SI 544 Final Project Report
Chen Huang, Yang Liu, Ning Wang

Introduction
The sports industry in the United States is a booming business that produced over $213 billion last year and is projected to continue its current trend [1]. To succeed in such business, companies should have a solid understanding of their target market. Unlike most sports market analysis reports that are based on private or long-term field surveys, we analyzed the sports market using a nationwide public survey “American Time Use Survey” from American Bureau of Labor Statistics [2]. Using such public surveys enable us to represent a complete population of the United States, yet save us significant cost and time to do field analysis. Our results show that even analyzing a public survey, we were able to gain many useful insights and verify interesting hypotheses in the sports market.

Data Description
The American Time Use Survey (ATUS) is an annual survey that “measures how people divide their time among life’s activities”. The ATUS data sample is a subset population of Current Population Survey (CPS). [3] Subjects are interviewed one time about “how they spent their time on the previous day, where they were, and whom they were with”. The survey resulted in six data tables containing different information, among which we used the activity summary data file for the year of 2007 (atussum_2007.dat). It contains respondents’ demographic information as well as their time use on various activities. Here are some examples for the data fields with explanations

Demographic (examples):
- TEAGE: age
- TESEX: gender
...

Time spent on different activities (we focus on sport activities)
- t130103: time spent on playing basketball
- t130104: time spent on biking
...

Below are some of the general statistics of our dataset
- 418 columns and 12248 rows
- 10001 out of 12248 spend no time on sports
- Age ranges from 15 to 85, mean is 46.22, sd is 17.77
- 6950 Female, 5298 male
- Average sport time use 20.35230 min, sd is 64.19
- Average weekly income is 830.25, sd is 627.44.
- 7572 employed VS 3863 unemployed
- Top ten most popular activities (Figure 1)
- Sports population distribution (Figure 2)

Hypotheizes:
To get a better understanding of people’s sporting behavior, we will test the following hypotheizes:

1. Whether factors such as gender, age, employment status and income level have impact on people’s time use on sports, exercises and recreations?
2. How do people manage to squeeze their time for sports, exercise and recreations?
3. Whether certain kinds of sports are more popular among people with higher income?
4. As people grow older, will they keep doing exercises, and whether this differs from sport to sport?

Hypothesize tests and summaries

1.1 $\chi^2$ test of whether employed population is more likely to attend sports regularly (Figure 3)
Results: p-value = 0.2802
We want to test if people’s decision in regular exercise depends on their employ status. Thus we applied $\chi^2$ tests. Here independent variables are employed or not and whether the person exercises regularly.
The bar plot (Figure 3) and chi square test shows that the two variables are independent with each other. Thus we cannot reject the null hypothesis that employment status does not have a significant effect on whether one would like to spend time on sports. The result means that even unemployed, people tend not to stop participating or attending sports.

1.2 $\chi^2$ test whether males are more likely to attend sports regularly (Figure 4)
Results: p-value < 2.2e-16
Same as the first test, $\chi^2$ test also applies here for the same reason.
The p-value is small enough to reject the null hypothesis. So the two variable are not independent of each other, instead, gender does has a great effect on whether one spends some time on sports. As shown in figure 4, in female group, the proportion of people who do not sports at all is much larger than the male group.

1.3 Pairwise t test of time use on sports for different income levels (Figure 5)

<table>
<thead>
<tr>
<th>Income Level</th>
<th>Time (min)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.236</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.989</td>
<td>0.989</td>
</tr>
<tr>
<td>2</td>
<td>0.989</td>
<td>0.989</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
We tested if the average time uses on sports are the same for different income levels. All p-values are not statistically significant. Therefore we cannot reject the null hypothesis that people of different income levels spend same amount of time on sports activities. Along with Figure 5, we can tell that people with different income levels do not have difference on time use on sports.

2. One-way ANOVA and t-test of activities time consumption between people who exercise regularly and those who do not (Figure 6)

<table>
<thead>
<tr>
<th></th>
<th>ANOVA test</th>
<th>T test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Watching TV</td>
<td>&lt; 2.2e-16</td>
<td>&lt; 2.2e-16</td>
</tr>
<tr>
<td>Working</td>
<td>7.939e-09</td>
<td>7.573e-10</td>
</tr>
<tr>
<td>Sleep</td>
<td>0.0007604</td>
<td>0.0002411</td>
</tr>
<tr>
<td>Cooking</td>
<td>2.587e-08</td>
<td>3.965e-12</td>
</tr>
<tr>
<td>Relaxing</td>
<td>0.0003681</td>
<td>7.581e-06</td>
</tr>
</tbody>
</table>

P-values of ANOVA test and T test

In order to compare groups of people who exercise regularly with those not, we picked 5 activities on which two groups of people spend significantly different mean time. People who exercise daily spend on average 80 minutes less on those activities than those don’t. Thus we hope these activities are the most representative. In order to indentify whether those activities are statically different between two groups, we used one-way ANOVA and t-test to verify. Here the independent variables are "exercise daily or not", and dependent variables are number of minutes spent on sports.

The ANOVA and t-test shows that all activities are significantly different in time. Our conclusion is that for people who exercise daily, they tend to squeeze time from working, watching TV, sleeping, cooking and relaxing.

We further tried to see if there is any correlation between time spend on doing exercise and on other activities, however we could not find significantly strong correlations among them, thus we believe that although people tend to utilize their time in watching TV, working and etc. on exercises, they might compensation their time on watching TV, working and etc. by sacrifice other activities’ time.

3. Two-way ANOVA test of how income and sport events affect sports’ popularity (Figure 9)

To exam the individual and interaction effect, we used two-way ANOVA test. Here independent variables are income and sports, and dependent variable is the number of people exercise on specific sports.

Our ANOVA test shows that each independent variable has effect on dependent variables and their interaction effect is also significant. From the plot and the ANOVA test, we can observe not all people would like to spend the same amount of time on doing exercise. We can also find out that different kinds of sports have different popularity in the population, e.g. walking is among the most popular sports in all population and Skiing is not so popular. Most importantly, we can find significant interaction effect; certain sports are more popular in low income population, e.g. basketball. Several sports are more popular in middle class, e.g. Cardiovascular, running, hiking, water Sports, weight-lifting. For high-income class, we can observe significant decrease in certain sports, e.g. working out, cardiovascular, weight-lifting, running.

4. Two-way ANOVA test of how age and sport events affect sports’ popularity (Figure 8)

The ANOVA test shows that age actually does not have a significant effect on sports popularity. The percentage of people who exercise daily is constant among different age groups. However, different age groups have their own preference on sports. For example, walking and working-out are more popular in all age groups than other sports activities. We can also observe age and sport’s interaction effect. For example, walking becomes much popular as people grow older, but working out is constantly popular. Certain sports are popular for young people, for example: play basketball, weight lifting and water sports. Certain sports are more popular for middle age, for example: cardiovascular, biking. Sports that are popular among old people are: walking, golfing. The rest categories’ popularities are relatively stable in all population.