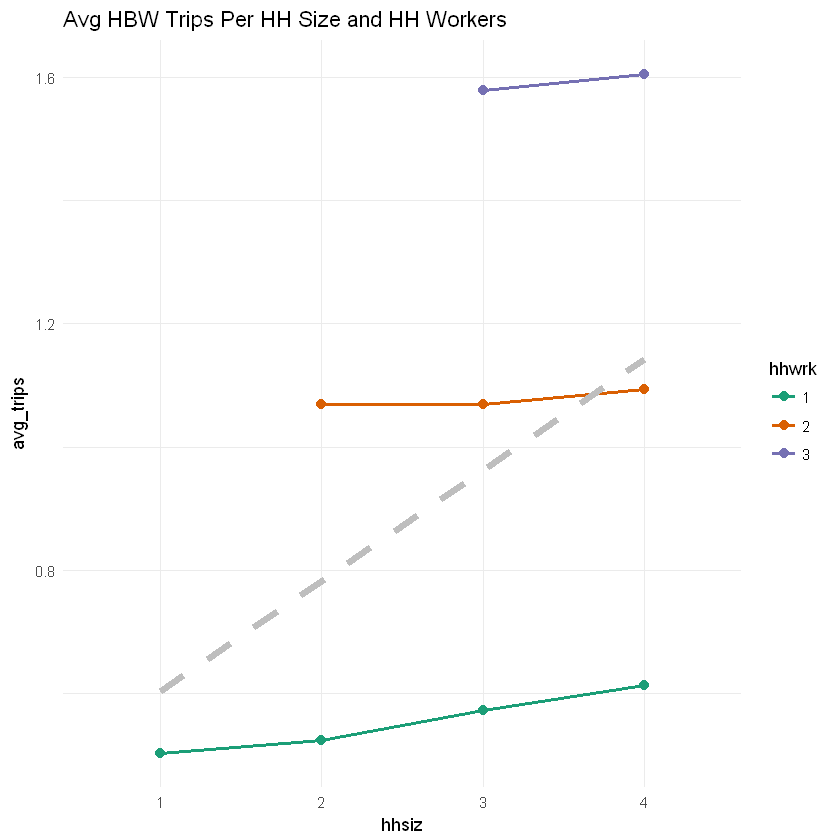
**Kevin Saavedra**

USP 587 – Homework 2

1. Develop Cross-Classification Trip Rate Tables for Household

This assignment was completed using R and Jupyter Notebook. The lm() function was used to calculate linear regression and checked against manual calculations for SSE, SST, R2, and F-statistics. A separate Appendix of these calculations is provided.

As this assignment involves HBW Trips, it was logical to begin with the number of workers per surveyed household size. While the provided dataset had counts of up to 5 workers per household, I decided to recategorize for 1, 2, or 3 or more workers, as this seems to be more in line with the average household. Initially, allowing a value of 0 for household workers led to a 0-cell issue, so the minimum worker count must be 1. I also opted to recategorize household size to 1, 2, 3, and 4+, again to better represent average household sizes.

*Table 1, Fig 1. CCA, Average HBW Trips by HH Size: HH Workers*

| **HH Size** | **HH Workers** | **Avg HBW Trips** |
| --- | --- | --- |
| 1 | 1 | 0.5025381 |
| 2 | 1 | 0.5240964 |
| 2 | 2 | 1.0689115 |
| 3 | 1 | 0.5731225 |
| 3 | 2 | 1.0696721 |
| 3 | 3 | 1.5786164 |
| 4 | 1 | 0.6123188 |
| 4 | 2 | 1.0927152 |
| 4 | 3 | 1.6047619 |

SSE = 3369.96520922889

SST = 3968.93572744015

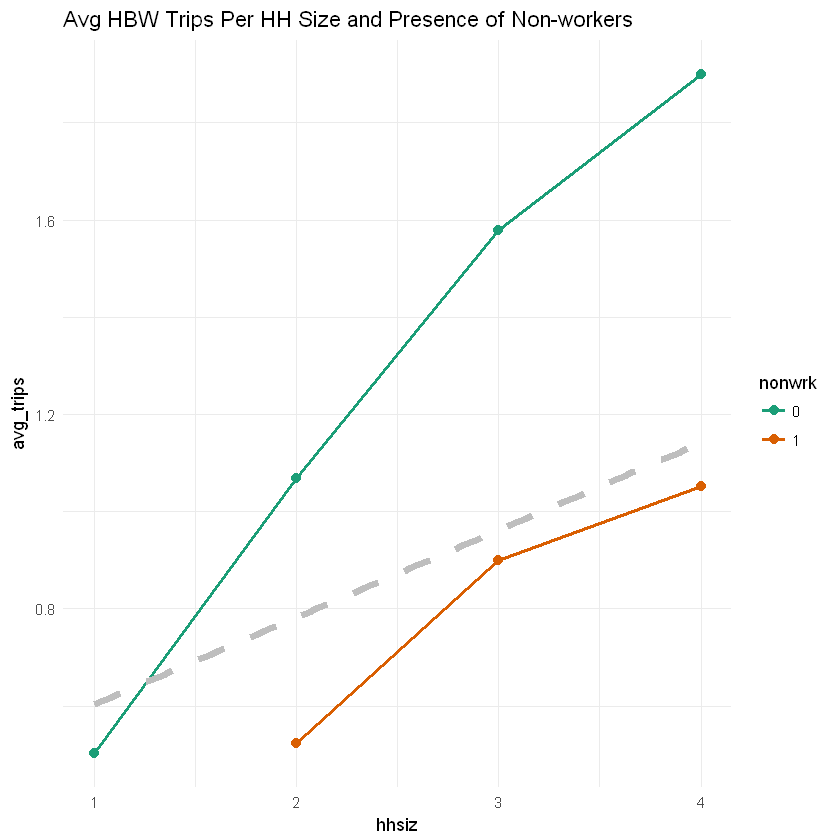
R2 = 0.150914642953307\*

Adj R2 = 0.149818240536903\*

*\*Note: could not generate all coefficients for factor*

*analysis due to collinearity errors.*

However, when attempting to calculate linear regression using R, several coefficients could not be calculated due to strong collinearity between variables. This makes sense, since there is often a 1:1 correlation between workers and household members. I then attempted a CCA using HH size and the presence of non-workers (Table 2, Fig 2). Non-workers were identified in a separate column with the dummy variable 1 if *n*HHSize > *n*HHWorkers.

*Table 2, Fig 2. CCA, Average HBW Trips by HH Size: Presence of Non-workers*

| **HH Size** | **Non-workers** | **Avg HBW Trips** |
| --- | --- | --- |
| 1 | 0 | 0.5025381 |
| 2 | 0 | 1.0689115 |
| 2 | 1 | 0.5240964 |
| 3 | 0 | 1.5786164 |
| 3 | 1 | 0.9001350 |
| 4 | 0 | 1.9000000 |
| 4 | 1 | 1.0520231 |

SSE = 3508.24607053155

SST = 3968.93572744015

R2 = 0.116073851668474

Adj R2 = 0.115258872737687

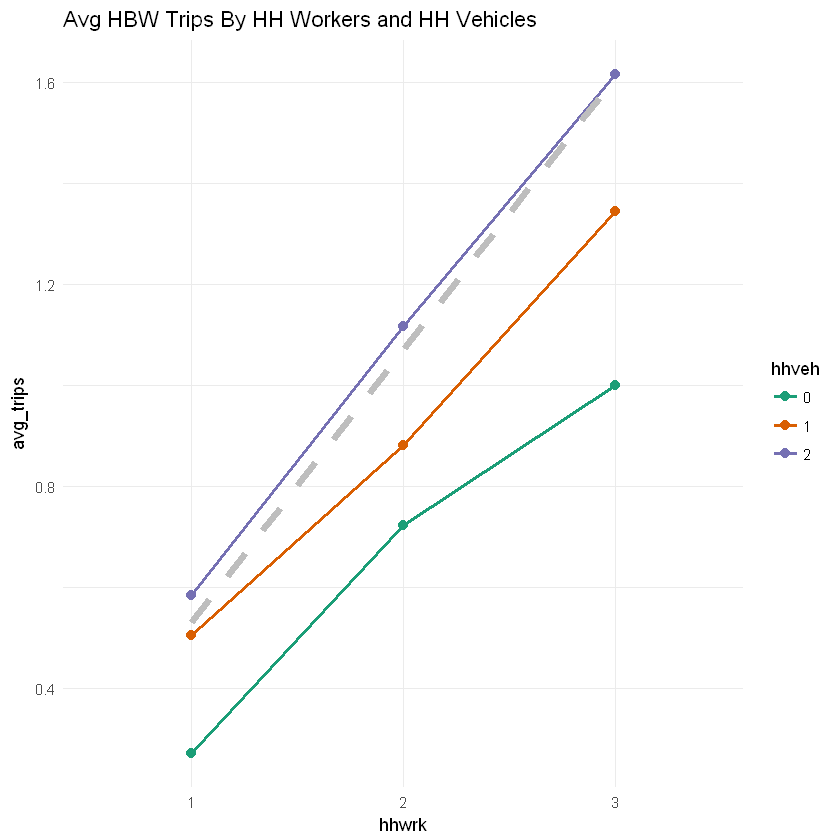
In this CCA, the presence of non-workers showed fewer HBW trips generated per household, which is intuitive. However, this model did not appear to have a very good predictive ability, indicated by the adjusted R2 value of 0.1152.

Next, I tried a CCA using the number of workers with the number of vehicles per household (Table 3, Fig 3), under the assumption that the presence of a household vehicle will lead to higher HBW trips. In a first attempt, I compared HH size to the *presence* of a household vehicle (dummy variable 0 or 1). This produced an adjusted R2 value of .1533. Next, I tried a CCA with HH workers against the *number* of vehicles. This actually produced a higher R2, .1587, which is reproduced below. The following section will compare the model using the presence of a household vehicle to the model using categorical variables representing the number of household vehicles.

As a check against HH vehicles, I performed a CCA against HH workers and the number of licensed drivers per household (Table 4, Fig 4). This did not produce as high of a value for R2 as HH workers and number of household vehicles.

*Table 3, Fig 3: CCA, Average HBW Trips by HH Workers: HH Vehicles*

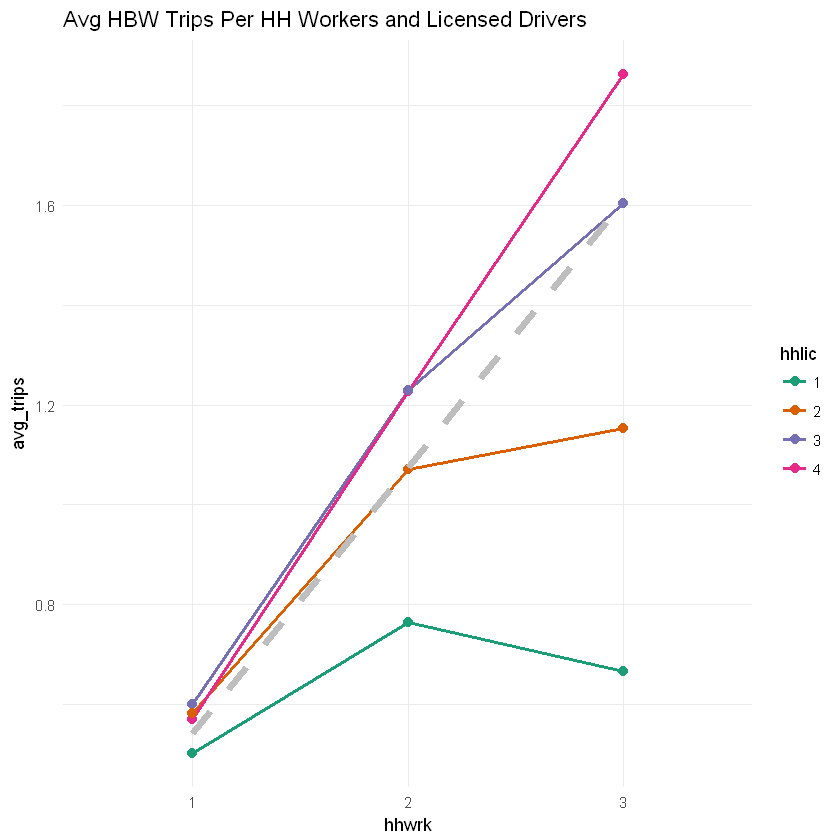
| **HH Workers** | **HH Vehicles** | **AVG HBW Trips** |
| --- | --- | --- |
| 1 | 0 | 0.2716049 |
| 1 | 1 | 0.5060554 |
| 1 | 2 | 0.5838103 |
| 2 | 0 | 0.7222222 |
| 2 | 1 | 0.8817204 |
| 2 | 2 | 1.1164773 |
| 3 | 0 | 1.0000000 |
| 3 | 1 | 1.3448276 |
| 3 | 2 | 1.6165192 |

SSE = 3334.23016798233

SST = 3968.93572744015

R2 = 0.159918326484764

adj R2 = 0.158678353534926

*Table 4, Fig 4: CCA, Average HBW Trips by HH Workers: Licensed Drivers*

| **HH Workers** | **Licenses** | **AVG HBW Trips** |
| --- | --- | --- |
| 1 | 1 | 0.5022388 |
| 1 | 2 | 0.5830904 |
| 1 | 3 | 0.5957447 |
| 2 | 1 | 0.7654321 |
| 2 | 2 | 1.0701999 |
| 2 | 3 | 1.2296820 |
| 3 | 1 | 0.6666667 |
| 3 | 2 | 1.1538462 |
| 3 | 3 | 1.6751592 |

SSE = 3313.09318881791

SST = 3917.58631008042

R2 = 0.154302438648786

Adj R2 = 0.153193430981221

1. Estimating a Linear Regression Model of Household Trip Generation

This section will compare the best-performing linear regression models in the preceding section. Linear regressions were calculated in R as with F-statistics calculated manually. Model 1 represents HH Workers: HH Drivers Licenses, Model 2 represents HH Workers: Presence of Household Vehicles, and Model 3 represents HH Workers: Number of Household Vehicles.

*Table 5: Comparison between Final Models.*



As models 2 and 3 showed the highest adjusted R-squared values, I computed the F-statistic with alpha=0.05 to compare both models. The result was an F-statistic of 26.26, exceeding the critical value of 2.6 for NR=3 and N-J=inf. This validates Model 3 as a better predictor of HBW trips.

As a result of this analysis, my proposed model is to use HH workers and number of household vehicles for a reasonable estimate of home-based-work trips. Full formulas and computations in R can be found in separate Appendix.