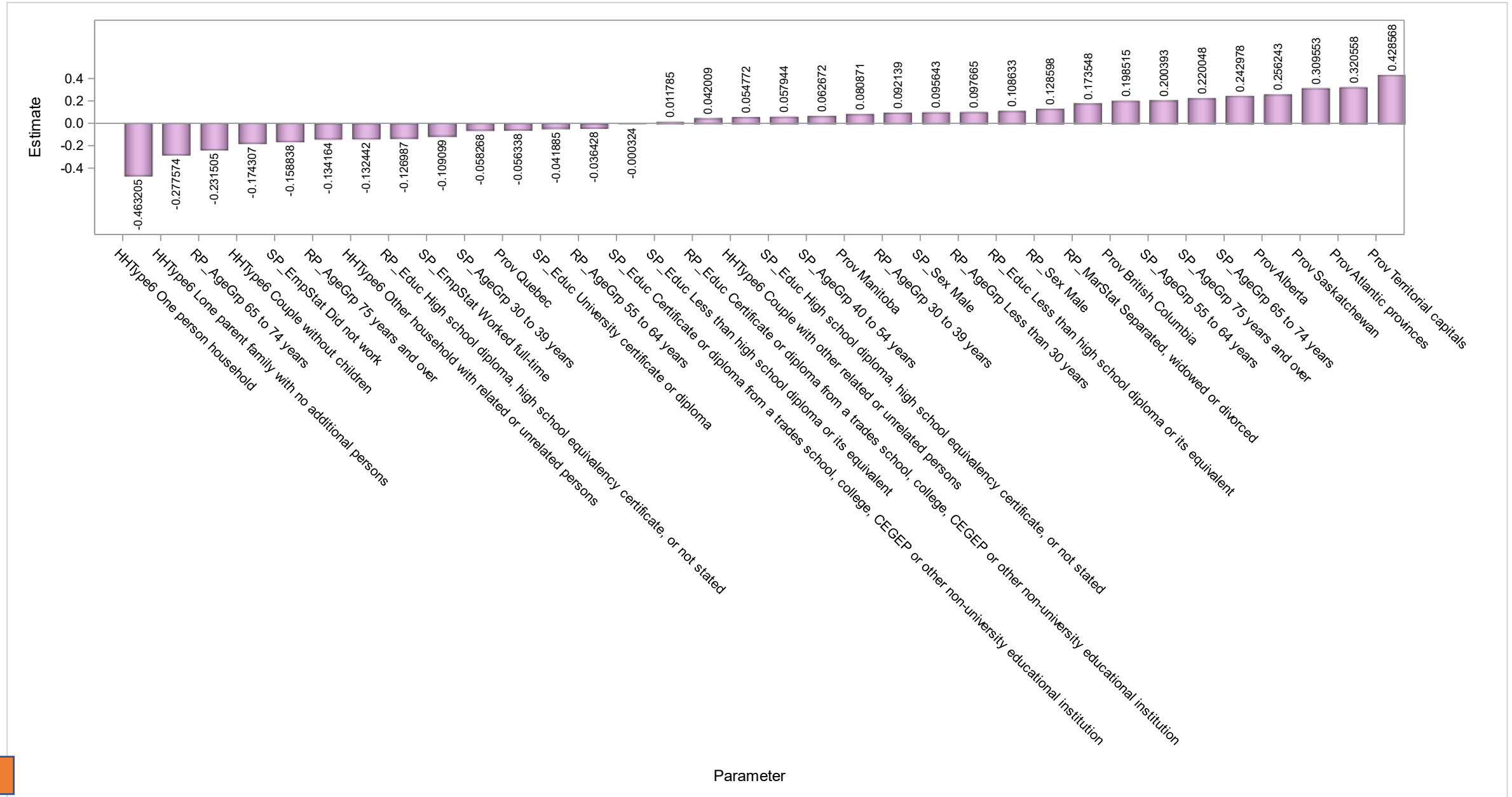
Stepwise Selection Summary

Step	Effect Entered	Effect Removed	Number Effects In	Number Parms In	F Value	Pr > F
0	Intercept		1	1	0.00	1.0000
1	HHType6		2	6	64259.8	<.0001
2	Prov		3	13	25475.3	<.0001
3	RP_Educ		4	16	13234.2	<.0001
4	RP_AgeGrp		5	21	5181.50	<.0001
5	SP_AgeGrp		6	26	4181.88	<.0001
6	SP_EmpStat		7	28	9299.10	<.0001
7	RP_MarStat		8	29	11909.7	<.0001
8	RP_Sex		9	30	9018.25	<.0001
9	SP_Educ		10	34	2151.78	<.0001
10	SP_Sex		11	35	4516.95	<.0001

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	5.400864	0.001956	2760.98
HHType6 Couple with other related or unrelated persons	1	0.062654	0.001377	45.50
HHType6 Couple without children	1	-0.175151	0.000991	-176.78
HHType6 Lone parent family with no additional persons	1	-0.069830	0.017390	-4.02
HHType6 One person household	1	-0.259681	0.017335	-14.98
HHType6 Other household with related or unrelated persons	1	0.115770	0.017362	6.67
RP_AgeGrp 40 to 54 years	1	-0.101091	0.001236	-81.82
RP_AgeGrp 55 to 64 years	1	-0.122411	0.001467	-83.44
RP_AgeGrp 65 to 74 years	1	-0.336182	0.001757	-191.29
RP_AgeGrp 75 years and over	1	-0.214953	0.002062	-104.25
RP_AgeGrp Less than 30 years	1	0.022457	0.001580	14.21
RP_Sex Male	1	0.113401	0.001137	99.70
RP_MarStat Separated, widowed or divorced	1	0.173442	0.001362	127.31
RP_MarStat Single, never married	0	0	.	.
RP_Educ High school diploma, high school equivalency certificate, or not stated	1	-0.164461	0.000935	-175.98
RP_Educ Less than high school diploma or its equivalent	1	0.027181	0.001178	23.08
RP_Educ Masked records (territorial capitals)	1	0.290023	0.010997	26.37
RP_Educ University certificate or diploma	1	-0.043579	0.000848	-51.39
SP_AgeGrp 40 to 54 years	1	0.129780	0.001419	91.47
SP_AgeGrp 55 to 64 years	1	0.259369	0.001673	155.03
SP_AgeGrp 65 to 74 years	1	0.319515	0.002208	144.69
SP_AgeGrp 75 years and over	1	0.241576	0.002736	88.29
SP_AgeGrp Less than 30 years	1	0.067964	0.002124	32.00
SP_AgeGrp No spouse	0	0	.	.
SP_Sex Male	1	0.075177	0.001437	52.32
SP_Sex No spouse	0	0	.	.
SP_Educ High school diploma, high school equivalency certificate, or not stated	1	0.041927	0.001093	38.35
SP_Educ Less than high school diploma or its equivalent	1	-0.004817	0.001534	-3.14
SP_Educ Masked records (territorial capitals)	0	0	.	.
SP_Educ No spouse	1	-0.009269	0.017269	-0.54
SP_Educ University certificate or diploma	1	-0.060028	0.001075	-55.82
SP_EmpStat No spouse	0	0	.	.
SP_EmpStat Worked full-time	1	0.060764	0.001223	49.67
SP_EmpStat Worked part-time	1	0.165830	0.001300	127.51

Parameters


Estimate
0.4
0.2
0.0
-0.2
-0.4
-0.463205
-0.277574
-0.231505
-0.174307
-0.158838
-0.134164
-0.132442
-0.126987
-0.109099
-0.058268
-0.056338
-0.041885
-0.036428
-0.000324
0.011785
0.042009
0.054772
0.057944
0.062672
0.080871
0.092139
0.095643
0.097665
0.108633
0.128598
0.173548
0.198515
0.200393
0.220048
0.242978
0.256243
0.309553
0.320558
0.428568
HHType6 One person household
HHType6 Lone parent family with no additional persons
RP_AgeGrp 65 to 74 years
HHType6 Couple without children
SP_EmpStat Did not work
RP_AgeGrp 75 years and over
HHType6 Other household with related or unrelated persons
RP_Educ High school diploma, high school equivalency certificate, or not stated
SP_EmpStat Worked full-time
SP_AgeGrp 30 to 39 years
Prov Quebec
SP_Educ University certificate or diploma
RP_AgeGrp 55 to 64 years
SP_Educ Certificate or diploma from a trades school, college, CEGEP or other non-university educational institution
SP_Educ Less than high school diploma or its equivalent
RP_Educ Certificate or diploma from a trades school, college, CEGEP or other non-university educational institution
HHType6 Couple with other related or unrelated persons
SP_Educ High school diploma, high school equivalency certificate, or not stated
SP_AgeGrp 40 to 54 years
Prov Manitoba
RP_AgeGrp 30 to 39 years
SP_Sex Male
RP_AgeGrp Less than 30 years
RP_Educ Less than high school diploma or its equivalent
RP_Sex Male
RP_MarStat Separated, widowed or divorced
Prov British Columbia
SP_AgeGrp 55 to 64 years
SP_AgeGrp 75 years and over
SP_AgeGrp 65 to 74 years
Prov Alberta
Prov Saskatchewan
Prov Atlantic provinces
Prov Territorial capitals
Parameter

Conclusions and Recommendations

Which product expense is the biggest regarding store purchase?(target)

Non-alcoholic beverages is the expense that **weighs the most on store purchases**.

Is the expense diverse regarding geography (Region and Province)?

Yes. **All groups have different expense behaviour** with **Territorial capitals** having the **biggest mean** expense despite having the **smallest population**.

Is the expense diverse regarding Household Type?

Yes. **All groups have different expense behaviour** with households having **Couple with other related or unrelated person** having the **biggest mean**.

Does higher Household Income mean higher expense?

No. There is **little to no association between Income and Expense** on non-alcoholic beverages.

What impacts the expense of the previously identified product?

We see that **residents of Territorial capitals have the biggest expense** and **households with only one person have the smallest despite being a meaningful proportion of the population**.

To maintain revenue, special attention should be applied for **stock keeping and market to retain consumers in** areas with **higher concentration** of:

- Households with **Couple with other related or unrelated persons**
- Households with **Couple with children**
- **Males**
- **Separated, widowed or divorced**
- Households with **couples that both have Less than high school diploma or its equivalent**
- And due to larger population of consumers: **Ontario Province**.

- ❖ **Customize products: flavours, package themes**
- ❖ **Target advertisement**

To increase sales, **target promotions to acquire consumers and increase spending of existing ones** in areas with higher concentration of:

- **One person households**
- Households with couples with a **combination of age group of with more than 65 years and over and spouse within 30 to 39 years**
- And due to larger population of consumers: **Quebec Province**.

- ❖ **Customize products: package size**
- ❖ **Target advertisement**

Next Steps

Territorial capitals (biggest expected expense) x **Quebec** (smallest expected expense)

- Is the price, availability, product characteristics or culture that motivates these regions to such contrast spending with non-alcoholic beverages?

Questions?

Thank You!

anaclarat@womenindata.ca

Appendix

SAS Script/Extra graphs

Tabular summaries

NATIONAL EXPENSE BY PRODUCT CATEGORY

The MEANS Procedure

Analysis Variable : Expense												
Product_Category	N Obs	N	N Miss	Mean	Median	Mode	Minimum	Maximum	Range	Quartile Range	Lower 95% CL for Mean	Upper 95% CL for Mean
Bakery products	14462249	14462249	0	562.62	431.17	0.00	0.00	9227.67	9227.67	636.81	562.35	562.89
Cereal grains and cereal products	14462249	14462249	0	320.71	199.94	0.00	0.00	4399.65	4399.65	450.84	320.50	320.91
Dairy products and eggs	14462249	14462249	0	839.36	664.32	0.00	0.00	7231.86	7231.86	901.68	838.98	839.75
Fish and seafood	14462249	14462249	0	218.83	0.00	0.00	-964.08	7534.21	8498.29	270.68	218.59	219.07
Fruit, fruit preparations and nuts	14462249	14462249	0	731.50	543.66	0.00	0.00	9225.84	9225.84	823.46	731.13	731.87
Meat	14462249	14462249	0	1109.95	752.44	0.00	0.00	17772.79	17772.79	1323.55	1109.31	1110.59
Non-alcoholic beverages and other food products	14462249	14462249	0	1452.85	1141.97	0.00	0.00	21199.40	21199.40	1397.03	1452.17	1453.52
Vegetables and vegetable preparations	14462249	14462249	0	702.13	536.43	0.00	0.00	5718.18	5718.18	771.10	701.79	702.48

REGIONAL EXPENSE BY PRODUCT CATEGORY

The MEANS Procedure

Analysis Variable : Expense

Region	Product_Category	N Obs	N	N Miss	Mean	Median	Mode	Minimum	Maximum	Range	Quartile Range	Lower 95% CL for Mean	Upper 95% CL for Mean
Atlantic provinces	Bakery products	1006001	1006001	0	576.37	439.40	0.00	0.00	4211.74	4211.74	628.74	575.29	577.44
	Cereal grains and cereal products	1006001	1006001	0	359.04	223.60	0.00	0.00	2990.62	2990.62	480.28	358.25	359.83
	Dairy products and eggs	1006001	1006001	0	834.65	676.26	0.00	0.00	4011.47	4011.47	901.41	833.30	836.01
	Fish and seafood	1006001	1006001	0	186.92	0.00	0.00	0.00	4109.82	4109.82	188.24	186.12	187.71
	Fruit, fruit preparations and nuts	1006001	1006001	0	623.50	417.56	0.00	0.00	5405.56	5405.56	707.08	622.23	624.77
	Meat	1006001	1006001	0	1084.85	766.48	0.00	0.00	9287.54	9287.54	1283.36	1082.57	1087.13
	Non-alcoholic beverages and other food products	1006001	1006001	0	1504.56	1247.74	0.00	0.00	8296.08	8296.08	1487.46	1502.24	1506.88
	Vegetables and vegetable preparations	1006001	1006001	0	595.16	443.82	0.00	0.00	4569.18	4569.18	675.74	594.06	596.27
Central	Bakery products	9011976	9011976	0	581.24	464.92	0.00	0.00	4525.01	4525.01	637.11	580.89	581.58
	Cereal grains and cereal products	9011976	9011976	0	279.47	181.41	0.00	0.00	2776.29	2776.29	415.59	279.25	279.68
	Dairy products and eggs	9011976	9011976	0	798.01	640.38	0.00	0.00	6418.48	6418.48	851.79	797.54	798.47
	Fish and seafood	9011976	9011976	0	234.62	0.00	0.00	0.00	6017.09	6017.09	286.13	234.29	234.95
	Fruit, fruit preparations and nuts	9011976	9011976	0	716.37	536.64	0.00	0.00	5360.06	5360.06	829.56	715.91	716.82
	Meat	9011976	9011976	0	1089.15	735.80	0.00	0.00	11400.07	11400.07	1272.43	1088.35	1089.96
	Non-alcoholic beverages and other food products	9011976	9011976	0	1399.70	1106.23	0.00	0.00	10862.66	10862.66	1293.40	1398.85	1400.55
	Vegetables and vegetable preparations	9011976	9011976	0	676.01	519.34	0.00	0.00	4917.51	4917.51	751.68	675.59	676.43
Prairie	Bakery products	2499883	2499883	0	534.22	374.10	0.00	0.00	4996.16	4996.16	657.30	533.56	534.88
	Cereal grains and cereal products	2499883	2499883	0	381.80	247.78	0.00	0.00	4228.45	4228.45	502.53	381.23	382.37
	Dairy products and eggs	2499883	2499883	0	929.44	706.42	0.00	0.00	7231.86	7231.86	1048.67	928.38	930.49
	Fish and seafood	2499883	2499883	0	170.57	0.00	0.00	-964.08	7534.21	8498.29	207.74	170.13	171.02
	Fruit, fruit preparations and nuts	2499883	2499883	0	770.20	603.69	0.00	0.00	5648.21	5648.21	838.70	769.30	771.11
	Meat	2499883	2499883	0	1224.91	856.77	0.00	0.00	17772.79	17772.79	1620.89	1223.22	1226.60
	Non-alcoholic beverages and other food products	2499883	2499883	0	1565.32	1234.74	0.00	0.00	9486.20	9486.20	1606.53	1563.57	1567.08
	Vegetables and vegetable preparations	2499883	2499883	0	736.18	542.36	0.00	0.00	4257.90	4257.90	876.53	735.33	737.04
Territorial capitals	Bakery products	21431	21431	0	618.65	482.86	0.00	0.00	9227.67	9227.67	527.05	608.93	628.37
	Cereal grains and cereal products	21431	21431	0	501.60	271.78	0.00	0.00	4399.65	4399.65	651.16	492.87	510.33
	Dairy products and eggs	21431	21431	0	1212.83	922.89	0.00	0.00	6067.74	6067.74	1330.42	1198.32	1227.34
	Fish and seafood	21431	21431	0	255.49	0.00	0.00	0.00	6901.12	6901.12	254.40	246.79	264.18
	Fruit, fruit preparations and nuts	21431	21431	0	1078.21	788.08	0.00	0.00	9225.84	9225.84	1147.38	1063.57	1092.85
	Meat	21431	21431	0	1597.11	899.43	0.00	0.00	11392.01	11392.01	2032.06	1571.81	1622.40
	Non-alcoholic beverages and other food products	21431	21431	0	2278.52	1721.97	0.00	0.00	21199.40	21199.40	1923.68	2249.64	2307.41
	Vegetables and vegetable preparations	21431	21431	0	1072.46	727.41	0.00	0.00	5642.18	5642.18	1027.05	1058.42	1086.50
West	Bakery products	1922958	1922958	0	504.48	369.43	0.00	0.00	3178.50	3178.50	594.36	503.81	505.16
	Cereal grains and cereal products	1922958	1922958	0	412.50	255.78	0.00	0.00	4163.64	4163.64	590.72	411.77	413.24
	Dairy products and eggs	1922958	1922958	0	914.39	732.16	0.00	0.00	4054.60	4054.60	954.62	913.31	915.47
	Fish and seafood	1922958	1922958	0	223.82	0.00	0.00	0.00	4503.37	4503.37	327.17	223.25	224.39
	Fruit, fruit preparations and nuts	1922958	1922958	0	804.75	564.98	0.00	0.00	6725.94	6725.94	816.66	803.61	805.89
	Meat	1922958	1922958	0	1065.67	664.04	0.00	0.00	9192.90	9192.90	1281.54	1063.99	1067.35
	Non-alcoholic beverages and other food products	1922958	1922958	0	1519.44	1182.14	0.00	0.00	8578.29	8578.29	1555.42	1517.65	1521.24
	Vegetables and vegetable preparations	1922958	1922958	0	832.12	667.68	0.00	0.00	5718.18	5718.18	874.74	831.06	833.18

PROVINCIAL/TERRITORIES EXPENSE BY PRODUCT CATEGORY

The MEANS Procedure

Analysis Variable : Expense

Region	Geography	Product_Category	N Obs	N	N Miss	Mean	Median	Mode	Minimum	Maximum	Range	Quartile Range	Lower 95% CL for Mean	Upper 95% CL for Mean
Atlantic provinces	Atlantic provinces	Bakery products	1006001	1006001	0	576.37	439.40	0.00	0.00	4211.74	4211.74	626.74	575.29	577.44
		Cereal grains and cereal products	1006001	1006001	0	359.04	223.60	0.00	0.00	2990.62	2990.62	480.26	358.25	359.83
		Dairy products and eggs	1006001	1006001	0	834.65	676.26	0.00	0.00	4011.47	4011.47	901.41	833.30	836.01
		Fish and seafood	1006001	1006001	0	186.92	0.00	0.00	0.00	4109.82	4109.82	188.24	186.12	187.71
		Fruit, fruit preparations and nuts	1006001	1006001	0	623.50	417.56	0.00	0.00	5405.56	5405.56	707.08	622.23	624.77
		Meat	1006001	1006001	0	1084.85	766.48	0.00	0.00	9287.54	9287.54	1283.36	1082.57	1087.13
		Non-alcoholic beverages and other food products	1006001	1006001	0	1504.56	1247.74	0.00	0.00	8296.06	8296.06	1487.46	1502.24	1506.88
		Vegetables and vegetable preparations	1006001	1006001	0	595.16	443.82	0.00	0.00	4569.18	4569.18	675.74	594.06	596.27
Central	Ontario	Bakery products	5401483	5401483	0	559.05	428.74	0.00	0.00	4404.03	4404.03	637.82	558.61	559.49
		Cereal grains and cereal products	5401483	5401483	0	286.21	190.53	0.00	0.00	2294.65	2294.65	415.74	285.93	286.50
		Dairy products and eggs	5401483	5401483	0	701.07	546.52	0.00	0.00	4400.50	4400.50	773.07	700.55	701.58
		Fish and seafood	5401483	5401483	0	185.64	0.00	0.00	0.00	2269.86	2269.86	246.46	185.34	185.94
		Fruit, fruit preparations and nuts	5401483	5401483	0	734.93	560.81	0.00	0.00	4253.95	4253.95	908.96	734.36	735.51
		Meat	5401483	5401483	0	1123.72	744.38	0.00	0.00	11400.07	11400.07	1323.71	1122.64	1124.79
		Non-alcoholic beverages and other food products	5401483	5401483	0	1483.99	1141.40	0.00	0.00	10862.66	10862.66	1346.85	1482.77	1485.21
		Vegetables and vegetable preparations	5401483	5401483	0	688.32	519.34	0.00	0.00	4208.57	4208.57	767.09	687.76	688.88
	Quebec	Bakery products	3610513	3610513	0	614.43	490.71	0.00	0.00	4525.01	4525.01	633.22	613.87	615.00
		Cereal grains and cereal products	3610513	3610513	0	269.37	178.97	0.00	0.00	2776.29	2776.29	410.08	269.04	269.70
		Dairy products and eggs	3610513	3610513	0	943.03	756.60	0.00	0.00	6418.48	6418.48	964.46	942.18	943.87
		Fish and seafood	3610513	3610513	0	307.91	0.00	0.00	0.00	6017.09	6017.09	327.31	307.23	308.59
		Fruit, fruit preparations and nuts	3610513	3610513	0	688.59	512.15	0.00	0.00	5360.06	5360.06	677.20	687.87	689.31
		Meat	3610513	3610513	0	1037.45	715.00	0.00	0.00	10026.44	10026.44	1201.20	1036.25	1038.65
		Non-alcoholic beverages and other food products	3610513	3610513	0	1273.59	1069.31	0.00	0.00	8375.62	8375.62	1182.10	1272.53	1274.66
		Vegetables and vegetable preparations	3610513	3610513	0	657.60	519.17	0.00	0.00	4917.51	4917.51	711.13	656.97	658.23
Prairie	Alberta	Bakery products	1574179	1574179	0	561.15	413.30	0.00	0.00	2878.83	2878.83	726.70	560.35	561.95
		Cereal grains and cereal products	1574179	1574179	0	374.91	260.00	0.00	0.00	2664.26	2664.26	492.96	374.26	375.55
		Dairy products and eggs	1574179	1574179	0	943.76	696.50	0.00	0.00	7231.86	7231.86	1120.08	942.37	945.15
		Fish and seafood	1574179	1574179	0	179.89	0.00	0.00	-964.08	1809.73	2773.81	248.30	179.16	180.22
		Fruit, fruit preparations and nuts	1574179	1574179	0	783.85	635.96	0.00	0.00	5147.60	5147.60	864.58	782.75	784.94
		Meat	1574179	1574179	0	1203.46	869.18	0.00	0.00	7724.34	7724.34	1629.28	1201.52	1205.39
		Non-alcoholic beverages and other food products	1574179	1574179	0	1582.31	1273.61	0.00	0.00	7615.07	7615.07	1574.40	1580.12	1584.49
		Vegetables and vegetable preparations	1574179	1574179	0	794.78	590.86	0.00	0.00	4060.49	4060.49	1039.21	793.68	795.89
	Manitoba	Bakery products	489946	489946	0	486.02	327.63	0.00	0.00	4996.16	4996.16	521.24	484.46	487.59
		Cereal grains and cereal products	489946	489946	0	378.60	195.16	0.00	0.00	4228.45	4228.45	501.26	377.18	380.02
		Dairy products and eggs	489946	489946	0	903.25	731.06	0.00	0.00	3933.20	3933.20	865.02	901.15	905.36
		Fish and seafood	489946	489946	0	158.52	0.00	0.00	0.00	3033.42	3033.42	138.84	157.44	159.59
		Fruit, fruit preparations and nuts	489946	489946	0	731.22	566.02	0.00	0.00	5648.21	5648.21	783.88	729.28	733.16
		Meat	489946	489946	0	1193.66	809.11	0.00	0.00	17772.79	17772.79	1374.62	1189.53	1197.78
		Non-alcoholic beverages and other food products	489946	489946	0	1397.89	1069.25	0.00	0.00	7699.33	7699.33	1458.86	1394.21	1401.17
		Vegetables and vegetable preparations	489946	489946	0	596.32	445.46	0.00	0.00	4044.54	4044.54	604.14	594.74	597.90
	Saskatchewan	Bakery products	435758	435758	0	491.13	335.43	0.00	0.00	3509.74	3509.74	599.82	489.50	492.76
		Cereal grains and cereal products	435758	435758	0	410.30	233.74	0.00	0.00	3083.25	3083.25	526.11	408.69	411.91
		Dairy products and eggs	435758	435758	0	907.14	683.02	0.00	0.00	5265.63	5265.63	1049.62	904.71	909.57
		Fish and seafood	435758	435758	0	151.19	0.00	0.00	0.00	7534.21	7534.21	84.00	150.00	152.39
		Fruit, fruit preparations and nuts	435758	435758	0	764.74	505.70	0.00	0.00	4968.59	4968.59	812.49	762.20	767.28
		Meat	435758	435758	0	1337.56	872.56	0.00	0.00	10596.04	10596.04	1747.93	1332.75	1342.36
		Non-alcoholic beverages and other food products	435758	435758	0	1692.44	1282.38	0.00	0.00	9496.20	9496.20	1804.38	1687.65	1697.23
		Vegetables and vegetable preparations	435758	435758	0	681.75	518.53	0.00	0.00	4257.90	4257.90	803.91	679.63	683.87
Territorial capitals	Territorial capitals	Bakery products	21431	21431	0	618.65	482.86	0.00	0.00	9227.67	9227.67	527.05	606.93	626.37
		Cereal grains and cereal products	21431	21431	0	501.60	271.78	0.00	0.00	4399.65	4399.65	651.16	492.67	510.33
		Dairy products and eggs	21431	21431	0	1212.83	922.89	0.00	0.00	6067.74	6067.74	1330.42	1196.32	1227.34
		Fish and seafood	21431	21431	0	255.49	0.00	0.00	0.00	6901.12	6901.12	254.40	246.79	264.18
		Fruit, fruit preparations and nuts	21431	21431	0	1078.21	788.08	0.00	0.00	9225.84	9225.84	1147.38	1063.57	1092.85
		Meat	21431	21431	0	1597.11	899.43	0.00	0.00	11392.01	11392.01	2032.06	1571.81	1622.40
		Non-alcoholic beverages and other food products	21431	21431	0	2278.52	1721.97	0.00	0.00	21199.40	21199.40	1923.68	2249.64	2307.41
		Vegetables and vegetable preparations	21431	21431	0	1072.46	727.41	0.00	0.00	5642.18	5642.18	1027.05	1058.42	1086.50
West	British Columbia	Bakery products	1922958	1922958	0	504.48	369.43	0.00	0.00	3178.50	3178.50	594.36	503.81	505.16
		Cereal grains and cereal products	1922958	1922958	0	412.50	255.78	0.00	0.00	4163.64	4163.64	590.72	411.77	413.24
		Dairy products and eggs	1922958	1922958	0	914.39	732.16	0.00	0.00	4054.60	4054.60	964.62	913.31	915.47
		Fish and seafood	1922958	1922958	0	223.82	0.00	0.00	0.00	4503.37	4503.37	327.17	223.25	224.39
		Fruit, fruit preparations and nuts	1922958	1922958	0	804.75	564.96	0.00	0.00	6725.94	6725.94	816.66	803.61	805.89
		Meat	1922958	1922958	0	1065.67	664.04	0.00	0.00	9192.90	9192.90	1281.54	1063.99	1067.35
		Non-alcoholic beverages and other food products	1922958	1922958	0	1519.44	1182.14	0.00	0.00	8578.29	8578.29	1555.42	1517.65	1521.24
		Vegetables and vegetable preparations	1922958	1922958	0	832.12	667.88	0.00	0.00	5718.18	5718.18	874.74	831.06	833.18

NATIONAL EXPENSE Non-alcoholic beverages and other food products

The MEANS Procedure

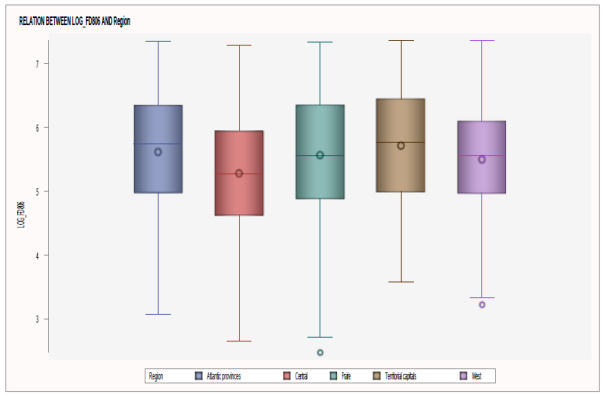
Analysis Variable : Expense												
Product_Sub_Category	N Obs	N	N Miss	Mean	Median	Mode	Minimum	Maximum	Range	Quartile Range	Lower 95% CL for Mean	Upper 95% CL for Mean
Non-alcoholic beverages and beverage mixes	14462249	14462249	0	429.15	194.48	0.00	0.00	11386.55	11386.55	569.04	428.77	429.52
Sugar and confectionery	14462249	14462249	0	252.64	124.32	0.00	0.00	6501.59	6501.59	353.34	252.44	252.84
Margarine, oils and fats (excluding butter)	14462249	14462249	0	54.20	0.00	0.00	-571.74	3037.84	3609.58	0.00	54.13	54.27
Condiments, spices and vinegars	14462249	14462249	0	231.98	139.94	0.00	0.00	2764.84	2764.84	347.62	231.84	232.13
Infant food	14462249	14462249	0	26.28	0.00	0.00	0.00	6350.24	6350.24	0.00	26.15	26.41
Frozen prepared food	14462249	14462249	0	118.77	0.00	0.00	0.00	4039.62	4039.62	156.00	118.65	118.90
Soup (except infant soup)	14462249	14462249	0	54.25	0.00	0.00	0.00	2098.72	2098.72	59.64	54.19	54.31
Ready-to-serve prepared food	14462249	14462249	0	86.34	0.00	0.00	0.00	1980.10	1980.10	89.70	86.23	86.45
Snack food	14462249	14462249	0	94.20	0.00	0.00	0.00	1761.24	1761.24	138.32	94.11	94.29

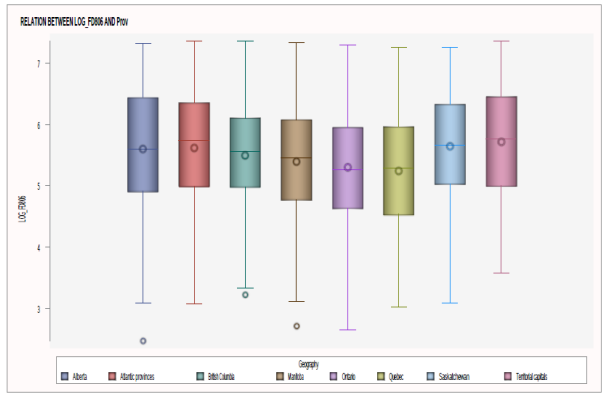
NATIONAL EXPENSE Non-alcoholic beverages and beverage mixes

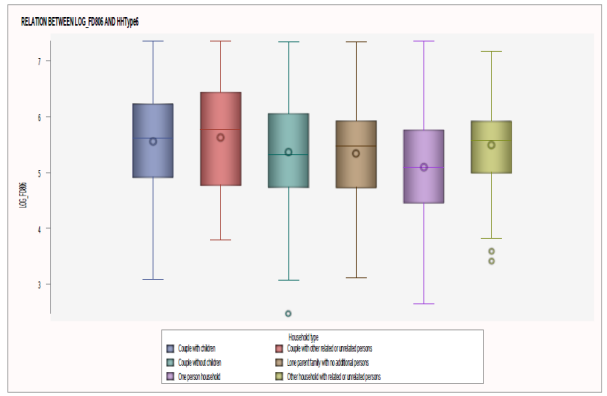
The MEANS Procedure

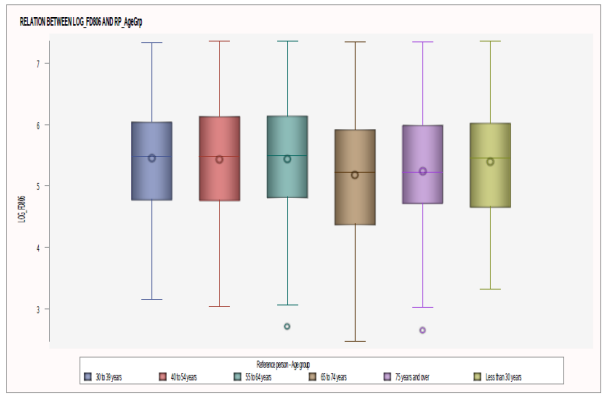
Analysis Variable : Expense												
Product	N Obs	N	N Miss	Mean	Median	Mode	Minimum	Maximum	Range	Quartile Range	Lower 95% CL for Mean	Upper 95% CL for Mean
Coffee and tea	14462249	14462249	0	174.53	0.00	0.00	0.00	5841.97	5841.97	205.92	174.33	174.73
Non-alcoholic beverages	14462249	14462249	0	254.62	76.13	0.00	0.00	8837.67	8837.67	293.05	254.32	254.91

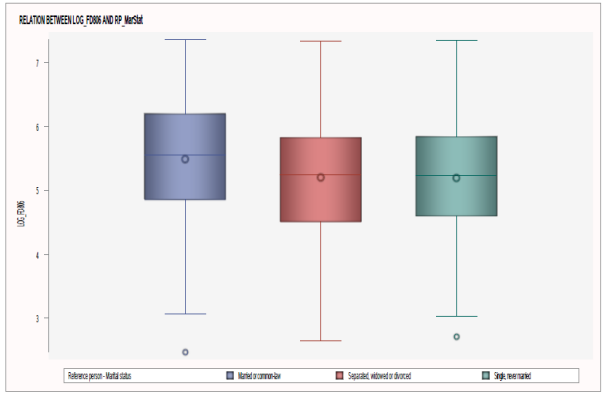
Bivariate After log transformation

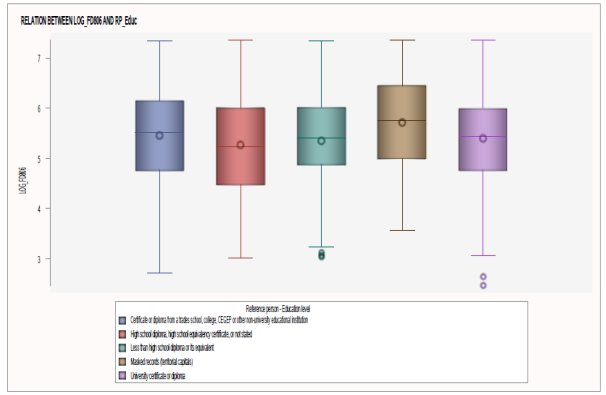


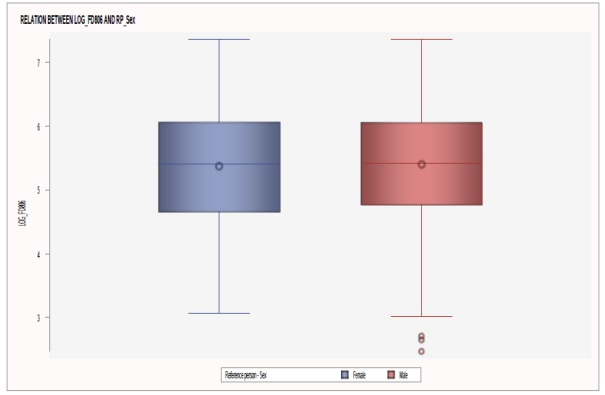


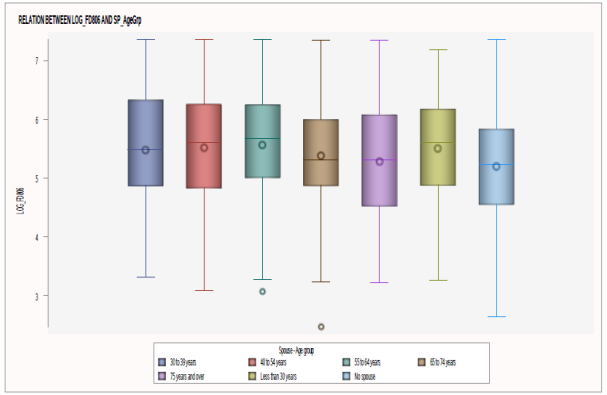


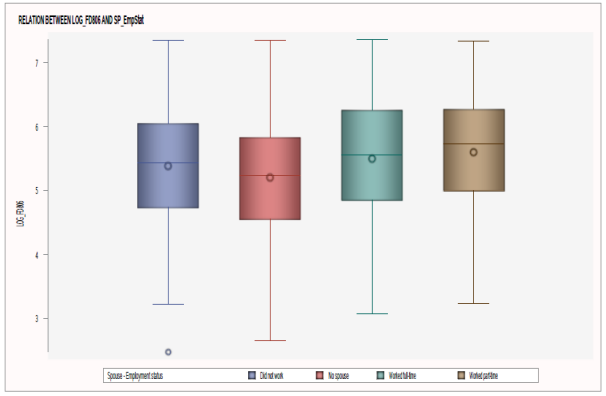


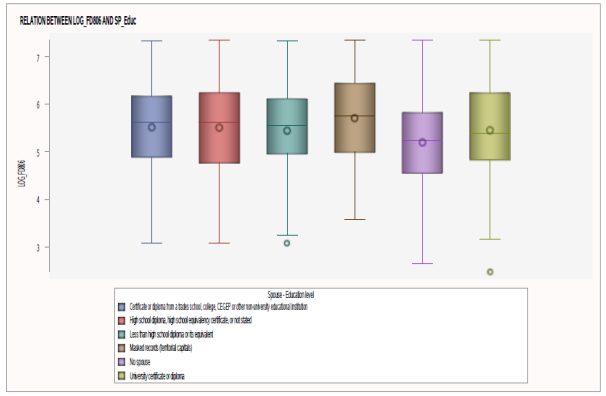


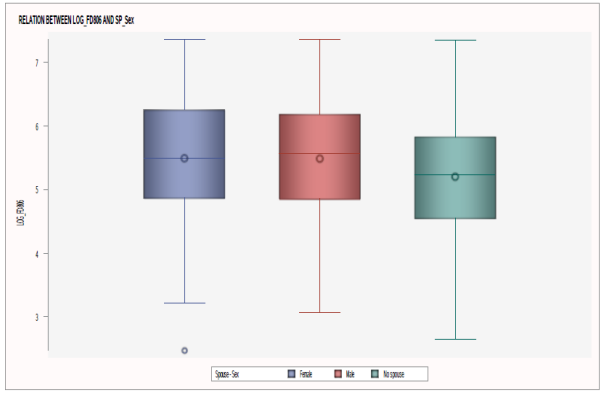




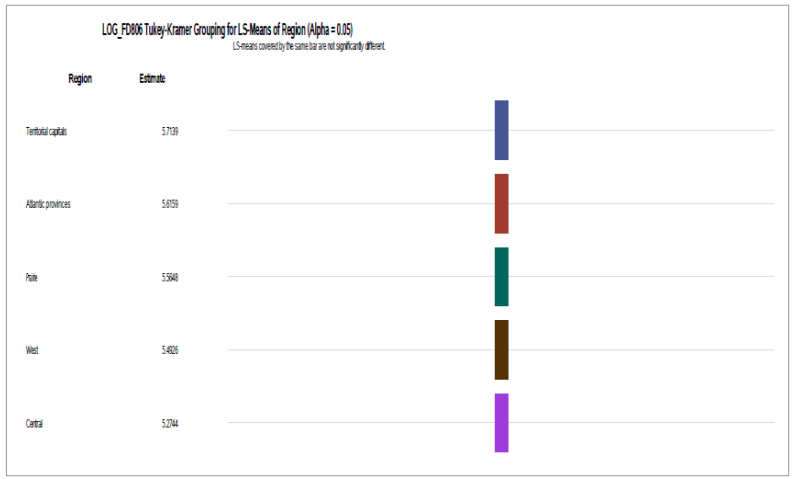


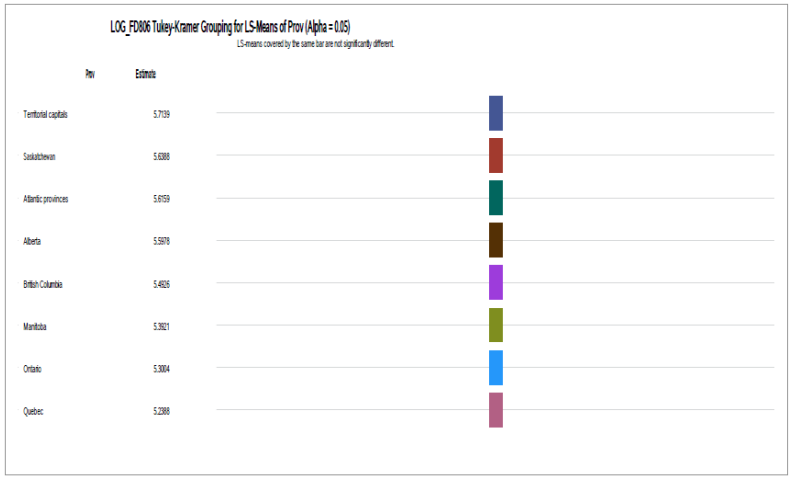


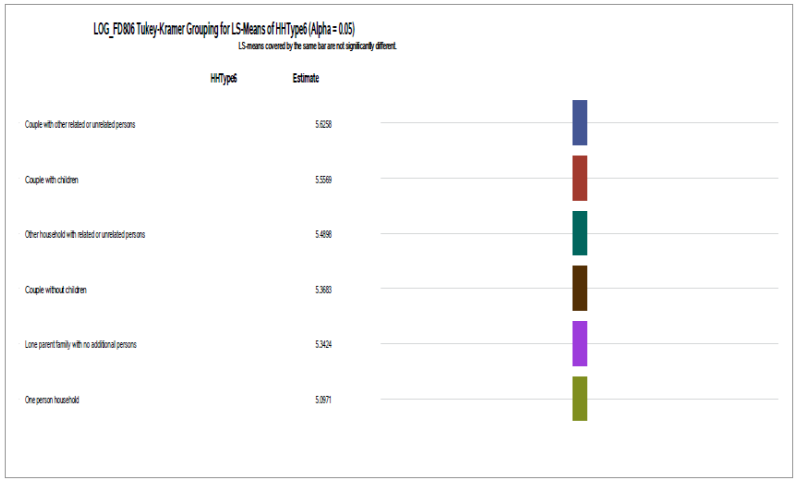


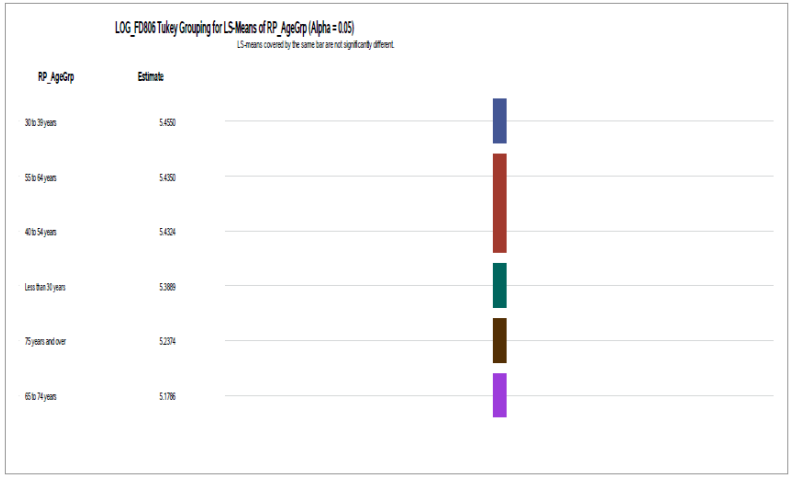


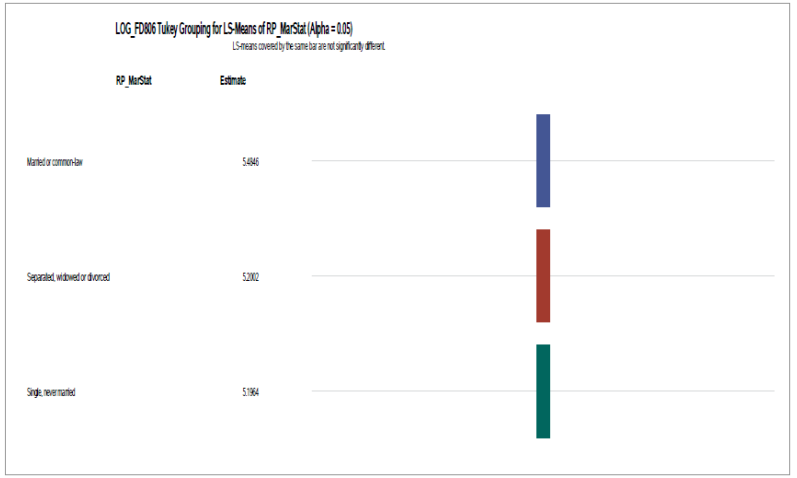
Which groups are different?

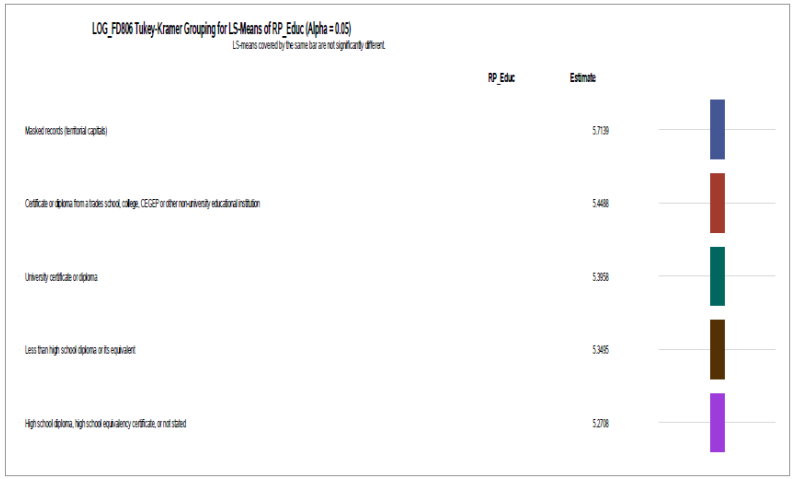


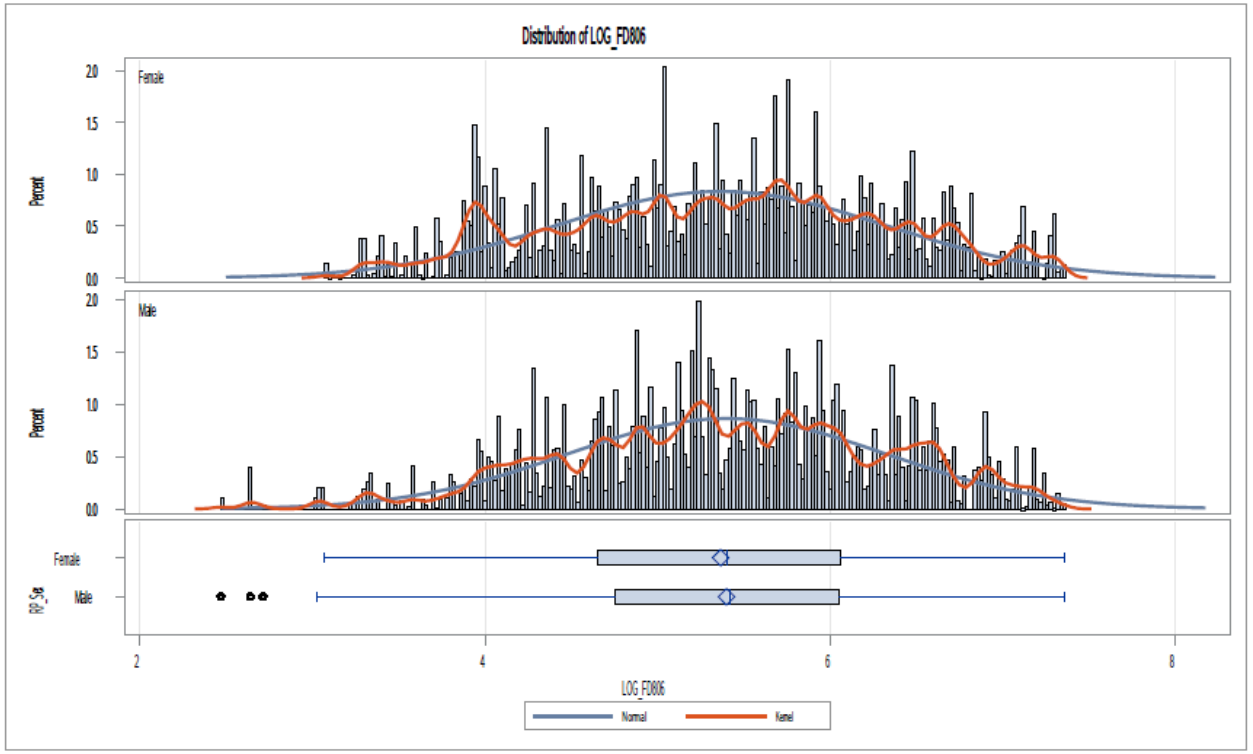










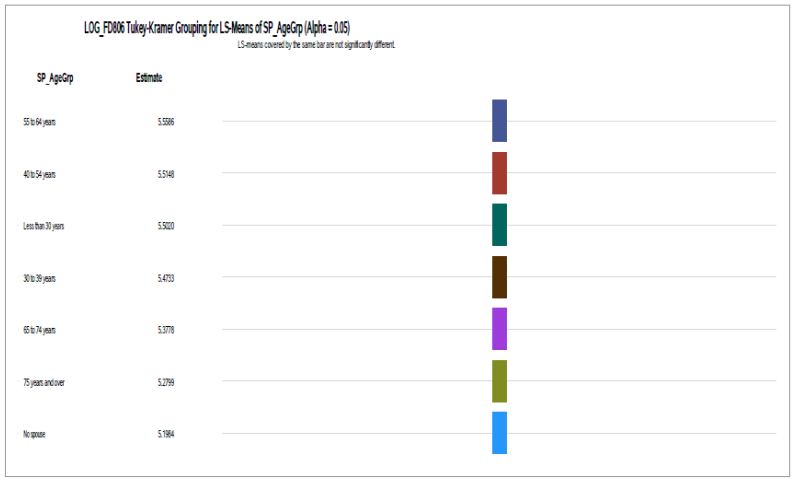


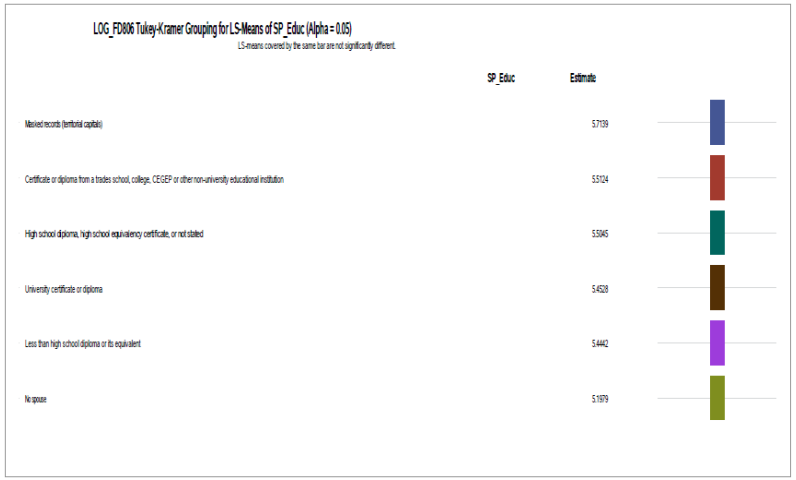
T-test for RP_Sex

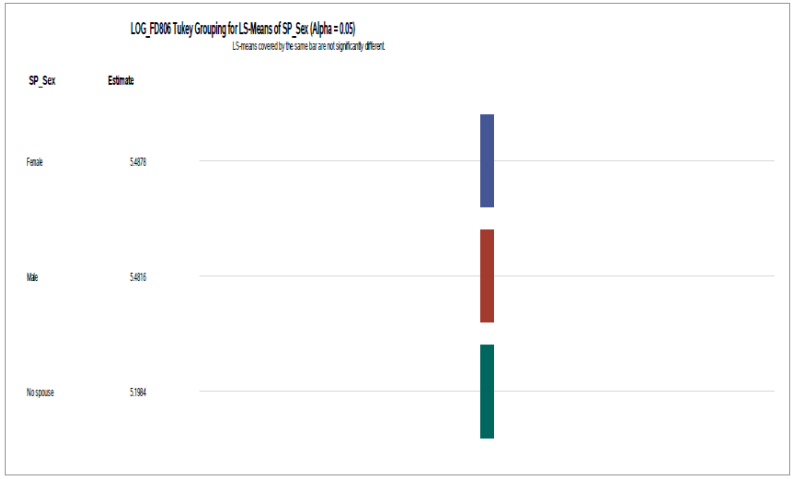
RP_Sex	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Female		5.3660	5.3651	5.3669	0.9557	0.9550	0.9563
Male		5.4000	5.3991	5.4010	0.9257	0.9250	0.9263
Diff (1-2)	Pooled	-0.0340	-0.0353	-0.0327	0.9415	0.9411	0.9420
Diff (1-2)	Satterthwaite	-0.0340	-0.0353	-0.0327			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	8.13E6	-51.47	<.0001
Satterthwaite	Unequal	8.09E6	-51.55	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	4.26E6	3.87E6	1.07	<.0001







Script

```
LIBNAME ANA "D:\1_Metro College\Courses\DSP-SAS\Project";

****************************************************** SAS Macros ************************************************;
PROC OPTIONS OPTION = MACRO;
RUN;

ODS GRAPHICS ON;


/*Auxiliar Information*/

*T-test - Road map:;
DATA TESTT;
INPUT @1 TEXT $CHAR125.;
INFILE DATALINES TRUNCOVER;
DATALINES;
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
;
RUN;

DATA TEST_ANOVA;
INPUT @1 TEXT $CHAR125.;
INFILE DATALINES TRUNCOVER;
DATALINES;
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.
;
```

```
*Levene's test:;
DATA LEVENE;
INPUT @1 TEXT $CHAR125. ;
INFILE DATALINES TRUNCOVER;
DATALINES;
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
;
RUN;

DATA SHAPIRO;
INPUT @1 TEXT $CHAR140.;
INFILE DATALINES TRUNCOVER;
DATALINES;
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.
;
RUN;

DATA TESTT1;
INPUT TEXT $160.;
DATALINES;
 "Null hypothesis: There's no difference in means"
;
RUN;

DATA CHI;
INPUT TEXT $140.;
DATALINES;
Chi - square (Road Map and Assumptions):
If condition of chi-square are satisfied and p-value is less than significant level (5%) reject null hypothesis:
There is a relationship between variables at the defined significant level.

Null hypothesis: Variables are independents.

1. N, the total frequency, should be reasonably large (greater than 50)
2. The sample observations should be independent. No individual item should be included twice or more in the sample
3. No expected frequencies should be small. Preferably each expected frequency should be larger than 10 but in any case not less than 5.
;
RUN;
```

```
DATA CORR;
INPUT TEXT $160.;
DATALINES;
Null hypothesis: there’s no association between variables.
1.Normal distribution for both variables for pearson
2.homoscedasticity assumes that data is equally distributed about the regression line.
3.Linear:
Linear: pearson
Monotonically related (not normal): spearman kendall hoeffding
;
RUN;

/*UNIVARIATE*/

%MACRO EDA(DATA,WEIGHT=,ALPHA=,OUTPUT_FILE_PATH=);
*PDF PAGE CONFIGURATION;
options   orientation=landscape;
ods listing close;

*gRAPH SIZE;
ODS GRAPHICS / HEIGHT=12CM WIDTH=20CM;

*CHECK IF A ALPHA VALUE WAS INFORMED;
%IF %SUPERQ(ALPHA) NE %STR() %THEN %LET ALPHA= &ALPHA; %ELSE %LET ALPHA = 0.05;

*SEPARATING CATEGORICAL AND NUMERICAL VALUES;
PROC CONTENTS DATA = &DATA. OUT= &DATA._PROJECT_VARS;RUN;
PROC SQL;
SELECT NAME INTO: NUM_ONLY SEPARATED BY " "
FROM &DATA._PROJECT_VARS
WHERE TYPE EQ 1
;
SELECT NAME INTO : CHAR_ONLY SEPARATED BY " "
FROM &DATA._PROJECT_VARS
WHERE TYPE EQ 2
;
QUIT;

/************************NUMERICAL ANALYSIS **********************/
%LET N1 = %SYSFUNC(COUNTW(&NUM_ONLY));
%DO I1 = 1 %TO &N1;
               %LET VAR = %SCAN(&NUM_ONLY,&I1);
```

```
*EXTRACTING LABEL;
DATA DEM2 ;
SET &DATA.;
KEEP &VAR.;
RUN;
proc sql
       noprint;
       select label into: label1
       from dictionary.columns
       where libname = 'WORK' and memname = 'DEM2';
       /* libname and memname values must be upper case  */
quit;
*******************************************************;
ODS PDF FILE = "&OUTPUT_FILE_PATH.\&DATA._PROFILING_&VAR._&SYSDATE9..PDF" STARTPAGE=NO UNIFORM;
*******************************************************;
TITLE "NUMERICAL UNIVARIATE ANALYSIS FOR &DATA";
PROC MEANS DATA=&DATA N NMISS MEAN MEDIAN MODE MIN MAX RANGE QRANGE CLM  MAXDEC=2 ALPHA = &ALPHA;
FREQ &WEIGHT;
VAR &VAR;RUN;

PROC SGPLOT DATA=&DATA;
TITLE J=LEFT "&LABEL1.(&VAR.)";
  HISTOGRAM &VAR / FILLATTRS =(COLOR = PLUM) DATASKIN=PRESSED FREQ=&WEIGHT;
  DENSITY &VAR/ LINEATTRS = (COLOR = BLACK);
  DENSITY &VAR/TYPE = KERNEL LINEATTRS = (COLOR = DARKBLUE) ;
    STYLEATTRS
    BACKCOLOR = SNOW
    WALLCOLOR = WHITESMOKE;
  KEYLEGEND / LOCATION = INSIDE POSITION = TOPRIGHT;
  XAXIS DISPLAY=(NOLABEL);
RUN;

PROC SGPLOT DATA=&DATA;
 TITLE J=LEFT "&LABEL1.(&VAR.)";
  HBOX &VAR / FILLATTRS =(COLOR = PLUM) DATASKIN=PRESSED FREQ=&WEIGHT;
    STYLEATTRS
    BACKCOLOR = SNOW
    WALLCOLOR = WHITESMOKE;
  KEYLEGEND / LOCATION = INSIDE POSITION = TOPRIGHT;
  XAXIS DISPLAY=(NOLABEL);
RUN;
QUIT;
TITLE;
RUN;

ODS PDF CLOSE;
%END;
```

```
/**********************CATEGORICAL ANALYSIS*****************/

%LET N2 = %SYSFUNC(COUNTW(&CHAR_ONLY));
%DO I2 = 1 %TO &N2;
               %LET VAR = %SCAN(&CHAR_ONLY,&I2);
*********************************************************;
*EXTRACTING LABEL;
DATA DEM2 ;
SET &DATA.;
KEEP &VAR.;
RUN;
proc sql
       noprint;
       select label into: label1
       from dictionary.columns
       where libname = 'WORK' and memname = 'DEM2';
       /* libname and memname values must be upper case  */
quit;
*********************************************************;
ODS PDF FILE = "&OUTPUT_FILE_PATH.\&DATA._PROFILING_&VAR._&SYSDATE9..PDF" STARTPAGE=NO UNIFORM;
*********************************************************;

 TITLE "CATEGORICAL UNIVARIATE ANALYSIS FOR &DATA";
  PROC FREQ DATA=&DATA NLEVELS ;
  TABLE &VAR / MISSING;
  WEIGHT &WEIGHT;
 RUN;

PROC SGPLOT DATA = &DATA ;
TITLE J=LEFT "&LABEL1.(&VAR.)";
 VBAR &VAR / CATEGORYORDER=RESPASC FREQ= &WEIGHT
               FILLATTRS=(COLOR = PLUM)
               DATASKIN=pressed;
               STYLEATTRS
               BACKCOLOR=snow
    WALLCOLOR=WhiteSmoke
               AXISEXTENT= DATA
     ;
                XAXIS DISPLAY=(NOLABEL);

 RUN;
TITLE;
QUIT;
ODS PDF CLOSE;
%END;
%MEND EDA ;
```

```
/*BIVARIATE*/

%MACRO BIVAR_CAT_CONT(DATA,VAR,ALPHA,WEIGHT=,OUTPUT_FILE_PATH= )/minoperator;
*PDF PAGE CONFIGURATION;
options   orientation=landscape;
ods listing close;

*GRAPH SIZE;
ODS GRAPHICS / HEIGHT=12CM WIDTH=20CM;

*CHECK IF A ALPHA VALUE WAS INFORMED;
%IF %SUPERQ(ALPHA) NE %STR() %THEN %LET ALPHA= &ALPHA; %ELSE %LET ALPHA = 0.05;

*SEPARATING CATEGORICAL AND NUMERICAL VALUES;
PROC CONTENTS DATA = &DATA. OUT= &DATA._PROJECT_VARS;RUN;

PROC SQL;
SELECT NAME INTO: NUM_ONLY SEPARATED BY " "
FROM &DATA._PROJECT_VARS
WHERE TYPE EQ 1
;

SELECT NAME INTO : CHAR_ONLY SEPARATED BY " "
FROM &DATA._PROJECT_VARS
WHERE TYPE EQ 2
;
QUIT;

/*CATEGORICAL VARIABLES*/
%LET N2 = %SYSFUNC(COUNTW(&CHAR_ONLY));
%DO I2 = 1 %TO &N2;*CATEGORICAL IF;
               %LET CLASS = %SCAN(&CHAR_ONLY,&I2);
*************************************************************;
*EXTRACTING LABEL -CAT;
DATA DEM2 ;
SET &DATA.;
KEEP &CLASS.;
RUN;
proc sql
       noprint;
       select label into: LABEL_CAT
       from dictionary.columns
       where libname = 'WORK' and memname = 'DEM2';
       /* libname and memname values must be upper case  */
quit;

*EXTRACTING LABEL -CONT;
DATA DEM3 ;
SET &DATA.;
KEEP &VAR.;
RUN;
```

```
proc sql
       noprint;
       select label into: LABEL_CONT
       from dictionary.columns
       where libname = 'WORK' and memname = 'DEM3';
       /* libname and memname values must be upper case  */
quit;

*********************************************************;
ODS PDF FILE = "&OUTPUT_FILE_PATH.\&DATA._BIVARIATE_&CLASS._AND_&VAR._&SYSDATE9..PDF" STARTPAGE=NO UNIFORM;
*********************************************************;

TITLE "BIVARIATE ANALYSIS OF &CLASS. AND &VAR. FOR &DATA";

PROC SORT DATA = &DATA;
BY &CLASS; RUN;

/*This presents summary for bivariate analysis*/
%LET N = %SYSFUNC(COUNTW(&VAR));
%DO I = 1 %TO &N;
               %LET X = %SCAN(&VAR,&I);
               PROC MEANS DATA = &DATA. N NMISS MIN Q1 MEDIAN MEAN Q3 MAX QRANGE CV CLM SKEW MAXDEC=2 ALPHA = &ALPHA ;
               FREQ &WEIGHT;
               TITLE2 " RELATION BETWEEN &X. AND &CLASS.";
               CLASS &CLASS. ;
               VAR &X.;
               OUTPUT OUT= OUT_&CLASS._&X. MIN =   MEAN=  STD = MAX = /AUTONAME ;
               RUN;
%END;

TITLE;TITLE2; RUN;


/*This presents visual for bivariate analysis*/
*BOXPLOT;
PROC SGPLOT DATA=&DATA;
TITLE J=LEFT " RELATION BETWEEN &X. AND &CLASS.";
VBOX &VAR/ GROUP = &CLASS DATASKIN=pressed FREQ=WEIGHTD;
               STYLEATTRS
               BACKCOLOR=snow
    WALLCOLOR=WhiteSmoke
               AXISEXTENT= DATA ;
RUN;
QUIT;

*HISTOGRAM;
/*PROC SGPANEL DATA=&DATA;*/
/*  PANELBY &CLASS. / layout=COLUMNLATTICE ONEPANEL NOVARNAME;*/
/*  HISTOGRAM &VAR. / DATASKIN=PRESSED FILLATTRS=(COLOR = PLUM);*/
/*RUN;*/
```

```
/*This presents test of independency for bivariate analysis*/

/*How many levels has CLASS variable?*/
PROC SQL NOPRINT;
SELECT COUNT(DISTINCT &CLASS.) INTO: LEVELS
FROM &DATA.
QUIT;

*CONDUCT T-TEST IF CAT HAS 2 LEVELS;
%IF &LEVELS. EQ 2 %THEN %DO;
TITLE "T-test - Road map:";
PROC REPORT DATA = TESTT NOWINDOWS NOHEADER NOCENTER
  STYLE(COLUMN) = { BACKGROUND= NONE }; RUN;
TITLE1;

/*Normality check*/
TITLE "THIS IS FOR NORMALITY CHECK OF &VAR. AND &CLASS";

proc report data= SHAPIRO nowindows noheader nocenter
  style(column) = { background = NONE }; RUN;
TITLE1;TITLE2;

PROC UNIVARIATE DATA=&DATA NORMAL ALPHA=&ALPHA;
FREQ  &WEIGHT;
VAR &VAR;
*pLOTS WERE COMMENTED BECAUSE OF COMPUTATIONAL POWER;
*QQPLOT / NORMAL(MU=EST SIGMA=EST) SQUARE ;
/*HISTOGRAM / NORMAL(COLOR=(RED BLUE) mu= est sigma= est);*/
BY &CLASS;
RUN;


/*Variance check*/
TITLE "THIS IS FOR VARIANCE CHECK OF &VAR. AND &CLASS";
TITLE2 "Levene's test";

proc report data= LEVENE nowindows noheader nocenter
  style(column) = { background = NONE }; RUN;
TITLE1;TITLE2;

PROC GLM data=&DATA ALPHA=&ALPHA PLOTS=NONE;
CLASS &CLASS;
FREQ  &WEIGHT;
MODEL &VAR = &CLASS;
MEANS &CLASS / hovtest=levene(type=abs) WELCH ALPHA=&ALPHA;
RUN;
QUIT;
```

```
/*TTEST*/

TITLE "THIS IS FOR TTEST OF &VAR. AND &CLASS";
TITLE2 "Levene's test";
proc report data= TESTT1 nowindows noheader nocenter
  style(column) = { background = NONE }; RUN;
TITLE1;

PROC TTEST DATA = &DATA ALPHA=0.05 PLOTS(ONLY)=SUMMARY;*(UNPACK);
VAR &VAR;
CLASS &CLASS;
FREQ  &WEIGHT;
RUN;
%END;

%ELSE %DO;*CONDUCT ANOVA IF CAT HAS MORE THAN 2 LEVELS;
TITLE "Anova - Roadmap";
PROC REPORT DATA = TEST_ANOVA NOWINDOWS NOHEADER NOCENTER
  STYLE(COLUMN) = { BACKGROUND= NONE }; RUN;
TITLE1;

/*Normality check*/
TITLE "THIS IS FOR NORMALITY CHECK OF &VAR. AND &CLASS";

proc report data= SHAPIRO nowindows noheader nocenter
  style(column) = { background = NONE }; RUN;
TITLE1;TITLE2;

PROC UNIVARIATE DATA=&DATA NORMAL ALPHA=&ALPHA;
FREQ  &WEIGHT;
VAR &VAR;
*pLOTS WERE COMMENTED BECAUSE OF COMPUTATIONAL POWER;
*QQPLOT / NORMAL(MU=EST SIGMA=EST) SQUARE ;
/*HISTOGRAM / NORMAL(COLOR=(RED BLUE) mu= est sigma= est);*/
BY &CLASS;
RUN;

/*Variance check*/
TITLE "THIS IS FOR ANOVA AND VARIANCE CHECK OF &VAR. AND &CLASS";
TITLE2 "Levene's test";

proc report data= LEVENE nowindows noheader nocenter
  style(column) = { background = NONE }; RUN;
TITLE1;TITLE2;
```

```
************************************************************;

/*
https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwiH7PaNjKf0AhWEZzABHaZiDXEQFnoECBgQAQ&url=https%3A%2F%2Fsupport.sas.com%2Fresources%2Fpapers%2Fproceedings%2Fproceedings%2Fsugi24%2FStats%2Fp264-24.pdf&usg=AOvVaw0bommMFk3-gjL7Yx5fwExW
*"For unequal sample sizes, the confidence coefficient is greater than 1 - a.
*In other words,
the Tukey method is conservative when there are unequal sample sizes."
 https://en.wikipedia.org/wiki/Tukey%27s_range_test
*Tukey's assumptions:
    The observations being tested are independent within and among the groups.
    The groups associated with each mean in the test are normally distributed.
    There is equal within-group variance across the groups associated with each
mean in the test (homogeneity of variance). */


PROC GLM data=&DATA ALPHA=&ALPHA PLOTS(ONLY)= CONTROLPLOT;
CLASS &CLASS;
FREQ  &WEIGHT;
MODEL &VAR = &CLASS;
MEANS &CLASS / hovtest=levene(type=abs) WELCH CLDIFF ALPHA=&ALPHA;*balanced one-way;
LSMEANS &CLASS /PDIFF ADJUST=TUKEY PLOT = MEANPLOT(CONNECT CL) LINES ALPHA=&ALPHA;*unbalanced one-way;
RUN;
QUIT;
%END;
ODS PDF CLOSE;
%END; *CATEGORICAL IF;


/*NUMERICAL VARIABLES*/

*EXTRACTING LABEL -CONT;
DATA DEM3 ;
SET &DATA.;
KEEP &VAR.;
RUN;


proc sql
       noprint;
       select label into: LABEL_CONT
       from dictionary.columns
       where libname = 'WORK' and memname = 'DEM3';
       /* libname and memname values must be upper case  */
quit;

%LET N3 = %SYSFUNC(COUNTW(&NUM_ONLY));
%DO I3 = 1 %TO &N3;*NUMERICAL IF;
               %LET VAR1 = %SCAN(&NUM_ONLY,&I3);
               %IF %SUPERQ(VAR1) NE %SUPERQ(VAR) %THEN %DO ;*VERIFY VAR AND VAR1;
```

```
*********************************************************;
*EXTRACTING LABEL -CAT;
DATA DEM2 ;
SET &DATA.;
KEEP &VAR1.;
RUN;
proc sql
       noprint;
       select label into: LABEL_CAT
       from dictionary.columns
       where libname = 'WORK' and memname = 'DEM2';
       /* libname and memname values must be upper case  */
quit;

*********************************************************;
ODS PDF FILE = "&OUTPUT_FILE_PATH.\&DATA._BIVARIATE_&VAR._AND_&VAR1._&SYSDATE9..PDF" STARTPAGE=NO UNIFORM;
*********************************************************;
TITLE "BIVARIATE ANALYSIS OF &VAR1. AND &VAR. FOR &DATA";

proc report data= CORR nowindows noheader nocenter
  style(column) = { background = NONE }; RUN;
TITLE1;TITLE2;

TITLE "CORRELATION OF &VAR1 AND &VAR. FOR &DATA";
PROC CORR DATA = &DATA  PEARSON SPEARMAN  PLOTS= (SCATTER MATRIX(HISTOGRAM));
TITLE "CORRELATION OF &VAR1 AND &VAR. FOR &DATA";
 VAR &VAR1. &VAR.;
 FREQ &WEIGHT.;
RUN;

ODS PDF CLOSE;
%END; *NUMERICAL IF;
%END;*VERIFY VAR AND VAR1;

%MEND BIVAR_CAT_CONT;

*DUMMY VARIABLES;
*SOURCE: https://blogs.sas.com/content/iml/2020/08/31/best-generate-dummy-variables-sas.html;

/* define a macro to create dummy variables */
%macro DummyVars(DSIn,    /* the name of the input data set */
                 VarList, /* the names of the categorical variables */
                 DSOut);  /* the name of the output data set */
   /* 1. add a fake response variable */
   data AddFakeY / view=AddFakeY;
      set &DSIn;
      _Y = 0;      /* add a fake response variable */
   run;
   /* 2. Create the design matrix. Include the original variables, if desired */
   proc glmselect data=AddFakeY NOPRINT outdesign(addinputvars)=&DSOut(drop=_Y);
      class      &VarList;
      model _Y = &VarList /  noint selection=none;
   run;
%mend;
```

```
************************************************** Beginning of Analysis ********************************************************;
* LOADING DATA;

PROC IMPORT OUT = ANA.DEM
            DATAFILE = "D:\1_Metro College\Courses\DSP-SAS\HouseHolding Spending\Data\SHS_EDM_2017-eng\SHS_EDM_2017\Data - Données\CSV\SHS-62M004X-E-2017-Diary_F1.csv"
            DBMS = CSV
            REPLACE;
            GETNAMES =YES;
            DATAROW=2;
            GUESSINGROWS=300;
RUN;

*DATA PROFILING;
PROC CONTENTS DATA = ANA.DEM;
RUN;

PROC PRINT DATA=ANA.DEM (OBS=10);RUN;

PROC CONTENTS DATA = ANA.DEM VARNUM SHORT; RUN;
*CaseID WeightD Prov HHType6 HHSize P0to4YN P5to17YN P18to24YN P25to64YN P65to74YN
P75plusYN NumFT NumPT RP_AgeGrp RP_Sex RP_MarStat RP_Educ RP_EmpStat SP_AgeGrp SP_Sex
SP_Educ SP_EmpStat DwelTyp Tenure CondoYN YearBuilt Repairs NumBedr NumBthr SecResYN
OthPropYN LandlineYN NumCell ComputerYN InternetYN IntCon_HSTel IntCon_Cable
IntCon_Wireless IntCon_Other TVCon_Cable CableTyp TVCon_SatDish TVCon_Phone VehicleYN
RecVehYN RP_TotInc RP_EarnInc RP_InvInc RP_GovInc RP_OthInc SP_TotInc SP_EarnInc
SP_InvInc SP_GovInc SP_OthInc OTH_TotInc OTH_EarnInc OTH_InvInc OTH_GovInc OTH_OthInc
HH_TotInc HH_EarnInc HH_InvInc HH_GovInc HH_OthInc HH_MajIncSrc CC001 CC001_C CC001_D
CF001 CI001 CL001 CL001_C CL001_D CL007 CL010 CL010_C CL010_D CM001 CS001 CS001_C CS001_D
CS003 CS004 CS005 CS007 CS008 CS010 CS011 CT010 ED002 ED002_C ED002_D ED003 ED010 ED020
ED030 ED030_C ED030_D EP001 FD001 FD003 FD100 FD1001 FD1002 FD1003 FD1004 FD101 FD102 FD103
FD104 FD105 FD106 FD107 FD108 FD112 FD200 FD201 FD202 FD203 FD204 FD205 FD206 FD207 FD208
FD209 FD212 FD300 FD301 FD302 FD303 FD304 FD305 FD308 FD309 FD315 FD316 FD330 FD331 FD350
FD380 FD381 FD382 FD400 FD401 FD402 FD403 FD404 FD405 FD406 FD407 FD408 FD409 FD410 FD411
FD412 FD418 FD421 FD440 FD441 FD442 FD447 FD470 FD471 FD478 FD479 FD500 FD501 FD502 FD503
FD504 FD505 FD520 FDD852 FD853 FD854 FD855 FD857 FD870 FD871 FD872 FD873 FD874 FD875 FD879 FD880
FD881 FD882 FD883 FD884 FD885 FD889 FD990 FD991 FD992 FD993 FD994 FD995 GC001 HC001 HC001_C
HC001_D HC002 HC002_C HC002_D HC022 HC023 HC024 HC025 HE001 HE001_C HE001_D HE002 HE002_C
HE002_D HE010 HE010_C HE010_D HE016 HE020 HF001 HF001_C HF001_D HF002 HF002_C HF002_D HO001
HO001_C HO001_D HO002 HO003 HO003_C HO003_D HO004 HO005 HO006 HO010 HO014 HO018 HO018_C
HO018_D HO022 ME001 ME001_C ME001_D ME002 ME010 ME010_C ME010_D MG001 PC001 PC001_C PC001_D
PC002 PC020 RE001 RE001_C RE001_D RE002 RE002_C RE002_D RE004 RE005 RE006 RE007 RE008 RE010
RE010_C RE010_D RE016 RE016_C RE016_D RE022 RE026 RE040 RE040_C RE040_D RE041 RE041_C
RE041_D RE052 RE060 RE060_C RE060_D RE061 RE061_C RE061_D RE062 RE063 RE066 RE067 RE070
RE074 RE078 RO001 RO001_C RO001_D RO002 RO003 RO004 RO005 RO006 RV001 RV001_C RV001_D
RV002 RV010 RV010_C RV010_D SH001 SH002 SH003 SH004 SH010 SH011 SH012 SH015 SH016 SH019
SH030 SH031 SH032 SH033 SH034 SH040 SH041 SH042 SH044 SH046 SH047 SH050 SH060 SH061 SH062
SH990 SH991 SH992 TA001 TA002 TA005 TA006 TA007 TA008 TC001 TC001_C TC001_D TE001 TE001_C
TE001_D TR001 TR001_C TR001_D TR002 TR002_C TR002_D TR003 TR004 TR008 TR010 TR020 TR020_C
TR020_D TR021 TR022 TR030 TR030_C TR030_D TR031 TR032 TR033 TR034 TR035 TR036 TR037 TR038
TR039 TR050 TX001;
```

```
/*CATEGORICAL OR NUMERICAL?*/


/*FORMATS*/
PROC FORMAT;
                VALUE Prov_F
                                14 = "Atlantic provinces"
                                24 = "Quebec"
                                35 = "Ontario"
                                46 = "Manitoba"
                                47 = "Saskatchewan"
                                48 = "Alberta"
                                59 = "British Columbia"
                                63 = "Territorial capitals"
                                ;

                VALUE HHType6_F
                                1 = "One person household"
                                2 = "Couple without children"
                                3 = "Couple with children"
                                4 = "Couple with other related or unrelated persons"
                                5 = "Lone parent family with no additional persons"
                                6 = "Other household with related or unrelated persons"
                                ;

                VALUE HHSize_F
                                1 = "1"
                                2 = "2"
                                3 = "3"
                                4 = "4 or more"
                                ;

                VALUE Yes_No_F
                                1 = "Yes"
                                2 = "No"
                                ;


                VALUE Num_F
                                0 = "0"
                                1 = "1"
                                2 = "2 or more"
                                ;

                VALUE RP_AgeGrp_F
                                01 = "Less than 30 years"
                                02 = "30 to 39 years"
                                03 = "40 to 54 years"
                                04 = "55 to 64 years"
                                05 = "65 to 74 years"
                                06 = "75 years and over"
                                ;
```

```
VALUE RP_Sex_F
                1 = "Male"
                2 = "Female"
                ;

VALUE MarStat_F
                1 = "Married or common-law"
                2 = "Single, never married"
                3 = "Separated, widowed or divorced"
                ;

VALUE RP_Educ_F
                1 = "Less than high school diploma or its equivalent"
                2 = "High school diploma, high school equivalency certificate, or not stated"
                3 = "Certificate or diploma from a trades school, college, CEGEP or other non-university educational institution"
                4 = "University certificate or diploma"
                9 = "Masked records (territorial capitals)"
                ;

VALUE RP_EmpStat_F
                1 = "Worked full-time"
                2 = "Worked part-time"
                3 = "Did not work"
                ;

VALUE SP_AgeGrp_F
                01 = "Less than 30 years"
                02 = "30 to 39 years"
                03 = "40 to 54 years"
                04 = "55 to 64 years"
                05 = "65 to 74 years"
                06 = "75 years and over"
                96 = "No spouse"
                ;


VALUE SP_Sex_F
                1 = "Male"
                2 = "Female"
                6 = "No spouse"
                ;

VALUE SP_Educ_F
                1 = "Less than high school diploma or its equivalent"
                2 = "High school diploma, high school equivalency certificate, or not stated"
                3 = "Certificate or diploma from a trades school, college, CEGEP or other non-university educational institution"
                4 = "University certificate or diploma"
                6 = "No spouse"
                9 = "Masked records (territorial capitals)"
                ;
```

```
VALUE SP_EmpStat_F
                1 = "Worked full-time"
                2 = "Worked part-time"
                3 = "Did not work"
                6 = "No spouse"
                ;

VALUE DwelTyp_F
                1 = "Single detached"
                2 = "Double, row, terrace or duplex"
                3 = "Apartment or other"
                ;

VALUE Tenure_F
                1 = "Owned with mortgage"
                2 = "Owned without mortgage"
                3 = "Rented"
                ;

VALUE YearBuilt_F
                1 = "1960 or before"
                2 = "1961 to 1980"
                3 = "1981 and after"
                9 = "Masked records (territorial capitals)"
                ;

VALUE Repairs_F
                1 = "Regular maintenance"
                2 = "Minor repairs"
                3 = "Major repairs"
                9 = "Masked records (territorial capitals)"
                ;

VALUE NumBedr_F
                1 = "0 or 1"
                2 = "2"
                3 = "3"
                4 = "4 or more"
                9 = "Masked records (territorial capitals)"
                ;

VALUE NumBthr_F
                1 = "0 or 1"
                2 = "2"
                3 = "3 or more"
                ;

VALUE NumCell_F
                0 = "0"
                1 = "1"
                2 = "2"
                3 = "3 or more"
                ;
```

```
                VALUE CableTyp_F
                                1 = "Basic cable service"
                                2 = "Extended cable service"
                                6 = "Not applicable"
                                ;

                VALUE HH_MajIncSrc_F
                                1 = "Earnings (employment income)"
                                2 = "Investment income"
                                3 = "Government transfer payments"
                                4 = "Other income"
                                5 = "All sources = 0"
                                ;

RUN;

DATA ANA.DEM1;
SET ANA.DEM;
/*CONVERTING DATA TYPE*/
Prov_1 = PUT(Prov,$Prov_F.);
HHType6_1 = PUT(HHType6,$HHType6_F.);
HHSize_1 = PUT(HHSize,$HHSize_F.);
P0to4YN_1 = PUT(P0to4YN,$Yes_No_F.);
P5to17YN_1 = PUT(P5to17YN,$Yes_No_F.);
P18to24YN_1 = PUT(P18to24YN,$Yes_No_F.);
P25to64YN_1 = PUT(P25to64YN,$Yes_No_F.);
P65to74YN_1 = PUT(P65to74YN,$Yes_No_F.);
P75plusYN_1 = PUT(P75plusYN,$Yes_No_F.);
NumFT_1 = PUT(NumFT,$Num_F.);
NumPT_1 = PUT(NumPT,$Num_F.);
RP_AgeGrp_1 = PUT(RP_AgeGrp,$RP_AgeGrp_F.);
RP_Sex_1 = PUT(RP_Sex,$RP_Sex_F.);
RP_MarStat_1 = PUT(RP_MarStat,$MarStat_F.);
RP_Educ_1 = PUT(RP_Educ,$RP_Educ_F.);
RP_EmpStat_1 = PUT(RP_EmpStat,$RP_EmpStat_F.);
SP_AgeGrp_1 = PUT(SP_AgeGrp,$SP_AgeGrp_F.);
SP_Sex_1 = PUT(SP_Sex,$SP_Sex_F.);
SP_Educ_1 = PUT(SP_Educ,$SP_Educ_F.);
SP_EmpStat_1 = PUT(SP_EmpStat,$SP_EmpStat_F.);
DwelTyp_1 = PUT(DwelTyp,$DwelTyp_F.);
Tenure_1 = PUT(Tenure,$Tenure_F.);
CondoYN_1 = PUT(CondoYN,$Yes_No_F.);
YearBuilt_1 = PUT(YearBuilt,$YearBuilt_F.);
Repairs_1 = PUT(Repairs,$Repairs_F.);
NumBedr_1 = PUT(NumBedr,$NumBedr_F.);
NumBthr_1 = PUT(NumBedr,$NumBthr_F.);
SecResYN_1 = PUT(SecResYN,$Yes_No_F.);
OthPropYN_1 = PUT(OthPropYN,$Yes_No_F.);
LandlineYN_1 = PUT(LandlineYN,$Yes_No_F.);
NumCell_1 = PUT(NumCell,$NumCell_F.);
ComputerYN_1 = PUT(ComputerYN,$Yes_No_F.);
InternetYN_1 = PUT(InternetYN,$Yes_No_F.);
IntCon_HSTel_1 = PUT(IntCon_HSTel,$Yes_No_F.);
```

```
IntCon_Cable_1 = PUT(IntCon_Cable,$Yes_No_F.);
IntCon_Wireless_1 = PUT(IntCon_Wireless,$Yes_No_F.);
IntCon_Other_1 = PUT(IntCon_Other,$Yes_No_F.);
TVCon_Cable_1 = PUT(TVCon_Cable,$Yes_No_F.);
CableTyp_1 = PUT(CableTyp,$CableTyp_F.);
TVCon_SatDish_1 = PUT(TVCon_SatDish,$Yes_No_F.);
TVCon_Phone_1 = PUT(TVCon_Phone,$Yes_No_F.);
VehicleYN_1 = PUT(VehicleYN,$Yes_No_F.);
RecVehYN_1 = PUT(RecVehYN,$Yes_No_F.);
HH_MajIncSrc_1 = PUT(RecVehYN,$HH_MajIncSrc_F.);

DROP /*OLD VARIABLES*/
Prov
HHType6
HHSize
P0to4YN
P5to17YN
P18to24YN
P25to64YN
P65to74YN
P75plusYN
NumFT
NumPT
RP_AgeGrp
RP_Sex
RP_MarStat
RP_Educ
RP_EmpStat
SP_AgeGrp
SP_Sex
SP_Educ
SP_EmpStat
DwelTyp
Tenure
CondoYN
YearBuilt
Repairs
NumBedr
NumBthr
SecResYN
OthPropYN
LandlineYN
NumCell
ComputerYN
InternetYN
IntCon_HSTel
IntCon_Cable
IntCon_Wireless
IntCon_Other
TVCon_Cable
CableTyp
TVCon_SatDish
```

```
TVCon_Phone
VehicleYN
RecVehYN
HH_MajIncSrc
;

RENAME /*NEW VARIABLES*/
Prov_1 = Prov
HHType6_1 = HHType6
HHSize_1 = HHSize
P0to4YN_1 = P0to4YN
P5to17YN_1 = P5to17YN
P18to24YN_1 = P18to24YN
P25to64YN_1 = P25to64YN
P65to74YN_1 = P65to74YN
P75plusYN_1 = P75plusYN
NumFT_1 = NumFT
NumPT_1 = NumPT
RP_AgeGrp_1 = RP_AgeGrp
RP_Sex_1 = RP_Sex
RP_MarStat_1 = RP_MarStat
RP_Educ_1 = RP_Educ
RP_EmpStat_1 = RP_EmpStat
SP_AgeGrp_1 = SP_AgeGrp
SP_Sex_1 = SP_Sex
SP_Educ_1 = SP_Educ
SP_EmpStat_1 = SP_EmpStat
DwelTyp_1 = DwelTyp
Tenure_1 = Tenure
CondoYN_1 = CondoYN
YearBuilt_1 = YearBuilt
Repairs_1 = Repairs
NumBedr_1 = NumBedr
NumBthr_1 = NumBthr
SecResYN_1 = SecResYN
OthPropYN_1 = OthPropYN
LandlineYN_1 = LandlineYN
NumCell_1 = NumCell
ComputerYN_1 = ComputerYN
InternetYN_1 = InternetYN
IntCon_HSTel_1 = IntCon_HSTel
IntCon_Cable_1 = IntCon_Cable
IntCon_Wireless_1 = IntCon_Wireless
IntCon_Other_1 = IntCon_Other
TVCon_Cable_1 = TVCon_Cable
CableTyp_1 = CableTyp
TVCon_SatDish_1 = TVCon_SatDish
TVCon_Phone_1 = TVCon_Phone
VehicleYN_1 = VehicleYN
RecVehYN_1 = RecVehYN
HH_MajIncSrc_1 = HH_MajIncSrc
;
RUN;
```

```
/* INSERTING LABELS */
DATA ANA.DEM1;
SET ANA.DEM1;
LABEL
CableTyp="Type of cable television service"
CC001="Child care"
CC001_C="Child care - Interview"
CC001_D="Child care - Diary"
CF001="Women's and girls' wear (4 years and over)"
CI001="Children's wear (under 4 years)"
CL001="Clothing and accessories"
CL001_C="Clothing and accessories - Interview"
CL001_D="Clothing and accessories - Diary"
CL007="Clothing fabric, yarn, thread, and other notions"
CL010="Clothing services"
CL010_C="Clothing services - Interview"
CL010_D="Clothing services - Diary"
CM001="Men's and boys' wear (4 years and over)"
ComputerYN="Home computer"
CondoYN="Condominium"
CS001="Communications"
CS001_C="Communications - Interview"
CS001_D="Communications - Diary"
CS003="Telephone services and equipment"
CS004="Landline telephone services"
CS005="Cell phone and pager services"
CS007="Internet access services"
CS008="Online services"
CS010="Postal, courier and other communication services"
CS011="Telephones and equipment"
CT010="Gifts of clothing for non-household members"
DwelTyp="Dwelling type"
ED002="Education"
ED002_C="Education - Interview"
ED002_D="Education - Diary"
ED003="Tuition fees"
ED010="School supplies"
ED020="Textbooks"
ED030="Textbooks and school supplies"
ED030_C="Textbooks and school supplies - Interview"
ED030_D="Textbooks and school supplies - Diary"
EP001="Personal insurance payments and pension contributions"
FD001="Food expenditures"
FD003="Food purchased from stores"
FD100="Bakery products"
FD1001="Frozen side dishes and other frozen prepared food"
FD1002="Other ready-to-serve prepared food"
FD1003="Cod, flounder, sole and haddock (fresh or frozen, uncooked)"
FD1004="Other oils and fats"
FD101="Bread and unsweetened rolls and buns"
FD102="Bread"
FD103="Unsweetened rolls and buns"
```

```
FD104="Cookies and crackers"
FD105="Cookies and sweet biscuits"
FD106="Crackers and crisp breads"
FD107="Other bakery products"
FD108="Other bakery products (except frozen)"
FD112="Frozen bakery products"
FD200="Cereal grains and cereal products"
FD201="Rice and rice mixes"
FD202="Rice"
FD203="Rice mixes"
FD204="Pasta products"
FD205="Pasta (fresh or dry)"
FD206="Pasta (canned)"
FD207="Pasta mixes"
FD208="Other cereal grains and cereal products"
FD209="Flour and flour-based mixes"
FD212="Breakfast cereal and other grain products (except infant)"
FD300="Fruit, fruit preparations and nuts"
FD301="Fresh fruit"
FD302="Apples (fresh)"
FD303="Bananas and plantains (fresh)"
FD304="Grapes (fresh)"
FD305="Peaches and nectarines (fresh)"
FD308="Pears (fresh)"
FD309="Berries (fresh)"
FD315="Citrus fruit (fresh)"
FD316="Other fruit (fresh)"
FD330="Preserved fruit and fruit preparations"
FD331="Fruit juice"
FD350="Other preserved fruit and fruit preparations"
FD380="Nuts and seeds"
FD381="Peanuts (shelled or unshelled)"
FD382="Other nuts and seeds"
FD400="Vegetables and vegetable preparations"
FD401="Fresh vegetables"
FD402="Potatoes (except sweet potatoes)"
FD403="Tomatoes (fresh)"
FD404="Lettuce (fresh)"
FD405="Cabbage (fresh)"
FD406="Carrots (fresh)"
FD407="Onions (fresh)"
FD408="Celery (fresh)"
FD409="Cucumber (fresh)"
FD410="Mushrooms (fresh)"
FD411="Broccoli (fresh)"
FD412="Other vegetables (fresh)"
FD418="Peppers (fresh)"
FD421="Fresh herbs"
FD440="Frozen and dried vegetables"
FD441="Potato products (frozen)"
FD442="Other frozen vegetables"
FD447="Dried vegetables and legumes"
FD470="Canned vegetables and other vegetable preparations"
```

```
FD471="Canned or bottled vegetables"
FD478="Ready-to-serve or ready-to-cook prepared salads and side dishes, fruit or vegetable based"
FD479="Vegetable juice (canned or bottled)"
FD500="Dairy products and eggs"
FD501="Cheese"
FD502="Cheddar cheese"
FD503="Mozzarella cheese"
FD504="Processed cheese"
FD505="Other cheeses"
FD520="Milk"
FD521="Fluid whole milk"
FD522="Fluid low-fat milk"
FD525="Skim and other fluid milk"
FD540="Butter"
FD541="Ice cream and ice milk (including novelties)"
FD550="Other dairy products"
FD551="Other processed milk products"
FD555="Other processed dairy products"
FD570="Eggs and other egg products"
FD571="Eggs"
FD572="Other egg products"
FD600="Meat"
FD601="Meat (except processed meat)"
FD602="Beef"
FD603="Pork"
FD604="Poultry"
FD607="Other meat and poultry"
FD650="Processed meat"
FD651="Bacon and ham"
FD660="Other processed meat"
FD700="Fish and seafood"
FD701="Fresh or frozen fish"
FD705="Salmon (fresh or frozen, uncooked)"
FD706="Other fish (fresh or frozen, uncooked)"
FD720="Canned fish or other preserved fish"
FD721="Tuna (canned)"
FD722="Salmon (canned)"
FD723="Other fish (canned or bottled)"
FD724="Cured fish"
FD730="Seafood and other marine products"
FD731="Shrimp and prawns"
FD732="Other seafood and marine products"
FD800="Non-alcoholic beverages and other food products"
FD801="Non-alcoholic beverages and beverage mixes"
FD802="Coffee and tea"
FD806="Non-alcoholic beverages"
FD814="Sugar and confectionery"
FD815="Sugar, syrups and sugar substitutes"
FD821="Candies and chocolates"
FD827="Margarine, oils and fats (excluding butter)"
FD828="Margarine"
FD829="Cooking and salad oils"
FD833="Condiments, spices and vinegars"
FD834="Mayonnaise, salad dressings and dips"
```

```
FD835="Pasta and pizza sauces (canned, bottled or dried)"
FD836="Other sauces and gravies (canned, bottled or dried)"
FD837="Dried herbs and spices"
FD838="Ketchup"
FD839="Other condiments (including vinegar)"
FD840="Pickled vegetables (including olives)"
FD841="Infant food"
FD842="Infant formula"
FD843="Infant cereals and biscuits"
FD844="Canned or bottled infant food"
FD845="Frozen prepared food"
FD846="Frozen dinners and entrees"
FD847="Frozen pizza"
FD850="Soup (except infant soup)"
FD851="Soup (chilled, frozen, canned or bottled)"
FD852="Soup (dried)"
FD853="Ready-to-serve prepared food"
FD854="Dinners and entrees (except frozen)"
FD855="Pizza (except frozen)"
FD857="Fish portions (pre-cooked and frozen)"
FD870="Other food preparations"
FD871="Peanut butter and other nut butters"
FD872="Honey"
FD873="Flavoured drink powders, crystals and syrups"
FD874="Non-dairy frozen ice treats"
FD875="Dessert powders"
FD879="Food seasonings (including table salt)"
FD880="Other materials for food preparation"
FD881="Tofu"
FD882="Other canned, bottled or dried meals"
FD883="Snack food"
FD884="Potato-based snack foods"
FD885="Other snack foods"
FD889="Other infant food (including frozen)"
FD990="Food purchased from restaurants"
FD991="Restaurant meals"
FD992="Restaurant dinners"
FD993="Restaurant lunches"
FD994="Restaurant breakfasts"
FD995="Restaurant snacks and beverages"
GC001="Games of chance"
HC001="Health care"
HC001_C="Health care - Interview"
HC001_D="Health care - Diary"
HC002="Direct costs to household for health care"
HC002_C="Direct costs to household - Interview"
HC002_D="Direct costs to household - Diary"
HC022="Private health insurance plan premiums"
HC023="Private health care plan premiums"
HC024="Dental plan premiums"
HC025="Premiums for accident or disability insurance"
HE001="Household equipment"
HE001_C="Household equipment - Interview"
```

HE001_D="Household equipment - Diary"
HE002="Household appliances"
HE002_C="Household appliances - Interview"
HE002_D="Household appliances - Diary"
HE010="Other household equipment"
HE010_C="Other household equipment - Interview"
HE010_D="Other household equipment - Diary"
HE016="Maintenance and repairs of household furnishings and equipment"
HE020="Services related to household furnishings and equipment"
HF001="Household furnishings and equipment"
HF001_C="Household furnishings and equipment - Interview"
HF001_D="Household furnishings and equipment - Diary"
HF002="Household furnishings"
HF002_C="Household furnishings - Interview"
HF002_D="Household furnishings - Diary"
HH_EarnInc="Household - Earnings"
HH_GovInc="Household - Government transfer payments"
HH_InvInc="Household - Investment income"
HH_MajIncSrc="HH_MajIncSrc"
HH_OthInc="Household - Other income"
HH_TotInc="Household - Total income"
HHSize="Household size"
HHType6="Household type"
HH_MajIncSrc="Household - Major source of income"
HO001="Household operations"
HO001_C="Household operations - Interview"
HO001_D="Household operations - Diary"
HO002="Domestic and other custodial services (excluding child care)"
HO003="Pet expenses"
HO003_C="Pet expenses - Interview"
HO003_D="Pet expenses - Diary"
HO004="Pet food"
HO005="Purchase of pets and pet-related goods"
HO006="Veterinarian and other services"
HO010="Household cleaning supplies and equipment"
HO014="Paper, plastic and foil supplies"
HO018="Garden supplies and services"
HO018_C="Garden supplies and services - Interview"
HO018_D="Garden supplies and services - Diary"
HO022="Other household supplies"
IntCon_Cable="Type of Internet connection - Cable"
IntCon_HSTel="Type of Internet connection - High-speed telephone"
IntCon_Other="Type of Internet connection - Other"
IntCon_Wireless="Type of Internet connection - Wireless"
InternetYN="Internet access"
LandlineYN="Landline telephone service"
ME001="Miscellaneous expenditures"
ME001_C="Miscellaneous expenditures - Interview"
ME001_D="Miscellaneous expenditures - Diary"
ME002="Financial services"
ME010="Other miscellaneous goods and services"
ME010_C="Other miscellaneous goods and services - Interview"
ME010_D="Other miscellaneous goods and services - Diary"

```
MG001="Gifts of money, support payments and charitable contributions"
NumBedr="Number of bedrooms"
NumBthr="Number of bathrooms"
NumCell="Number of cell phones"
NumFT="Number of full-time earners"
NumPT="Number of part-time earners"
OTH_EarnInc="Other persons - Earnings"
OTH_GovInc="Other persons - Government transfer payments"
OTH_InvInc="Other persons - Investment income"
OTH_OthInc="Other persons - Other income"
OTH_TotInc="Other persons - Total income"
OthPropYN="Other property"
P0to4YN="Presence of persons aged 0 to 4 years"
P18to24YN="Presence of persons aged 18 to 24 years"
P25to64YN="Presence of persons aged 25 to 64 years"
P5to17YN="Presence of persons aged 5 to 17 years"
P65to74YN="Presence of persons aged 65 to 74 years"
P75plusYN="Presence of persons aged 75 years and over"
PC001="Personal care"
PC001_C="Personal care - Interview"
PC001_D="Personal care - Diary"
PC002="Personal care products"
PC020="Personal care services"
Prov="Geography"
RE001="Recreation"
RE001_C="Recreation - Interview"
RE001_D="Recreation - Diary"
RE002="Recreation equipment and related services"
RE002_C="Recreation equipment and related services - Interview"
RE002_D="Recreation equipment and related services - Diary"
RE004="Outdoor play equipment and accessories"
RE005="Children's toys"
RE006="Video game systems and accessories"
RE007="Art and craft materials"
RE008="Sports, athletic and recreation equipment and related services"
RE010="Computer equipment and supplies"
RE010_C="Computer equipment and supplies - Interview"
RE010_D="Computer equipment and supplies - Diary"
RE016="Photographic goods and services"
RE016_C="Photographic goods and services - Interview"
RE016_D="Photographic goods and services - Diary"
RE022="Collectors' items"
RE026="Other recreational equipment and related services"
RE040="Home entertainment equipment and services"
RE040_C="Home entertainment equipment and services - Interview"
RE040_D="Home entertainment equipment and services - Diary"
RE041="Home entertainment equipment"
RE041_C="Home entertainment equipment - Interview"
RE041_D="Home entertainment equipment - Diary"
RE052="Home entertainment services"
RE060="Recreation services"
RE060_C="Recreation services - Interview"
RE060_D="Recreation services - Diary"
```

RE061="Entertainment"
RE061_C="Entertainment - Interview"
RE061_D="Entertainment - Diary"
RE062="Movie theatres"
RE063="Live sporting and performing arts events"
RE066="Admission fees to museums, zoos, and other sites"
RE067="Television and satellite radio services"
RE070="Use of recreation facilities"
RE074="Package trips"
RE078="Other recreational activities and services"
RecVehYN="Owned or operated a recreational vehicle"
Repairs="Dwelling condition"
RO001="Reading materials and other printed matter"
RO001_C="Reading materials and other printed matter - Interview"
RO001_D="Reading materials and other printed matter - Diary"
RO002="Newspapers"
RO003="Magazines and periodicals"
RO004="Books and E-Books"
RO005="Maps, sheet music and other printed matter"
RO006="Services related to reading materials"
RP_AgeGrp="Reference person - Age group"
RP_EarnInc="Reference person - Earnings"
RP_Educ="Reference person - Education level"
RP_EmpStat="Reference person - Employment status"
RP_GovInc="Reference person - Government transfer payments"
RP_InvInc="Reference person - Investment income"
RP_MarStat="Reference person - Marital status"
RP_OthInc="RP_OthInc"
RP_Sex="Reference person - Sex"
RP_TotInc="Reference person - Total income"
RV001="Recreational vehicles and associated services"
RV001_C="Recreational vehicles and associated services - Interview"
RV001_D="Recreational vehicles and associated services - Diary"
RV002="Purchase of recreational vehicles"
RV010="Operation of recreational vehicles"
RV010_C="Operation of recreational vehicles - Interview"
RV010_D="Operation of recreational vehicles - Diary"
SecResYN="Secondary residence"
SH001="Shelter"
SH002="Principal accommodation"
SH003="Rented principal residence"
SH004="Rent"
SH010="Owned principal residence"
SH011="Mortgage paid on the principal residence"
SH012="Repairs and maintenance of owned principal residence"
SH015="Homeowners’ property insurance for owned principal residence"
SH016="Other expenditures for owned principal residence"
SH019="Premiums for mortgage-related insurance for owned principal residence"
SH030="Water, fuel and electricity for principal accommodation"
SH031="Water and sewage‟
SH032="Electricity"
SH033="Natural gas"

```
SH034="Other fuel"
SH040="Other accommodation"
SH041="Owned secondary residences"
SH042="Mortgage paid on secondary residences"
SH044="Property insurance for owned secondary residences"
SH046="Other expenses for owned secondary residences"
SH047="Other owned properties"
SH050="Accommodation away from home"
SH060="Communication and home security services, satellite radio and Internet for owned secondary residences"
SH061="Property and school taxes, water and sewage charges for owned secondary residences"
SH062="Electricity and fuel for owned secondary residences"
SH990="Other expenses for rented principal residence"
SH991="Condominium fees, property taxes and school taxes for owned principal residence"
SH992="All other expenses for the principal residence"
SP_AgeGrp="Spouse - Age group"
SP_EarnInc="Spouse - Earnings"
SP_Educ="Spouse - Education level"
SP_EmpStat="Spouse - Employment status"
SP_GovInc="Spouse - Government transfer payments"
SP_InvInc="Spouse - Investment income"
SP_OthInc="Spouse - Other income"
SP_Sex="Spouse - Sex"
SP_TotInc="Spouse - Total income"
TA001="Tobacco products and alcoholic beverages"
TA002="Tobacco products and smokers' supplies"
TA005="Alcoholic beverages"
TA006="Alcoholic beverages served on licensed premises and in restaurants"
TA007="Alcoholic beverages purchased from stores"
TA008="Self-made alcoholic beverages"
TC001="Total current consumption"
TC001_C="Total current consumption - Interview"
TC001_D="Total current consumption - Diary"
TE001="Total expenditure"
TE001_C="Total expenditure - Interview"
TE001_D="Total expenditure - Diary"
Tenure="Dwelling tenure"
TR001="Transportation"
TR001_C="Transportation - Interview"
TR001_D="Transportation - Diary"
TR002="Private transportation"
TR002_C="Private transportation - Interview"
TR002_D="Private transportation - Diary"
TR003="Private use vehicles"
TR004="Purchase of vehicles"
TR008="Accessories for vehicles"
TR010="Fees for leased vehicles"
TR020="Rented vehicles"
TR020_C="Rented vehicles - Interview"
TR020_D="Rented vehicles - Diary"
TR021="Rental fees for vehicles"
```

```
TR022="Other expenses for rented vehicles"
TR030="Vehicle operations"
TR030_C="Vehicle operations - Interview"
TR030_D="Vehicle operations - Diary"
TR031="Vehicle registration fees"
TR032="Vehicle insurance premiums for owned and leased vehicles"
TR033="Tires, batteries, and other parts and supplies for vehicles"
TR034="Maintenance and repairs of vehicles"
TR035="Vehicle security and communication services"
TR036="Gas and other fuels"
TR037="Other vehicle services"
TR038="Parking costs"
TR039="Drivers' licences and tests, and driving lessons"
TR050="Public transportation"
TVCon_Cable="Type of television services - Cable"
TVCon_Phone="Type of television services - Phone line"
TVCon_SatDish="Type of television services - Satellite dish"
TX001="Income taxes"
VehicleYN="Owned, leased or operated a vehicle"
YearBuilt="Period of construction of the dwelling"
;
RUN;


*DATA PROFILING;
PROC CONTENTS DATA=ANA.DEM1;
RUN;

PROC  PRINT DATA=ANA.DEM1 (OBS=10)LABEL ;
RUN;

%LET PATH=D:\1_Metro College\Courses\DSP-SAS\Project\PROFILING\ORIGINAL_FORMATED_DATASET;
/*%MACRO EDA(DATA,WEIGHT=,ALPHA=,OUTPUT_FILE_PATH=);*/
%EDA(ANA.DEM1,WEIGHT=WeightD,ALPHA=0.05,OUTPUT_FILE_PATH=&PATH.);


* DATA TRANSFORMATION/VARIABLE SELECTION/FEATURING ENGINEERING FOR STUDY:
Expense Behaviour;
*SETTING SCOPE SUb categories;
DATA FOOD1;
SET ANA.DEM1;
KEEP CaseID Prov WeightD FD003 FD100 FD1001 FD1002 FD1003 FD1004 FD101 FD102 FD103 FD104 FD105 FD106 FD107 FD108 FD112 FD200 FD201 FD202 FD203
FD204 FD205 FD206 FD207 FD208 FD209 FD212 FD300 FD301 FD302 FD303 FD304 FD305 FD308 FD309 FD315 FD316 FD330 FD331 FD350 FD380 FD381 FD382 FD400
FD401 FD402 FD403 FD404 FD405 FD406 FD407 FD408 FD409 FD410 FD411 FD412 FD418 FD421 FD440 FD441 FD442 FD447 FD470 FD471 FD478 FD479 FD500 FD501
FD502 FD503 FD504 FD505 FD520 FD521 FD522 FD525 FD540 FD541 FD550 FD551 FD555 FD570 FD571 FD572 FD600 FD601 FD602 FD603 FD604 FD607 FD650 FD651
FD660 FD700 FD701 FD705 FD706 FD720 FD721 FD722 FD723 FD724 FD730 FD731 FD732 FD800 FD801 FD802 FD806 FD814 FD815 FD821 FD827 FD828 FD829 FD833
FD834 FD835 FD836 FD837 FD838 FD839 FD840 FD841 FD842 FD843 FD844 FD845 FD846 FD847 FD850 FD851 FD852 FD853 FD854 FD855 FD857 FD870 FD871 FD872
FD873 FD874 FD875 FD879 FD880 FD881 FD882 FD883 FD884 FD885 FD889  ;

RUN;
PROC CONTENTS DATA=FOOD1;RUN;
PROC PRINT DATA=WORK.FOOD1(OBS=10);RUN;
```

```
*https://communities.sas.com/t5/New-SAS-User/creating-BOXPLOT-for-multiple-variables/td-p/519688;
data long;
    set food1;
    array FDS(*) FD003--FD889;

    do I=1 to dim(FDS);
        FD=compress(vname(FDS(I)), 'kd');
        Value=FDS(i);
                            output;
                            DROP FD003--FD889 I;
    end;
run;

PROC CONTENTS DATA=LONG;RUN;

PROC PRINT DATA=LONG (OBS=15) LABEL; RUN;

* ADDING CATEGORIES;
PROC SQL;
CREATE TABLE ANA.FOOD AS
SELECT  CASEID, WEIGHTD,
(CASE
WHEN PROV = 'Atlantic provinces' THEN 'Atlantic provinces'
WHEN PROV = 'Quebec' THEN 'Central'
WHEN PROV = 'Ontario' THEN 'Central'
WHEN PROV = 'Manitoba' THEN 'Prairie'
WHEN PROV = 'Saskatchewan' THEN 'Prairie'
WHEN PROV = 'Alberta' THEN 'Prairie'
WHEN PROV = 'British Columbia' THEN 'West'
WHEN PROV = 'Territorial capitals' THEN 'Territorial capitals'
END) AS Region,
Prov,
FD,
(CASE
WHEN FD LIKE 'FD100' THEN 'Bakery products'
WHEN FD LIKE 'FD200' THEN 'Cereal grains and cereal products'
WHEN FD LIKE 'FD300' THEN 'Fruit, fruit preparations and nuts'
WHEN FD LIKE 'FD400' THEN 'Vegetables and vegetable preparations'
WHEN FD LIKE 'FD500' THEN 'Dairy products and eggs'
WHEN FD LIKE 'FD600' THEN 'Meat'
WHEN FD LIKE 'FD700' THEN 'Fish and seafood'
WHEN FD LIKE 'FD800' THEN 'Non-alcoholic beverages and other food products'
END) AS Product_Category,
(CASE
WHEN FD  LIKE 'FD1001' THEN 'Frozen side dishes and other frozen prepared food'
WHEN FD  LIKE 'FD104' THEN 'Cookies and crackers'
WHEN FD  LIKE 'FD107' THEN 'Other bakery products'
WHEN FD  LIKE 'FD201' THEN 'Rice and rice mixes'
WHEN FD  LIKE 'FD204' THEN 'Pasta products'
WHEN FD  LIKE 'FD208' THEN 'Other cereal grains and cereal products'
WHEN FD  LIKE 'FD301' THEN 'Fresh fruit'
WHEN FD  LIKE 'FD330' THEN 'Preserved fruit and fruit preparations'
WHEN FD  LIKE 'FD380' THEN 'Nuts and seeds'
```

```
WHEN FD  LIKE 'FD401' THEN 'Fresh vegetables'
WHEN FD  LIKE 'FD440' THEN 'Frozen and dried vegetables'
WHEN FD  LIKE 'FD470' THEN 'Canned vegetables and other vegetable preparations'
WHEN FD  LIKE 'FD501' THEN 'Cheese’
WHEN FD  LIKE 'FD520' THEN 'Milk'
WHEN FD  LIKE 'FD540' THEN 'Butter'
WHEN FD  LIKE 'FD541' THEN 'Ice cream and ice milk (including novelties)'
WHEN FD  LIKE 'FD550' THEN 'Other dairy products'
WHEN FD  LIKE 'FD570' THEN 'Eggs and other egg products'
WHEN FD  LIKE 'FD601' THEN 'Meat (except processed meat)'
WHEN FD  LIKE 'FD650' THEN 'Processed meat'
WHEN FD  LIKE 'FD701' THEN 'Fresh or frozen fish'
WHEN FD  LIKE 'FD720' THEN 'Canned fish or other preserved fish'
WHEN FD  LIKE 'FD730' THEN 'Seafood and other marine products'
WHEN FD  LIKE 'FD801' THEN 'Non-alcoholic beverages and beverage mixes'
WHEN FD  LIKE 'FD814' THEN 'Sugar and confectionery'
WHEN FD  LIKE 'FD827' THEN 'Margarine, oils and fats (excluding butter)'
WHEN FD  LIKE 'FD833' THEN 'Condiments, spices and vinegars'
WHEN FD  LIKE 'FD841' THEN 'Infant food'
WHEN FD  LIKE 'FD845' THEN 'Frozen prepared food'
WHEN FD  LIKE 'FD850' THEN 'Soup (except infant soup)'
WHEN FD  LIKE 'FD853' THEN 'Ready-to-serve prepared food'
WHEN FD  LIKE 'FD883' THEN 'Snack food'
END) AS Product_Sub_Category,
(CASE
WHEN FD  LIKE 'FD1002' THEN 'Other ready-to-serve prepared food'
WHEN FD  LIKE 'FD1003' THEN 'Cod, flounder, sole and haddock (fresh or frozen, uncooked)'
WHEN FD  LIKE 'FD1004' THEN 'Other oils and fats'
WHEN FD  LIKE 'FD101' THEN 'Bread and unsweetened rolls and buns'
WHEN FD  LIKE 'FD102' THEN 'Bread'
WHEN FD  LIKE 'FD103' THEN 'Unsweetened rolls and buns'
WHEN FD  LIKE 'FD105' THEN 'Cookies and sweet biscuits'
WHEN FD  LIKE 'FD106' THEN 'Crackers and crisp breads’
WHEN FD  LIKE 'FD108' THEN 'Other bakery products (except frozen)'
WHEN FD  LIKE 'FD112' THEN 'Frozen bakery products'
WHEN FD  LIKE 'FD202' THEN 'Rice'
WHEN FD  LIKE 'FD203' THEN 'Rice mixes'
WHEN FD  LIKE 'FD205' THEN 'Pasta (fresh or dry)'
WHEN FD  LIKE 'FD206' THEN 'Pasta (canned)'
WHEN FD  LIKE 'FD207' THEN 'Pasta mixes'
WHEN FD  LIKE 'FD209' THEN 'Flour and flour-based mixes'
WHEN FD  LIKE 'FD212' THEN 'Breakfast cereal and other grain products (except infant)'
WHEN FD  LIKE 'FD302' THEN 'Apples (fresh)'
WHEN FD  LIKE 'FD303' THEN 'Bananas and plantains (fresh)'
WHEN FD  LIKE 'FD304' THEN 'Grapes (fresh)'
WHEN FD  LIKE 'FD305' THEN 'Peaches and nectarines (fresh)'
WHEN FD  LIKE 'FD308' THEN 'Pears (fresh)'
WHEN FD  LIKE 'FD309' THEN 'Berries (fresh)'
WHEN FD  LIKE 'FD315' THEN 'Citrus fruit (fresh)'
WHEN FD  LIKE 'FD316' THEN 'Other fruit (fresh)'
WHEN FD  LIKE 'FD331' THEN 'Fruit juice'
WHEN FD  LIKE 'FD350' THEN 'Other preserved fruit and fruit preparations'
WHEN FD  LIKE 'FD381' THEN 'Peanuts (shelled or unshelled)'
```

```
WHEN FD  LIKE 'FD382' THEN 'Other nuts and seeds'
WHEN FD  LIKE 'FD402' THEN 'Potatoes (except sweet potatoes)'
WHEN FD  LIKE 'FD403' THEN 'Tomatoes (fresh)’
WHEN FD  LIKE 'FD404' THEN 'Lettuce (fresh)'
WHEN FD  LIKE 'FD405' THEN 'Cabbage (fresh)'
WHEN FD  LIKE 'FD406' THEN 'Carrots (fresh)'
WHEN FD  LIKE 'FD407' THEN 'Onions (fresh)'
WHEN FD  LIKE 'FD408' THEN 'Celery (fresh)'
WHEN FD  LIKE 'FD409' THEN 'Cucumber (fresh)'
WHEN FD  LIKE 'FD410' THEN 'Mushrooms (fresh)'
WHEN FD  LIKE 'FD411' THEN 'Broccoli (fresh)'
WHEN FD  LIKE 'FD412' THEN 'Other vegetables (fresh)'
WHEN FD  LIKE 'FD418' THEN 'Peppers (fresh)'
WHEN FD  LIKE 'FD421' THEN 'Fresh herbs'
WHEN FD  LIKE 'FD441' THEN 'Potato products (frozen)'
WHEN FD  LIKE 'FD442' THEN 'Other frozen vegetables'
WHEN FD  LIKE 'FD447' THEN 'Dried vegetables and legumes'
WHEN FD  LIKE 'FD471' THEN 'Canned or bottled vegetables'
WHEN FD  LIKE 'FD478' THEN 'Ready-to-serve or ready-to-cook prepared salads and side dishes, fruit or vegetable based'
WHEN FD  LIKE 'FD479' THEN 'Vegetable juice (canned or bottled)'
WHEN FD  LIKE 'FD502' THEN 'Cheddar cheese'
WHEN FD  LIKE 'FD503' THEN 'Mozzarella cheese'
WHEN FD  LIKE 'FD504' THEN 'Processed cheese'
WHEN FD  LIKE 'FD505' THEN 'Other cheeses'
WHEN FD  LIKE 'FD521' THEN 'Fluid whole milk'
WHEN FD  LIKE 'FD522' THEN 'Fluid low-fat milk'
WHEN FD  LIKE 'FD525' THEN 'Skim and other fluid milk'
WHEN FD  LIKE 'FD551' THEN 'Other processed milk products'
WHEN FD  LIKE 'FD555' THEN 'Other processed dairy products'
WHEN FD  LIKE 'FD571' THEN 'Eggs'
WHEN FD  LIKE 'FD572' THEN 'Other egg products'
WHEN FD  LIKE 'FD602' THEN 'Beef'
WHEN FD  LIKE 'FD603' THEN 'Pork'
WHEN FD  LIKE 'FD604' THEN 'Poultry'
WHEN FD  LIKE 'FD607' THEN 'Other meat and poultry'
WHEN FD  LIKE 'FD651' THEN 'Bacon and ham'
WHEN FD  LIKE 'FD660' THEN 'Other processed meat'
WHEN FD  LIKE 'FD705' THEN 'Salmon (fresh or frozen, uncooked)'
WHEN FD  LIKE 'FD706' THEN 'Other fish (fresh or frozen, uncooked)'
WHEN FD  LIKE 'FD721' THEN 'Tuna (canned)'
WHEN FD  LIKE 'FD722' THEN 'Salmon (canned)'
WHEN FD  LIKE 'FD723' THEN 'Other fish (canned or bottled)'
WHEN FD  LIKE 'FD724' THEN 'Cured fish'
WHEN FD  LIKE 'FD731' THEN 'Shrimp and prawns’
WHEN FD  LIKE 'FD732' THEN 'Other seafood and marine products'
WHEN FD  LIKE 'FD802' THEN 'Coffee and tea'
WHEN FD  LIKE 'FD806' THEN 'Non-alcoholic beverages'
WHEN FD  LIKE 'FD815' THEN 'Sugar, syrups and sugar substitutes'
WHEN FD  LIKE 'FD821' THEN 'Candies and chocolates'
WHEN FD  LIKE 'FD828' THEN 'Margarine'
WHEN FD  LIKE 'FD829' THEN 'Cooking and salad oils'
WHEN FD  LIKE 'FD834' THEN 'Mayonnaise, salad dressings and dips'
WHEN FD  LIKE 'FD835' THEN 'Pasta and pizza sauces (canned, bottled or dried)'
```

```
WHEN FD  LIKE 'FD836' THEN 'Other sauces and gravies (canned, bottled or dried)'
WHEN FD  LIKE 'FD837' THEN 'Dried herbs and spices'
WHEN FD  LIKE 'FD838' THEN 'Ketchup'
WHEN FD  LIKE 'FD839' THEN 'Other condiments (including vinegar)'
WHEN FD  LIKE 'FD840' THEN 'Pickled vegetables (including olives)'
WHEN FD  LIKE 'FD842' THEN 'Infant formula'
WHEN FD  LIKE 'FD843' THEN 'Infant cereals and biscuits'
WHEN FD  LIKE 'FD844' THEN 'Canned or bottled infant food'
WHEN FD  LIKE 'FD846' THEN 'Frozen dinners and entrees'
WHEN FD  LIKE 'FD847' THEN 'Frozen pizza'
WHEN FD  LIKE 'FD851' THEN 'Soup (chilled, frozen, canned or bottled)'
WHEN FD  LIKE 'FD852' THEN 'Soup (dried)'
WHEN FD  LIKE 'FD854' THEN 'Dinners and entrees (except frozen)'
WHEN FD  LIKE 'FD855' THEN 'Pizza (except frozen)'
WHEN FD  LIKE 'FD857' THEN 'Fish portions (pre-cooked and frozen)'
WHEN FD  LIKE 'FD870' THEN 'Other food preparations'
WHEN FD  LIKE 'FD871' THEN 'Peanut butter and other nut butters'
WHEN FD  LIKE 'FD872' THEN 'Honey'
WHEN FD  LIKE 'FD873' THEN 'Flavoured drink powders, crystals and syrups'
WHEN FD  LIKE 'FD874' THEN 'Non-dairy frozen ice treats'
WHEN FD  LIKE 'FD875' THEN 'Dessert powders'
WHEN FD  LIKE 'FD879' THEN 'Food seasonings (including table salt)'
WHEN FD  LIKE 'FD880' THEN 'Other materials for food preparation'
WHEN FD  LIKE 'FD881' THEN 'Tofu'
WHEN FD  LIKE 'FD882' THEN 'Other canned, bottled or dried meals'
WHEN FD  LIKE 'FD884' THEN 'Potato-based snack foods'
WHEN FD  LIKE 'FD885' THEN 'Other snack foods'
WHEN FD  LIKE 'FD889' THEN 'Other infant food (including frozen)'
END) AS Product,
VALUE AS Expense
FROM LONG;
QUIT;


* SORTING DATA;
PROC SORT DATA=ANA.FOOD;
BY Region Prov Product_Category Product_Category;
RUN;

*FORMATTING DATA;
DATA ANA.FOOD;
SET ANA.FOOD;
FORMAT EXPENSE DOLLAR13.2;
RUN;

PROC CONTENTS DATA=ANA.FOOD;RUN;
PROC PRINT DATA=ANA.FOOD (OBS=10);RUN;
```

```
*************************************************
* VISUALS AND SUMMARIES;
%LET PATH_IMG = "D:\1_Metro College\Courses\DSP-SAS\Project\IMAGES";

ODS HTML PATH= &PATH_IMG.;
TITLE "NATIONAL EXPENSE BY PRODUCT CATEGORY";
PROC MEANS DATA=ANA.FOOD N NMISS MEAN MEDIAN MODE MIN MAX RANGE QRANGE CLM MAXDEC=2 ALPHA = 0.05 ORDER=DATA;
WHERE Product_Category IS NOT NULL;
FREQ WeightD;
CLASS Product_Category;
VAR Expense;
RUN;
ODS HTML CLOSE;

ODS HTML PATH= &PATH_IMG.;
ODS GRAPHICS / IMAGENAME="NAT_PROD" HEIGHT=6IN WIDTH=10IN ;
PROC SGPLOT DATA=ANA.FOOD;
WHERE Product_Category IS NOT NULL;
TITLE J=CENTER  "NATIONAL EXPENSE BY PRODUCT CATEGORY";
VBOX Expense/FREQ=WeightD GROUP = Product_Category DATASKIN=pressed ;
               STYLEATTRS
               BACKCOLOR=snow
    WALLCOLOR=WhiteSmoke
               AXISEXTENT=DATA;
               XAXIS DISCRETEORDER=DATA;
RUN;
ODS HTML CLOSE;
QUIT;

ODS HTML PATH= &PATH_IMG.;
TITLE "REGIONAL EXPENSE BY PRODUCT CATEGORY";
PROC MEANS DATA=ANA.FOOD N NMISS MEAN MEDIAN MODE MIN MAX RANGE QRANGE  CLM MAXDEC=2 ALPHA = 0.05 ORDER=DATA;
WHERE Product_Category IS NOT NULL;
FREQ WeightD;
CLASS REGION Product_Category;
VAR Expense;
RUN;
ODS HTML CLOSE;

ODS HTML PATH= &PATH_IMG.;
ODS GRAPHICS / IMAGENAME="REG_PROD" HEIGHT=7IN WIDTH=13IN;
PROC SGPANEL DATA=ANA.FOOD;
WHERE Product_Category IS NOT NULL;
TITLE J=CENTER "REGIONAL EXPENSE BY PRODUCT CATEGORY";
PANELBY REGION / COLUMNS = 5 SPACING=3 NOVARNAME;
VBOX Expense/ GROUP = Product_Category DATASKIN=pressed;
               STYLEATTRS
               BACKCOLOR=snow
    WALLCOLOR=WhiteSmoke;
RUN;
ODS HTML CLOSE;
QUIT;
```

```
ODS HTML PATH= &PATH_IMG.;
ODS GRAPHICS / IMAGENAME="REG_PROV_PROD_MEANS" HEIGHT=7IN WIDTH=10IN;
TITLE "PROVINCIAL/TERRITORIES EXPENSE BY PRODUCT CATEGORY";
PROC MEANS DATA=ANA.FOOD N NMISS MEAN MEDIAN MODE MIN MAX RANGE QRANGE CLM MAXDEC=2 ALPHA = 0.05 ORDER=DATA;
WHERE Product_Category IS NOT NULL;
FREQ WeightD;
CLASS REGION PROV Product_Category;
VAR Expense;
RUN;
ODS HTML CLOSE;


ODS HTML PATH= &PATH_IMG.;
ODS GRAPHICS / IMAGENAME="REG_PROV_PROD" HEIGHT=7IN WIDTH=15IN;
PROC SGPLOT DATA=ANA.FOOD;
WHERE Product_Category IS NOT NULL;
TITLE "PROVINCIAL/TERRITORIES EXPENSE BY PRODUCT CATEGORY";
BLOCK X=PROV BLOCK=Region / POSITION=BOTTOM ;
VBOX Expense/ GROUP = Product_Category CATEGORY=Prov
                DATASKIN=pressed;
                STYLEATTRS
                BACKCOLOR=snow
    WALLCOLOR=WhiteSmoke;
                XAXIS DISCRETEORDER=DATA;
RUN;
ODS HTML CLOSE;
QUIT;


ODS HTML PATH= &PATH_IMG.;
TITLE "NATIONAL EXPENSE Non-alcoholic beverages and other food products";
PROC MEANS DATA=ANA.FOOD N NMISS MEAN MEDIAN MODE MIN MAX RANGE QRANGE CLM MAXDEC=2 ALPHA = 0.05 ORDER=DATA;
WHERE FD LIKE "FD801" OR FD LIKE "FD814" OR
FD LIKE "FD827" OR FD LIKE "FD833" OR
FD LIKE "FD841" OR FD LIKE "FD845" OR
FD LIKE "FD850" OR FD LIKE "FD853" OR
FD LIKE "FD883";
FREQ WeightD;
CLASS Product_Sub_Category;
VAR Expense;
RUN;
ODS HTML CLOSE;
```

```
ODS HTML PATH= &PATH_IMG.;
ODS GRAPHICS / IMAGENAME="NAT_NONALC_PROD" HEIGHT=6IN WIDTH=10IN;
PROC SGPLOT DATA=ANA.FOOD;
WHERE FD LIKE "FD801" OR FD LIKE "FD814" OR
FD LIKE "FD827" OR FD LIKE "FD833" OR
FD LIKE "FD841" OR FD LIKE "FD845" OR
FD LIKE "FD850" OR FD LIKE "FD853" OR
FD LIKE "FD883";
TITLE J=CENTER  "NATIONAL EXPENSE Non-alcoholic beverages and other food products";
VBOX Expense/FREQ=WeightD GROUP = Product_Sub_Category DATASKIN=pressed ;
               STYLEATTRS
               BACKCOLOR=snow
    WALLCOLOR=WhiteSmoke
               AXISEXTENT=DATA;
               XAXIS DISCRETEORDER=DATA;
RUN;
ODS HTML CLOSE;
QUIT;


ODS HTML PATH= &PATH_IMG.;
TITLE "NATIONAL EXPENSE Non-alcoholic beverages and beverage mixes";
PROC MEANS DATA=ANA.FOOD N NMISS MEAN MEDIAN MODE MIN MAX RANGE QRANGE CLM MAXDEC=2 ALPHA = 0.05 ORDER=DATA;
WHERE FD LIKE "FD802" OR FD LIKE "FD806";
FREQ WeightD;
CLASS Product;
VAR Expense;
RUN;
ODS HTML CLOSE;

ODS HTML PATH= &PATH_IMG.;
ODS GRAPHICS / IMAGENAME="NAT_BEV_PROD" HEIGHT=6IN WIDTH=10IN;
PROC SGPLOT DATA=ANA.FOOD;
WHERE FD LIKE "FD802" OR FD LIKE "FD806";
TITLE J=CENTER  "NATIONAL EXPENSE Non-alcoholic beverages and beverage mixes";
VBOX Expense/FREQ=WeightD GROUP = Product DATASKIN=pressed ;
               STYLEATTRS
               BACKCOLOR=snow
    WALLCOLOR=WhiteSmoke
               AXISEXTENT=DATA;
               XAXIS DISCRETEORDER=DATA;
RUN;
ODS HTML CLOSE;
QUIT;
```

```
* how much does expense of FD806 weights nationally?;
PROC SQL;
CREATE TABLE EXPS AS
SELECT "STORE PURCHASE",SUM(EXPENSE * WEIGHTD )
FROM ANA.FOOD
WHERE FD LIKE "FD003"
UNION
SELECT 'FD806 - Non-alcoholic beverages',SUM(EXPENSE * WEIGHTD)
FROM ANA.FOOD
WHERE FD LIKE "FD806" ;
QUIT;

DATA PERC_EXPS;
SET WORK.EXPS;
RENAME
_TEMA002=CATEGORY
_TEMA004=EXPENSE;
RUN;

PROC PRINT DATA=WORK.PERC_EXPS;RUN;


**********************************************************************
*Non-alcoholic beverages IS THE BIGGEST EXPENSE IN THE CATEGORY AND SUB CATEGORY WITH THE BIGGEST EXPENSES;
**********************************************************************

*SUBSETTING DATASET AND FEATURE ENGINEERING TO CONSTRUCT THE MODEL:
-  ONLY THE OBSERVATIONS WITH EXPENSES WILL BE MAINTAINED, SINCE THE GOAL IS TO MODEL THE CONSUMERS BEHAVIOR;


DATA ANA.MODEL;
SET ANA.DEM1;
WHERE FD806 NE 0;
KEEP WeightD  Prov HHType6 RP_AgeGrp RP_Sex RP_MarStat RP_Educ SP_AgeGrp SP_Sex SP_Educ SP_EmpStat HH_TotInc FD806;
LABEL FD806 = "Expense Non-alcoholic beverages";
RUN;

PROC SQL;
CREATE TABLE ANA.MODEL AS
SELECT  WeightD, Prov, HHType6, RP_AgeGrp, RP_Sex, RP_MarStat,
RP_Educ, SP_AgeGrp, SP_Sex, SP_Educ, SP_EmpStat, HH_TotInc, FD806,
(CASE
WHEN PROV = 'Atlantic provinces' THEN 'Atlantic provinces'
WHEN PROV = 'Quebec' THEN 'Central'
WHEN PROV = 'Ontario' THEN 'Central'
WHEN PROV = 'Manitoba' THEN 'Prairie'
WHEN PROV = 'Saskatchewan' THEN 'Prairie'
WHEN PROV = 'Alberta' THEN 'Prairie'
WHEN PROV = 'British Columbia' THEN 'West'
WHEN PROV = 'Territorial capitals' THEN 'Territorial capitals'
END) AS Region
From ANA.MODEL;
QUIT;
```

```
DATA ANA.MODEL;
SET ANA.MODEL;
LABEL Region="Region";
RUN;

PROC CONTENTS DATA=ANA.MODEL;RUN;

PROC CONTENTS DATA=ANA.MODEL VARNUM SHORT;RUN;
*WeightD HH_TotInc HHType6 RP_AgeGrp RP_Sex RP_MarStat RP_Educ SP_AgeGrp SP_Sex SP_Educ SP_EmpStat FD806;

*UNIVARIATE ANALYSIS;
%LET PATH=D:\1_Metro College\Courses\DSP-SAS\Project\PROFILING\MODEL_DATASET_BEFORE_TREAT;
%EDA(ANA.MODEL,WEIGHT=WeightD,ALPHA=0.05,OUTPUT_FILE_PATH=&PATH.);

*PLOTTING RP_EDUCATION;
%LET PATH_IMG = "D:\1_Metro College\Courses\DSP-SAS\Project\IMAGES";
ODS HTML PATH= &PATH_IMG.;
ODS GRAPHICS / IMAGENAME="RP_Educ1" HEIGHT=6IN WIDTH=10IN;
PROC SGPLOT DATA = ANA.MODEL ;
TITLE J=LEFT "Reference person - Education level(RP_Educ)";
 VBAR RP_Educ / CATEGORYORDER=RESPASC FREQ= WEIGHTD
                FILLATTRS=(COLOR = Plum)
                DATASKIN=pressed;
                STYLEATTRS
                BACKCOLOR=snow
    WALLCOLOR=WhiteSmoke
                AXISEXTENT= DATA ;
                XAXIS DISPLAY=(NOLABEL);
RUN;
ODS HTML CLOSE;
QUIT;

*TREATING OUTLIERS;
%LET VAR = FD806;
%LET WEIGHT = WEIGHTD;

PROC MEANS DATA=ANA.MODEL N Q1 Q3 QRANGE;
VAR &VAR.;
FREQ &WEIGHT.;
OUTPUT OUT=TEMP Q1 =Q1 Q3=Q3 QRANGE=IQR;
RUN;

DATA TEMP;
                SET TEMP;
                LOWER_LIMIT = Q1 - (3*IQR);
                UPPER_LIMIT = Q3 + (3*IQR);
                RUN;

PROC SQL NOPRINT;* CREATE SQL TO EXCLUDE OUTLIERS;
CREATE TABLE TEMP2 AS
SELECT A.*,B.LOWER_LIMIT,B.UPPER_LIMIT
FROM ANA.MODEL AS A, TEMP AS B
WHERE &VAR. BETWEEN LOWER_LIMIT AND UPPER_LIMIT;*ADDED TO CREATE FINAL DATASET HERE;
RUN;
QUIT;
```

```
DATA ANA.MODEL1;
SET TEMP2;
DROP LOWER_LIMIT UPPER_LIMIT;
RUN;

*EDA;
PROC CONTENTS DATA=ANA.MODEL1;RUN;

*UNIVARIATE ANALYSIS;
%LET PATH=D:\1_Metro College\Courses\DSP-SAS\Project\PROFILING\MODEL_DATASET_AFTER_TREAT_OUT;

%EDA(ANA.MODEL1,WEIGHT=WeightD,ALPHA=0.05,OUTPUT_FILE_PATH=&PATH.);

*PLOTTING RP_EDUCATION;
ODS HTML PATH= &PATH_IMG.;
ODS GRAPHICS / IMAGENAME="RP_Educ2" HEIGHT=6IN WIDTH=10IN;
PROC SGPLOT DATA = ANA.MODEL ;
TITLE J=LEFT "Reference person - Education level(RP_Educ)";
 VBAR RP_Educ / CATEGORYORDER=RESPASC FREQ= WEIGHTD
                FILLATTRS=(COLOR = Plum)
                DATASKIN=pressed;
                STYLEATTRS
                BACKCOLOR=snow
    WALLCOLOR=WhiteSmoke
                AXISEXTENT= DATA ;
                XAXIS DISPLAY=(NOLABEL);
RUN;
ODS HTML CLOSE;
QUIT;

*BIVARIATE ANALYSIS;

%BIVAR_CAT_CONT(ANA.MODEL1,FD806,WEIGHT=WeightD,OUTPUT_FILE_PATH=&PATH.);

*WE SEE THAT Y VARIABLE IS VERY SKEWED;
*CHECKING RECOMMENDATION FOR TRANSFORMATION;
%DummyVars(ANA.MODEL1, HHType6 RP_AgeGrp RP_Sex RP_MarStat RP_Educ SP_AgeGrp SP_Sex SP_Educ SP_EmpStat, ClassDummy);

PROC PRINT DATA=CLASSDUMMY (OBS=10);RUN;

DATA TESTE;
SET WORK.CLASSDUMMY;
DROP HHType6 RP_AgeGrp RP_Sex RP_MarStat RP_Educ SP_AgeGrp SP_Sex SP_Educ SP_EmpStat ;
RUN;
```

```
PROC CONTENTS DATA=WORK.TESTE VARNUM SHORT; RUN;

*'HHType6 Couple with children'n 'HHType6 Couple with other relate'n
'HHType6 Couple without children'n 'HHType6 Lone parent family with'n
'HHType6 One person household'n 'HHType6 Other household with rel'n
'RP_AgeGrp 30 to 39 years'n 'RP_AgeGrp 40 to 54 years'n 'RP_AgeGrp 55 to 64 years'n
'RP_AgeGrp 65 to 74 years'n 'RP_AgeGrp 75 years and over'n 'RP_AgeGrp Less than 30 years'n
'RP_Sex Female'n 'RP_Sex Male'n 'RP_MarStat Married or common-law'n 'RP_MarStat Separated,
widowed or'n 'RP_MarStat Single, never married'n 'RP_Educ Certificate or diploma f'n
'RP_Educ High school diploma, hig'n 'RP_Educ Less than high school di'n
'RP_Educ Masked records (territor'n 'RP_Educ University certificate o'n
'SP_AgeGrp 30 to 39 years'n 'SP_AgeGrp 40 to 54 years'n 'SP_AgeGrp 55 to 64 years'n
'SP_AgeGrp 65 to 74 years'n 'SP_AgeGrp 75 years and over'n 'SP_AgeGrp Less than 30 years'n
'SP_AgeGrp No spouse'n 'SP_Sex Female'n 'SP_Sex Male'n 'SP_Sex No spouse'n
'SP_Educ Certificate or diploma f'n 'SP_Educ High school diploma, hig'n
'SP_Educ Less than high school di'n 'SP_Educ Masked records (territor'n
'SP_Educ No spouse'n 'SP_Educ University certificate o'n 'SP_EmpStat Did not work'n
'SP_EmpStat No spouse'n 'SP_EmpStat Worked full-time'n 'SP_EmpStat Worked part-time'n
WeightD HH_TotInc LOG_PERC_EXPENSE;

proc transreg data = WORK.TESTE;
              model boxcox(FD806)=identity('HHType6 Couple with children'n 'HHType6 Couple with other relate'n 'HHType6 Couple without children'n 'HHType6 Lone parent family with'n 'HHType6
One person household'n 'HHType6 Other household with rel'n 'RP_AgeGrp 30 to 39 years'n 'RP_AgeGrp 40 to 54 years'n 'RP_AgeGrp 55 to 64 years'n 'RP_AgeGrp 65 to 74 years'n 'RP_AgeGrp 75 years
and over'n 'RP_AgeGrp Less than 30 years'n 'RP_Sex Female'n 'RP_Sex Male'n 'RP_MarStat Married or common-law'n 'RP_MarStat Separated, widowed or'n 'RP_MarStat Single, never married'n 'RP_Educ
Certificate or diploma f'n 'RP_Educ High school diploma, hig'n 'RP_Educ Less than high school di'n 'RP_Educ Masked records (territor'n 'RP_Educ University certificate o'n 'SP_AgeGrp 30 to 39
years'n 'SP_AgeGrp 40 to 54 years'n 'SP_AgeGrp 55 to 64 years'n 'SP_AgeGrp 65 to 74 years'n 'SP_AgeGrp 75 years and over'n 'SP_AgeGrp Less than 30 years'n 'SP_AgeGrp No spouse'n 'SP_Sex
Female'n 'SP_Sex Male'n 'SP_Sex No spouse'n 'SP_Educ Certificate or diploma f'n 'SP_Educ High school diploma, hig'n 'SP_Educ Less than high school di'n 'SP_Educ Masked records (territor'n
'SP_Educ No spouse'n 'SP_Educ University certificate o'n 'SP_EmpStat Did not work'n 'SP_EmpStat No spouse'n 'SP_EmpStat Worked full-time'n 'SP_EmpStat Worked part-time'n HH_TotInc);
FREQ WeightD;
run;

*AS EXPECTED LOG IS THE RECCOMENDATION;

DATA ANA.MODEL2;
SET ANA.MODEL1;
LOG_FD806 = LOG(FD806);
DROP FD806;
RUN;


*EDA;
%LET PATH=D:\1_Metro College\Courses\DSP-SAS\Project\PROFILING\MODEL_DATASET_AFTER_TRANSF;
*UNIVARIATE ANALYSIS;
%EDA(ANA.MODEL2,WEIGHT=WeightD,ALPHA=0.05,OUTPUT_FILE_PATH=&PATH.);

*BIVARIATE ANALYSIS;
%BIVAR_CAT_CONT(ANA.MODEL2,LOG_FD806,WEIGHT=WeightD,OUTPUT_FILE_PATH=&PATH.);
```

```
*MODELLING HAVING AS SELECTION PARAMETER FOR STEPWISE
METHOD THE SIGNIFICANCE OF FEATURES AS 0.05 FOR BOTH ENTRY
AND STAY FOR EXPLAINING PURPOSES;
PROC GLMSELECT DATA = ANA.MODEL2  PLOTS=ALL;
CLASS Region Prov HHType6 RP_AgeGrp RP_Sex
RP_MarStat RP_Educ SP_AgeGrp SP_Sex SP_Educ SP_EmpStat/PARAM=REFERENCE REF=FIRST ORDER=FREQ;
FREQ WeightD;
 MODEL LOG_FD806 =  Region Prov HHType6 RP_AgeGrp RP_Sex
RP_MarStat RP_Educ SP_AgeGrp SP_Sex SP_Educ
SP_EmpStat/SELECTION = STEPWISE DETAILS = STEPS
                              SELECT =SL SLSTAY=0.05 SLENTRY=0.05;
RUN;
QUIT;
TITLE;

*PLOTTING PARAMETERS;
proc print data=work.robparamest;

PROC SORT DATA=WORK.ROBPARAMEST;
BY Estimate;
RUN;

ODS GRAPHICS / HEIGHT=8IN WIDTH=15IN;
proc sgplot data=RobParamEst noautolegend;
TITLE "PARAMETERS";
   where (Parameter not in ("Intercept" "Scale")) AND (Step EQ 10) AND Estimate NE 0;
   vbarparm category=Parameter response=Estimate /
DISCRETEOFFSET=0.4 DATALABEL = ESTIMATE
GROUPORDER=ascending BARWIDTH=0.7
FILLATTRS =(COLOR = PLUM) DATASKIN=PRESSED BASELINE=0 ;
XAXIS DISPLAY=ALL;
RUN;
```