The MEANS Procedure

					Lower			Upper		Quartile	Coeff of	Lower 95%	Upper 95%	
Region	N Obs	N	N Miss	Minimum	Quartile	Median	Mean	Quartile	Maximum	Range	Variation	CL for Mean	CL for Mean	Skewness
Atlantic provinces	629768	629768	0	3.07	4.97	5.74	5.62	6.35	7.35	1.38	16.90	5.61	5.62	-0.41
Central	4987062	4987062	0	2.64	4.62	5.27	5.27	5.95	7.29	1.33	17.70	5.27	5.28	-0.07
Prairie	1530428	1530428	0	2.47	4.88	5.56	5.56	6.36	7.34	1.48	16.94	5.56	5.57	-0.28
Territorial capitals	11711	11711	0	3.57	4.99	5.76	5.71	6.45	7.36	1.47	16.60	5.70	5.73	-0.28
West	969907	969907	0	3.22	4.97	5.55	5.49	6.10	7.36	1.14	16.17	5.49	5.49	-0.29

RELATION BETWEEN LOG_FD806 AND Region
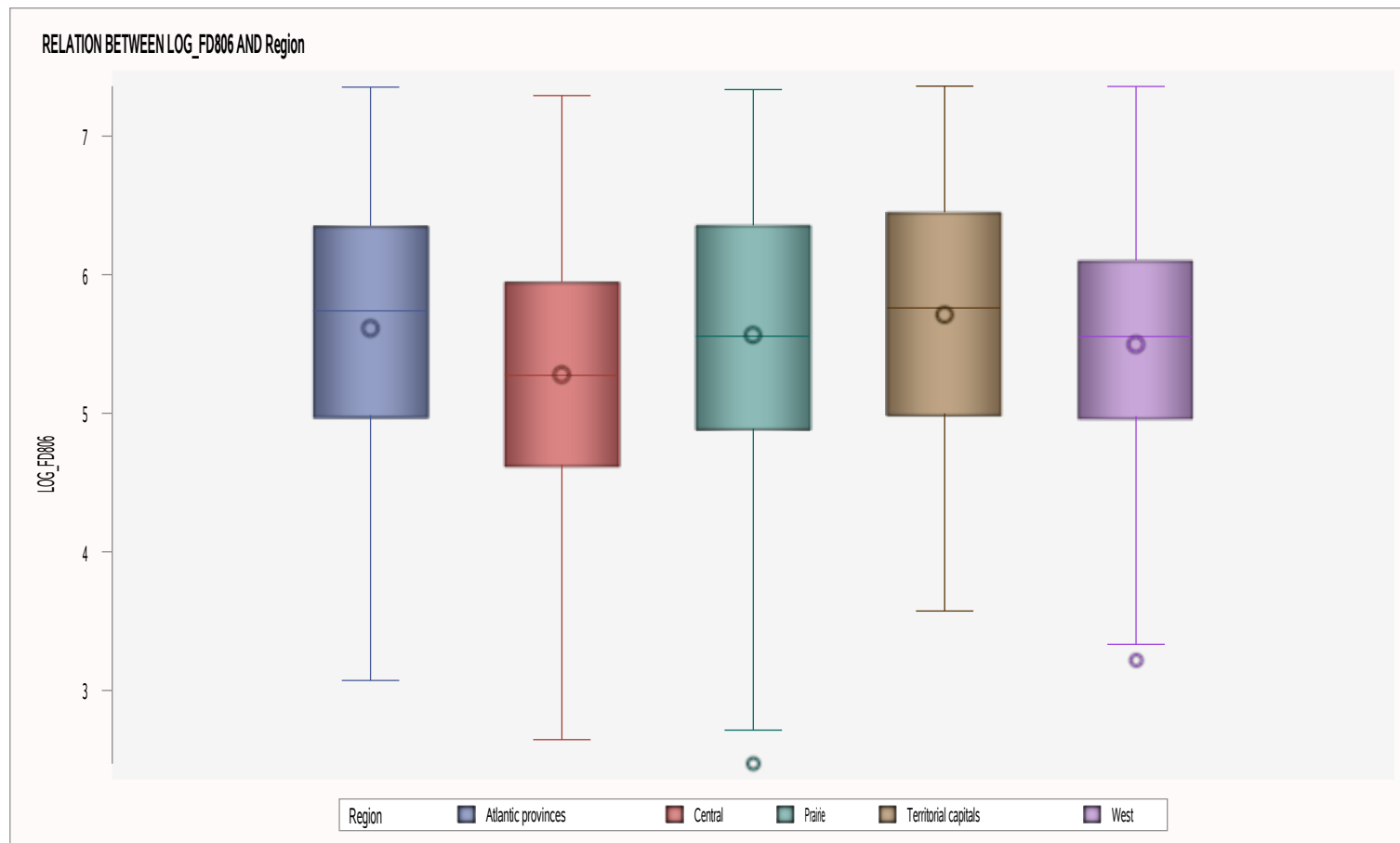
One-way ANOVA Assumptions
In order to run a one-way ANOVA the following assumptions must be met:
1.The response of interest is continuous and normally distributed for each treatment group:
Normality test: PROC UNIVARIATE NORMAL and QQPlot for each group.
2.Treatment groups are independent of one another. Experimental units only receive one treatment,and they do not overlap.
3.There are no major outliers.
4.A check for unequal variances will help determine which version of a one-way ANOVA is most appropriate
(Levene's test, Null hypothesis: variances are equal between groups):
A .If variances are equal, then the assumptions of a standard one-way ANOVA are met.
B. If variances are unequal, then a Welch's one-way ANOVA is appropriate.

Normal Distribution?
Null hypothesis: sample has a normal distribution
CLT :
a.If it looks normal and each group have more than 30 observations
b.If moderately skewed, each group must have more than 100 observations
*rule of thumb: If skewness is between -1 and -0.5 or between 0.5 and 1, the distribution is moderately skewed.
*if the sample size is over 2000, the Kolmgorov test should be used. If the sample size is less than 2000, the Shapiro test is better.

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Region=Atlantic provinces

Moments			
N	629768	Sum Weights	629768
Mean	5.61588116	Sum Observations	3536702.25
Std Deviation	0.94909276	Variance	0.90077707
Skewness	-0.4111892	Kurtosis	-0.3415584

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Region=Atlantic provinces

Moments			
Uncorrected SS	20428979.2	Corrected SS	567279.674
Coeff Variation	16.9001575	Std Error Mean	0.00119596

Basic Statistical Measures			
Location		Variability	
Mean	5.615881	Std Deviation	0.94909
Median	5.738828	Variance	0.90078
Mode	4.333099	Range	4.28184
		Interquartile Range	1.37771

Tests for Location: Mu0=0				
Test		Statistic	p Value	
Student's t	t	4695.691	Pr > |t|	<.0001
Sign	M	314884	Pr >= |M|	<.0001
Signed Rank	S	9.915E10	Pr >= |S|	<.0001

Tests for Normality				
Test		Statistic	p Value	
Kolmogorov-Smirnov	D	0.052851	Pr > D	<0.0100
Cramer-von Mises	W-Sq	365.0699	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	2531.387	Pr > A-Sq	<0.0050

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Region=Atlantic provinces

Quantiles (Definition 5)	
Level	Quantile
100% Max	7.35361
99%	7.32214
95%	7.03277
90%	6.73863
75% Q3	6.35141
50% Median	5.73883
25% Q1	4.97369
10%	4.32665
5%	3.89182
1%	3.25810
0% Min	3.07177

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
3.07177	840	331	7.34003	776	214
3.07177	347	246	7.34357	669	19
3.15274	132	187	7.35014	1360	637
3.21165	209	688	7.35109	759	717
3.21727	186	527	7.35361	491	225

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Region=Central

Moments			
N	4987062	Sum Weights	4987062
Mean	5.27439207	Sum Observations	26303720.3
Std Deviation	0.93379405	Variance	0.87197133
Skewness	-0.0677122	Kurtosis	-0.5844914
Uncorrected SS	143084708	Corrected SS	4348574.21
Coeff Variation	17.7042973	Std Error Mean	0.00041815

Basic Statistical Measures			
Location		Variability	
Mean	5.274392	Std Deviation	0.93379
Median	5.273820	Variance	0.87197
Mode	4.276944	Range	4.64785
		Interquartile Range	1.33043

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	12613.73	Pr > |t|	<.0001
Sign	M	2493531	Pr >= |M|	<.0001
Signed Rank	S	6.218E12	Pr >= |S|	<.0001

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Region=Central

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	D	0.039316	Pr > D	<0.0100
Cramer-von Mises	W-Sq	1237.889	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	9225.814	Pr > A-Sq	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	7.29261
99%	7.17749
95%	6.80870
90%	6.51514
75% Q3	5.94694
50% Median	5.27382
25% Q1	4.61651
10%	3.99046
5%	3.87287
1%	3.27790
0% Min	2.64476

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Region=Central

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
2.64476	15313	1259	7.23335	5785	1060
3.02237	4062	1240	7.24642	6213	1142
3.03495	8060	1283	7.25157	2957	887
3.05777	7665	1297	7.28909	12702	903
3.27790	16121	1013	7.29261	12196	793

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Region=Prairie

Moments			
N	1530428	Sum Weights	1530428
Mean	5.56479741	Sum Observations	8516521.77
Std Deviation	0.94258654	Variance	0.88846939
Skewness	-0.2784877	Kurtosis	-0.4230712
Uncorrected SS	48752455.8	Corrected SS	1359737.54
Coeff Variation	16.9383802	Std Error Mean	0.00076193

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Region=Prairie

Basic Statistical Measures			
Location		**Variability**	
Mean	5.564797	Std Deviation	0.94259
Median	5.555901	Variance	0.88847
Mode	6.439478	Range	4.86516
		Interquartile Range	1.47808

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	t	7303.56	Pr > |t|	<.0001
Sign	M	765214	Pr >= |M|	<.0001
Signed Rank	S	5.856E11	Pr >= |S|	<.0001

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	D	0.053018	Pr > D	<0.0100
Cramer-von Mises	W-Sq	637.2333	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	4755.851	Pr > A-Sq	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	7.33580
99%	7.30953
95%	7.06397
90%	6.73285

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Region=Prairie

Quantiles (Definition 5)	
Level	**Quantile**
75% Q3	6.35580
50% Median	5.55590
25% Q1	4.87771
10%	4.32968
5%	4.01241
1%	3.38608
0% Min	2.47064

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
2.47064	3940	1855	7.27891	505	1585
2.71337	1520	1916	7.30361	5912	1353
3.08374	1372	1344	7.30953	8411	1436
3.08374	3250	1318	7.32095	5649	1693
3.10727	381	1617	7.33580	1566	1781

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Region=Territorial capitals

Moments			
N	11711	Sum Weights	11711
Mean	5.71392282	Sum Observations	66915.7502
Std Deviation	0.94859878	Variance	0.89983964
Skewness	-0.2810241	Kurtosis	-0.9449775
Uncorrected SS	392888.554	Corrected SS	10537.1222
Coeff Variation	16.6015329	Std Error Mean	0.00876568

Basic Statistical Measures			
Location		Variability	
Mean	5.713923	Std Deviation	0.94860
Median	5.759879	Variance	0.89984
Mode	4.574092	Range	3.78783
		Interquartile Range	1.46558

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	651.8517	Pr > |t|	<.0001
Sign	M	5855.5	Pr >= |M|	<.0001
Signed Rank	S	34289808	Pr >= |S|	<.0001

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Region=Territorial capitals

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	D	0.082104	Pr > D	<0.0100
Cramer-von Mises	W-Sq	16.42393	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	115.1312	Pr > A-Sq	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	7.36074
99%	7.30971
95%	7.08256
90%	6.93674
75% Q3	6.45063
50% Median	5.75988
25% Q1	4.98504
10%	4.25135
5%	4.13772
1%	3.78328
0% Min	3.57291

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Region=Territorial capitals

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
3.57291	18	2077	7.27783	46	2131
3.78328	115	2058	7.28773	45	2097
3.85757	157	2064	7.30971	58	2104
3.91562	58	2130	7.32818	59	1960
3.94119	33	2065	7.36074	43	2084

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Region=West

Moments			
N	969907	Sum Weights	969907
Mean	5.49258536	Sum Observations	5327296.99
Std Deviation	0.88839624	Variance	0.78924789
Skewness	-0.2858383	Kurtosis	-0.3704013
Uncorrected SS	30026129.7	Corrected SS	765496.26
Coeff Variation	16.174464	Std Error Mean	0.00090207

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Region=West

Basic Statistical Measures			
Location		**Variability**	
Mean	5.492585	Std Deviation	0.88840
Median	5.553541	Variance	0.78925
Mode	6.246301	Range	4.14116
		Interquartile Range	1.13596

Tests for Location: Mu0=0				
Test		**Statistic**	**p Value**	
Student's t	t	6088.848	Pr > |t|	<.0001
Sign	M	484953.5	Pr >= |M|	<.0001
Signed Rank	S	2.352E11	Pr >= |S|	<.0001

Tests for Normality				
Test		**Statistic**	**p Value**	
Kolmogorov-Smirnov	D	0.051237	Pr > D	<0.0100
Cramer-von Mises	W-Sq	491.1901	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	3618.11	Pr > A-Sq	<0.0050

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	7.35843
99%	7.20510
95%	7.04319
90%	6.60932

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Region=West

Quantiles (Definition 5)	
Level	**Quantile**
75% Q3	6.10216
50% Median	5.55354
25% Q1	4.96620
10%	4.14804
5%	3.87867
1%	3.36246
0% Min	3.21727

Extreme Observations					
Lowest			**Highest**		
Value	**Freq**	**Obs**	**Value**	**Freq**	**Obs**
3.21727	1553	2301	7.10462	9539	2207
3.33220	6907	2300	7.19090	3791	2255
3.36246	3183	2272	7.20510	5467	2235
3.44042	4303	2210	7.21127	1683	2307
3.63627	3779	2260	7.35843	4907	2168

Null hypothesis: equal variances
a.If variances are equal, then a pooled t-test is appropriate
b.If variances are unequal, then a Satterthwaite (also known as Welch's) test is appropriate

The GLM Procedure

Class Level Information		
Class	**Levels**	**Values**
Region	5	Atlantic provinces Central Prairie Territorial capitals West

Number of Observations Read	2327
Number of Observations Used	2327
Sum of Frequencies Read	8128876
Sum of Frequencies Used	8128876

The GLM Procedure

Dependent Variable: LOG_FD806

Frequency: WeightD

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	156486.214	39121.553	45098.0	<.0001
Error	8.13E6	7051624.800	0.867		
Corrected Total	8.13E6	7208111.014			

R-Square	Coeff Var	Root MSE	LOG_FD806 Mean
0.021710	17.30495	0.931386	5.382190

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Region	4	156486.2138	39121.5534	45098.0	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Region	4	156486.2138	39121.5534	45098.0	<.0001

The GLM Procedure

Levene's Test for Homogeneity of LOG_FD806 Variance ANOVA of Absolute Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Region	4	3169.5	792.4	2785.20	<.0001
Error	8.13E6	2312645	0.2845		

Welch's ANOVA for LOG_FD806			
Source	DF	F Value	Pr > F
Region	4.0000	44621.4	<.0001
Error	91152.3		

The GLM Procedure

Level of Region	N	LOG_FD806	
		Mean	Std Dev
Atlantic provinces	629768	5.61588116	0.94909276
Central	4987062	5.27439207	0.93379405
Prairie	1530428	5.56479741	0.94258654
Territorial capitals	11711	5.71392282	0.94859878
West	969907	5.49258536	0.88839624

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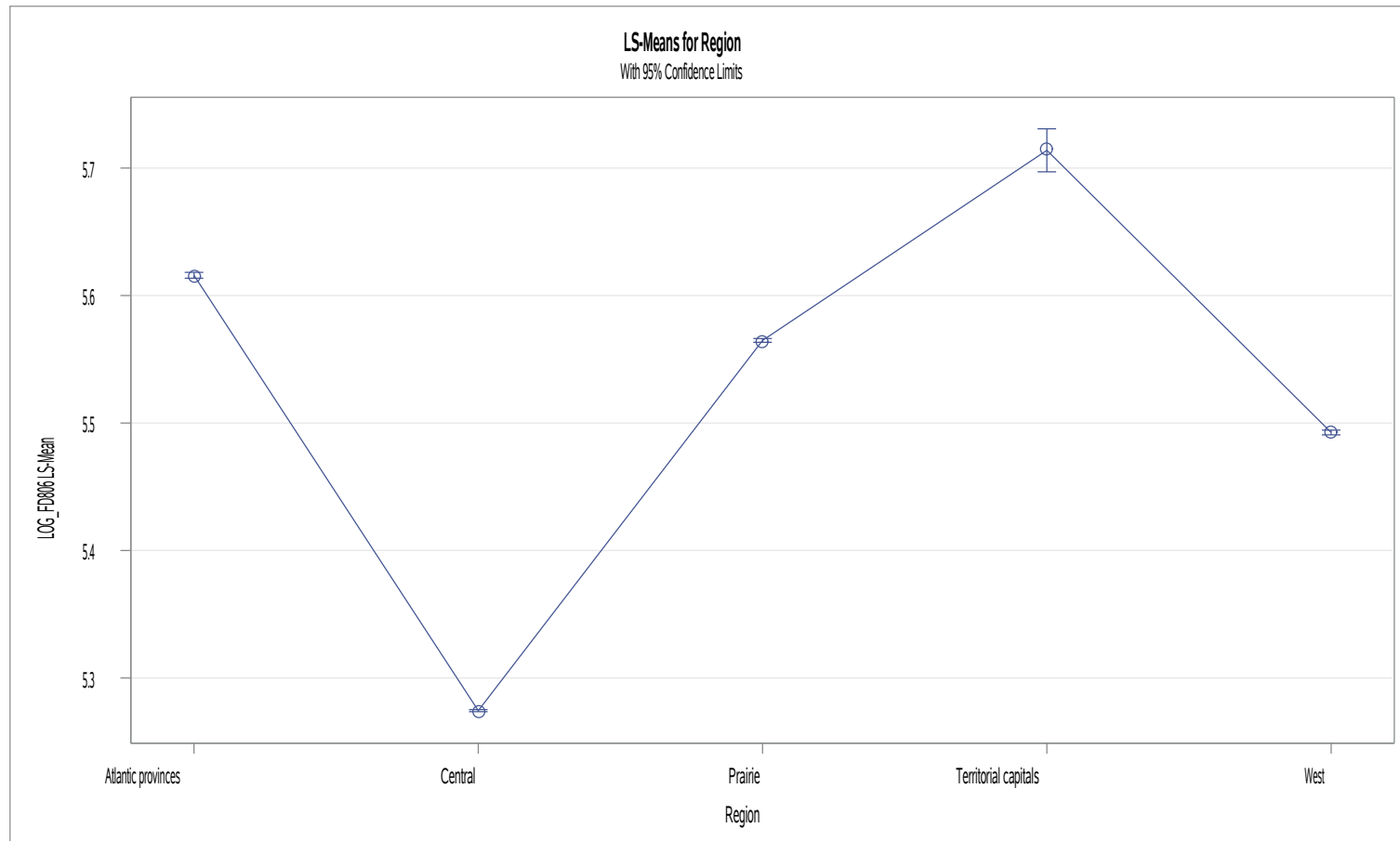## Adjustment for Multiple Comparisons: Tukey-Kramer

Region	LOG_FD806 LSMEAN	LSMEAN Number
Atlantic provinces	5.61588116	1
Central	5.27439207	2
Prairie	5.56479741	3

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

Region	LOG_FD806 LSMEAN	LSMEAN Number
Territorial capitals	5.71392282	4
West	5.49258536	5

Least Squares Means for effect Region Pr > |t| for H0: LSMean(i)=LSMean(j) Dependent Variable: LOG_FD806					
i/j	**1**	**2**	**3**	**4**	**5**
1		<.0001	<.0001	<.0001	<.0001
2	<.0001		<.0001	<.0001	<.0001
3	<.0001	<.0001		<.0001	<.0001
4	<.0001	<.0001	<.0001		<.0001
5	<.0001	<.0001	<.0001	<.0001	

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

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

LOG_FD806 Tukey-Kramer Grouping for LS-Means of Region (Alpha = 0.05)

LS-means covered by the same bar are not significantly different.

Region	Estimate
Territorial capitals	5.7139
Atlantic provinces	5.6159
Prairie	5.5648
West	5.4926
Central	5.2744