Factors Affecting Americans Engagement in Physical Activities — An Application of Logistic Regression Model to ATUS Survey Data

Khairul Islam

Department of Mathematics and Statistics, Eastern Michigan University, Ypsilanti, MI, USA

Abstract Engagement in physical activity (PA) contributes to improved health outcomes and benefits. The American Time Use Survey (ATUS) is an indispensable data resource for Americans engagement in PA due to engagement in sports and exercise activity. This paper investigates factors of Americans engagement in PA and measures the effect of these factors by implementing Chi-square test and logistic regression analysis with appropriate survey weight. The study reveals that only 18.8% Americans 15 years and older are engaged in PA. Bivariate analysis suggests that socio-demographic factors such as gender, age, family income, education, ethnicity, marital status, geographical region and living quarters are significant factors (chi-squared p-value<0.05) of Americans engagement in PA. Adjusted and unadjusted logistic regression analyses reveal that gender, income, education, ethnicity and geographical region are significant (chi-squared and F-tests p-value<0.05) factors of engagement in PA. Adjusted (multiple logistic regression) analysis reveals the significance and risk of non-engagement in PA for various factor levels compared to a reference level adjusted for other factors via 95% confidence interval (CI) for related odds ratios (OR) estimates. In adjusted analysis, the OR estimates suggest that females are significantly 1.19 times more likely to be non-engaged in PA compared to males (95% CI: 1.02-1.39) since the CI estimate does not include 1. Using the reference income of $75,000+, Americans with income <$25,000 and $25,000-$49,999 are significantly about 1.4 times (95% CI: 1.09-1.83) and 1.5 times (95% CI: 1.18-1.83) more likely to be non-engaged in PA. Using Master+ (Master's degree and above) as a reference category, Americans with education <HS are significantly 1.86 times (95% CI: 1.48-2.33) and those having some college or associate degrees are 1.66 times (95% CI: 1.34-2.05) more likely to be non-engaged in PA.

Keywords American Time Use Survey, Physical activity, Chi-squared test, Logistic regression model, Risk analysis

1. Introduction

As per the Physical Activity Guidelines for Americans [1], regular physical activity (PA) is one of the most important actions people can do to improve their health. It improves physical function among individuals of all ages, enabling them to conduct their daily lives with energy and without undue fatigue. In addition to improving physical function, it may improve cognitive function among youth and adults which includes memory, attention, executive function (the ability to plan and organize; monitor, inhibit, or facilitate behaviors; initiate tasks; and control emotions), and academic performance among youth [1]. Physical activity results in many health benefits such as lowering risk and severity of chronic diseases (including heart disease, diabetes, and some cancers), lowering mortality rates, and improving mental health and physical well-being [2]. Yet

many Americans do not comply with the guideline recommendations [1-2]. In this study, we investigate what socio-demographic factors affect Americans engagement in physical activity due to engagement in sports and exercise activities using the American Time Use Survey (ATUS) 2019 data. ATUS collects activity information for each reported activity in a 24-hour day in 17 broad categories along with their coding information [3-6]. These categories are (1) Personal care, (2) Household activities; (3) Caring for and helping household members; (4) Caring for and helping non-household members; (5) Work and work related activities; (6) Education; (7) Consumer purchases; (8) Professional and personal care services; (9) Household services; (10) Government services and civil obligations; (11) Eating and drinking; (12) Socializing, relaxing, and leisure; (13) Sports, exercise, and recreation; (14) Religious and spiritual activities; (15) Volunteer activities; (16) Telephone calls; and 17) Traveling. The ATUS is a significant data source of physical activity information of Americans 15 years and older due to various socio-demographic factors of the participants.

The physical activity refers to any bodily movement produced by the contraction of skeletal muscle that increases energy expenditure above a basal level ([1], [7]). It includes any subset of physical activity that enhances health. Physical activity time includes active (not spectator) sports, exercise, and recreational activities [2].

In this study, physical activity will be measured from the participation of respondents in active sports, exercise and recreational activities, by excluding any utilitarian physical activity such as physical effort for work or household tasks. We extract physical activity information of Americans 15 years and older and the distribution of such activity due to socio-demographic factors, and investigate how age, gender, ethnicity, income and other socio-demographic factors affect engagement in sports and exercise activity.

As analytical tools, we incorporate chi-squared tests for independence between response and socio-demographic factors. In order to assess the risk and significance of non-engagement in physical activity for various factor levels compared to a reference level, we employ simple and multiple factor logistic regression analysis in the presence of survey weights.

2. Methods and Models

The utilization of the ATUS data requires merging multiple data sets, recoding data values and the weighted data analyses techniques to accommodate appropriate survey weight. In this section, we introduce the data description together with data merging, recoding of various socio-demographic factors and statistical models that are being utilized in this paper.

2.1. Data Description

We utilize ATUS 2019 data for analysing significance of socio-demographic factors of Americans 15 years and above affecting Americans engagement in physical activity due to engagement in sports and exercise activity. For the purpose of the desired analysis, we merge three ATUS 2019 datasets—activity file, activity summary file and ATUS-CPS file by using the common identification characteristic called TUCASEID. The response and various socio-demographic factors in this study are generated from the merged data file as follows:

Response: We create a binary response variable called the physical activity (PA) engagement due to engagement in sports and exercise activities. The response PA takes a value of 0 for non-engagement and 1 for engagement in any sports and exercise activity evaluated by using the six-digit activity code (TRCODE, from the activity data file). The TRCODE with values between 130100 and 130136 defines the physical activity engagement status of 1 due to engagement in sports and exercise activity, and the engagement status 0 is defined otherwise. The details about the ATUS assigned 6-digit

activity code are available in the ATUS Activity Lexicons Coding Manuals [8].

Socio-demographic factors: In order to investigate the effect of various socio-demographic factors, we extract potential factors from the ATUS data files. The activity summary data file has been used to recreate gender and age group variables. The sex is created using the TESEX variable with values 1=male and 2=female. The variable age (age group) is created by recoding the values of the variable TEAGE with values 15-29, 30-44, 45-69 and 70+ as distinct groups.

Most of the factors in the study are derived from the ATUS-CPS data file. The education (educational status) is created by recoding the values of the variable PEEDUCA with values < HS, Some col./ass. (Some college or Associate's degree), Bachelor (Bachelor's degree) and Master+ (Master or Higher) as distinct groups. The factor income (family income) is created by recoding the values of HEFAMINC with values <25,000, 25,000-49,999, 50,000-74,999, 75,000+ for four groups. The factor ethnicity is defined from PTDTRACE with values 1=White only, 2=Black only, 3=Hispanic only and 4=Others. The region (geographical region) is derived from GEREG with recoded values 1=Northeast, 2=Midwest, 3=South, 4=West. The variable LFP (Labor Force Participation) is created using PRCIVLF with values Yes=In civilian labor force, No=Not in civilian labor force. The variable marital (marital status) is derived from PEMARITL with values 1=Married, 2=Widowed, 3=Divorced/Separated and 4=Never married. The nativity is derived from PRCITSHP with values 1=Native, 2=Naturalized citizen (Nat-citizen) and 3=Not a citizen (non-citizen). The variable quarters (living quarters) is taking values 1=Owned, 2=Rented (Rented for cash) and 3=Occupied (Occupied without cash). The categorization of factors utilized in the study is consistent with those used by the U.S. Bureau of Labor Statistics [9].

Survey weight: The ATUS uses weights to compensate for the important aspects of the sampling and data collection process due to stratification, non-uniformity of sample across the days of week or demographic groups, so as to represent each group correctly [5]. The ATUS weight variable TUFINLWGT appears in the activity summary data file has been utilized in all analyses in this study.

2.2. Association of Factors

Initially, we test the null hypothesis of independence or no association between the engagement in physical activity (due to engagement in sports and exercise activities) and various socio-demographic factors. For the test of no association, we analyse two-way cross-classified table with I levels or categories (rows) of an underlying socio-demographic factor and 2 levels or categories (column) of the response PA, with level 1 referring to PA=0 (non-engaged in PA) and 2 referring to PA=1 (engaged in PA). The null and alternative hypotheses involved for the association test are of the form:

H_0: There is no association between physical activity engagement status and a given socio-demographic factor

H_1: There is an association between physical activity engagement status and a given socio-demographic factor

We test the hypothesis of no association or independence between physical activity and each socio-demographic factor via chi-squared tests [10]. Due to the survey weight associated with the ATUS data, we performed weighted chi-squared test for the independence of two-way table generated by the cross-classification of an underlying factor due to the dichotomous response. The form of cross-classified table of $I \times 2$ cells in this study is as presented in Table 1. In Table 1, $\widehat{N}_{ik}$ is the estimate of total weights of all subjects falling in (i, k) cell, with subject j receiving the survey weight W_j available due to the survey weight variable TUFINLWGT.

Table 1. Distribution of subjects by factor and response levels

	Response		
Factor levels	Level 1	Level 2	Total
1	$\widehat{N}_{11}$	$\widehat{N}_{12}$	$\widehat{N}_{1\cdot}$
2	$\widehat{N}_{21}$	$\widehat{N}_{22}$	$\widehat{N}_{2\cdot}$
⋮	⋮	⋮	⋮
I	$\widehat{N}_{I1}$	$\widehat{N}_{I2}$	$\widehat{N}_{I\cdot}$
Total	$\widehat{N}_{\cdot1}$	$\widehat{N}_{\cdot2}$	$\widehat{N}$

Under the null hypothesis of no association or independence, the expected weighted frequency (total) in cell (i, k) is given by $E_{ik} = \frac{\widehat{N}_{i\cdot} \times \widehat{N}_{\cdot k}}{\widehat{N}}$. Then, given a sample of n subjects, the test of no association or independence between a response and a given socio-demographic factor is given by

$$Q_p = \frac{n}{\widehat{N}} \sum_{i=1}^{I} \sum_{k=1}^{2} \frac{(\widehat{N}_{ik} - E_{ik})^2}{E_{ik}} \qquad (1)$$

The statistic Q_p follows a Chi-squared distribution with degrees of freedom $(I - 1) \times (2 - 1)$. We carry out this test of no association by implementing the SAS Proc Surveyfreq procedure [11]. The basic syntax of Proc Surveyfreq procedure is as follows:

Proc Surveyfreq Data=DatasetName;
Weight Weight Variable;
*Tables factor*response/chisq;*
Run;

The Proc Surveyfreq is a required statement that invokes the procedure and identifies the data set to be analyzed. The Tables statement specifies frequency or crosstabulation tables and requests tests. The WEIGHT statement names the sampling weight variable. The Run statement completes the execution of the procedure. Other statements and options that are irrelevant for this study have been suppressed and can be found in [11].

The set of Proc SurveyFreq commands such as above generates analysis results as in Table 2 for factor Eth (ethnicity) and response PA (physical activity). Note that the

Table 2 reports unweighted and weighted frequency, along with percent, standard error and results of chi-squared test for a cross-classified table. This test provides a chi-squared value of 33.09 with df=3 and p-value<0.0001, which provides the evidence of significant relationship between physical activity and ethnicity. Given the test is significant, we evaluate the risk of Americans non-engagement in sports and exercise activity due to various socio-demographic factors by logistic regression models.

Table 2. Analysis of Ethnicity (Eth) by PA*

Eth	PA	Freq	Wgt Freq	Std Err Wgt Freq	Percent	Std Err Percent
1	0	5164	4.9E+10	7.5E+08	51.5	0.70
	1	1239	1.3E+10	4.8E+08	13.2	0.48
	Total	6403	6.2E+10	8.2E+08	64.7	0.69
2	0	1030	9.8E+09	4.0E+08	10.2	0.41
	1	159	1.7E+09	1.8E+08	1.8	0.19
	Total	1189	1.2E+10	4.4E+08	12.0	0.45
3	0	1118	1.4E+10	5.7E+08	14.7	0.55
	1	170	2.1E+09	2.3E+08	2.2	0.23
	Total	1288	1.6E+10	6.1E+08	16.9	0.58
4	0	425	4.7E+09	3.1E+08	4.9	0.31
	1	130	1.5E+09	1.7E+08	1.6	0.18
	Total	555	6.2E+09	3.5E+08	6.4	0.36
Total	0	7737	7.8E+10	8.9E+08	81.2	0.56
	1	1698	1.8E+10	5.8E+08	18.8	0.56
	Total	9435	9.6E+10	9.1E+08	100	

*PA=1 (0) refers to engagement (non-engagement status)
Eth=1 (only white), 2 (only black), 3 (only Hispanic) and 4 (others)

2.3. Models of Assessment of Risks

Given the binary response, with values 0 for non-engagement in physical activity and 1 for engagement in activity, we analyze the risk of non-engagement due to various levels of the socio-demographic factors. Let $\Pr\{Y = 0\} = \pi$ be the probability of non-engagement in the sports and exercise activity. The risk of non-engagement in PA is measured by the odds of non-engagement of an individual to sports and exercise activity due to various socio-demographic factors x, which includes age group, gender, ethnicity, family income, level of education, labor force participation status, marital status, nativity, geographical region, etc. Since all the socio-demographic factors involved in this study are polytomous, one factor level will be considered as a base-level category, which enables us to evaluate risk of any given factor label as compared to the base-level category. For the analysis of risk, we employ simple and multiple logistic regression models [12-13]. Simple model provides significance of the factor without adjustment for other factors, while multiple regression model provides the significance of an underlying factor adjusted for the other factors in the model.

An *I*-category single factor logistic regression model is specified by

$$log\left(\frac{\pi}{1-\pi}\right) = \alpha + \sum_{i=1}^{I-1}\beta_i x_i \qquad (2)$$

where α is the intercept and β_i is the effect of ith level of a given socio-demographic factor such as age group, ethnicity, family income, etc. It is to be noted that we arbitrarily choose one factor label as a reference group so as to make the interpretation easy and relevant to our interest and study.

A p-factor multiple logistic regression model with factor f having I_f labels is specified by

$$log\left(\frac{\pi}{1-\pi}\right) = \alpha + \sum_{f=1}^{p}\sum_{i=1}^{I_f-1}\beta_{fi} x_{fi} \qquad (3)$$

where the parameter α is the intercept and β_{fi} is the effect of ith level of the factor f. The parameters of the model are estimated by incorporating survey weight W_j assigned to subject j by implementing SAS Proc Surveylogistic method [14]. The basic coding syntax of Proc Surveylogistic is of the form:

Proc Surveylogistic Data= DatasetName;
Class predictor_1 predictor_2 ... predictor_p;
Model response=predictor_1 predector_2... predictor_p;
Weight Weight Variable;
run;

The Proc Surveylogistic statement invokes the Surveylogistic procedure and identifies the data set to be analysed. The Proc Surveylogistic and Model statements are required. The Model statement specifies response and predictors, each separated by space(s). The response and predictors are separated by an equal sign as specified in the code. The Class statement identifies predictors which are categorical or factor. Each statement in SAS ends with a semicolon (;). Other statements and options irrelevant to this study are suppressed and can be found in [14].

3. Results and Discussions

The ATUS 2019 had 9,435 respondents with only 18.8% engaged in PA due engagement in sports and exercise activity and 82.2% with no PA engagement. In Table 3, we present the results of bivariate analysis of Americans engagement in sports and exercise activity due to different socio-demographic factors.

Table 3. Test of independence between PA and various factors

	PA=0		PA=1			
Factors	%	StdErr	%	StdErr	chisq	pvalue
Gender						
Male	38.72	0.68	9.72	0.43		
Female	42.48	0.68	9.08	0.41	4.71	**0.030**
Age						
15-29	19.20	0.67	5.43	0.43		
30-44	19.27	0.52	4.62	0.26		
45-69	31.52	0.62	6.44	0.30		

	PA=0		PA=1			
	%	StdErr	%		chisq	pvalue
70+	11.21	0.36	2.32	0.16	14.64	**0.002**
Income						
<25,000	13.67	0.45	2.45	0.24		
25,000-49,999	18.79	0.53	3.13	0.24		
50,000-74,999	15.86	0.51	3.45	0.27		
75,000+	32.88	0.67	9.77	0.42	43.11	**<.0001**
<HS	25.04	0.63	3.78	0.32		
Some col/asso	20.51	0.55	3.67	0.27		
Bachelor	15.67	0.46	5.09	0.28		
Master+	19.98	0.56	6.27	0.35	80.05	**<.0001**
Ethnicity						
White only	51.47	0.70	13.23	0.48		
Black only	10.20	0.41	1.77	0.19		
Hispanic only	14.67	0.55	2.23	0.23		
Others	4.85	0.31	1.57	0.18	33.09	**<.0001**
Marital						
Married	41.53	0.67	9.36	0.38		
Widowed	4.78	0.23	0.94	0.10		
Div./sep.	10.23	0.35	1.89	0.15		
Never married	24.66	0.68	6.61	0.43	13.14	**0.004**
Nativity						
Native	68.17	0.67	15.66	0.52		
Nat-citizen	5.69	0.32	1.57	0.16		
Non-citizen	7.34	0.39	1.57	0.22	2.03	0.362
Region						
Midwest	13.68	0.49	3.39	0.27		
South	19.09	0.53	4.12	0.29		
West	31.47	0.64	6.61	0.36		
Northeast	16.96	0.54	4.68	0.29	9.27	**0.026**
LFP						
Yes	51.79	0.70	11.83	0.47		
No	29.41	0.63	6.97	0.37	0.25	0.621
Quarters						
Owned	57.71	0.70	13.94	0.49		
Rented	22.82	0.60	4.81	0.33		
Occupied	0.67	0.10	0.06	0.02	7.23	**0.027**

From the results in Table 3, of bivariate analysis of the test of no association or independence, between physical activity engagement and socio-demographic factors it follows that gender, age group, family income, ethnicity, marital status, geographical region and living quarters are all significantly associated with physical activity engagement. However, the labor force status and nativity are not associated with physical activity engagement.

In Tables 4.1-4.2, we provide the results of test of significance of various sociodemographic factors resulting from a single factor logistic regression and multiple factor logistic regression (LoR) model implemented via the Proc Surveylogistic method.

Table 4.1. Chi2 tests of significance of effects with LoR model

		Unadjusted		Adjusted	
Effect	DF	Chi2	P-value	Chi2	P-value
Gender	1	4.703	**0.0301**	5.09	**0.024**
Age	3	11.477	**0.0094**	5.34	0.149
Income	3	44.724	**<0.001**	13.86	**0.003**
Education	3	75.566	**<0.001**	50.80	**<0.001**
Ethnicity	3	31.349	**<0.001**	22.03	**<0.001**
Marital	3	11.403	**0.0097**	3.17	0.367
Nativity	2	2.518	0.2839	8.92	**0.012**
Region	3	9.675	**0.0215**	8.45	**0.038**
LFP	1	0.247	0.6194	4.71	**0.030**
Quarters	2	8.476	**0.0144**	5.11	0.078

Table 4.2. F tests of significance of effects with LoR model

			Unadjusted F		Adjusted F	
Effect	DF1	DF2	Value	p-value	Value	p-value
Gender	1	9434	4.7	**0.030**	5.09	**0.024**
Age	3	9432	3.82	**0.009**	1.78	0.149
Income	3	9432	14.9	**<.0001**	4.62	**0.003**
Education	3	9432	25.18	**<.0001**	16.93	**<.0001**
Ethnicity	3	9432	10.45	**<.0001**	7.34	**<.0001**
Marital	3	9432	3.8	**0.010**	1.06	0.367
Nativity	2	9433	1.26	0.284	4.46	**0.012**
Region	3	9432	3.22	**0.022**	2.81	**0.038**
LFP	1	9434	0.25	0.619	4.71	**0.030**
Quarters	2	9433	4.24	**0.015**	2.56	0.078

As appears in Tables 4.1-4.2, SAS Proc Surveylogistic provides significance of various sociodemographic factors via Chi-squared test and F-test for single factor (unadjusted) and multiple factor (adjusted) LoR models. It appears that gender, income (family income), education (level of education), ethnicity and region (geographical region) are significant factors of physical activity engagement for both unadjusted and adjusted analysis. Furthermore, the unadjusted analysis suggests that age (age), marital (marital status) and quarters (living quarters) also significantly affect the physical activity engagement, whereas adjusted analysis suggests that nativity and labor force participation status are significant factors of physical activity engagement.

Table 5 provide risk of non-engagement in physical activity measured via odds ratio for each factor compared to the baseline or reference category of the factor (unadjusted for a single factor logistic regression model and adjusted for multiple logistic regression model). The results of logistic regression analysis presented in Table 5 suggest that, for example, using unadjusted analysis, females are significantly 1.17 times more likely to be non-engaged in sports and exercise activities compared to males with 95% confidence interval (**1.02, 1.36**).

Table 5. Unadjusted and Adjusted OR (UOR and AOR) and 95% CI for LoR models

Factors	UOR est (95% CI)	AOR est (95% CI)
Gender (ref: Male)		
Female	1.17 (**1.02, 1.36**)	1.19 (**1.02, 1.39**)
Age (Ref: 15-29)		
30-44	1.18 (0.95, 1.47)	1.13 (0.88, 1.46)
45-69	1.38 (**1.12, 1.70**)	1.29 (**1.00, 1.67**)
70+	1.37 (**1.08, 1.73**)	1.35 (0.99, 1.83)
Income (ref: 75,000+)		
<25,000	1.66 (**1.32, 2.10**)	1.41 (**1.09, 1.83**)
25,000-49,999	1.78 (**1.47, 2.16**)	1.47 (**1.18, 1.83**)
50,000-74,999	1.37 (**1.12, 1.67**)	1.22 (0.99, 1.50)
Education (ref: Master+)		
<HS	2.08 (**1.67, 2.60**)	1.86 (**1.48, 2.33**)
Some col/asso	1.75 (**1.43, 2.15**)	1.66 (**1.34, 2.05**)
Bachelor	0.97 (0.81, 1.16)	0.99 (0.83, 1.19)
Ethnicity (ref: White only)		
Black only	1.48 (**1.16, 1.89**)	1.38 (**1.07, 1.80**)
Hispanic only	1.69 (**1.33, 2.15**)	1.93 (**1.43, 2.60**)
Others	0.79 (0.60, 1.04)	1.14 (0.83, 1.56)
Marital (ref: Married)		
Widowed	1.14 (0.89, 1.47)	0.87 (0.66, 1.15)
Div/sep.	1.22 (**1.00, 1.48**)	0.97 (0.78, 1.19)
Never married	0.84 (0.71, 1.00)	0.84 (0.68, 1.04)
Nativity (ref: Native)		
Nat-citizen	0.83 (0.65, 1.06)	0.66 (0.50, 0.87)
Non-citizen	1.07 (0.79, 1.46)	0.86 (0.59, 1.26)
Region (ref: South)		
Northeast	0.85 (0.68, 1.05)	0.95 (0.76, 1.18)
Midwest	0.97 (0.80, 1.18)	1.06 (0.87, 1.29)
West	0.76 (0.63, 0.92)	0.79 (0.65, 0.96)
LFP (ref: No)		
Yes	1.04 (0.89, 1.21)	1.23 (**1.02, 1.48**)
Quarters (ref: Owned)		
Rented	1.15 (0.97, 1.36)	1.00 (0.83, 1.21)
Occupied	2.82 (**1.26, 6.32**)	2.54 (**1.13, 5.69**)

Confidence interval estimates of OR are of particular interests to researchers due to the fact that any confidence interval estimate that does not include 1 is significant. Because the 95% confidence interval of female non-engagement likelihood does not include 1, gender is a significant factor for engaging in PA. Similarly, using adjusted analysis via multiple logistic regression suggests that females are significantly 1.192 times more likely to be non-engaged in sports and exercise activities compared to males with 95% confidence interval (**1.02, 1.39**). The 95% confidence intervals of all effects with significantly higher risk of non-engagement to the physical activity as measured by odds ratio estimates are printed in bold. With unadjusted analysis (Table 5) and using reference income of $75,000+,

Americans in income groups of <$25,000, $25,000-$49,999 and $50,000-$74,999 are about 1.5, 1.8 and 1.4 times more likely to be non-engaged in sports and exercise activities with corresponding 95% confidence intervals (1.32, 2.10), (1.47, 2.16) and (1.12, 1.67), all of which are significant. However, adjusted analysis suggests that with the reference income of $75,000+, Americans in income groups of <$25,000 and $25,000-$49,999 are about 1.4 and 1.5 times more likely to be non-engaged in sports and exercise activities with corresponding 95% confidence intervals (1.09, 1.83) and (1.18, 1.83) which are significant. Overall, from the results in Table 5, it is very clear that income, education and ethnicity are very significant factors for non-engagement in sports and exercises. Regarding the education, Americans with lower education have higher likelihood of non-engagement in sports and exercise activities. It is also evident from the adjusted analysis that compared to the White, Black and Hispanic ethnic Americans are about 1.38 and 1.93 times more likely to be non-engaged in physical activity, with respective 95% confidence interval (1.07,1.80) and (1.43, 2.60), which are significant. Unadjusted analysis also implies similar significance regarding the ethnicity of PA engagement. Living quarters are partially significant for both analyses referring higher likelihood of being non-engaged of Americans living in occupied apartment compared to those living in owned apartment.

4. Concluding Remarks

Physical activity is significantly associated with improved health outcomes of individuals of all ages. It contributes in living a life with energy and without undue fatigue, along with improving cognitive function among youth and adults, with and without any chronic health conditions. Many studies reveal significant benefits of engagement in physical activity, for example [1,3,4,7,9]. Yet, Americans do not comply with the guideline recommendations of engagement in physical activity [1-2]. In this study, we undertake an investigation of what socio-demographic factors affect Americans engagement in physical activity due to the participation in sports and activity excluding any utilitarian physical activity such as physical effort for work or household tasks. The ATUS survey data provides a great opportunity of engaging in such investigation. We extract information of Americans engagement in sports and exercise activities along with their socio-demographic factors such as age, gender, ethnicity, income, labor force participation, education, etc. to explicitly study the relationship between physical activity engagement and these factors. Initially, this study considers bi-variate analysis to determine independence or no association between physical activity due to engagement in sports and exercise and various socio-demographic factors. Most of the factors investigated in this study via bivariate analysis appear to affect Americans engagement in sports and exercise. For example, socio-demographic factors such as gender, age group, family income, ethnicity, marital status, geographical region and living quarters are all significantly affecting PA due to engagement in sports and exercise. The bivariate analysis suggest that the labor force status and nativity do not affect PA significantly. Overall, the results of Tables 2 provide the significance of relationship between physical activity and socio-demographic factors of Americans 15 years and older who participated in the survey. In addition to knowing the significance of relationship between physical activity and socio-demographic factors, we also investigate the level-specific impact of a factor compared to any reference category by employing logistic regression analysis. Logistic regression is of great use in real-life for assessing category specific impact by studying the OR estimates. The confidence interval estimates of OR enables us to evaluate if any level-specific effect is more or less likely than a reference category. In this study, we employ logistic regression with single factor and multiple factors by incorporating survey weights. A single factor model provides unadjusted OR estimates, while the multiple factor model provides adjusted OR estimates, they both convey significant thoughts with study outcomes to researchers and policymakers. In reality, for examples, if multiple factors are unavailable, simple model is of great use. On the other hand, in the availability of multiple factors, it is of great importance to compare the impact of any factor level adjusted for other factors in the model. In this study, it appears that both unadjusted and adjusted analyses agree on the statistical significance of socio-demographic factors gender, family income, level of education, ethnicity and geographical region in the engagement status of Americans in the physical activity. Results of Table 5 provide quantification of Americans risk of non-engagement to PA. The 95% confidence interval estimates of OR reported in Table 5 provide level-specific evidence of factors affecting Americans engagement in sports and exercise activities since the corresponding confidence interval estimates of OR does not include 1.

This study is significant for many reasons, for examples, it contributes in the understanding of a complex survey data which would lead to applied and databased research. ATUS is a significant data source for pedagogical reasons because it could be explored and utilized in academy to provide students with real-life exposure to big data, complex analysis techniques and applications in the phase of reality. It provides a proper understanding of ATUS participants which characterizes Americans 15 years and older in engagement of physical activity due to their participations in sports and exercise.

REFERENCES

[1]   U.S. Department of Health and Human Services. Physical activity guidelines for Americans. Second edition. Washington (DC), 2018.

[2] R. Sturm and D. A. Cohen, Free time and physical activity among Americans 15 years or older: cross-sectional analysis of the American time use survey. Preventing Chronic Disease Public Health Research, Practice, and Policy. vol. 16, E133, Sept. 2019.

[3] The American Time Use Survey, the U.S. Bureau of Labor Statistics. Available: https://www.bls.gov/tus/.

[4] American Time Use Survey (ATUS) Data Dictionary: 2019 Interview data variables collected in ATUS, June, 2020.

[5] June 2020. American Time Use Survey User's Guide. Understanding ATUS 2003 to 2019. Available: https://www.bls.gov/tus/atususersguide.pdf.

[6] American Time Use Survey (ATUS) Coding Rules 2018.

[7] World Health Organization Physical Activity Key Facts, 2018. Available:https://www.who.int/news-room/fact-sheets /detail/physical-activity.

[8] American Time Use Survey — Activity Coding Lexicons and Coding Rules Manuals. Available: https://www.bls.gov/tus/l exicons.htm.

[9] U.S. Bureau of Labor Statistics, Spotlight on Statistics, 2017. Available: https://www.bls.gov/spotlight/2017/sports-and-ex ercise/pdf/sports-and-exercise.pdf.

[10] R. E. Walpole, R. H. Myers, S. L. Meyers and K. Ye, Probability & Statistics for Engineers & Scientists, Ninth Edition, Prentice Hall, 2012.

[11] SAS/STAT® 13.1 User's Guide The SURVEYFREQ Procedure. Available: https://support.sas.com/documentation /onlinedoc/stat/131/surveyfreq.pdf.

[12] A. Agresti, Categorical data analysis, Third Edition, Wiley, 2013.

[13] M. H. Kutner, C. J. Nachtsheim and J. Neter. Applied Linear Regression Models, 4th Edition, McGraw-Hill, 2004.

[14] SAS/STAT® 14.2 User's Guide The SURVEYLOGISTIC Procedure. Available: https://support.sas.com/documentation /onlinedoc/stat/142/surveylogistic.pdf.