

Technical Supplement
to
Factors in Canadians' Cultural Activities

**Demographics and Cultural Crossovers Involved
in Book Reading, Performing Arts Attendance,
Art Gallery Attendance and Movie-Going**



<http://www.hillstrategies.com>

Statistical Insights on the Arts, Vol. 6 No. 3
© Hill Strategies Research Inc., February 2008

ABSTRACT

Data drawn from Statistic Canada's Social Survey of 2005 is used to determine which factors, demographic and otherwise, can predict various cultural activities. Canadians 15 years of age or older were asked about their participation in various cultural activities during the 12 months prior to the survey. The specific cultural activities examined in this report are: reading a book, attending a live performance, visiting an art gallery and watching a movie. Logistic regression analysis is utilized to determine the odds ratio for each of the independent variables included in the study. Variable selection, variable coding and model validity are discussed in this technical supplement.

INTRODUCTION

Binary logistic regression is used to analyze the data from the 2005 General Social Survey in order to identify the best predictors of four cultural activities: book reading, performing arts attendance, art gallery attendance and movie-going.

Preliminary exploration is done to discover trends in the dataset for each of the four cultural activities under investigation. The four cultural activities are analyzed separately. A different model is constructed where each of the four cultural activities is represented as the dependent variable. Frequency distributions are constructed to examine the percentage of respondents who participated (or not) in the dependent variable based on their demographic characteristics and other cultural activities (denoted as independent variables). Using frequency distributions, the percentage of respondents in each independent variable who participated in the cultural activity under investigation is calculated (percentages for those who did not participate are also calculated). The results are evaluated and a subset is constructed of the independent variables most likely to predict if respondents participated in each of the four cultural activities.

Logistic regression is used to predict the probability of an event occurring due to the occurrence of other events. More specifically, binary logistic regression is used when the single dependent variable (the event that is being predicted) is dichotomous and the independent variables (other events that have already occurred) are of any type (binary, continuous, categorical, etc.). This method is used in order to categorize predictor variables based on relative significance, to gauge multicollinearity of the independent variables and to understand the effect of the predictor variables on the dependent variable.

Binary (or binomial) logistic regression is a form of analysis where logistic regression equations are solved iteratively, a contrast from other forms of regression analysis where a mathematical equation is solved explicitly. In binary logistic regression a trial equation is fitted and adjusted in order to improve the fit:

$$\text{logit}[p] = \log [p/1-p] = \alpha + \beta_1x_1 + \beta_2x_2 + \dots + \beta_ix_i \quad [\text{Equation 1}]$$

where p is the probability of the event occurring, α is the constant of the equation and β_i are the coefficients of the predictor values. The user determines when the iterations stop. That is, either an explicit number of iterations are performed or a cut-value is selected so that the iterations stop when the improvement from step to the next is less than the cut-value.

Once the equation is fitted one can determine the odds ratio. The odds ratio is a prediction about how likely the respondent is to have participated in the activity in question. This is calculated by raising the base of the natural logarithm to the β th power, where β is the coefficient of the predictor variable from the trial equation. That is:

$$\text{ODDS} = e^{\beta} = p/1-p \quad [\text{Equation 2}]$$

where β is the slope of the logistic regression equation for a specific predictor variable. Once the odds of an event occurring have been calculated, they can be converted to probabilities:

$$p = \text{ODDS} / (1 + \text{ODDS})$$

[Equation 3]

Probabilities can be more useful when describing the effects of the predictor variables on the dependent variable.

Predictive analytic software, SPSS, is used to perform the iterations necessary to derive the logistic regression equation. Once the equation is fitted, the output presents the user with the coefficients of the predictor variables and also includes a 95% confidence interval used for error analysis. SPSS also has intrinsic tests to examine the validity of the created models. These tests are discussed more thoroughly in a subsequent section.

METHODS

Variable Selection

The goal of this study is to examine predictors of various cultural activities. The predictor variables can be divided into two categories: demographic variables and cultural variables. Preliminary exploration of the dataset was done to examine the percentages of respondents having read a book (for example), given their demographic characteristics and cultural activities. The independent variables with the highest and lowest percentages are considered as potential primary predictors. Careful consideration is given to the selection of the variables to be included in the analysis. All of the cultural variables that are included in previous Hill Strategies publications are also included in this study. Many of the demographic variables are chosen based on previous research in Canada and elsewhere that attribute a certain level of importance to each variable. Other demographic variables are included that could contribute to someone participating in the arts.

Thirty-one variables are selected in total plus an additional eight reconfigurations of some demographic variables. Of the 31 chosen variables, 14 are demographic and 17 are cultural variables. Brief explanations for each variable are provided in Appendix 2.

Once the base set of variables to be included in the study is constructed, exploratory tests are done to discover which predictor variables have large influences on each of the dependent variables. This is done using a forward stepwise conditional logistic regression model for three different configurations: only demographic variables, only cultural variables and all variables. Further investigation is done with each of the variables to determine, purely from base percentages, which variables are the best indicators of each of the four dependent variables.

A model is constructed for each dependent variable using the enter procedure for a logistic regression. A list is compiled of variables that appear to have an influence based on preliminary research and variables that could be taken out of the model while keeping the same or a higher level of predictive accuracy.

Three main statistical tests are used to test the validity of the models: -2 log likelihood, the Hosmer and Lemeshow test and tests for possible multicollinearity. These methods will be discussed in the section “Assessment of Study Quality”.

The variables that the preliminary research indicates may be of primary importance are included in the initial drafts of the models, while the remaining variables are added and removed in order to improve the accuracy of the model. This is done without preference for any of the variables. For example, if adding a variable does not cause the -2 log likelihood to change significantly, then the variable in question does not have much importance in predicting the dependent variable and can be left out of the model. These steps of removing and adding different combinations of variables is done until the -2 log likelihood is as small as possible, while the Hosmer and Lemeshow test show that the data fit the model well and the tests for multicollinearity do not show multicollinearity of the predictor variables.

Variable Coding

The 31 base variables and eight additional reconfigurations are included in the model as independent variables with the internal value “0” being set as the reference level. More specifically, each variable has internal numerical coding that reflects a specific answer. For example, variables that have two possible elements in the answer set are labeled dichotomous. These are fairly straightforward to analyze, as their coding would result in one element being the reference level and the other being compared to it. For example, if the respondent answered “no” to question LCA_Q360, then they did not attend any other kind/type of cultural performance during the past 12 months. This result is classified as the reference variable with coding “0”. The only other available response would be that the respondent had attended any other kind/type of cultural performance during the past 12 months (coded “1”).

For most of the dichotomous variables, “no” was coded “0” and “yes” was coded “1”. As a consequence, the results show that there is a direct effect on the dependent variable. For example, if the odds given in the output is 1.414 (Table 4: Results of the movie attendance model) then the implication is that someone who has attended any other kind/type of cultural performance during the past 12 months is 1.414 times as likely to have also gone to a movie or drive-in during the past 12 months.

Many of the questions asked by the surveyors include the possible responses: “don’t know”, “not asked” and “not stated”. For example, the question regarding the country of birth of the respondent has four possible responses: Canada, country outside of Canada, not stated and don’t know. Noticeably the responses “not stated” and “don’t know” would not be useful in predicting if the respondent has read a book (for example). For this reason the data was not used in the analysis and the independent variable has been altered to become a dichotomous variable. This variable was not coded in the manner explained previously for dichotomous variables as the two remaining responses are not “yes” and “no”. Instead the variable is coded “0” if Canada was the country of birth of the respondent and “1” for a country outside of Canada.

Another example of this is the question regarding the household language of the respondent. The respondent's possible answers are English only, French only, other language, not stated and don't know. In this case, when the two responses "not stated" and "don't know" are removed, a categorical variable remains. Coding of the variables in the categorical case is more complicated than for dichotomous variables. The question regarding the household language of the respondent is coded in the following manner: English only = 0, French only = 1, other language = 2. In this case, English represents the reference point, and each of the other two possible responses would display an influence on the dependent variable relative to the reference point.

Determining the importance of the elements of a categorical variable is not as simple as the dichotomous case. Looking at the SPSS output in Table 1 for the dependent variable "live performing arts", there are three possible responses for a household language of the respondent. With English as the only language being the reference element, the two other scenarios display odds relative to it. Respondents with French as their only household language are 1.332 times as likely to have attended a theatrical performance as respondents speaking English in their household (from Table 1: Exp(B) column). Respondents with a household language other than English or French are 0.629 times as likely to have attended a theatrical performance in the last 12 months as respondents who speak English in their household (from Table 1: Exp(B) column).

But what if a comparison between those who speak French in their household and those who speak a language other than French or English in their household is necessary? Since the ratios are: 1:1.332 and 1:0.629 for French speaking households and households speaking a language other than English or French, respectively, one can compare the two directly; 1.332:0.629. To help with the ease of comparison one could divide each side by one of the two values (the value chosen being the new reference variable). Dividing each side by 0.629 gives that respondents speaking French in their household are 2.18 times more likely to have gone to a theatrical performance during the past 12 months than respondents speaking a language other than French or English in their household.

RESULTS

Assessment of Study Quality

The validity of the models was determined using three fundamental tests: the likelihood ratio test, the Hosmer and Lemeshow test and a test for multicollinearity.

The log likelihood procedure is used to determine the difference between a model and a nested model, which is a subset of the original model. If there is no significant difference between the log likelihood of the model and the nested model, then the variable is dropped to create a more concise model. This is used to determine which variables are necessary to improve the prediction of the dependent variable.

The Hosmer and Lemeshow test is used to determine if the model's approximations fit the data adequately. If the Hosmer and Lemeshow test statistic is not significant, that is, greater than .05, we satisfy the hypothesis that the actual and model predicted values are analogous. As a result,

for the model to be good fitting, non-significant results on the Hosmer and Lemeshow test are required. This test is only an interim step in ensuring validity of the overall model.

Another problem that a model could have is multicollinearity. This occurs when there are independent variables in the model that have a strong correlation. The degree of the correlation is described by a correlation coefficient, a value between -1 and 1 that measures the degree of relationship between two values: 1 being completely related and -1 being completely oppositely related. A problem occurs when two or more variables have high correlation coefficients. In this case, the variance in the dependent variable due to some of the independents will not be legitimate. Consequently, as each of the models is analyzed, a matrix of correlation coefficients is constructed and each entry analyzed. If one element is greater than 0.90 or if several are greater than 0.70, multicollinearity could be a problem and the model should be discarded. None of the independent variables in our models have strong multicollinearity.

Exp(B) is the odds ratio for each independent variable and is a positive real number. If exp(B) is less than one, then the independent variable predicts a lower likelihood of participating in the cultural event than those in other in the other reference group. Whereas, if exp(B) is greater than one, it predicts a higher likelihood of participating. A useful error indicator of the value exp(B) is the 95% confidence interval. The 95% confidence interval is a range around the value exp(B) in the SPSS output. It provides a scope of uncertainty for the odds ratio for each independent variable. That is, it accounts for various types of error (sample, non-sampling) that could have an effect on the results. The 95% confidence interval helps to determine how useful each independent variable is as a predictor of the dependent variable. A problem arises when the 95% confidence interval includes both predictions: an exp(B) with both a value less than one and greater than one. If this occurs, the independent variable is considered to not be a useful predictor of the dependent variable.

Data Interpretation

The meaningful results obtained from the SPSS output are provided in Appendix 1 for each of the four models (book reading, art gallery attendance, performing arts attendance and movie theatre attendance). In each of the tables, the value in the column labeled “Exp(B)” represents the odds ratio for each of the independent variables. For dichotomous variables the odds ratio (exp(B)) indicates that someone who responded “yes” to the given question is exp(B) times as likely to participate in the cultural event than someone who responded “no” to the same question. For example, in Table 1, (Results of the live performance model) the “ACTLIMIT” variable is dichotomous. The reference level for this variable is the respondent not having any activity limitations. Using this methodology, the exp(B) value located in the table, 0.769, is the odds ratio identifying that someone who has activity limitations is 0.769 times as likely as someone with no limitations to attend a live performance.

For categorical variables, if the respondent fits into one of the classes of the variable, then they are exp(B) times as likely to participate compared to someone in the reference class. The reference class is the first option listed and always has an empty entry for exp(B). For instance, the variable “AGEGR10” in Table 1 has 7 classes. The first variable listed, 15 to 24, is the

reference level. The results are read as follows: the $\exp(B)$ for someone who is 75 or older is 1.49. Therefore, someone who is 75 or older is 1.49 times as likely to attend a live performance as someone who is 15 to 24.

Further, for each of the classes that has a value for $\exp(B)$, there are columns entitled “Lower” and “Upper”. These represent the 95% confidence interval for the odds ratio, more specifically the error in the odds ratio. If the 95% confidence interval contains the value “1”, then it contains both a prediction of “yes” and “no” with regards to the dependent variable. This corresponds to a value larger than .05 in the “sig.” column indicating that the independent variable is not an effective predictor in the model.

All dichotomous variables that are not considered significant predictors are excluded from the models. Care must be used when examining the results for the categorical variables since not all classes are necessarily significant. Categorical variables with at least one class that is useful are included in the model. Looking at the results for the categorical variable AGEGR10, there are several groupings that have values larger than .05 in the “sig.” column. The classes are: 25 to 34, 35 to 44 and 45 to 54. Therefore, the useful predictors in the model are the age groupings 15 to 24 and all groups 55 and over.

APPENDIX 1: SPSS OUTPUT

Table 1: Results of the live performing arts model (SPSS output for binary logistic regression model)

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I. for EXP(B)	
								Lower	Upper
ACTLIMIT- "respondent is limited in the amount of physical activity that he/she can do" (Reference is not being limited)		-0.263	0.087	9.1	1	0.003	0.769	0.648	0.912
AGEGR10	15 to 24			18.984	6	0.004			
	25 to 34	0.056	0.115	0.235	1	0.628	1.057	0.844	1.325
	35 to 44	-0.014	0.118	0.014	1	0.906	0.986	0.783	1.242
	45 to 54	0.094	0.119	0.617	1	0.432	1.098	0.869	1.388
	55 to 64		0.129	6.104	1	0.013	1.376	1.068	1.773
	65 to 74	0.439	0.15	8.622	1	0.003	1.552	1.157	2.081
	75 or older	0.398	0.178	5.022	1	0.025	1.49	1.051	2.111
EDU5	College diploma or some college			18.073	2	0			
	High school or less	-0.246	0.07	12.42	1	0	0.782	0.682	0.896
	Degree	0.091	0.073	1.556	1	0.212	1.095	0.95	1.262
INCMHSD*	Less than \$40,000			58.865	2	0			
	\$40,000 to \$79,999	0.255	0.072	12.51	1	0	1.29	1.12	1.486
	\$80,000 or more	0.648	0.085	57.678	1	0	1.912	1.618	2.26
REGION	Ontario			9.914	4	0.042			
	Atlantic region	-0.146	0.085	2.959	1	0.085	0.864	0.731	1.021
	Quebec	0.01	0.148	0.005	1	0.946	1.01	0.756	1.349
	Prairie region	-0.243	0.082	8.83	1	0.003	0.784	0.668	0.921
	British Columbia	-0.076	0.098	0.605	1	0.437	0.927	0.765	1.123
SEX- "sex of respondent" (Reference is female)		-0.278	0.06	21.852	1	0	0.757	0.674	0.851
LCA_Q130- "did you read a book?" (Reference is not reading)		0.187	0.067	7.714	1	0.005	1.205	1.056	1.375
LCA_Q210- "did you go to a movie or drive-in?" (Reference is not going)		0.801	0.066	147.402	1	0	2.227	1.957	2.535

LCA_Q240- "did you listen to music on CD, etc?" (Reference is not listening)	0.668	0.09	54.606	1	0	1.95	1.633	2.327
LCA_Q340- "did you go to a cultural or artistic festival?" (Reference is not going)	0.813	0.069	140.475	1	0	2.255	1.971	2.579
LCA_Q410- "did you go to a public art gallery or art museum?" (Reference is not going)	0.604	0.069	76.304	1	0	1.83	1.598	2.096
CHRFLAG- "single child(ren) of the respondent living in the household" (Reference is having child(ren) in the household)	0.312	0.073	18.09	1	0	1.366	1.183	1.577
MARSTAT*								
Married or common law			14.931	2	0.001			
Divorced, separated or widowed	0.176	0.085	4.253	1	0.039	1.192	1.009	1.409
Single (never married)	0.316	0.086	13.579	1	0	1.372	1.16	1.623
LCA_Q360- "did you attend any other kind/type of cultural performance?" (Reference is not attending)	0.586	0.094	39.261	1	0	1.798	1.496	2.16
LCA_Q420- "did you visit a museum other than public art galleries or art museums?" (Reference is not visiting)	0.219	0.074	8.85	1	0.003	1.245	1.078	1.438
LCA_Q450- "did you go to a conservation area or nature park?" (Reference is not going)	0.249	0.062	16.045	1	0	1.282	1.135	1.448
LCA_Q440- "did you go to a zoo, aquarium, botanical gardens, planetarium or observatory?" (Reference is not going)	0.209	0.063	10.942	1	0.001	1.232	1.089	1.395
LCA_Q120- "did you read a magazine?" (Reference is not reading)	0.369	0.077	23.213	1	0	1.447	1.245	1.681
LCA_Q430- "did you go to an historic site?" (Reference is not going)	0.342	0.065	27.854	1	0	1.408	1.24	1.599
LANHSDC								
English only			22.326	2	0			
French only	0.287	0.145	3.94	1	0.047	1.332	1.004	1.769
Other language	-0.463	0.124	13.899	1	0	0.629	0.493	0.803
Constant	-3.103	0.188	273.699	1	0	0.045		

Table 2: Results of the book reading model (SPSS output for binary logistic regression model)

		B	S.E.	Wald	Df	Sig.	Exp(B)	95.0% C.I. for EXP(B)	
								Lower	Upper
EDU5	College diploma or some college			149.691	2	0			
	High school or less	-0.316	.055	32.539	1	0	.729	0.654	0.813
	Degree	0.648	.07	71.619	1	0	1.912	1.646	2.222
LANHSDC	English only			6.490	2	.04			
	French only	-0.107	.073	2.173	1	.14	.898	0.779	1.036
	Other language	-0.218	.096	5.118	1	.02	.804	0.666	0.971
SEX- "sex of respondent" (Reference is being female)		-0.803	.051	246.002	1	0	.448	0.405	0.495
RLR_Q105*- "how important are your spiritual beliefs?" (Reference is very important spiritual beliefs)		-0.321	.054	35.253	1	0	.725	0.652	0.806
CHRFLAG- "single child(ren) of the respondent living in the household" (Reference is having child(ren) in the household)		0.255	.059	18.830	1	0	1.290	1.15	1.447
RELIG6	No religion			18.310	4	0			
	Roman Catholic	-0.069	.074	0.883	1	0.35	.933	0.808	1.078
	United Church	0.072	.097	0.543	1	0.46	1.074	0.888	1.3
	Protestant	0.190	.078	5.922	1	0.02	1.210	1.038	1.41
	Other	-0.194	.124	2.450	1	.12	0.824	0.646	1.05
MARSTAT*	Married or common law			16.615	2	0			
	Divorced, separated or widowed	0.199	.068	8.475	1	0	1.220	1.067	1.394
	Single (never married)	0.237	.066	12.913	1	0	1.268	1.114	1.442
LCA_Q110- "did you read a newspaper?" (Reference is not reading)		0.466	.072	41.831	1	0	1.593	1.384	1.384
LCA_Q120- "did you read a magazine?" (Reference is not reading)		0.613	.059	107.81	1	0	1.847	1.645	1.645
LCA_Q210- "did you go to a movie or drive-in?" (Reference is not going)		0.246	.055	19.935	1	0	1.279	1.148	1.148
LCA_Q240- "did you listen to music on CD, etc?" (Reference is not listening)		0.317	.064	24.197	1	0	1.373	1.21	1.21

LCA_Q300- "did you attend a concert or performance by professional artists of music, dance, etc?" (Reference is not attending)	0.241	.058	17.215	1	0	1.272	1.135	1.135
LCA_Q410- "did you go to a public art gallery or museum?" (Reference is not going)	0.630	.073	74.961	1	0	1.879	1.629	1.629
LCA_Q420- "did you visit a museum other than public art galleries or art museums?" (Reference is not visiting)	0.288	.077	14.19	1	0	1.334	1.148	1.148
LCA_Q430- "did you go to an historic site?" (Reference is not going)	0.158	.063	6.28	1	.01	1.171	1.035	1.035
LCA_Q450- "did you go to a conservation area or nature park?" (Reference is not going)	0.224	.056	16.072	1	0	1.251	1.121	1.121
LCA_Q360- did you attend any other kind/type of cultural performance?" (Reference is not attending)	0.250	.082	9.314	1	0	1.284	1.094	1.094
Constant	-0.614	.124	24.486	1	0	0.541		

Table 3: Results of the art gallery attendance model (SPSS output for binary logistic regression model)

		B	S.E.	Wald	Df	Sig.	Exp(B)	95.0% C.I. for EXP(B)	
								Lower	Upper
ACTLIMIT- "respondent is limited to the amount of physical activity that he/she can do" (Reference is not being limited)		-0.216	.103	44.06	1	0.036	0.806	0.658	0.986
AGEGR10	15 to 24			69.531	6	0			
	25 to 34	-0.325	.142	5.236	1	0.022	0.722	0.547	0.954
	35 to 44	-0.192	.143	1.798	1	0.180	0.825	0.623	1.093
	45 to 54	0.357	.143	6.234	1	0.013	1.430	1.08	1.893
	55 to 64	0.362	.154	5.501	1	0.019	1.436	1.061	1.944
	65 to 74	0.576	.176	10.688	1	0.001	1.779	1.26	2.513
	75 or older	0.721	.205	12.364	1	0	2.056	1.376	3.073
EDU5	College diploma or some college			98.775	2	0			
	High school or less	-0.367	.088	17.489	1	0	0.692	0.583	0.823
	Degree	0.585	.079	54.396	1	0	1.796	1.537	2.098
LUC_RST- "urban/rural indicator" (Reference is urban areas)		-0.338	.089	14.387	1	0	0.713	0.599	0.849
REGION	Ontario			14.504	4	0.006			
	Atlantic region	-0.120	.104	1.326	1	0.249	0.887	0.724	1.088
	Quebec	0.219	.098	5.02	1	0.025	1.245	1.028	1.508
	Prairie region	-0.128	.096	1.784	1	0.182	0.880	0.729	1.062
	British Columbia	0.113	.113	0.996	1	0.318	1.120	0.897	1.399
MARSTAT*	Married or common law			5.794	2	0.055			
	Divorced, separated or widowed	0.233	.097	5.786	1	0.016	1.263	1.044	1.527
	Single (never married)	0.057	.103	0.304	1	0.0581	1.059	0.864	1.297
RELIG6*- "religion of respondent in two categories" (Reference is not being affiliated with a religion)		-0.445	.103	18.499	1	0	0.641	0.524	0.785
CHRFLAG- "single child(ren) of the respondent living in the household" (Reference is having child(ren) in the household)		0.222	.087	6.584	1	0.010	1.249	1.054	1.48
LCA_Q440- "did you go to a zoo, aquarium, botanical gardens, planetarium or observatory?" (Reference is not going)		0.695	.069	99.967	1	0	2.003	1.748	2.295

LCA_Q350- "did you go to a performance of cultural/heritage music, theatre or dance?" (Reference is not going)	0.236	.086	7.498	1	0.006	1.266	1.069	1.498	
LCA_Q300- "did you attend a concert or performance by professional artists of music, dance, theatre or opera?" (Reference is not attending)	0.628	.071	77.194	1	0	1.873	1.628	2.155	
LCA_Q340- "did you go to a cultural or artistic festival?" (Reference is not going)	0.734	.076	93.637	1	0	2.084	1.796	2.418	
LCA_Q140- "did you use library services as a leisure activity?" (Reference is not using)	0.387	.071	29.6	1	0	1.473	1.281	1.693	
LCA_Q130- "did you read a book?" (Reference is not reading)	0.576	.086	44.855	1	0	1.780	1.503	2.107	
LCA_Q210- "did you go to a movie or drive-in?" (Reference is not going)	0.362	.081	20.087	1	0	1.436	1.226	1.682	
LCA_Q430- "did you go to an historic site?" (Reference is not going)	0.955	.069	189.831	1	0	2.600	2.269	2.978	
LCA_Q120- "did you read a magazine?" (Reference is not reading)	0.350	.097	13.03	1	0	1.419	1.173	1.716	
INCMHSD*	Less than \$40,000		2.295	2	0.318				
	\$40,000 to \$79,999	0.090	.086	1.107	1	0.293	1.095	0.925	1.296
	\$80,000 or more	0.150	.100	2.259	1	0.133	1.162	0.955	1.414
RELIGATT	At least once a week		4.535	4	0.338				
	At least once a month	0.139	.115	1.462	1	0.227	1.149	0.917	1.44
	A few times a year	-0.048	.101	0.226	1	0.634	0.953	0.781	1.162
	At least once a year	-0.138	.132	1.089	1	0.297	0.871	0.673	1.128
	Not at all	0.010	.101	0.01	1	0.922	1.010	0.828	1.231
Constant		-3.323	.236	197.407	1	0	0.036		

Table 4: Results of the movie going model (SPSS output for binary logistic regression model)

	B	S.E.	Wald	Df	Sig.	Exp(B)	95.0% C.I. for EXP(B)		
							Lower	Upper	
LCA_Q120- "did you read a magazine?" (Reference is not reading)	0.332	.077	18.451	1	0	1.394	1.198	1.622	
LCA_Q130- "did you read a book?" (Reference is not reading)	0.277	.069	16.072	1	0	1.319	1.152	1.511	
LCA_Q220- "did you watch a video?" (Reference is not watching)	0.845	.079	115.677	1	0	2.328	1.996	2.716	
LCA_Q240- "did you listen to music on CD, etc?" (Reference is not listening)	0.428	.086	24.704	1	0	1.535	1.296	1.817	
LCA_Q300- "did you attend a concert or performance by professional artists of music, dance, theatre or opera?" (Reference is not attending)	0.740	.070	111.168	1	0	2.096	1.827	2.406	
LCA_Q340- "did you go to a cultural or artistic festival?" (Reference is not going)	.207	.086	5.804	1	0.016	1.230	1.039	1.455	
LCA_Q360- "did you attend any other kind/type of cultural performance?" (Reference is not attending)	0.346	.110	9.917	1	0.002	1.414	1.14	1.754	
LCA_Q410- "did you go to a public art gallery or art museum?" (Reference is not going)	0.270	.083	10.558	1	0.001	1.310	1.113	1.541	
LCA_Q430- "did you go to an historic site?" (Reference is not going)	0.147	.074	3.941	1	0.047	1.158	1.002	1.338	
LCA_Q440- "did you go to a zoo, aquarium, botanical gardens, planetarium or observatory?" (Reference is not going)	0.325	.071	20.696	1	0	1.384	1.203	1.592	
AGEGR10	15 to 24		257.806	6	0				
	25 to 34	-0.788	.154	26.368	1	0	0.455	0.336	0.614
	35 to 44	-1.099	.150	53.902	1	0	0.333	0.248	0.447
	45 to 54	-1.476	.151	95.417	1	0	0.229	0.17	0.307
	55 to 64	-1.816	.159	130.862	1	0	0.163	0.119	0.222
	65 to 74	-2.173	.175	153.909	1	0	0.114	0.081	0.16
	75 years and older	-2.445	.202	146.601	1	0	0.087	0.058	0.129

EDU5*	College diploma or some college			19.455	2	0			
	High school or less	-0.217	.072	8.956	1	0.003	0.805	0.699	0.928
	Degree	0.189	.085	4.99	1	0.025	1.208	1.023	1.426
BRTHCAN- "country of birth of respondent" (Reference is birth in Canada)		-0.278	.099	7.822	1	0.005	0.757	0.623	0.92
LUC_RST- "urban/rural indicator" (Reference is living in an urban area)		-0.550	.074	55.309	1	0	0.577	0.499	0.667
INCMHSD	Less than \$40k			48.593	2	0			
	\$40k to \$80k	0.388	.076	25.9	1	0	1.475	1.27	1.712
	\$80k or more	0.644	.095	45.862	1	0	1.903	1.58	2.293
LANHSDC	English only			6.045	2	0.049			
	French only	0.040	.154	0.067	1	0.796	1.041	0.77	1.407
	Other language	-0.319	.136	5.493	1	0.019	0.727	0.556	0.949
REGION	Ontario			13.493	4	0.009			
	Atlantic region	0.169	.096	3.112	1	0.078	1.184	0.981	1.428
	Quebec	0.471	.157	8.95	1	0.003	1.601	1.176	2.179
	Prairie region	0.191	.089	4.575	1	0.032	1.210	1.016	1.441
	British Columbia	-0.018	.106	0.028	1	0.868	.982	0.798	1.21
RELIGATT	At least once a week			11.994	4	0.017			
	At least once a month	-0.102	.108	0.898	1	0.343	0.903	0.731	1.115
	A few times a year	-0.020	.094	0.045	1	0.833	0.980	0.815	1.179
	At least once a year	0.023	.123	0.034	1	0.855	1.023	0.804	1.301
	Not at all	-0.242	.088	7.545	1	0.006	0.785	0.661	0.933
MARSTAT*	Married or common law			37.024	2	0			
	Divorced, separated or widowed	0.467	.089	27.521	1	0	1.595	1.34	1.9
	Single (never married)	0.400	.092	19.066	1	0	1.491	1.246	1.784
LCA_Q420- "did you go to a public art gallery or art museum?" (Reference is not going)		0.212	.087	5.932	1	0.015	1.236	1.042	1.465
LCA_Q350- "did you go to a performance of cultural/heritage music, theatre or dance?" (Reference is not going)		0.157	.097	2.632	1	0.105	1.170	0.968	1.413
Constant		-0.672	.215	9.81	1	0.002	0.511		

Table 5: Hosmer and Lemeshow significance values for the four models

	Book Reading	Art Gallery Attendance	Movie Going	Live Performance Attendance
Hosmer and Lemeshow Test (sig.)	0.849	0.293	0.488	0.181

APPENDIX 2: VARIABLES INCLUDED IN STUDY

In all cases, the reference class is listed first.

*indicates that two or more of the available answers for the specific question were grouped together for analysis purposes

† indicates that some of the available responses were excluded from analysis (most commonly the responses “don’t know”, “not stated” and “not asked”)

DEMOGRAPHIC VARIABLES

Age

Age is thought to be an important factor in cultural activities. The variable was transformed several times in order to determine if some age groups were better predictors of the various dependent variables. This was done to distinguish groups that were not distributed evenly compared to others. For example, if the weighted percentage of individuals in some of the various classes of the variable were approximately equal, then the overall range of the age groups would be a better estimate than the smaller arrays.

- Age group of respondent (AGEGR10, Position 48 in the GSS dataset)
 - 15 to 24
 - 25 to 34
 - 35 to 44
 - 45 to 54
 - 55 to 64
 - 65 to 74
 - 75 and older
- Age group of respondent (based on AGEGR10, Position 48)*
 - 25 or older
 - 15 to 24
- Age of respondent (based on AGEGR10, Position 48)*
 - Under 75
 - 75 or older
- Age group of the respondent (based on AGEGR10, Position 48)*
 - 15 to 44

- 45 to 74
- 75 years and over

Sex

Sex is thought to be an important factor in cultural activities.

- Sex of respondent (SEX, Position 50)
 - Female
 - Male

Marital Status

Does the relationship status of the respondent have an effect on how culturally inclined they are? It could be that people who are married or in a common law relationship are less likely to attend cultural events as frequently as those who are single since single individuals may have more free time.

- Marital status of respondent (MARSTAT, Position 51) †
 - Married
 - Living common-law
 - Widowed
 - Separated
 - Divorced
 - Single (Never married)
- Marital status of respondent (based on MARSTAT, Position 51) † *
 - Married or common law
 - Separated, divorced or widowed
 - Single (Never married)

Single Children in the Household

Having a single child or children living in the household could represent an extra expenditure in the household from both financial and time perspectives. This additional expense could affect the funds allocated to “entertainment” in the household and could therefore affect have a negative impact on cultural activities. On the other hand, people may be more likely to take their children to cultural activities than they would be on their own.

- Single child(ren) of the respondent living in the household (CHRFLAG, Position 58)
 - Yes
 - No

Region

The region of residence of the respondent could be an important demographic factor, given differing levels of availability and interest in cultural activities.

- Region of residence of the respondent (REGION, Position 72)
 - Ontario
 - Atlantic region
 - Quebec
 - Prairie region
 - British Columbia

Urban/Rural Indicator

The size of the location in which the respondent lives could affect their cultural activities. For example, a large city is more likely to have many art galleries and performing arts venues than a rural area. So for purely reasons of convenience, a respondent living in a rural area could attend cultural events less frequently than if they were living in an urban area.

- Urban/Rural indicator (LUC_RST, Position 73) †
 - Larger Urban Centres (CMA/CA)
 - Rural and Small Town (non-CMA/CA)

Activity Limitations

The participation of disabled people in cultural activities is an important accessibility factor.

- Respondent is limited in the amount of physical activity he/she can do at home, at work, at school or in other activities because of a physical condition or a mental condition or health problem (ACTLIMIT, Position 1640) †
 - No
 - Yes

Education

Education is thought to be an important factor in cultural activities. The question regarding the amount of education acquired by the respondent was altered into three categories: no post-secondary education, some post secondary education and completion of a university degree or more.

- Highest level of education obtained by the respondent (EDU5, Position 1737) † *
 - Diploma/certificate from community college or trade/technical or some university/community college
 - High school diploma or some secondary/elementary/no schooling
 - Doctorate/masters/bachelor's degree

Country of Birth

This variable was included in the study because it is the best available indicator of whether early childhood cultural impacts could potentially affect the cultural interests of the respondent in adulthood. Immigrants' participation in cultural activities is an important accessibility question.

- Country of birth of respondent (BRTHCAN, Position 2222)†
 - Canada
 - Country outside Canada

Household Language

Language and official language minorities may have an impact on interest and availability of cultural activities.

- Respondent's household language (LANHSDC)† Position:2239
 - English only
 - French only
 - Other language

Religion

The US report "The Diversity of Cultural Participation" found that cultural attendees were more likely than non-attendees to attend religious services. We wanted to examine this relationship with Canadian data. This variable was also recalibrated into two categories, having no religion and being affiliated with a specific religious denomination. This was done to find out if the religious morals taught to the respondent at some point in their life has some sort of influence on cultural attendance. This variable was recalibrated several times to find out what aspects of spiritual importance have the largest impact on cultural activities.

- Religion of respondent In six categories (RELIG6, Position 2240)†
 - No religion
 - Roman Catholic
 - United Church
 - Protestant
 - Other
- Religion of respondent In six categories (based on RELIG6, Position 2240)†*
 - No religion
 - Roman Catholic, United Church, Protestant or other
- How important are your (religious or) spiritual beliefs to the way you live your life? Would you say they are: (RLR_Q105, Position 2241)†
 - Very important
 - Somewhat important
 - Not very important
 - Not at all important

- How important are your (religious or) spiritual beliefs to the way that you live your life?
Would you say they are: (based on RLR_Q105, Position 2241)†*
→ Very important
→ Somewhat important or not very important or not at all important
- How important are your (religious or) spiritual beliefs to the way that you live your life?
Would you say they are: (based on RLR_Q105, Position 2241)†**
→ Very important or somewhat important or not very important
→ Not at all important
- How important are your (religious or) spiritual beliefs to the way you live your life?
Would you say they are: (based on RLR_Q105Position 2241)†***
→ Very important
→ Somewhat important or not very important
→ Not at all important
- Religious attendance of the respondent (RELIGATT, Position 2242)†
→ At least once a week
→ At least once a month
→ A few times a year
→ At least once a year
→ Not at all

Household Income

Similar to education, household income is thought to be an important factor in cultural activities.

- Total household income (INCMHSD, Position 2248)†*
→ Less than \$40,000
→ \$40,000 to \$79,999
→ \$80,000 or more

CULTURAL VARIABLES

Preliminary research showed that all cultural activities appeared to have a positive effect on the percentage of Canadians who participated in other cultural activities.

- During the past 12 months, as a leisure activity (not for paid work or studies) did you read a newspaper? (LCA_Q110, Position 1802) †
→No
→Yes
- During the past 12 months, as a leisure activity (not for paid work or studies) did you read a magazine? (LCA_Q120, Position 1804) †
→No
→Yes
- During the past 12 months, as a leisure activity (not for paid work or studies) did you read a book? (LCA_Q130, Position 1806) †
→No
→Yes
- During the past 12 months, did you use library services as a leisure activity (including accessing the internet)? (LCA_Q140, Position 1808) †
→No
→Yes
- During the past 12 months, did you: go to a movie or drive-in? (LCA_Q210, Position 1815) †
→No
→Yes
- During the past 12 months did you: watch a video, rented or purchased, VHS or DVD? (LCA_Q220, Position 1817) †
→No
→Yes
- During the past 12 months did you: listen to downloaded music on your computer, MP3 player, etc.? (LCA_Q230, Position 1819) †
→No
→Yes
- During the past 12 months, did you: listen to music on CD, cassette tapes, DVD audio discs, records, etc.? (LCA_Q240, Position 1821) †
→No
→Yes

- During the past 12 months, did you: attend a concert or performance by professional artists of music, dance, theatre, or opera, excluding cultural festivals? (LCA_Q300, Position 1833) †
 →No
 →Yes
- During the past 12 months, did you: go to a cultural or artistic festival (such as film, fringe, dance, jazz, folk, rock, buskers or comedy)? (LCA_Q340, Position 1840) †
 →No
 →Yes
- During the past 12 months, did you: go to a performance of cultural/heritage music, theatre or dance (e.g. Aboriginal peoples, Chinese, Ukrainian)? (LCA_Q350, Position 1842) †
 →No
 →Yes
- During the past 12 months, did you: attend any other kind/type of cultural performance? (LCA_Q360, 1844) †
 →No
 →Yes
- During the past 12 months, did you: go to a public art gallery or art museum (including attendance at special art exhibits)? (LCA_Q410, Position 1846)†
 →No
 →Yes
- During the past 12 months, did you: visit a museum other than public art galleries or art museums? (LCA_Q420, Position 1848)†
 →No
 →Yes
- During the past 12 months, did you: go to an historic site? (LCA_Q430, Position 1850)†
 →No
 →Yes
- During the past 12 months, did you: go to a zoo, aquarium, botanical gardens, planetarium or observatory? (LCA_Q440, Position 1852)†
 →No
 →Yes
- During the past 12 months, did you: go to a conservation area or nature park? (LCA_Q450, Position 1854)†
 →No
 →Yes

Bibliography

Logistic Regression, G. David Garson. 2006,
<http://www2.chass.ncsu.edu/garson/PA765/logistic.htm>

Logistic Regression, 2002,
<http://userwww.sfsu.edu/~efc/classes/biol710/logistic/logisticreg.htm>

Logistic Regression, Gerard E. Dallal, Ph.D. 2001, <http://www.tufts.edu/~gdallal/logistic.htm>

General Social Survey- Cycle 19: Time Use (2005) User's Guide of the Public Use Microdata File, Statistics Canada, Social and Aboriginal Statistics Division. December 2006,
<http://data.library.ubc.ca/datalib/survey/statscan/gss/cycle19/gssc19gidv2.pdf>

The Diversity of Cultural Participation, Ostrower, Francie. Urban Institute, November 2005,
<http://www.wallacefoundation.org/NR/rdonlyres/687A0402-4FAC-478B-B6ED-DC96EA0788A8/0/TheDiversityofCulturalParticipation.pdf>