**STAT 504 proposal**

Yuanjie (Tukey) Tu

**Topic: Linking health and travel behaviors**

Despite the wide known benefits of physical activities, “only half of Americans reported meeting guidelines for aerobic physical activity in 2017” (Le and Dannenberg, 2020). The health outcomes such as obesity associated with insufficient physical activities have brought huge economic burden in the US. And the obesity rate in the US has surpassed 42% in 2020 (Rhea Farberman, 2020). To promote physical activities, it’s recommended that adults conduct moderate-intensity physical activities of at least 150 minutes per week (HHS, 2018). Among many ways of stimulating physical activities, local governments and policy makers are increasingly interested in promoting transit use and reduce the use of private cars. Since transit is no door-to-door service, it requires a certain amount of walking or biking from and to the transit stations. Thus. it’s fun to investigate the role of transit and private cars in promoting physical activities and health!

The 2017 National household travel survey (NHTS) data[[1]](#footnote-1) are selected in this project (FHWA, 2017). Collected by Federal Highway Administration (FHWA), “the NHTS is the authoritative source on the travel behavior of the American public”. This dataset has a great number of attributes about travel. Questions 1-3 are a series of questions that can be answered by using this dataset. For our course project, we can choose one or multiple questions; we can also target on particular population groups such as females, teenagers, elderly people, etc. Potential variables including health status, physical activities, travel characteristics, and socio-demographics are listed in (but not limited to) Table 1.

**Question 1:** Will transit use induce more physical activities?

**Question 2:** Will private car use reduce physical activities?

**Question 3:** Are more physical activities related to a healthier life?

For Questions 1 and 2, the dependent variables can be *Light/moderate physical activity* or *levels of physical activity*. Independent variables may include travel characteristics (e.g., transit use, private car use) and socio-demographics. For Question 3, the dependent variables can be opinion of health (or others). As for independent variables, we may include physical activities besides travel characteristics and socio-demographics. It’s also possible to find other related variables in the dataset.

The candidate models can be

* Multinomial logit regression model
* Logistic regression model if we recategorize the dependent variables into binary ones
* For Questions 1 and 2, we may also choose linear regressions

**Table 1. Variable description**

|  |  |
| --- | --- |
| **Variable name** | **Definition** |
| ***Health status*** | |
| Opinion of health (HEALTH) | If -9, not ascertained; if -8, I don’t know; if -7, prefer not to answer; if 01, excellent; if 02, very good; if 03, good; if 04, fair; if 05, poor |
| ***Physical activities*** | |
| Light/moderate physical activity (LPACT) | Count of times of light or moderate physical activity in past week |
| Vigorous physical activity (VPACT) | Count of times of vigorous physical activity in past week |
| Levels of physical activity (PHYACT) | If -9, not ascertained; if -8, I don’t know; if -7, prefer not to answer; if 01, rarely or never conduct any physical activities; if 2, some light or moderate physical activities; 3. Some vigorous physical activities |
| ***Travel characteristics*** | |
| Driver status (DRIVER) | If -1, skip; if 1, yes; if 2, no |
| Car use (YEARMILE) | Miles personally driven in all vehicles |
| Car share use (CARSHARE) | Count of carshare program usage |
| Bike use (NBIKETRP) | Count of bike trips |
| Bike share (BIKESHARE) | Count of Bike Share Program Usage |
| Walk use (NWALKTRP) | Count of walk trips |
| Transit use (PTUSED) | Count of bus trips |
| Rideshare use (RIDESHARE) | Count of rideshare app usage |
| Trip rates (CNTTDTR) | Count of person trips on a travel day |
| Primary activity in previous week (PRMACT) | If -8, I don’t know; if -7, prefer not to answer; if 1, working; if 2, temporarily absent from a job or business; if 3, looking for work/unemployed; if 4, a homemaker; if 5, going to school; if 6, retired |
| Medical condition results in limiting driving to daytime (CONDNIGH) | If -9, not ascertained; if -1, prefer not to answer; If 1, yes; If 2, no. |
| Medical condition results in asking others for rides (CONDRIDE) | If -9, not ascertained; if -1, prefer not to answer; If 1, yes; If 2, no. |
| Medical conditions in giving up driving (CONDRIVE) | If -9, not ascertained; if -1, prefer not to answer; If 1, yes; If 2, no. |
| Frequency of online shopping (DELIVER) | Count of times purchased online for delivery in last 30 days |
| ***Socio-demographics*** | |
| Household car ownership (HHVEHCNT) | Number of household vehicles |
| Education (EDUC) | If -8, I don’t know; if -7, prefer not to answer; if 1, appropriate skip; if 1, less than high school; if 2, high school graduate; if 3, college or associates degree; if 4, bachelor’s degree; if 5, graduate degree or professional degree |
| Hispanic or Latino origin (R\_HISP) | If -8, I don’t know; if -7, prefer not to answer; if 1, yes; if 2, no |
| Home location (HBHUR) | If C, second city; if R, rural; if T, small town; if U, urban |
| Population density (HTPPOPDN) | If -9, not ascertained; if 50, 0-99; if 300, 100-499; if 750, 500-999; if 1500, 1000-1999; if 3000, 2000-3999; if 7000, 4000-9999; if 17000, 10000-24999; if 30000, 25000-999999 |
| Born in the US (BORNINUS) | If -9, not ascertained; if -1, prefer not to answer; If 1, yes; If 2, no. |
| Age (R\_AGE) | The age of the respondent, ranging from 5 to 92 |
| Gender (R\_SEX) | If -8, I don’t know; if -7, prefer not to answer; if 1, male; if 2, female |
| Work status (WKFTPT) | If -9, not ascertained; if -8, I don’t know; if -7, prefer not to answer; if 1, full-time; if 2, part-time |
| Household income (HHFAMINC) | If -9, not ascertained; if -8, I don’t know; if -7, prefer not to answer; if 01, less than $10,000; if 02, $10,000 to $14,999; if 03, $15,000 to $24,999; if 04, $25,000 to $34,999; if 05, $35,000 to $49,999; if 06, $50,000 to $74,999; if 07, $75,000 to $99,999; if 08, $100,000 to $124,999; if 09, $125,000 to 149,999; if 10, $150,000 to $199,999; if 11, $200,000 or more |
| Race | If -8, I don’t know; if -7, prefer not to answer; if 1, white; if 2, Black or African American; if 3, Asian; if 4, American Indian or Alaska Native; if 5, Native Hawaiian or other Pacific Islander |

**Reference**

FHWA, 2017. National Household Travel Survey [WWW Document]. URL https://nhts.ornl.gov/ (accessed 1.15.21).

HHS, 2018. Physical Activity Guidelines for Americans, 2nd edition 118.

Le, V.T., Dannenberg, A.L., 2020. Moving Toward Physical Activity Targets by Walking to Transit: National Household Transportation Survey, 2001–2017. Am. J. Prev. Med. 59, e115–e123. https://doi.org/10.1016/j.amepre.2020.02.023

Rhea Farberman, 2020. The State of Obesity 2020: Better Policies for a Healthier America [WWW Document]. URL https://www.tfah.org/report-details/state-of-obesity-2020/ (accessed 1.15.21).

**Schedule**

1. Which question(s)?

* **Question 1:** Will transit use/private car use induce more physical activities?- Tukey and Yiyang
* **Question 2:** Are more physical activities related to a healthier life? – Wushuang and Houjie

1. Data extraction and cleaning - Tukey
2. Descriptive analysis (summary table, distribution of each variable, pair plots, etc.)
3. Check multi-collinearity, outliers, and variable transformations (optional)
4. Modeling
5. Model diagnostics
6. Presentation (slides + video)
7. Report

Next meeting: 7 pm Friday – Descriptive analysis + Model results

1. Only personal travel data are selected. For simplicity, household travel and vehicle data are excluded from this project. [↑](#footnote-ref-1)