Advanced Transportation Modeling and Statistics

Homework #3

Hossein Zamani Saghazadeh

29423388

University of Hawaii at Manoa

Department of Civil and Environmental Engineering

Initial Estimation

First, we'll catalog every attribute in the dataset by name, description and by type - Alternative-Varying (Generic) Variables, Individual-Specific Variables, and Mode Dummies - since this classification drives our Maximum Likelihood Estimation setup:

- **Alternative-Varying (Generic) Variables**: They are vary by alternative for the same person. Usually share the same parameter across modes. Examples could be Travel Time and Travel Cost.
- **Individual-Specific Variables**: They are Constant across alternatives for the same person, such as Household Size and Age. To include them, we interact them with mode dummies, converting them into Alternative-Specific Variables.
- **Mode Dummies**: These are a special case, like UNOTRN and UNOSHR. Technically, they are Alternative-Specific Variables because they vary across alternatives. In modeling terms, they're most often used to estimate Alternative-Specific Constants (ASCs) or to interact them with Individual-Specific Variables to create Alternative-Specific Variables.

- **PERSID**: Person Identification Number (not an attribute)
- **UNO**: 1 for all observations
- **IVTT**: In-Vehicle Travel Time (minutes) (Alternative-Varying (Generic) Variable, aka Level-of-Service (LOS) Variable)
- **OVTT**: Out-of-Vehicle Travel Time (minutes) (Alternative-Varying (Generic) Variable, aka Level-of-Service (LOS) Variable)
- **TOTTIME**: Total Travel Time (minutes) (Alternative-Varying (Generic) Variable, aka Level-of-Service (LOS) Variable) (TOTTIME = IVTT + OVTT)
- **TOTCOST**: Total Travel Cost ($) (Alternative-Varying (Generic) Variable, aka Level-of-Service (LOS) Variable)
- **WEIGHT**: Weight to be given to the observation.
- **UNOTRN**: 1 if record represents transit alternative, 0 otherwise (Mode Dummy, aka Alternative Specific Constant (ASC))
- **UNOSHR**: 1 if record represents shared alternative, 0 otherwise. (Mode Dummy, aka Alternative Specific Constant (ASC))
- **UNOCAR**: 1 if record represents car alternative, 0 otherwise. (Mode Dummy, aka Alternative Specific Constant (ASC))
- **ALTS**: Car if UNOCAR == 1, Shared if UNOSHR == 1, Train if UNOTRN == 1
- **CHOSEN**: Mode Chosen (Dependent Variable)
- **TDUMMY**: 1 if use of transit entails one or more transfers, 0 otherwise (Individual-Specific Variable)
- **TDUMMY_TRN**: Alternative-Specific version of TDUMMY for Train mode (Alternative-Specific Variable)
- **TDUMMY_SHR**: Alternative-Specific version of TDUMMY for Shared mode (Alternative-Specific Variable)
- **TDUMMY_CAR**: Alternative-Specific version of TDUMMY for Car mode (Alternative-Specific Variable)
- **HHSIZE**: Household Size (Individual-Specific Variable)
- **HHSIZE_TRN**: Alternative-Specific version of HHSIZE for Train mode (Alternative-Specific Variable)

- **HHSIZE_SHR**: Alternative-Specific version of HHSIZE for Shared mode (Alternative-Specific Variable)
- **HHSIZE_CAR**: Alternative-Specific version of HHSZIE for Car mode (Alternative-Specific Variable)
- **MALE**: 1 if individual is male, 0 otherwise (Individual-Specific Variable)
- **MALE_TRN**: Alternative-Specific version of MALE for Train mode (Alternative-Specific Variable)
- **MALE_SHR**: Alternative-Specific version of MALE for Shared mode (Alternative-Specific Variable)
- **MALE_CAR**: Alternative-Specific version of MALE for Car mode (Alternative-Specific Variable)
- **NUMVEH**: Number of vehicles in worker's household (Individual-Specific Variable)
- **NUMVEH_TRN**: Alternative-Specific version of NUMVEH for Train mode (Alternative-Specific Variable)
- **NUMVEH_SHR**: Alternative-Specific version of NUMVEH for Shared mode (Alternative-Specific Variable)
- **NUMVEH_CAR**: Alternative-Specific version of NUMVEH for Car mode (Alternative-Specific Variable)
- **WORKERS**: Number of workers in worker's household (Individual-Specific Variable)
- **WORKERS_TRN**: Alternative-Specific version of WORKERS for Train mode (Alternative-Specific Variable)
- **WORKERS_SHR**: Alternative-Specific version of WORKERS for Shared mode (Alternative-Specific Variable)
- **WORKERS_CAR**: Alternative-Specific version of WORKERS for Car mode (Alternative-Specific Variable)
- **VEHWORK**: Vehicles per worker in worker's household (Individual-Specific Variable) (VEHWORK = NUMVEH/WORKERS)
- **VEHWORK_TRN**: Alternative-Specific version of VEHWORK for Train mode (Alternative-Specific Variable)
- **VEHWORK_SHR**: Alternative-Specific version of VEHWORK for Shared mode (Alternative-Specific Variable)
- **VEHWORK_CAR**: Alternative-Specific version of VEHWORK for Car mode (Alternative-Specific Variable)
- **NEWINC**: Household income ($/year) (Individual-Specific Variable)
- **NEWINC_TRN**: Alternative-Specific version of NEWINC for Train mode (Alternative-Specific Variable)
- **NEWINC_SHR**: Alternative-Specific version of NEWINC for Shared mode (Alternative-Specific Variable)
- **NEWINC_CAR**: Alternative-Specific version of NEWINC for Car mode (Alternative-Specific Variable)
- **AEMPDENS**: Employment density at worker's employment place (Individual-Specific Variable)
- **AEMPDENS_TRN**: Alternative-Specific version of AEMPDENS for Train mode (Alternative-Specific Variable)

- **AEMPDENSC_SHR**: Alternative-Specific version of AEMPDENS for Shared mode (Alternative-Specific Variable)
- **AEMPDENS_CAR**: Alternative-Specific version of AEMPDENS for Car mode (Alternative-Specific Variable)
- **POPDENS**: Population density at worker's residence (Individual-Specific Variable)
- **POPDENS_TRN**: Alternative-Specific version of POPDENS for Train mode (Alternative-Specific Variable)
- **POPDENS_SHR**: Alternative-Specific version of POPDENS for Shared mode (Alternative-Specific Variable)
- **POPDENS_CAR**: Alternative-Specific version of POPDENS for Car mode (Alternative-Specific Variable)
- **DIST**: Travel distance to work (miles) (Individual-Specific Variable)
- **DIST_TRN**: Alternative-Specific version of DIST for Train mode (Alternative-Specific Variable)
- **DIST_SHR**: Alternative-Specific version of DIST for Shared mode (Alternative-Specific Variable)
- **DIST_CAR**: Alternative-Specific version of DIST for Car mode (Alternative-Specific Variable)
- **AGE**: Age of traveler (Individual-Specific Variable)
- **AGE_TRN**: Alternative-Specific version of AGE for Train mode (Alternative-Specific Variable)
- **AGE_SHR**: Alternative-Specific version of AGE for Shared mode (Alternative-Specific Variable)
- **AGE_CAR**: Alternative-Specific version of AGE for Car mode (Alternative-Specific Variable)

Base Model (Constants Only Model) Without Weights (m1)

For the base case, we consider the Constants-Only Model as the base case. The Constants-Only Model only uses Mode Dummy Variables. So we only use UNOTRN and UNOSHR for this model (UNOCRN is not used since we consider Car to be the base mode in our entire model). The deterministic part of the utilities will be:

$$V_{Car} = 0$$
$$V_{Train} = \beta_{UNOTRN}(UNOTRN)$$
$$V_{Shared} = \beta_{UNOSHR}(UNOSHR)$$

```
> ### Base Model (Constants Only Model) Without Weights (m1) ###
> m1 <- mlogit(CHOSEN~UNOTRN + UNOSHR|0, data = D)

Frequencies of alternatives:choice
    Car  Shared   Train
0.41689 0.17600 0.40711

Coefficients :
        Estimate Std. Error  z-value Pr(>|z|)
UNOTRN -0.023734   0.065693  -0.3613   0.7179
UNOSHR -0.862336   0.084751 -10.1750   <2e-16 ***

Log-Likelihood: -1165.9

> ## Determine Choice Probabilities ##
> Pr_Choice_m1<- predict(m1, newdata = D)
> apply(Pr_Choice_m1, 2, mean)
      Car    Shared     Train
0.4168889 0.1760000 0.4071111
```

As it is shown in the output of model **m1**, the model-predicted market shares are:

- **Car**: 0.417
- **Train**: 0.407
- **Shared**: 0.0.176

These match exactly with the observed shares in the dataset (Frequencies of alternatives:choice), confirming that the base model (**m1**) reproduces the sample proportions.

The estimated values for model parameters are as follows:

- $\beta_{UNOCAR} = 0$: This is the base mode.
- $\beta_{UNOTRN} = -0.023734$: This is the ASC for Train relative to Car. Train is less preferred than Car by 0.023734 utility units if all other variables are held constant. ($p = 0.7179$) suggests that it is statistically insignificant.
- $\beta_{UNOSHR} = -0.862336$: This is the ASC for Shared relative to Car. Shared is less preferred than Car by 0.862336 utility units if all other variables are held constant. $\left(p < 2e^{-16}\right)$ suggests that it is statistically highly significant.

Base Model (Constants Only Model) With Weights (m2)

Now we consider another base model (**m2**) with the same configuration as **m1**, but this time we enter Weights. Entering the weights will adjust for sampling bias and produce population-representative results. Using weights is for policy analysis and forecasting.

$$V_{Car} = 0$$
$$V_{Train} = \beta_{UNOTRN}(UNOTRN)$$
$$V_{Shared} = \beta_{UNOSHR}(UNOSHR)$$

```
> ### Base Model (Constants Only Model) With Weights (m2) ###
> m2 <- mlogit(CHOSEN~UNOTRN + UNOSHR|0, data = D, weights = WEIGHT)

Frequencies of alternatives:choice
    Car  Shared   Train
0.41689 0.17600 0.40711

Coefficients :
        Estimate Std. Error z-value  Pr(>|z|)
UNOTRN -1.841420   0.092104 -19.993 < 2.2e-16 ***
UNOSHR -1.912267   0.094977 -20.134 < 2.2e-16 ***

Log-Likelihood: -795.44

> ## Determine Choice Probabilities ##
> Pr_Choice_m2<- predict(m2, newdata = D)
> apply(Pr_Choice_m2, 2, mean)
      Car    Shared     Train
0.7654992 0.1130987 0.1214021
```

In Table 1 a short comparison between model **m1** and model **m2** is presented.

Aspect	**Without Weights (m1)**	**With Weights (m2)**
Log-Likelihood	−1165.9	−795.44 (higher = better fit)
UNOTRN (Train ASC)	−0.024 (not significant)	−1.84 (very significant)
UNOSHR (Shared ASC)	−0.86 (very significant)	−1.91 (very significant)
Implied Mode Shares	Car: 41.7%, Train: 40.7%, Shared: 17.6%	Car: 76.5%, Train: 12.1%, Shared: 11.3%
Observed Frequencies	41.7%, 17.6%, 40.7% (matches unweighted model)	Same — but weights reweight these

Table 1: Comparison of model results with and without weights

It can be seen that Log-likelihood is much better (-795 vs. -1165), indicating better model fit after weighting. Both ASCs are strongly negative and significant — showing that Train and Shared are much less preferred than Car when adjusted for population representation. Mode shares shift significantly toward Car (76.5%) — this suggests the population uses Car much more than the sample alone would suggest. For the rest of the modeling we stick to the weighted models as they are a better representative of real data.

Include LOS Attributes

Now, in addition to ASCs, we want to add LOS variables. There are four types of LOS variables in our dataset: IVTT, OVTT, TOTTIME, and TOTCOST. In this part we want to only use IVTT, OVTT, and TOTCOST.

Constants + LOS Model With Weights (IVTT + OVTT + TOTCOST) (m3_1)

For this model, the deterministic part of the utilities will be:

$$V_{Car} = \beta_{IVTT}(IVTT_{Car}) + \beta_{OVTT}(OVTT_{Car}) + \beta_{TOTCOST}(TOTCOST_{Car})$$

$$V_{Train} = \beta_{UNOTRN}(UNOTRN)+\beta_{IVTT}(IVTT_{Train})+\beta_{OVTT}(OVTT_{Train})+\beta_{TOTCOST}(TOTCOST_{Train})$$

$$V_{Shared} = \beta_{UNOSHR}(UNOSHR)+\beta_{IVTT}(IVTT_{Shared})+\beta_{OVTT}(OVTT_{Shared})+\beta_{TOTCOST}(TOTCOST_{Shared})$$

```
> ### Constants + LOS Model With Weights (IVTT + OVTT + TOTCOST) (m3_1) ###
> m3_1 <- mlogit(CHOSEN~UNOTRN + UNOSHR + IVTT + OVTT + TOTCOST|0, data = D, weights = WEIGHT)

Frequencies of alternatives:choice
    Car  Shared   Train
0.41689 0.17600 0.40711

Coefficients :
          Estimate Std. Error  z-value  Pr(>|z|)
UNOTRN  -0.4224333  0.2069273  -2.0415 0.0412053 *
UNOSHR  -2.1869327  0.1221828 -17.8989 < 2.2e-16 ***
IVTT    -0.0357600  0.0101001  -3.5405 0.0003993 ***
OVTT    -0.0579188  0.0099527  -5.8194 5.906e-09 ***
TOTCOST -0.7255516  0.0545064 -13.3113 < 2.2e-16 ***

Log-Likelihood: -630.97

> ## Determine Choice Probabilities ##
> Pr_Choice_m3_1<- predict(m3_1, newdata = D)
> apply(Pr_Choice_m3_1, 2, mean)
      Car    Shared     Train
0.6557455 0.1109085 0.2333460
```

The estimated values for model parameters are as follows:

- $\beta_{UNOCAR} = 0$: This is the base mode.
- $\beta_{UNOTRN} = -0.4224333$: This is the ASC for Train relative to Car. Train is less preferred than Car by 0.4224333 utility units if all other variables are held constant. $(p = 0.0412053)$ suggests that it is statistically significant.
- $\beta_{UNOSHR} = -2.1869327$: This is the ASC for Shared relative to Car. Shared is less preferred than Car by 2.1869327 utility units if all other variables are held constant. $\left(p < 2e^{-16}\right)$ suggests that it is statistically highly significant.
- $\beta_{IVTT} = -0.0357600$: Each additional minute in-vehicle decreases utility by 0.0357600 units. $(p = 0.0003993)$ suggests that it is statistically highly significant.
- $\beta_{OVTT} = -0.0579188$: Each additional minute out-vehicle decreases utility by 0.0579188 units. $\left(p = 5.906e^{-9}\right)$ suggests that it is statistically highly significant.
- $\beta_{TOTCOST} = -0.7255516$: Each additional dollar of travel cost decreases utility by 0.7255516 units. $\left(p < 2.2e^{-16}\right)$ suggests that it is statistically highly significant.

Now we go through relative insights between OVTT, IVTT, and TOTCOST.

- **OVTT and IVTT**: OVTT is approximately 62% more disutulity per minute than IVTT which completely make sense since waiting and walking feel worse than riding.

$$\frac{\beta_{OVTT}}{\beta_{IVTT}} = \frac{-0.0579188}{-0.0357600} = 1.62$$

- **Value of Time(VoT)**: The Value of Time is the marginal rate of substitution between time and money.
  - **In-Vehicle VoT**: How much people value in-vehicle time.

$$VoT_{IVTT} = \frac{\beta_{IVTT}}{\beta_{TOTCOST}} = \frac{-0.0357600}{-0.7255516} = 0.0493\ \$/min = 2.96\ \$/hour$$

  - **Out-of-Vehicle VoT**: How much people value out-of-vehicle time.

$$VoT_{OVTT} = \frac{\beta_{OVTT}}{\beta_{TOTCOST}} = \frac{-0.0579188}{-0.7255516} = 0.0798\ \$/min = 4.79\ \$/hour$$

To compare **m3_1** and **m2** we use three metric of **Log-Likelihood**, **Rho-Squared**, and **Log-Likelihood Ratio Test**.

- **Log-Likelihood**: In discrete choice models like multinomial logit, the log-likelihood is a measure of how well your model explains the observed choices. The likelihood measures how probable it is to observe your actual data given the model's predicted probabilities. Since likelihood values are very small, we take the log of it — hence, log-likelihood.

$$LL = \sum_{i=1}^{N} \log P_{chosen,i}$$

$$P_{chosen,i} = \text{probability our model assigns to the mode actually chosen by person } i$$

  Higher LL (closer to zero) means better model fit and predicts observed choices.

- **Rho-Squared**: Rho-squared is a common goodness-of-fit measure in discrete choice models like multinomial logit.

$$\rho^2 = 1 - \frac{LL_{full}}{LL_{null}}$$

  It tells us how much better the full model fits the data compared to a model with only constants. Values typically range between 0 and 0.4 (though values > 0.2 are often considered quite good in practice). Higher ρ^2 means better model fit.

- **Likelihood Ratio Test (LRT)**: The Likelihood Ratio Test is used to formally test whether adding new variables significantly improves model fit.

$$LR = -2\,(LL_{restricted} - LL_{full})$$

$$LL_{restricted} = \text{log-likelihood of the simpler model}$$

$$LL_{full} = \text{log-likelihood of the more complex model}$$

  Large LR value means big improvement in fit, so new variables are important. We should compare the LR statistic to a critical chi-squared value:

$$df = \text{number of added parameters}$$

$$\text{Compare} : LR \ \text{ vs } \ \chi^2_{0.95}(\text{df})$$

We can see that all metrics confirm that model **m3_1** is much better than model **m2**. In this case the value of Chi-Squared for 95% confidence and $df = 3$ is 7.814728. So if the value of LR is greater than this, then we are at the Reject Area. In Table 4 these two models are compared. Moreover, the result of LRT is displayed in Table 5

```
> ### Comparing m3_1 and m2 ###
> ## Log-Likelihood ##
> LL_m2 = m2$logLik[1]
> print(LL_m2)
[1] -795.4357
>
> LL_m3_1 = m3_1$logLik[1]
> print(LL_m3_1)
[1] -630.9664
>
> ## RHO-Squred ##
> RHO_SQ_m3_1 = 1-(LL_m3_1/LL_m2)
> print(RHO_SQ_m3_1)
[1] 0.2067663
>
> ## LRT ##
> LR_m3_1 = -2*(LL_m2 - LL_m3_1)
> print(LR_m3_1)
[1] 328.9386
> qchisq(0.95, df = 3)
[1] 7.814728
> lrtest(m2,m3_1)
Likelihood ratio test

Model 1: CHOSEN ~ UNOTRN + UNOSHR | 0
Model 2: CHOSEN ~ UNOTRN + UNOSHR + IVTT + OVTT + TOTCOST | 0
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   2 -795.44
2   5 -630.97  3 328.94  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Variable	**Model m2**			**Model m3_1**		
	Coefficient	t-statistic	Significance	Coefficient	t-statistic	Significance
UNOTRN	-1.8414	-19.9930	***	-0.4224	-2.0415	*
UNOSHR	-1.9123	-20.1340	***	-2.1869	-17.8989	***
IVTT	—	—	—	-0.0358	-3.5405	***
OVTT	—	—	—	-0.0579	-5.8194	***
TOTCOST	—	—	—	-0.7256	-13.3113	***
N – Sample Size	1125			1125		
LL – Constants	-795.4357			-795.4357		
LL – Convergence	-795.4357			-630.9664		
Rho-Squared	—			0.2068		

Table 2: Model Estimation Results for Model m2 and Model m3_1

Model m2	`CHOSEN ~ UNOTRN + UNOSHR	0`
Model m3_1	`CHOSEN ~ UNOTRN + UNOSHR + IVTT + OVTT + TOTCOST	0`

Model	**#Df**	**LogLik**	**Df**	**Chisq**	**Pr(> Chisq)**
m2	2	-795.44	—	—	—
m3_1	5	-630.97	3	328.94	$< 2.2 \times 10^{-16}$ ***

Table 3: Likelihood Ratio Test Comparing Model m2 and Model m3_1

Estimation and Interpretation of other Model Specifications (Part A)

There are 5 different ways that we can introduce LOS variables:

- **IVTT + OVTT + TOTCOST (m3_1)**: This model can be introduced as follow:

```
### Constants + LOS Model With Weights (IVTT + OVTT + TOTCOST) (m3_1) ###
m3_1 <- mlogit(CHOSEN~UNOTRN + UNOSHR + IVTT + OVTT + TOTCOST|0, data = D, weights = WEIGHT)
```

- **TOTCOST (m3_2)**: This model can be introduced as follow:

```
### Constants + LOS Model With Weights (TOTCOST) (m3_2) ###
m3_2 <- mlogit(CHOSEN~UNOTRN + UNOSHR + TOTCOST|0, data = D, weights = WEIGHT)
```

- **TOTTIME (m3_3)**: This model can be introduced as follow:

```
### Constants + LOS Model With Weights (TOTTIME) (m3_3) ###
m3_3 <- mlogit(CHOSEN~UNOTRN + UNOSHR + TOTTIME|0, data = D, weights = WEIGHT)
```

- **IVTT + OVTT (m3_4)**: This model can be introduced as follow:

```
### Constants + LOS Model With Weights (IVTT + OVTT) (m3_4) ###
m3_4 <- mlogit(CHOSEN~UNOTRN + UNOSHR + IVTT + OVTT|0, data = D, weights = WEIGHT)
```

- **TOTTIME + TOTCOST (m3_5)**: This model can be introduced as follow:

```
### Constants + LOS Model With Weights (TOTTIME + TOTCOST) (m3_5) ###
m3_5 <- mlogit(CHOSEN~UNOTRN + UNOSHR + TOTTIME + TOTCOST|0, data = D, weights = WEIGHT)
```

Constants + LOS Model With Weights (IVTT + OVTT + TOTCOST) (m3_1)

For this model, the deterministic part of the utilities will be:

$$V_{Car} = \beta_{IVTT}(IVTT_{Car}) + \beta_{OVTT}(OVTT_{Car}) + \beta_{TOTCOST}(TOTCOST_{Car})$$

$$V_{Train} = \beta_{UNOTRN}(UNOTRN)+\beta_{IVTT}(IVTT_{Train})+\beta_{OVTT}(OVTT_{Train})+\beta_{TOTCOST}(TOTCOST_{Train})$$

$$V_{Shared} = \beta_{UNOSHR}(UNOSHR)+\beta_{IVTT}(IVTT_{Shared})+\beta_{OVTT}(OVTT_{Shared})+\beta_{TOTCOST}(TOTCOST_{Shared})$$

```
> ### Constants + LOS Model With Weights (IVTT + OVTT + TOTCOST) (m3_1) ###
> m3_1 <- mlogit(CHOSEN~UNOTRN + UNOSHR + IVTT + OVTT + TOTCOST|0, data = D, weights = WEIGHT)

Frequencies of alternatives:choice
    Car  Shared   Train
0.41689 0.17600 0.40711

Coefficients :
          Estimate Std. Error  z-value  Pr(>|z|)
UNOTRN  -0.4224333  0.2069273  -2.0415 0.0412053 *
UNOSHR  -2.1869327  0.1221828 -17.8989 < 2.2e-16 ***
IVTT    -0.0357600  0.0101001  -3.5405 0.0003993 ***
OVTT    -0.0579188  0.0099527  -5.8194 5.906e-09 ***
TOTCOST -0.7255516  0.0545064 -13.3113 < 2.2e-16 ***
```

```
Log-Likelihood: -630.97

> ## Determine Choice Probabilities ##
> Pr_Choice_m3_1<- predict(m3_1, newdata = D)
> apply(Pr_Choice_m3_1, 2, mean)
      Car    Shared     Train
0.6557455 0.1109085 0.2333460
```

The estimated values for model parameters are as follows:

- $\beta_{UNOCAR} = 0$: This is the base mode.
- $\beta_{UNOTRN} = -0.4224333$: This is the ASC for Train relative to Car. Train is less preferred than Car by 0.4224333 utility units if all other variables are held constant. ($p = 0.0412053$) suggests that it is statistically significant.
- $\beta_{UNOSHR} = -2.1869327$: This is the ASC for Shared relative to Car. Shared is less preferred than Car by 2.1869327 utility units if all other variables are held constant. $\left(p < 2e^{-16}\right)$ suggests that it is statistically highly significant.
- $\beta_{IVTT} = -0.0357600$: Each additional minute in-vehicle decreases utility by 0.0357600 units. ($p = 0.0003993$) suggests that it is statistically highly significant.
- $\beta_{OVTT} = -0.0579188$: Each additional minute out-vehicle decreases utility by 0.0579188 units. $\left(p = 5.906e^{-9}\right)$ suggests that it is statistically highly significant.
- $\beta_{TOTCOST} = -0.7255516$: Each additional dollar of travel cost decreases utility by 0.7255516 units. $\left(p < 2.2e^{-16}\right)$ suggests that it is statistically highly significant.

Constants + LOS Model With Weights (TOTCOST) (m3_2)

For this model, the deterministic part of the utilities will be:

$$V_{Car} = \beta_{TOTCOST}(TOTCOST_{Car})$$
$$V_{Train} = \beta_{UNOTRN}(UNOTRN) + \beta_{TOTCOST}(TOTCOST_{Train})$$
$$V_{Shared} = \beta_{UNOSHR}(UNOSHR) + \beta_{TOTCOST}(TOTCOST_{Shared})$$

```
> ### Constants + LOS Model With Weights (TOTCOST) (m3_2) ###
> m3_2 <- mlogit(CHOSEN~UNOTRN + UNOSHR + TOTCOST|0, data = D, weights = WEIGHT)

Frequencies of alternatives:choice
    Car  Shared   Train
0.41689 0.17600 0.40711

Coefficients :
         Estimate Std. Error z-value  Pr(>|z|)
UNOTRN  -1.881799   0.111647 -16.855 < 2.2e-16 ***
UNOSHR  -2.573797   0.107503 -23.942 < 2.2e-16 ***
TOTCOST -0.883108   0.051295 -17.216 < 2.2e-16 ***

Log-Likelihood: -660.34

> ## Determine Choice Probabilities ##
> Pr_Choice_m3_2<- predict(m3_2, newdata = D)
```

```
> apply(Pr_Choice_m3_2, 2, mean)
      Car    Shared     Train
0.6659157 0.1183354 0.2157490
```

The estimated values for model parameters are as follows:

- $\beta_{UNOCAR} = 0$: This is the base mode.
- $\beta_{UNOTRN} = -1.881799$: This is the ASC for Train relative to Car. Train is less preferred than Car by 1.881799 utility units if all other variables are held constant. $\left(p < 2.2e^{-16}\right)$ suggests that it is statistically highly significant.
- $\beta_{UNOSHR} = -2.573797$: This is the ASC for Shared relative to Car. Shared is less preferred than Car by 2.573797 utility units if all other variables are held constant. $\left(p < 2e^{-16}\right)$ suggests that it is statistically highly significant.
- $\beta_{TOTCOST} = -0.883108$: Each additional dollar of travel cost decreases utility by 0.883108 units. $\left(p < 2.2e^{-16}\right)$ suggests that it is statistically highly significant.

Constants + LOS Model With Weights (TOTTIME) (m3_3)

For this model, the deterministic part of the utilities will be:

$$V_{Car} = \beta_{TOTTIME}(TOTTIME_{Car})$$
$$V_{Train} = \beta_{UNOTRN}(UNOTRN) + \beta_{TOTTIME}(TOTTIME_{Train})$$
$$V_{Shared} = \beta_{UNOSHR}(UNOSHR) + \beta_{TOTTIME}(TOTTIME_{Shared})$$

```
> ### Constants + LOS Model With Weights (TOTTIME) (m3_3) ###
> m3_3 <- mlogit(CHOSEN~UNOTRN + UNOSHR + TOTTIME|0, data = D, weights = WEIGHT)

Frequencies of alternatives:choice
    Car  Shared   Train
0.41689 0.17600 0.40711

Coefficients :
          Estimate Std. Error z-value Pr(>|z|)
UNOTRN   0.0965759  0.1364098   0.708    0.479
UNOSHR  -1.5332553  0.1049739 -14.606   <2e-16 ***
TOTTIME -0.0674244  0.0066052 -10.208   <2e-16 ***

Log-Likelihood: -702.93

> ## Determine Choice Probabilities ##
> Pr_Choice_m3_3<- predict(m3_3, newdata = D)
> apply(Pr_Choice_m3_3, 2, mean)
      Car    Shared     Train
0.7215586 0.1060082 0.1724332
```

The estimated values for model parameters are as follows:

- $\beta_{UNOCAR} = 0$: This is the base mode.
- $\beta_{UNOTRN} = 0.0965759$: This is the ASC for Train relative to Car. Train is almost as preferable as Car. $(p = 0.479)$ suggests that it is statistically insignificant.

- $\beta_{UNOSHR} = -1.5332553$: This is the ASC for Shared relative to Car. Shared is less preferred than Car by 1.5332553 utility units if all other variables are held constant. $\left(p < 2e^{-16}\right)$ suggests that it is statistically highly significant.

- $\beta_{TOTTIME} = -0.0674244$: Each additional minute of total travel time decreases utility by 0.0674244 units. $\left(p < 2.2e^{-16}\right)$ suggests that it is statistically highly significant.

Constants + LOS Model With Weights (IVTT + OVTT) (m3_4)

For this model, the deterministic part of the utilities will be:

$$V_{Car} = \beta_{IVTT}(IVTT_{Car}) + \beta_{OVTT}(OVTT_{Car})$$
$$V_{Train} = \beta_{UNOTRN}(UNOTRN) + \beta_{IVTT}(IVTT_{Train}) + \beta_{OVTT}(OVTT_{Train})$$
$$V_{Shared} = \beta_{UNOSHR}(UNOSHR) + \beta_{IVTT}(IVTT_{Shared}) + \beta_{OVTT}(OVTT_{Shared})$$

```
> ### Constants + LOS Model With Weights (IVTT + OVTT) (m3_4) ###
> m3_4 <- mlogit(CHOSEN~UNOTRN + UNOSHR + IVTT + OVTT|0, data = D, weights = WEIGHT)

Frequencies of alternatives:choice
    Car  Shared   Train
0.41689 0.17600 0.40711

Coefficients :
         Estimate Std. Error  z-value  Pr(>|z|)
UNOTRN  0.4127083  0.1629371   2.5329   0.01131 *
UNOSHR -1.5831470  0.1059236 -14.9461 < 2.2e-16 ***
IVTT   -0.0425227  0.0097606  -4.3566 1.321e-05 ***
OVTT   -0.0876252  0.0093369  -9.3848 < 2.2e-16 ***

Log-Likelihood: -698.55

> ## Determine Choice Probabilities ##
> Pr_Choice_m3_4<- predict(m3_4, newdata = D)
> apply(Pr_Choice_m3_4, 2, mean)
      Car    Shared     Train
0.7191072 0.1054292 0.1754636
```

The estimated values for model parameters are as follows:

- $\beta_{UNOCAR} = 0$: This is the base mode.

- $\beta_{UNOTRN} = 0.4127083$: This is the ASC for Train relative to Car. Train is more preferred than Car by 0.41270833 utility units if all other variables are held constant. $(p = 0.0412053)$ suggests that it is statistically significant.

- $\beta_{UNOSHR} = -1.5831470$: This is the ASC for Shared relative to Car. Shared is less preferred than Car by 1.5831470 utility units if all other variables are held constant. $\left(p < 2e^{-16}\right)$ suggests that it is statistically highly significant.

- $\beta_{IVTT} = -0.0425227$: Each additional minute in-vehicle decreases utility by 0.0425227 units. $\left(p = 1.321e^{-5}\right)$ suggests that it is statistically highly significant.

- $\beta_{OVTT} = -0.0876252$: Each additional minute out-vehicle decreases utility by 0.0876252 units. $\left(p < 2e^{-16}\right)$ suggests that it is statistically highly significant.

Constants + LOS Model With Weights (TOTTIME + TOTCOST) (m3_5)

For this model, the deterministic part of the utilities will be:

$$V_{Car} = \beta_{TOTTIME}(TOTTIME_{Car}) + \beta_{TOTCOST}(TOTCOST_{Car})$$
$$V_{Train} = \beta_{UNOTRN}(UNOTRN) + \beta_{TOTTIME}(TOTTIME_{Train}) + \beta_{TOTCOST}(TOTCOST_{Train})$$
$$V_{Shared} = \beta_{UNOSHR}(UNOSHR) + \beta_{TOTTIME}(TOTTIME_{Shared}) + \beta_{TOTCOST}(TOTCOST_{Shared})$$

```
> ### Constants + LOS Model With Weights (TOTTIME + TOTCOST) (m3_5) ###
> m3_5 <- mlogit(CHOSEN~UNOTRN + UNOSHR + TOTTIME + TOTCOST|0, data = D, weights = WEIGHT)

Frequencies of alternatives:choice
    Car  Shared   Train
0.41689 0.17600 0.40711

Coefficients :
          Estimate Std. Error  z-value  Pr(>|z|)
UNOTRN  -0.5981484  0.1790184  -3.3413  0.000834 ***
UNOSHR  -2.1690444  0.1211993 -17.8965 < 2.2e-16 ***
TOTTIME -0.0472162  0.0071324  -6.6199 3.594e-11 ***
TOTCOST -0.7391662  0.0547492 -13.5009 < 2.2e-16 ***

Log-Likelihood: -631.83

> ## Determine Choice Probabilities ##
> Pr_Choice_m3_1<- predict(m3_1, newdata = D)
> apply(Pr_Choice_m3_1, 2, mean)
      Car    Shared     Train
0.6557455 0.1109085 0.2333460
```

The estimated values for model parameters are as follows:

- $\beta_{UNOCAR} = 0$: This is the base mode.
- $\beta_{UNOTRN} = -0.5981484$: This is the ASC for Train relative to Car. Train is less preferred than Car by 0.5981484 utility units if all other variables are held constant. $(p = 0.000834)$ suggests that it is statistically highly significant.
- $\beta_{UNOSHR} = -2.1690444$: This is the ASC for Shared relative to Car. Shared is less preferred than Car by 2.1690444 utility units if all other variables are held constant. $\left(p < 2e^{-16}\right)$ suggests that it is statistically highly significant.
- $\beta_{TOTTIME} = -0.0472162$: Each additional minute of total travel time decreases utility by 0.0472162 units. $\left(p = 3.594e^{-11}\right)$ suggests that it is statistically highly significant.
- $\beta_{TOTCOST} = -0.7391662$: Each additional dollar of travel cost decreases utility by 0.7391662 units. $\left(p < 2.2e^{-16}\right)$ suggests that it is statistically highly significant.

Comparing All Models (m2, m3_1, m3_2, m3_3, m3_4, m3_5)

Up to now, we have a total of 6 different models:

- **Constants Only Model(m2)**: This the base model and can be introduced as follow:

```
### Base Model (Constants Only Model) With Weights (m2) ###
m2 <- mlogit(CHOSEN~UNOTRN + UNOSHR|0, data = D, weights = WEIGHT)
```

- **IVTT + OVTT + TOTCOST (m3_1)**: This model can be introduced as follow:

```
### Constants + LOS Model With Weights (IVTT + OVTT + TOTCOST) (m3_1) ###
m3_1 <- mlogit(CHOSEN~UNOTRN + UNOSHR + IVTT + OVTT + TOTCOST|0, data = D, weights = WEIGHT)
```

- **TOTCOST (m3_2)**: This model can be introduced as follow:

```
### Constants + LOS Model With Weights (TOTCOST) (m3_2) ###
m3_2 <- mlogit(CHOSEN~UNOTRN + UNOSHR + TOTCOST|0, data = D, weights = WEIGHT)
```

- **TOTTIME (m3_3)**: This model can be introduced as follow:

```
### Constants + LOS Model With Weights (TOTTIME) (m3_3) ###
m3_3 <- mlogit(CHOSEN~UNOTRN + UNOSHR + TOTTIME|0, data = D, weights = WEIGHT)
```

- **IVTT + OVTT (m3_4)**: This model can be introduced as follow:

```
### Constants + LOS Model With Weights (IVTT + OVTT) (m3_4) ###
m3_4 <- mlogit(CHOSEN~UNOTRN + UNOSHR + IVTT + OVTT|0, data = D, weights = WEIGHT)
```

- **TOTTIME + TOTCOST (m3_5)**: This model can be introduced as follow:

```
### Constants + LOS Model With Weights (TOTTIME + TOTCOST) (m3_5) ###
m3_5 <- mlogit(CHOSEN~UNOTRN + UNOSHR + TOTTIME + TOTCOST|0, data = D, weights = WEIGHT)
```

Our approach is to choose the most complex LOS model and compare it to all available models (base and other LOS models). If we find a better model, then we repeat this process. Finally, we choose the simplest model that has the best performance. Notice that **Rho-Squared** can be used to compare any model with the Constant Only Mode, and **LRT** can be used to compare restricted and unrestricted models. For comparing any two models, **Log-Likelihood** can be used. To summarize:

- **Log-Likelihood**: It can be used to compare:
  - **m2** and **m3_1**, **m3_2**, **m3_3**, **m3_4**, or **m3_5**
  - **m3_1** and **m3_2**, **m3_3**, **m3_4**, or **m3_5**
  - **m3_2** and **m3_3**, **m3_4**, or **m3_5**
  - **m3_3** and **m3_4**, or **m3_5**
  - **m3_4** and **m3_5**
- **Rho-Squared**: It can be used to compare:
  - **m2** and **m3_1**, **m3_2**, **m3_3**, **m3_4**, or **m3_5**
- **LRT**: It can be used to compare:
  - **m2** and **m3_1**, **m3_2**, **m3_3**, **m3_4**, or **m3_5**
  - **m3_3** and **m3_5**
  - **m3_1** and **m3_4**

m3_1 vs m2

We can see that all metrics (Log-Likelihood, Rho-Squared, and LRT) confirm that model **m3_1** is much better than model **m2**. In this case the value of Chi-Squared for 95% confidence and $df = 3$ is 7.814728. So if the value of LR is greater than this, then we are at the Reject Area. In Table 4 these two models are compared. Moreover, the result of LRT is displayed in Table 5

Variable	**Model m2**			**Model m3_1**		
	Coefficient	t-statistic	Significance	Coefficient	t-statistic	Significance
UNOTRN	-1.8414	-19.9930	***	-0.4224	-2.0415	*
UNOSHR	-1.9123	-20.1340	***	-2.1869	-17.8989	***
IVTT	—	—	—	-0.0358	-3.5405	***
OVTT	—	—	—	-0.0579	-5.8194	***
TOTCOST	—	—	—	-0.7256	-13.3113	***
N – Sample Size	1125			1125		
LL – Constants	-795.4357			-795.4357		
LL – Convergence	-795.4357			-630.9664		
Rho-Squared	—			0.2068		

Table 4: Model Estimation Results for Model m2 and Model m3_1

Model m2	`CHOSEN ~ UNOTRN + UNOSHR	0`
Model m3_1	`CHOSEN ~ UNOTRN + UNOSHR + IVTT + OVTT + TOTCOST	0`

Model	**#Df**	**LogLik**	**Df**	**Chisq**	**Pr(> Chisq)**
m2	2	-795.44	—	—	—
m3_1	5	-630.97	3	328.94	$< 2.2 \times 10^{-16}$ ***

Table 5: Likelihood Ratio Test Comparing Model m2 and Model m3_1

Final Chosen Model from This Comparison is m3_1

m3_1 vs m3_2

We can see that all metrics (Log-Likelihood) confirm that model **m3_1** is much better than model **m3_2**.In Table 6, these two models are compared.

Model m3_1	`CHOSEN ~ UNOTRN + UNOSHR + IVTT + OVTT + TOTCOST| 0`
Model m3_2	`CHOSEN ~ UNOTRN + UNOSHR + TOTCOST	0`

Variable	**Model m3_1**			**Model m3_2**		
	Coefficient	t-statistic	Significance	Coefficient	t-statistic	Significance
UNOTRN	-0.4224333	-2.0415	*	-1.881799	-16.855	***
UNOSHR	-2.1869327	-17.8989	***	-2.573797	-23.942	***
TOTCOST	-0.7255516	-13.3113	***	-0.883108	-17.216	***
IVTT	-0.0357600	-3.5405	***	—	—-	—
OVTT	-0.0579188	-5.8194	***	—	—	—
N – Sample Size	1125			1125		
LL – Convergence	-630.9664			-660.34		

Table 6: Model Estimation Results for Model m3_1 and Model m3_2

Final Chosen Model from This Comparison is m3_1

m3_1 vs m3_3

We can see that all metrics (Log-Likelihood) confirm that model **m3_1** is much better than model **m3_3**.In Table 7, these two models are compared.

Model m3_1	`CHOSEN ~ UNOTRN + UNOSHR + IVTT + OVTT + TOTCOST| 0`
Model m3_3	`CHOSEN ~ UNOTRN + UNOSHR + TOTTIME	0`

Variable	**Model m3_1**			**Model m3_3**		
	Coefficient	t-statistic	Significance	Coefficient	t-statistic	Significance
UNOTRN	-0.4224333	-2.0415	*	0.0965759	0.708	
UNOSHR	-2.1869327	-17.8989	***	-1.5332553	-14.606	***
TOTCOST	-0.7255516	-13.3113	***	—	—	***
IVTT	-0.0357600	-3.5405	***	—	—	—
OVTT	-0.0579188	-5.8194	***	—	—	—
TOTTIME	—	—	—	-0.0674244	-10.208	—
N – Sample Size	1125			1125		
LL – Convergence	-630.97			-702.93		

Table 7: Model Estimation Results for Model m3_1 and Model m3_3

Final Chosen Model from This Comparison is m3_1

m3_1 vs m3_4

We can see that all metrics (Log-Likelihood and LRT) confirm that model **m3_1** is much better than model **m3_4**. In this case, the value of Chi-Squared for 95% confidence and $df = 1$ is 3.841459. So if the value of LR is greater than this, then we are at the Reject Area. In Table 8, these two models are compared. Moreover, the result of LRT is displayed in Table 9

Variable	**Model m3_1**			**Model m3_4**		
	Coefficient	t-statistic	Significance	Coefficient	t-statistic	Significance
UNOTRN	-0.4224333	-2.0415	*	0.4127083	2.5329	*
UNOSHR	-2.1869327	-17.8989	***	-1.5831470	-14.9461	***
TOTCOST	-0.7255516	-13.3113	***	—	—	—
IVTT	-0.0357600	-3.5405	***	-0.0425227	-4.3566	***
OVTT	-0.0579188	-5.8194	***	-0.0876252	-9.3848	***
N – Sample Size	1125			1125		
LL – Convergence	-630.97			-698.55		

Table 8: Model Estimation Results for Model m3_1 and Model m3_4

Model m3_4	`CHOSEN ~ UNOTRN + UNOSHR + IVTT + OVTT	0`
Model m3_1	`CHOSEN ~ UNOTRN + UNOSHR + IVTT + OVTT + TOTCOST	0`

Model	**#Df**	**LogLik**	**Df**	**Chisq**	**Pr(> Chisq)**
m3_4	4	-698.55	—	—	—
m3_1	5	-630.97	1	135.16	$< 2.2 \times 10^{-16}$ ***

Table 9: Likelihood Ratio Test Comparing Model m3_1 and Model m3_4

Final Chosen Model from This Comparison is m3_1

m3_1 vs m3_5

We can see that all metrics (Log-Likelihood) confirm that model **m3_1** is as good as model **m3_5** (**m3_1** is slightly better than **m3_5**), meaning that splitting travel time does not improve our model and only adds complexity to it. In Table 10, these two models are compared.

Model m3_1	`CHOSEN ~ UNOTRN + UNOSHR + IVTT + OVTT + TOTCOST| 0`
Model m3_5	`CHOSEN ~ UNOTRN + UNOSHR + TOTTIME + TOTCOST | 0`

Variable	**Model m3_1**			**Model m3_5**		
	Coefficient	t-statistic	Significance	Coefficient	t-statistic	Significance
UNOTRN	-0.4224333	-2.0415	*	-0.5981484	-3.3413	***
UNOSHR	-2.1869327	-17.8989	***	-2.1690444	-17.8965	***
TOTCOST	-0.7255516	-13.3113	***	-0.7391662	-13.5009	***
IVTT	-0.0357600	-3.5405	***	—	—	—
OVTT	-0.0579188	-5.8194	***	—	—	—
TOTTIME	—	—	—	-0.0472162	-6.6199	***
N – Sample Size	1125			1125		
LL – Convergence	-630.97			-631.83		

Table 10: Model Estimation Results for Model m3_1 and Model m3_5

Final Chosen Model from This Comparison is m3_5

Final Chosen Model

So far, we have seen that **m2** is worse than **m3_1**, **m3_2**, **m3_3**, **m3_4**, and **m3_5**. Also, **m3_1** outplayed **m3_2**, **m3_3**, and **m3_4**. Making a comparison between **m3_2**, **m3_3**, and **m3_4** is pointless since no matter what the result might be, eventually they will be outplayed by **m3_1**. **m3_1** is slightly better than **m3_5** but has more complexity comparing to **m3_5** as it has more variables. Since **m3_1** and **m3_5** are equally good, **m3_5** will outperform other models, too. So, our final decision is to choose **m3_5** as the final model for **Constants + LOS Variables** model.

```
### Constants + LOS Model With Weights (TOTTIME + TOTCOST) (m3_5) ###
m3_5 <- mlogit(CHOSEN~UNOTRN + UNOSHR + TOTTIME + TOTCOST|0, data = D, weights = WEIGHT)
```

Estimation and Interpretation of other Model Specifications (Part B)

Our final model from the previous part is **m3_5** which is defined as follows:

```
### Constants + LOS Model With Weights (TOTTIME + TOTCOST) (m3_5) ###
m3_5 <- mlogit(CHOSEN~UNOTRN + UNOSHR + TOTTIME + TOTCOST|0, data = D, weights = WEIGHT)
```

In this part, we want to investigate the effect of automobile availability on mode choice. Two Alternative-Specific Variables of (**NUMVEH_TRN** & **NUMVEH_SHR**) and (**VEHWORK_TRN** & **VEHWORK_SHR**) should be added to our model separately. Since the relationship of $\left(VEHWORK = \frac{NUMVEH}{WORKERS}\right)$ exists between these two sets of variables, so we cannot add them simultaneously in one model. We need to create two new models based on these and compare those models to see which one has better performance. We create these two models based on **m3_5**.

Constants + LOS Model With Weights (TOTTIME + TOTCOST) + ASV (NUMVEH_TRN + NUMVEH_SHR) (m4)

```
> ### Constants + LOS Model With Weights (TOTTIME + TOTCOST)
+ ASV (NUMVEH_TRN + NUMVEH_SHR) (m4) ###
> m4 <- mlogit(CHOSEN~UNOTRN + UNOSHR + TOTTIME + TOTCOST
+ NUMVEH_TRN + NUMVEH_SHR|0, data = D, weights = WEIGHT)

Frequencies of alternatives:choice
    Car  Shared   Train
0.41689 0.17600 0.40711

Coefficients :
            Estimate Std. Error  z-value  Pr(>|z|)
UNOTRN      0.518886   0.263341   1.9704   0.04879 *
UNOSHR     -1.948394   0.246233  -7.9128 2.442e-15 ***
TOTTIME    -0.041332   0.006951  -5.9461 2.746e-09 ***
TOTCOST    -0.740227   0.055438 -13.3523 < 2.2e-16 ***
NUMVEH_TRN -0.704828   0.132303  -5.3274 9.963e-08 ***
NUMVEH_SHR -0.127087   0.113853  -1.1162   0.26432

Log-Likelihood: -619.17

> ## Determine Choice Probabilities ##
> Pr_Choice_m4<- predict(m4, newdata = D)
> apply(Pr_Choice_m4, 2, mean)
      Car    Shared     Train
0.6484332 0.1116444 0.2399224
```

The estimated values for model parameters are as follows:

- $\beta_{UNOCAR} = 0$: This is the base mode.
- $\beta_{UNOTRN} = 0.518886$: This is the ASC for Train relative to Car. Train is more preferred than Car by 0.518886 utility units if all other variables are held constant. ($p = 0.04879$) suggests that it is statistically significant.
- $\beta_{UNOSHR} = -1.948394$: This is the ASC for Shared relative to Car. Shared is less preferred than Car by 1.948394 utility units if all other variables are held constant. $\left(p = 2.442e^{-15}\right)$ suggests that it is statistically highly significant.
- $\beta_{TOTTIME} = -0.041332$: Each additional minute of total travel time decreases utility by 0.041332 units. $\left(p = 2.746e^{-9}\right)$ suggests that it is statistically highly significant.

- $\beta_{TOTCOST} = -0.740227$: Each additional dollar of travel cost decreases utility by 0.740227 units. $\left(p < 2.2e^{-16}\right)$ suggests that it is statistically highly significant.
- $\beta_{NUMVEH_{TRN}} = -0.704828$: Each additional vehicle in a household decreases the utility of choosing Train by 0.740227 units. $\left(p = 9.963e^{-8}\right)$ suggests that it is statistically highly significant.
- $\beta_{NUMVEH_{SHR}} = -0.127087$: Each additional vehicle in a household decreases the utility of choosing Shared by 0.127087 units. $(p = 0.26432)$ suggests that it is statistically insignificant.

Constants + LOS Model With Weights (TOTTIME + TOTCOST) + ASV (VEHWORK_TRN + VEHWORK_SHR) (m5)

```
> ### Constants + LOS Model With Weights (TOTTIME + TOTCOST)
+ ASV (VEHWORK_TRN + VEHWORK_SHR) (m5) ###
> m5 <- mlogit(CHOSEN~UNOTRN + UNOSHR + TOTTIME + TOTCOST
+ VEHWORK_TRN + VEHWORK_SHR|0, data = D, weights = WEIGHT)

Frequencies of alternatives:choice
    Car  Shared   Train
0.41689 0.17600 0.40711

Coefficients :
              Estimate Std. Error  z-value  Pr(>|z|)
UNOTRN       0.2195975  0.2500251   0.8783 0.3797800
UNOSHR      -1.3473335  0.2424297  -5.5576 2.735e-08 ***
TOTTIME     -0.0433935  0.0070787  -6.1301 8.780e-10 ***
TOTCOST     -0.7667358  0.0564504 -13.5825 < 2.2e-16 ***
VEHWORK_TRN -0.9520546  0.2208039  -4.3118 1.620e-05 ***
VEHWORK_SHR -0.8403298  0.2332052  -3.6034 0.0003141 ***

Log-Likelihood: -621.22

> ## Determine Choice Probabilities ##
> Pr_Choice_m5<- predict(m5, newdata = D)
> apply(Pr_Choice_m5, 2, mean)
      Car    Shared     Train
0.6486749 0.1152260 0.2360991
```

The estimated values for model parameters are as follows:

- $\beta_{UNOCAR} = 0$: This is the base mode.
- $\beta_{UNOTRN} = 0.2195975$: This is the ASC for Train relative to Car. Train is more preferred than Car by 0.2195975 utility units if all other variables are held constant. $(p = 0.3797800)$ suggests that it is statistically insignificant.
- $\beta_{UNOSHR} = -1.34733354$: This is the ASC for Shared relative to Car. Shared is less preferred than Car by 1.34733354 utility units if all other variables are held constant. $\left(p = 2.735e^{-8}\right)$ suggests that it is statistically highly significant.
- $\beta_{TOTTIME} = -0.0433935$: Each additional minute of total travel time decreases utility by 0.0433935 units. $\left(p = 28.780e^{-10}\right)$ suggests that it is statistically highly significant.
- $\beta_{TOTCOST} = -0.7667358$: Each additional dollar of travel cost decreases utility by 0.7667358 units. $\left(p < 2.2e^{-16}\right)$ suggests that it is statistically highly significant.
- $\beta_{VEHWORK_{TRN}} = -0.9520546$: Each additional vehicle per worker decreases the utility of choosing Train by 0.9520546 units. $\left(p = 1.620e^{-5}\right)$ suggests that it is statistically highly significant.

- $\beta_{VEHWORK_{SHR}} = -0.8403298$: Each additional vehicle per worker decreases the utility of choosing Shared by 0.8403298 units. ($p = 0.0003141$) suggests that it is statistically highly significant.

Now we can compare these models to decide which one to choose as our final model for this part. First, we compare each of **m_4** and **m_5** with **m3_5** using **Log-Likelihood** and **LRT** metrics. Then we can make a comparison between **m_4** and **m_5** using **Log-Likelihood** metric (note that we cannot use **LRT** for this comparison).

m3_5 vs m4

We can see that all metrics (Log-Likelihood and LRT) confirm that model **m4** is much better than model **m3_5**. In this case, the value of Chi-Squared for 95% confidence and $df = 2$ is 5.991465. So if the value of LR is greater than this, then we are at the Reject Area. In Table 11, these two models are compared. Moreover, the result of LRT is displayed in Table 12

Variable	**Model m3_5**			**Model m4**		
	Coefficient	t-statistic	Significance	Coefficient	t-statistic	Significance
UNOTRN	-0.5981484	-3.3413	***	0.518886	1.9704	*
UNOSHR	-2.1690444	-17.8965	***	-1.948394	-7.9128	***
TOTTIME	-0.0472162	-6.6199	***	-0.041332	-5.9461	***
TOTCOST	-0.7391662	-13.5009	***	-0.740227	-13.3523	***
NUMVEH_TRN	—	—	—	-0.704828	-5.3274	***
NUMVEH_SHR	—	—	—	-0.127087	-1.1162	
N – Sample Size	1125			1125		
LL – Convergence	-631.83			-619.17		

Table 11: Model Estimation Results for Model m3_5 and Model m4

Model m3_5	`CHOSEN ~ UNOTRN + UNOSHR + TOTTIME + TOTCOST | 0`
Model m4	`CHOSEN ~ UNOTRN + UNOSHR + TOTTIME + TOTCOST + NUMVEH_TRN + NUMVEH_SHR | 0`

Model	**#Df**	**LogLik**	**Df**	**Chisq**	**Pr(> Chisq)**
m3_5	4	-631.83	—	—	—
m4	6	-619.17	2	25.326	$3.167e^{-6}$ ***

Table 12: Likelihood Ratio Test Comparing Model m3_5 and Model m4

Final Chosen Model from This Comparison is m4

m3_5 vs m5

We can see that all metrics (Log-Likelihood and LRT) confirm that model **m5** is much better than model **m3_5**. In this case, the value of Chi-Squared for 95% confidence and $df = 2$ is 5.991465. So if the value of LR is greater than this, then we are at the Reject Area. In Table 13, these two models are compared. Moreover, the result of LRT is displayed in Table 14

Variable	Model m3_5			Model m5		
	Coefficient	t-statistic	Significance	Coefficient	t-statistic	Significance
UNOTRN	-0.5981484	-3.3413	***	0.2195975	0.8783	
UNOSHR	-2.1690444	-17.8965	***	1.3473335	-5.5576	***
TOTTIME	-0.0472162	-6.6199	***	-0.0433935	-6.1301	***
TOTCOST	-0.7391662	-13.5009	***	-0.7667358	-13.5825	***
VEHWORK_TRN	—	—	—	-0.9520546	-4.3118	***
VEHWORK_SHR	—	—	—	-0.8403298	-3.6034	***
N – Sample Size	1125			1125		
LL – Convergence	-631.83			-621.22		

Table 13: Model Estimation Results for Model m3_5 and Model m5

Model m3_5	`CHOSEN ~ UNOTRN + UNOSHR + TOTTIME + TOTCOST	0`
Model m5	`CHOSEN ~ UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR	0`

Model	#Df	LogLik	Df	Chisq	Pr(> Chisq)
m3_5	4	-631.83	—	—	—
m5	6	-621.22	2	21.226	$2.467e^{-5}$ ***

Table 14: Likelihood Ratio Test Comparing Model m3_5 and Model m5

Final Chosen Model from This Comparison is m5

Final Result

We can see comparing to the base model of **m3_5**, both **m4** and **m5** have better performance. The only metric that can be compared between these two models is **Log-Likelihood**. The value of log-likelihood for model **m4** is better than **m5** by the value of 2.05, meaning that considering this metric, these two models' performance are equal. In model **m4** NUMVEH_SHR is insignificant, which means that each additional vehicle in the household has no effect on the likelihood of choosing Shared mode, which seems to be unrealistic. On the other hand, in model **m5** both VEHWORK_TRN and VEHWORK_SHR are highly significant meaning that number of vehicles per worker highly influence the likelihood of choosing Train and Shared mode. So we choose **m5** since it provides a more behaviorally rich and interpretable representation of car availability.

Estimation and Interpretation of other Model Specifications (Part C)

So far, we have the following model (**m5**):

```
### Constants + LOS Model With Weights (TOTTIME + TOTCOST) +
ASV (VEHWORK_TRN + VEHWORK_SHR) (m5) ###
m5 <- mlogit(CHOSEN~UNOTRN + UNOSHR + TOTTIME + TOTCOST
+ VEHWORK_TRN + VEHWORK_SHR|0, data = D, weights = WEIGHT)
```

Recall from before that all the variables we have are as follows:

- **ASC**
  - **UNOTRN**
  - **UNOSHR**
- **LOS**
  - **TOTTIME**
  - **TOTCOST**
- **ASV**
  - **TDUMMY_TRN**
  - **TDUMMY_SHR**
  - **HHSIZE_TRN**
  - **HHSIZE_SHR**
  - **MALE_TRN**
  - **MALE_SHR**
  - **NUMVEH_TRN**
  - **NUMVEH_SHR**
  - **WORKERS_TRN**
  - **WORKERS_SHR**
  - **VEHWORK_TRN**
  - **VEHWORK_SHR**
  - **NEWINC_TRN**
  - **NEWINC_SHR**
  - **AEMPDENS_TRN**
  - **AEMPDENSC_SHR**
  - **POPDENS_TRN**
  - **POPDENS_SHR**
  - **DIST_TRN**
  - **DIST_SHR**
  - **AGE_TRN**
  - **AGE_SHR**

So far we have reached to the best model (**m5**) by considering **ASC** and **LOS** variables as well as **NUMVEH**, **WORKERS**, and **VEHWORK** variables from the set of **ASVs**. Now, we want to improve our model by adding the rest of Alternative-Specific Variables. We do this by using the One-Factor-at-a-Time (OFAT) approach, meaning that at each time we only add one new variable to our base model and compare the new model with the base model. If we see a conspicuous improvement (through using **LRT**), we keep that variable; if not, we ignore it. Then we repeat this process by adding two new variables, and so on, until we reach the best model with the fewest possible number of variables (the simplest best model).

m5 vs m6_1

We can see that all metrics (Log-Likelihood and LRT) confirm that model **m6_1** is relatively better than **m5** (but not very significant). In this case, the value of Chi-Squared for 95% confidence and $df = 2$ is 5.991465. So if the value of LR is greater than this, then we are at the Reject Area. In Table 15, these two models are compared. Moreover, the result of LRT is displayed in Table 16

```
### Constants + LOS Model + ASV With Weights (m6_1) ###
m6_1 <- mlogit(CHOSEN~UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR
+ TDUMMY_TRN + TDUMMY_SHR|0, data = D, weights = WEIGHT)
```

Variable	**Model m5**			**Model m6_1**		
	Coefficient	t-statistic	Significance	Coefficient	t-statistic	Significance
UNOTRN	0.2195975	0.8783		0.2297407	0.9030	
UNOSHR	-1.3473335	-5.5576	***	-0.9677447	-3.8430	***
TOTTIME	-0.0433935	-6.1301	***	-0.0477759	-5.9949	***
TOTCOST	-0.7667358	-13.5825	***	-0.7998940	-13.7265	***
VEHWORK_TRN	-0.9520546	-4.3118	***	-0.9807941	-4.3718	***
VEHWORK_SHR	-0.8403298	-3.6034	***	-0.7878326	-3.4743	***
TDUMMY_TRN	—	—	—	0.3036804	1.3530	
TDUMMY_SHR	—	—	—	-0.6620641	-3.3671	***
N – Sample Size	1125			1125		
LL – Convergence	-621.22			-614.40		

Table 15: Model Estimation Results for Model m5 and Model m6_1

Model m5	`CHOSEN ~ UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR | 0`
Model m6_1	`CHOSEN ~ UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR + TDUMMY_TRN + TDUMMY_SHR | 0`

Model	**#Df**	**LogLik**	**Df**	**Chisq**	**Pr(> Chisq)**
m5	6	-621.22	—	—	—
m6_1	8	-614.40	2	13.634	0.001095 **

Table 16: Likelihood Ratio Test Comparing Model m5 and Model m6_1

Adding a transfer has no strong effect on Train utility in our sample, but significantly reduces the utility of choosing Shared Ride. Each required transfer on a Shared Ride alternative reduces utility by 0.662 units, and it is statistically significant. For Train, the effect is positive but not statistically significant, meaning there's no clear evidence that transfers affect Train utility in this model. Even though TDUMMY_TRN is not significant on its own, the overall model improves significantly, and the effect of TDUMMY_SHR is clearly meaningful. It's reasonable to retain both for behavioral completeness and model fit. **Final Chosen Model from This Comparison is m6_1**

m5 vs m6_2

We can see that all metrics (Log-Likelihood and LRT) confirm that model **m6_2** is not better than **m5**. In this case, the value of Chi-Squared for 95% confidence and $df = 2$ is 5.991465. So if the value of LR is greater than this, then we are at the Reject Area. In Table 17, these two models are compared. Moreover, the result of LRT is displayed in Table 18

```
### Constants + LOS Model + ASV With Weights (m6_2) ###
m6_2 <- mlogit(CHOSEN~UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR
+ HHSIZE_TRN + HHSIZE_SHR|0, data = D, weights = WEIGHT)
```

Variable	**Model m5**			**Model m6_2**		
	Coefficient	t-statistic	Significance	Coefficient	t-statistic	Significance
UNOTRN	0.2195975	0.8783		0.4767933	1.4169	
UNOSHR	-1.3473335	-5.5576	***	-1.7339242	-5.0704	***
TOTTIME	-0.0433935	-6.1301	***	-0.0419619	-5.9360	***
TOTCOST	-0.7667358	-13.5825	***	-0.7675185	-13.5779	***
VEHWORK_TRN	-0.9520546	-4.3118	***	-0.9957913	-4.4249	***
VEHWORK_SHR	-0.8403298	-3.6034	***	-0.7986889	-3.4807	***
HHSIZE_TRN	—	—	—	-0.0857726	-1.1335	
HHSIZE_SHR	—	—	—	0.1077539	1.5010	
N – Sample Size	1125			1125		
LL – Convergence	-621.22			-619.2		

Table 17: Model Estimation Results for Model m5 and Model m6_2

Model m5	`CHOSEN ∼ UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR	0`
Model m6_2	`CHOSEN ∼ UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR + HHSIZE_TRN + HHSIZE_SHR	0`

Model	**#Df**	**LogLik**	**Df**	**Chisq**	**Pr(> Chisq)**
m5	6	-621.22	—	—	—
m6_2	8	-619.2	2	4.0482	0.1321

Table 18: Likelihood Ratio Test Comparing Model m5 and Model m6_2

Adding a household size has no strong effect on Train and Shared utility in our sample. Neither HH-SIZE_TRN and HHSIZE_SHR are not significant on their own, and the overall model does not improve significantly. It's reasonable to exclude both from our model. **Final Chosen Model from This Comparison is m5**

m5 vs m6_3

We can see that all metrics (Log-Likelihood and LRT) confirm that model **m6_3** is not better than **m5**. In this case, the value of Chi-Squared for 95% confidence and $df = 2$ is 5.991465. So if the value of LR is greater than this, then we are at the Reject Area. In Table 19, these two models are compared. Moreover, the result of LRT is displayed in Table 20

```
### Constants + LOS Model + ASV With Weights (m6_3) ###
m6_3 <- mlogit(CHOSEN~UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR
+ MALE_TRN + MALE_SHR|0, data = D, weights = WEIGHT)
```

Variable	Model m5			Model m6_3		
	Coefficient	t-statistic	Significance	Coefficient	t-statistic	Significance
UNOTRN	0.2195975	0.8783		0.202348	0.7743	
UNOSHR	-1.3473335	-5.5576	***	-1.264412	-4.8691	***
TOTTIME	-0.0433935	-6.1301	***	-0.043417	-6.0850	***
TOTCOST	-0.7667358	-13.5825	***	-0.770975	-13.6160	***
VEHWORK_TRN	-0.9520546	-4.3118	***	-0.953650	-4.3230	***
VEHWORK_SHR	-0.8403298	-3.6034	***	-0.833421	-3.5577	***
MALE_TRN	—	—	—	0.036771	0.1873	
MALE_SHR	—	—	—	-0.166839	-0.8523	
N – Sample Size	1125			1125		
LL – Convergence	-621.22			-620.81		

Table 19: Model Estimation Results for Model m5 and Model m6_3

Model m5	`CHOSEN ∼ UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR	0`
Model m6_3	`CHOSEN ∼ UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR + MALE_TRN + MALE_SHR	0`

Model	#Df	LogLik	Df	Chisq	Pr(> Chisq)
m5	6	-621.22	—	—	—
m6_3	8	-620.81	2	0.8194	0.6639

Table 20: Likelihood Ratio Test Comparing Model m5 and Model m6_3

Adding a gender has no strong effect on Train and Shared utility in our sample. Neither MALE_TRN and MALE_SHR are not significant on their own, and the overall model does not improve significantly. It's reasonable to exclude both from our model. **Final Chosen Model from This Comparison is m5**

m5 vs m6_4

We can see that all metrics (Log-Likelihood and LRT) confirm that model **m6_4** is not better than **m5**. In this case, the value of Chi-Squared for 95% confidence and $df = 2$ is 5.991465. So if the value of LR is greater than this, then we are at the Reject Area. In Table 21, these two models are compared. Moreover, the result of LRT is displayed in Table 22

```
### Constants + LOS Model + ASV With Weights (m6_4) ###
m6_4 <- mlogit(CHOSEN~UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR
+ NEWINC_TRN + NEWINC_SHR|0, data = D, weights = WEIGHT)
```

Variable	Model m5			Model m6_4		
	Coefficient	t-statistic	Significance	Coefficient	t-statistic	Significance
UNOTRN	0.2195975	0.8783		0.6300993	1.8068	
UNOSHR	-1.3473335	-5.5576	***	-1.0622932	-3.1407	**
TOTTIME	-0.0433935	-6.1301	***	-0.0431437	-6.0866	***
TOTCOST	-0.7667358	-13.5825	***	-0.7713592	-13.6396	***
VEHWORK_TRN	-0.9520546	-4.3118	***	-0.9916113	-4.4119	***
VEHWORK_SHR	-0.8403298	-3.6034	***	-0.8616490	-3.6863	***
NEWINC_TRN	—	—	—	-0.0636713	-1.7197	
NEWINC_SHR	—	—	—	-0.0443942	-1.2140	
N – Sample Size	1125			1125		
LL – Convergence	-621.22			-619.58		

Table 21: Model Estimation Results for Model m5 and Model m6_4

Model m5	`CHOSEN ~ UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR	0`
Model m6_4	`CHOSEN ~ UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR + NEWINC_TRN + NEWINC_SHR	0`

Model	#Df	LogLik	Df	Chisq	Pr(> Chisq)
m5	6	-621.22	—	—	—
m6_4	8	-619.58	2	3.275	0.1945

Table 22: Likelihood Ratio Test Comparing Model m5 and Model m6_4

Adding an income has no strong effect on Train and Shared utility in our sample. Neither NEWINC_TRN and NEWINC_SHR are not significant on their own, and the overall model does not improve significantly. It's reasonable to exclude both from our model. **Final Chosen Model from This Comparison is m5**

m5 vs m6_5

We can see that all metrics (Log-Likelihood and LRT) confirm that model **m6_5** is relatively better than **m5** (but not very significant). In this case, the value of Chi-Squared for 95% confidence and $df = 2$ is 5.991465. So if the value of LR is greater than this, then we are at the Reject Area. In Table 23, these two models are compared. Moreover, the result of LRT is displayed in Table 24

```
### Constants + LOS Model + ASV With Weights (m6_5) ###
m6_5 <- mlogit(CHOSEN~UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR
+ AEMPDENS_TRN + AEMPDENS_SHR|0, data = D, weights = WEIGHT)
```

Variable	Model m5			Model m6_5		
	Coefficient	t-statistic	Significance	Coefficient	t-statistic	Significance
UNOTRN	0.2195975	0.8783		0.13450538	0.5324	
UNOSHR	-1.3473335	-5.5576	***	-1.30934544	-5.2689	***
TOTTIME	-0.0433935	-6.1301	***	-0.04075787	-5.6897	***
TOTCOST	-0.7667358	-13.5825	***	-0.70400054	-10.8364	***
VEHWORK_TRN	-0.9520546	-4.3118	***	-1.09377299	-4.7392	***
VEHWORK_SHR	-0.8403298	-3.6034	***	-0.80782028	-3.3623	***
AEMPDENS_TRN	—	—	—	0.00099798	2.6948	**
AEMPDENS_SHR	—	—	—	-0.00045362	-0.9828	
N – Sample Size		1125			1125	
LL – Convergence		-621.22			-617.41	

Table 23: Model Estimation Results for Model m5 and Model m6_5

Model m5	`CHOSEN ~ UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR	0`
Model m6_5	`CHOSEN ~ UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR + AEMPDENS_TRN + AEMPDENS_SHR	0`

Model	#Df	LogLik	Df	Chisq	Pr(> Chisq)
m5	6	-621.22	—	—	—
m6_5	8	-617.41	2	7.6189	0.02216 *

Table 24: Likelihood Ratio Test Comparing Model m5 and Model m6_5

Adding a population density at the workplace has a strong effect on Train utility in our sample, which makes sense as dense job centers are well-served by transit, but does not significantly affect the utility of choosing Shared Ride. Each unit of population density increase the utility of using Train by 0.00099798 units, and it is statistically significant. For sahred, the effect is negative but not statistically significant, meaning there's no clear evidence that population density at the workplace affects Shared utility in this model. Even though AEMPDENS_SHR is not significant on its own, the overall model improves significantly, and the effect of AEMPDENS_TRN is clearly meaningful. It's reasonable to retain both for behavioral completeness and model fit. **Final Chosen Model from This Comparison is m6_5**

m5 vs m6_6

We can see that all metrics (Log-Likelihood and LRT) confirm that model **m6_6** is not better than **m5**. In this case, the value of Chi-Squared for 95% confidence and $df = 2$ is 5.991465. So if the value of LR is greater than this, then we are at the Reject Area. In Table 25, these two models are compared. Moreover, the result of LRT is displayed in Table 26

```
### Constants + LOS Model + ASV With Weights (m6_6) ###
m6_6 <- mlogit(CHOSEN~UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR
+ POPDENS_TRN + POPDENS_SHR|0, data = D, weights = WEIGHT)
```

Variable	Model m5			Model m6_6		
	Coefficient	t-statistic	Significance	Coefficient	t-statistic	Significance
UNOTRN	0.2195975	0.8783		0.3221458	1.0589	
UNOSHR	-1.3473335	-5.5576	***	-1.4061190	-5.1539	***
TOTTIME	-0.0433935	-6.1301	***	-0.0443016	-6.0291	***
TOTCOST	-0.7667358	-13.5825	***	-0.7718101	-13.5402	***
VEHWORK_TRN	-0.9520546	-4.3118	***	-0.9830730	-4.3592	***
VEHWORK_SHR	-0.8403298	-3.6034	***	-0.8168522	-3.4549	***
POPDENS_TRN	—	—	—	-0.0024535	-0.5461	
POPDENS_SHR	—	—	—	0.0027342	0.5755	
N – Sample Size		1125			1125	
LL – Convergence		-621.22			-620.92	

Table 25: Model Estimation Results for Model m5 and Model m6_6

Model m5	`CHOSEN ∼ UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR │ 0`
Model m6_6	`CHOSEN ∼ UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR + POPDENS_TRN + POPDENS_SHR │ 0`

Model	#Df	LogLik	Df	Chisq	Pr(> Chisq)
m5	6	-621.22	—	—	—
m6_6	8	-620.92	2	0.6059	0.7387

Table 26: Likelihood Ratio Test Comparing Model m5 and Model m6_6

Adding the population density at workers' residence has no strong effect on Train and Shared utility in our sample. Neither POPDENS_TRN and POPDENS_SHR are not significant on their own, and the overall model does not improve significantly. It's reasonable to exclude both from our model. **Final Chosen Model from This Comparison is m5**

m5 vs m6_7

We can see that all metrics (Log-Likelihood and LRT) confirm that model **m6_7** is relatively better than **m5** (but not very significant). In this case, the value of Chi-Squared for 95% confidence and $df = 2$ is 5.991465. So if the value of LR is greater than this, then we are at the Reject Area. In Table 27, these two models are compared. Moreover, the result of LRT is displayed in Table 28

```
### Constants + LOS Model + ASV With Weights (m6_7) ###
m6_7 <- mlogit(CHOSEN~UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR
+ DIST_TRN + DIST_SHR|0, data = D, weights = WEIGHT)
```

Variable	Model m5			Model m6_7		
	Coefficient	t-statistic	Significance	Coefficient	t-statistic	Significance
UNOTRN	0.2195975	0.8783		0.1656519	0.6596	
UNOSHR	-1.3473335	-5.5576	***	-1.2076902	-4.8026	***
TOTTIME	-0.0433935	-6.1301	***	-0.0521663	-6.7808	***
TOTCOST	-0.7667358	-13.5825	***	-0.7890058	-12.7728	***
VEHWORK_TRN	-0.9520546	-4.3118	***	-1.1739554	-5.1268	***
VEHWORK_SHR	-0.8403298	-3.6034	***	-0.7906384	-3.3468	***
DIST_TRN	—	—	—	0.0625074	4.1947	**
DIST_SHR	—	—	—	-0.0116079	-1.0606	
N – Sample Size	1125			1125		
LL – Convergence	-621.22			-613.07		

Table 27: Model Estimation Results for Model m5 and Model m6_7

Model m5	`CHOSEN ~ UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR | 0`
Model m6_7	`CHOSEN ~ UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR + DIST_TRN + DIST_SHR | 0`

Model	#Df	LogLik	Df	Chisq	Pr(> Chisq)
m5	6	-621.22	—	—	—
m6_7	8	-613.07	2	16.307	0.0002878 ***

Table 28: Likelihood Ratio Test Comparing Model m5 and Model m6_7

Adding distance to workplace has a strong effect on Train utility in our sample, which makes sense as Train is more appealing (time- and cost-wise) as the trip gets longer, but does not significantly affect the utility of choosing Shared Ride. Each unit of distance increases the utility of using Train by 0.0625074 unit, and it is statistically significant. For Shared, the effect is negative but not statistically significant, meaning there's no clear evidence that distance to the workplace affects Shared utility in this model. Even though DIST_SHR is not significant on its own, the overall model improves significantly, and the effect of DIST_TRN is clearly meaningful. It's reasonable to retain both for behavioral completeness and model fit. **Final Chosen Model from This Comparison is m6_7**

m5 vs m6_8

We can see that all metrics (Log-Likelihood and LRT) confirm that model **m6_8** is not better than **m5**. In this case, the value of Chi-Squared for 95% confidence and $df = 2$ is 5.991465. So if the value of LR is greater than this, then we are at the Reject Area. In Table 29, these two models are compared. Moreover, the result of LRT is displayed in Table 30

```
### Constants + LOS Model + ASV With Weights (m6_8) ###
m6_8 <- mlogit(CHOSEN~UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR
+ AGE_TRN + AGE_SHR|0, data = D, weights = WEIGHT)
```

Variable	Model m5			Model m6_8		
	Coefficient	t-statistic	Significance	Coefficient	t-statistic	Significance
UNOTRN	0.2195975	0.8783		0.7641121	2.0858	*
UNOSHR	-1.3473335	-5.5576	***	-1.1550654	3.0344	***
TOTTIME	-0.0433935	-6.1301	***	-0.0429191	-6.0467	***
TOTCOST	-0.7667358	-13.5825	***	-0.7598282	-13.4077	***
VEHWORK_TRN	-0.9520546	-4.3118	***	-0.8942475	-4.0186	***
VEHWORK_SHR	-0.8403298	-3.6034	***	-0.8271379	-3.5230	***
AGE_TRN	—	—	—	-0.0148916	-1.9320	
AGE_SHR	—	—	—	-0.0049136	-0.6278	
N – Sample Size	1125			1125		
LL – Convergence	-621.22			-619.79		

Table 29: Model Estimation Results for Model m5 and Model m6_8

Model m5	`CHOSEN ~ UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR | 0`
Model m6_8	`CHOSEN ~ UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR + AGE_TRN + AGE_SHR | 0`

Model	#Df	LogLik	Df	Chisq	Pr(> Chisq)
m5	6	-621.22	—	—	—
m6_8	8	-619.79	2	2.8691	0.2382

Table 30: Likelihood Ratio Test Comparing Model m5 and Model m6_8

Adding the age has no strong effect on Train and Shared utility in our sample. Neither AGE_TRN and AGE_SHR are not significant on their own, and the overall model does not improve significantly. It's reasonable to exclude both from our model. **Final Chosen Model from This Comparison is m5**

Final Result

We have seen that the following ASVs can be ignored: **HHSIZE**, **MALE**, **NEWINC**, **POPDENS**, and **AGE**. Also, the following variables have a significant effect on our base model (arranged from the highest significance to the lowest): **DIST**, **TDUMMY**, and **AEMPDENS**. Now, first we add **DIST** to our base model **m5**, which we already did, and create the new model **m6_7**, then we add **TDUMMY** to this model and create a new model **m6_7_1**, and finally we add **AEMPDENS** to **m6_7_1**, and create the last model **m6_7_1_1** to see which model is the best to choose.

m6_7 vs m6_7_1: We can see that all metrics (Log-Likelihood and LRT) confirm that model **m6_7_1** is better than **m6_7**. In this case, the value of Chi-Squared for 95% confidence and $df = 2$ is 5.991465. So if the value of LR is greater than this, then we are at the Reject Area. In Table 31, these two models are compared. Moreover, the result of LRT is displayed in Table 32

```
### Constants + LOS Model + ASV With Weights (m6_7_1) ###
m6_7_1 <- mlogit(CHOSEN~UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR
+ DIST_TRN + DIST_SHR + TDUMMY_TRN + TDUMMY_SHR|0, data = D, weights = WEIGHT)
```

Variable	Model m6_7			Model m6_7_1		
	Coefficient	t-statistic	Significance	Coefficient	t-statistic	Significance
UNOTRN	0.1656519	0.6596		0.1988650	0.7811	
UNOSHR	-1.2076902	-4.8026	***	-0.9034440	-3.5401	***
TOTTIME	-1.2076902	-6.7808	***	-0.0560085	-6.7689	***
TOTCOST	-0.7890058	-12.7728	***	-0.7849730	-12.6730	***
VEHWORK_TRN	-1.1739554	-5.1268	***	-1.1792527	-5.1305	***
VEHWORK_SHR	-0.7906384	-3.3468	***	-0.7843766	-3.3874	***
DIST_TRN	0.0625074	4.1947	**	0.0634342	4.1595	***
DIST_SHR	-0.0116079	-1.0606		0.0009521	0.0814	
TDUMMY_TRN	—	—	—	0.1746148	0.7548	
TDUMMY_SHR	—	—	—	-0.6687029	-3.0938	**
N – Sample Size	1125			1125		
LL – Convergence	-613.07			-607.60		

Table 31: Model Estimation Results for Model m6_7 and Model m6_7_1

Model m6_7	`CHOSEN ~ UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR + DIST_TRN + DIST_SHR | 0`
Model m6_7_1	`CHOSEN ~ UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR + DIST_TRN + DIST_SHR + TDUMMY_TRN + TDUMYY_SHR| 0`

Model	#Df	LogLik	Df	Chisq	Pr(> Chisq)
m6_7	8	-613.07	—	—	—
m6_7_1	10	-607.60	2	10.932	0.004228 **

Table 32: Likelihood Ratio Test Comparing Model m6_7 and Model m6_7_1

Adding a transfer has no strong effect on Train utility in our sample, but significantly reduces the utility of choosing Shared Ride. Each required transfer on a Shared Ride alternative reduces utility by 0.6687029 units, and it is statistically significant. For Train, the effect is positive but not statistically significant, meaning there's no clear evidence that transfers affect Train utility in this model. Even though TDUMMY_TRN is not significant on its own, the overall model improves significantly, and the effect of TDUMMY_SHR is clearly meaningful. It's reasonable to retain both for behavioral completeness and model fit.**Final Chosen Model from This Comparison is m6_7_1**

m6_7_1 vs m6_7_1_1: We can see that all metrics (Log-Likelihood and LRT) confirm that model **m6_7_1_1** is not better than **m6_7_1**. In this case, the value of Chi-Squared for 95% confidence and $df = 2$ is 5.991465. So if the value of LR is greater than this, then we are at the Reject Area. In Table 33, these two models are compared. Moreover, the result of LRT is displayed in Table 34

```
### Constants + LOS Model + ASV With Weights (m6_7_1_1) ###
m6_7_1_1 <- mlogit(CHOSEN ~ UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN +
VEHWORK_SHR + DIST_TRN + DIST_SHR + TDUMMY_TRN + TDUMMY_SHR +
AEMPDENS_TRN + AEMPDENS_SHR | 0, data = D, weights = WEIGHT)
```

Variable	**Model m6_7_1**			**Model m6_7_1_1**		
	Coefficient	t-statistic	Significance	Coefficient	t-statistic	Significance
UNOTRN	0.1988650	0.7811		0.14980619	0.5831	
UNOSHR	-0.9034440	-3.5401	***	-0.88805906	-3.3963	***
TOTTIME	-0.0560085	-6.7689	***	-0.05229113	-6.2191	***
TOTCOST	-0.7849730	-12.6730	***	-0.74595440	-10.4556	***
VEHWORK_TRN	-1.1792527	-5.1305	***	-1.24556320	-5.2681	***
VEHWORK_SHR	-0.7843766	-3.3874	***	-0.76596560	-3.2032	**
DIST_TRN	0.0634342	4.1595	***	0.05731883	3.6858	***
DIST_SHR	0.0009521	0.0814		0.00277688	0.2355	
TDUMMY_TRN	0.1746148	0.7548		0.11216541	0.4816	
TDUMMY_SHR	-0.6687029	-3.0938	**	-0.68822614	-3.1336	**
AEMPDENS_TRN	—	—	—	0.00055440	1.4418	
AEMPDENS_SHR	—	—	—	-0.00050569	-1.0956	
N – Sample Size		1125			1125	
LL – Convergence		-607.6			-605.9	

Table 33: Model Estimation Results for Model m6_7_1 and Model m6_7_1_1

Model m6_7_1	`CHOSEN ∼ UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR + DIST_TRN + DIST_SHR + TDUMMY_TRN + TDUMYY | 0`
Model m6_7_1_1	`CHOSEN ∼ UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR + DIST_TRN + DIST_SHR + TDUMMY_TRN + TDUMYY_SHR + AEMPDENS_TRN + AEMPDENS | 0`

Model	**#Df**	**LogLik**	**Df**	**Chisq**	**Pr(> Chisq)**
m6_7_1	10	-607.6	—	—	—
m6_7_1_1	12	-607.60	2	3.3992	0.1828

Table 34: Likelihood Ratio Test Comparing Model m6_7_1 and Model m6_7_1_1

Adding a population density at the workplace has no strong effect on Train and Shared utility in our sample. Neither AEMPDENS_TRN and AEMPDENS_SHR are not significant on their own, and the overall model does not improve significantly. It's reasonable to exclude both from our model.**Final Chosen Model from This Comparison is m6_7_1**

Our Final Model for Part C is m6_7_1

```
### Constants + LOS Model + ASV With Weights (m6_7_1) ###
m6_7_1 <- mlogit(CHOSEN~UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR
+ DIST_TRN + DIST_SHR + TDUMMY_TRN + TDUMMY_SHR|0, data = D, weights = WEIGHT)
```

Estimation and Interpretation of other Model Specifications (Part D)

Our final model is **m6_7_1** which is like follows:

```
### Constants + LOS Model + ASV With Weights (m6_7_1) ###
m6_7_1 <- mlogit(CHOSEN~UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR
+ DIST_TRN + DIST_SHR + TDUMMY_TRN + TDUMMY_SHR|0, data = D, weights = WEIGHT)
```

However, in this part it is asked for the value of in-vehicle travel time and the value of out-of-vehicle travel time. Previously, we saw that substituting TOTTIME with IVTT and OVTT does not change the performance of our model. So we make this substitution in our current model, and introduce a new model of **m6_7_1_2**.

```
### Constants + LOS Model + ASV With Weights (m6_7_1) ###
m6_7_1 <- mlogit(CHOSEN~UNOTRN + UNOSHR + TOTTIME + TOTCOST + VEHWORK_TRN + VEHWORK_SHR
+ DIST_TRN + DIST_SHR + TDUMMY_TRN + TDUMMY_SHR|0, data = D, weights = WEIGHT)



### Constants + LOS Model + ASV With Weights (m6_7_1_2) ###
m6_7_1 <- mlogit(CHOSEN~UNOTRN + UNOSHR + IVTT + OVTT + TOTCOST + VEHWORK_TRN + VEHWORK_SHR
+ DIST_TRN + DIST_SHR + TDUMMY_TRN + TDUMMY_SHR|0, data = D, weights = WEIGHT)
```

The summary of this model is as follows:

```
### Constants + LOS Model + ASV With Weights (m6_7_1_2) ###
m6_7_1_2 <- mlogit(CHOSEN~UNOTRN + UNOSHR + IVTT + OVTT + TOTCOST + VEHWORK_TRN + VEHWORK_SHR
+ DIST_TRN + DIST_SHR + TDUMMY_TRN + TDUMMY_SHR|0, data = D, weights = WEIGHT)

Frequencies of alternatives:choice
    Car  Shared   Train
0.41689 0.17600 0.40711

Coefficients :
             Estimate Std. Error  z-value  Pr(>|z|)
UNOTRN       0.545615   0.268532   2.0318 0.0421698 *
UNOSHR      -0.882613   0.257364  -3.4294 0.0006048 ***
IVTT        -0.031162   0.010844  -2.8736 0.0040579 **
OVTT        -0.086684   0.012729  -6.8101 9.753e-12 ***
TOTCOST     -0.756753   0.061568 -12.2913 < 2.2e-16 ***
VEHWORK_TRN -1.123747   0.227291  -4.9441 7.650e-07 ***
VEHWORK_SHR -0.785010   0.233090  -3.3678 0.0007576 ***
DIST_TRN     0.080802   0.016059   5.0315 4.867e-07 ***
DIST_SHR    -0.002385   0.011773  -0.2026 0.8394672
TDUMMY_TRN   0.111414   0.233219   0.4777 0.6328460
TDUMMY_SHR  -0.667455   0.215603  -3.0958 0.0019631 **

Log-Likelihood: -603.49

> ## Determine Choice Probabilities ##
> Pr_Choice_m6_7_1_2<- predict(m6_7_1_2, newdata = D)
> apply(Pr_Choice_m6_7_1_2, 2, mean)
      Car    Shared     Train
0.6377651 0.1153879 0.2468469
```

Value of Time (VoT): The Value of Time is the marginal rate of substitution between time and money.

- **In-Vehicle VoT**: How much people value in-vehicle time.

$$VoT_{IVTT} = \frac{\beta_{IVTT}}{\beta_{TOTCOST}} = \frac{-0.031162}{-0.756753} = 0.0412\ \$/min = 2.472\ \$/hour$$

- **Out-of-Vehicle VoT**: How much people value out-of-vehicle time.

$$VoT_{OVTT} = \frac{\beta_{OVTT}}{\beta_{TOTCOST}} = \frac{-0.086684}{-0.756753} = 0.1145\ \$/min = 6.87\ \$/hour$$

Marginal Effect: The marginal effect of a variable on the probability of choosing a mode tells us how much the probability of choosing that mode changes when the variable increases by one unit, all else equal. We know that For a given alternative j, the choice probability in a multinomial logit model is:

$$P_j = \frac{e^{V_j}}{\sum_k e^{V_k}}$$

V_j is the systematic utility of alternative j and usually is linear in parameters: $V_j = \beta_1 X_1 + \beta_2 X_2 + \dots$

Margina effect of a variable X on P_j is defined as:

$$\frac{\partial P_j}{\partial X} = \beta_X \cdot P_j \cdot (1 - P_j) \quad \text{(if } X \text{ is generic)}$$

- β_X: Coefficient for the variable X
- P_j: Probability of choosing alternative j

This is derived using the chain rule applied to the multinomial logit probability formula.

```
> ## Determine Choice Probabilities ##
> Pr_Choice_m6_7_1_2<- predict(m6_7_1_2, newdata = D)
> apply(Pr_Choice_m6_7_1_2, 2, mean)
      Car    Shared     Train
0.6377651 0.1153879 0.2468469
>
> ## Computing Marginal Effect of LOS Variables ##
> Beta_IVTT <- coef(m6_7_1_2)["IVTT"]
> Beta_OVTT <- coef(m6_7_1_2)["OVTT"]
> Beta_TOTCOST <- coef(m6_7_1_2)["TOTCOST"]
>
> # Compute marginal effect per alternative:  × P × (1 - P)
> mfx_IVTT <- (Pr_Choice_m6_7_1_2 * (1 - Pr_Choice_m6_7_1_2)) * Beta_IVTT
> mfx_OVTT <- (Pr_Choice_m6_7_1_2 * (1 - Pr_Choice_m6_7_1_2)) * Beta_OVTT
> mfx_TOTCOST <- (Pr_Choice_m6_7_1_2 * (1 - Pr_Choice_m6_7_1_2)) * Beta_TOTCOST
>
> # Average marginal effect for each mode
> marginal_effect_IVTT <- colMeans(mfx_IVTT)
> marginal_effect_OVTT <- colMeans(mfx_OVTT)
> marginal_effect_TOTCOST <- colMeans(mfx_TOTCOST)
>
> # Display result
> print(round(marginal_effect_IVTT, 6))
      Car    Shared     Train
-0.004554 -0.003089 -0.003047
> print(round(marginal_effect_OVTT, 6))
      Car    Shared     Train
```

```
-0.012668 -0.008594 -0.008476
> print(round(marginal_effect_TOTCOST, 6))
      Car    Shared     Train
-0.110589 -0.075022 -0.073997
```

- **IVTT**: All modes are negatively affected by longer in-vehicle time, as expected — but Car is most sensitive.
  - **Car, mfx = -0.004554**: Increasing in-vehicle time by 1 minute decreases Car choice probability by 0.4554 percentage points.
  - **Shared, mfx = -0.003089**: Increasing in-vehicle time by 1 minute decreases Shared choice probability by 0.3089 percentage points.
  - **Train, mfx = -0.003047**: Increasing in-vehicle time by 1 minute decreases Train choice probability by 0.3047 percentage points.
- **OVTT**: All modes are negatively affected by longer out-of-vehicle time, as expected, but Car is most sensitive. Also, OVTT has a much larger effect than IVTT — this is consistent with behavioral expectations (people dislike walking, waiting more than sitting in a vehicle).
  - **Car, mfx = 0.01267**: Increasing out-of-vehicle time by 1 minute decreases Car choice probability by 1.267 percentage points.
  - **Shared, mfx = -0.008594**: Increasing out-of-vehicle time by 1 minute decreases Shared choice probability by 0.8594 percentage points.
  - **Train, mfx = -0.008476**: Increasing out-of-vehicle time by 1 minute decreases Train choice probability by 0.8476 percentage points.
- **TOTCOST**: All modes are negatively affected by high amount of cost, as expected, but Car is most sensitive.
  - **Car, mfx = 0.11059**: A $1 increase in cost decreases Car choice by 11.059 percentage points.
  - **Shared, mfx = 0.07502**: A $1 increase in cost decreases Shared choice by 7.502 percentage points.
  - **Train, mfx = 0.07399**: A $1 increase in cost decreases Train choice by 7.399 percentage points.

Elasticity: An elasticity measures the percentage change in the probability of choosing a mode when a variable changes by 1%. Elasticities give scale-free, interpretable insights — useful when comparing variables that are in different units. For alternative j, the elasticity of the choice probability P_j with respect to variable X is:

$$E_j = \frac{\partial P_j}{\partial X} \cdot \frac{X}{P_j}$$

- $\frac{\partial P_j}{\partial X}$: Marginal effect
- X: Mean value of the variable
- P_j: Probability of choosing alternative j

This gives:

- **Negative elasticity** $\Rightarrow$ Increasing X reduces choice probability (P_j decreases)
- **Positive elasticity** $\Rightarrow$ Increasing X increases choice probability (P_j increases)

```
> ## Computing Elasticity
> # Means of the variables
> mean_IVTT <- mean(D$IVTT)
> mean_OVTT <- mean(D$OVTT)
> mean_TOTCOST <- mean(D$TOTCOST)
>
> # Mean predicted probabilities for each mode
> mean_Pr <- colMeans(Pr_Choice_m6_7_1_2)
>
> # Elasticities
> elasticity_IVTT <- (marginal_effect_IVTT * mean_IVTT) / mean_Pr
> elasticity_OVTT <- (marginal_effect_OVTT * mean_OVTT) / mean_Pr
> elasticity_TOTCOST <- (marginal_effect_TOTCOST * mean_TOTCOST) / mean_Pr
>
> # Display results
> print(round(elasticity_IVTT, 6))
      Car    Shared     Train
-0.174139 -0.652944 -0.301045
> print(round(elasticity_OVTT, 6))
      Car    Shared     Train
-0.294928 -1.105850 -0.509860
> print(round(elasticity_TOTCOST, 6))
      Car    Shared     Train
-0.284182 -1.065557 -0.491283
```

- **IVTT**: All modes are negatively affected by longer in-vehicle time, as expected — but Shared is most sensitive.
  - **Car, E = -0.174139**: A 1% increase in IVTT leads to a 0.174139% decrease in the probability of choosing Car.
  - **Shared, mfx = -0.652944**: A 1% increase in IVTT leads to a 0.652944% decrease in the probability of choosing Shared.
  - **Train, mfx = -0.301045**: A 1% increase in IVTT leads to a 0.301045% decrease in the probability of choosing Train.
- **OVTT**: All modes are negatively affected by longer out-of-vehicle time, as expected, but Shared is most sensitive. Also, OVTT has a much larger effect than IVTT — this is consistent with behavioral expectations (people dislike walking, waiting more than sitting in a vehicle).
  - **Car, E = -0.294928**: A 1% increase in OVTT leads to a 0.294928% drop in Car choice probability.
  - **Shared, E = -1.105850**: A 1% increase in OVTT leads to a 1.105850% drop in Shared choice probability.
  - **Train, E = -0.509860**: A 1% increase in OVTT leads to a 0.509860% drop in Train choice probability.
- **TOTCOST**: All modes are negatively affected by the high amount of cost, as expected, but Shared is most sensitive.
  - **Car, E = -0.284182**: A 1% increase in travel cost leads to a 0.284182% decrease in Car choice probability.
  - **Shared, E = -1.065557**: A 1% increase in travel cost leads to a 1.065557% decrease in Shared choice probability.
  - **Train, E = -0.491283**: A 1% increase in travel cost leads to a 0.491283% decrease in Train choice probability.

Variable	Mode	Marginal Effect	Elasticity
IVTT	Car	-0.004554	-0.174139
IVTT	Train	-0.003047	-0.652944
IVTT	Shared	-0.003089	-0.301045
OVTT	Car	-0.012668	-0.294928
OVTT	Train	-0.008476	-1.105850
OVTT	Shared	-0.008594	-0.509860
TOTCOST	Car	-0.110589	-0.284182
TOTCOST	Train	-0.073997	-1.065557
TOTCOST	Shared	-0.075022	-0.491283

Table 35: Marginal Effects and Elasticities by Mode and Variable

Aggregate Effect on Mode shares Due to Level-of-Service Changes

The goal is to determine how mode shares (choice probabilities) change if Trains and Shared IVTT are reduced by 25% and Car IVTT is increased by 30%. Then re-estimate predicted probabilities with the adjusted IVTT values should be compared to the base case.

```
> # Reduce IVTT by 25% for Train and Shared alternatives
> D_policy$IVTT[D_policy$alt %in% c("Train", "Shared")] <-
+   D_policy$IVTT[D_policy$alt %in% c("Train", "Shared")] * 0.75
>
> # Increase IVTT by 30% for Auto (Car)
> D_policy$IVTT[D_policy$alt == "Car"] <-
+   D_policy$IVTT[D_policy$alt == "Car"] * 1.30
>
> ## Determine Choice Probabilities ##
> Pr_Choice_m6_7_1_2_policy <- predict(m6_7_1_2, newdata = D_policy)
>
> # Original (base case)
> base_shares <- colMeans(Pr_Choice_m6_7_1_2)
>
> # New (after policy)
> policy_shares <- colMeans(Pr_Choice_m6_7_1_2_policy)
>
> # Change in shares
> delta_shares <- policy_shares - base_shares
>
> # Combine into a data frame
> share_comparison <- data.frame(
+   Mode = names(base_shares),
+   Base = round(base_shares, 4),
+   Policy = round(policy_shares, 4),
+   Change = round(delta_shares, 4)
+ )
>
> print(share_comparison)
         Mode   Base Policy  Change
Car       Car 0.6378 0.5801 -0.0577
Shared Shared 0.1154 0.1485  0.0331
Train   Train 0.2468 0.2715  0.0246
```

Mode	Base Policy	Change Policy	Change
Car	0.6378	0.5801	-0.0577
Shared	0.1154	0.1485	0.0331
Train	0.2468	0.2715	0.0246

Table 36: Mode Shares Under Base and Change Policy Scenarios

Drive-alone usage drops significantly by nearly 6 percentage points, reflecting the increased disutility from longer in-vehicle times. Shared and Train usage increase. This shift suggests that the policy effectively incentivizes more sustainable modes, particularly shared and train use.

Nested Logit Estimation (Preliminary Nested Logit Estimation)

In the previous part (Multinomial Logit), we assumed that there was no nesting in the choice procedure of the people, meaning that people choose between Car, Train, and Shared in one step. Now, we want to create three different nesting structures to see whether our data supports this nesting behavior of mode choice or not. We compare these three nested model with the base Multinomial Logit Model (m7).

```
> ## Multinomial Logit Model
> m7 <- mlogit(CHOSEN~IVTT + OVTT + TOTCOST|VEHWORK + HHSIZE + TDUMMY, data = D,
reflevel = 'Car', weights = WEIGHT)
```

- **Structure 1:** At the first level, we have two nests of Solo and Group. At the second level, under the Solo nest we have only Car, and under the Group nest we have Train and shared. The variable we used are those used by the class example for this part: **IVTT**, **OVTT**, **TOTCOST**, **VEHWORK**, **HHSIZE**, and **TDUMMY**.

```
> n1 <- mlogit(CHOSEN~IVTT + OVTT + TOTCOST|VEHWORK + HHSIZE + TDUMMY, data = D,
reflevel = 'Car', weights = WEIGHT, nests = list(Solo = c('Car'),
Group = c('Train','Shared')), un.nest.el = TRUE)
```

- **Structure 2:** At the first level, we have two nests of Not-shared and Shared. At the second level, under the Not-shared nest, we have Car and Train, and under the Shared nest, we have Shared. The variable we used are those used by the class example for this part: **IVTT**, **OVTT**, **TOTCOST**, **VEHWORK**, **HHSIZE**, and **TDUMMY**.

```
> n1 <- mlogit(CHOSEN~IVTT + OVTT + TOTCOST|VEHWORK + HHSIZE + TDUMMY, data = D,
reflevel = 'Car', weights = WEIGHT, nests = list(Not_Shared = c('Car', 'Train'),
Shared = c('Shared')), un.nest.el = TRUE)
```

- **Structure 3:** At the first level, we have two nests of Auto and Transit. At the second level, under the Auto nest, we have Car and Shared, and under the Transit nest, we have Train. The variable we used are those used by the class example for this part: **IVTT**, **OVTT**, **TOTCOST**, **VEHWORK**, **HHSIZE**, and **TDUMMY**.

```
> n1 <- mlogit(CHOSEN~IVTT + OVTT + TOTCOST|VEHWORK + HHSIZE + TDUMMY, data = D,
reflevel = 'Car', weights = WEIGHT, nests = list(Auto = c('Car', 'Shared'),
Transit = c('Train')), un.nest.el = TRUE)
```

Structure 1 (n1) Model vs MNL (m7) Model

First we take a look at the summary of model m7 and model n1.

Variable	Model m7			Model n1		
	Coefficient	t-statistic	Significance	Coefficient	t-statistic	Significance
(Intercept):Shared	-1.331716	-3.8672	***	-1.129715	-2.7599	**
(Intercept):Train	0.540643	1.5646		0.481565	0.9114	
IVTT	-0.038364	-3.5809	***	-0.035866	-2.3842	*
OVTT	-0.053697	-4.9962	***	-0.045015	-3.0300	**
TOTCOST	-0.788570	-13.5329	***	-0.755468	-13.1030	***
VEHWORK:Shared	-0.755700	-3.3792	***	-0.781749	-3.2074	**
VEHWORK:Train	-0.987458	-4.2903	***	-0.982411	-2.7907	**
HHSIZE:Shared	0.099726	1.4010		0.087108	1.1565	
HHSIZE:Train	-0.070828	-0.9198		-0.065734	-0.6099	
TDUMMY:Shared	-0.659177	-3.3392	***	-0.721477	-3.4906	***
TDUMMY:Train	0.273643	1.2163		0.204284	0.6347	
iv	—	—	—	0.799046	5.1045	***
N – Sample Size		1125			1125	
LL – Convergence		-612.28			-611.64	

Table 37: Model Estimation Results for Model m7 and Model n1

In this problem we have two levels so we have two different scale parameters ($\mu_{upper\ level}$ and $\mu_{lower\ level}$). The lower-level scale parameter is the base case and is considered to be 1 for identification. The value we obtain from the software is actually an upper-level scale factor. The model is only valid if this value is between 0 and 1. If the scale parameter is greater than or equal to 1, then our assumption of nesting logit is not valid, and multinomial logit is good enough.

We can see in this case that the value of the scale parameter is 0.799046, which suggests that our nesting structure is valid at first glance. However, we must keep in mind that the t-test is testing whether the value of a parameter is better to be equal to zero or not. On the other hand, here we want to check whether the value of the scale parameter is equal to or greater than 1 or not. So we need to do the t-test manually as well.

$$t = \frac{\hat{\beta} - 1}{\mathrm{SE}(\hat{\beta})}$$

```
> (coef(n1)['iv'] - 1) / sqrt(vcov(n1)['iv', 'iv'])
       iv
-1.283731
```

We can see that the t-value is in the acceptance region (between -1.96 and 1.96), so we accept the null hypothesis that the value of the scale parameter coefficient is equal to 1. So the nested logit is not valid, and MNL is sufficient in this case. Moreover, we double-check our deduction by performing an LRT between these two models. In this case, the value of Chi-Squared for 95% confidence and $df = 1$ is 3.841459. So if the value of LR is greater than this, then we are at the Reject Area.

Model	#Df	LogLik	Df	Chisq	Pr(> Chisq)
m7	11	-612.28	—	—	—
n1	12	-611.64	1	1.2828	0.2574

Table 38: Likelihood Ratio Test Comparing Model m7 and Model n1

Summary: Based on the model comparison results, the data does not support the nested logit model (n1) over the simpler multinomial logit model (m7).

- **LRT:** Since $p = 0.2574 > 0.05$, the improvement in model fit from m7 to n1 is not statistically significant. Consequently, we fail to reject the null hypothesis that the simpler model (m7) fits just as well as the more complex nested model (n1). Thus, there is no strong statistical support for preferring the nested structure.

- **t-test:** Estimate: 0.799 and $p < 0.001$ (statistically significant). This value of the inclusive value (IV) parameter is significantly different from zero, which is required for the nested logit to be valid. Test for difference from 1: $t = -1.28$. This t-value tells us that we fail to reject the hypothesis that $IV = 1$.. This z-score is not significant, so we do not reject the hypothesis that IV = 1, meaning the nested structure is not statistically distinguishable from a multinomial logit. The nested structure is theoretically valid (IV is significantly different from 0). But IV is not significantly different from 1, and the LRT shows no significant gain in fit. Therefore, the data do not support using the nested logit model over the multinomial logit.

Structure 2 (n2) Model vs MNL (m7) Model

First we take a look at the summary of model m7 and model n2.

Variable	**Model m7**			**Model n2**		
	Coefficient	t-statistic	Significance	Coefficient	t-statistic	Significance
(Intercept):Shared	-1.331716	-3.8672	***	-1.538068	-3.5464	***
(Intercept):Train	0.540643	1.5646		0.283587	0.7664	
IVTT	-0.038364	-3.5809	***	-0.019450	-1.2659	
OVTT	-0.053697	-4.9962	***	-0.030893	-1.7910	
TOTCOST	-0.788570	-13.5329	***	-0.481013	-2.3519	*
VEHWORK:Shared	-0.755700	-3.3792	***	-0.618891	-2.3663	*
VEHWORK:Train	-0.987458	-4.2903	***	-0.565957	-1.6583	
HHSIZE:Shared	0.099726	1.4010		0.107613	1.2942	
HHSIZE:Train	-0.070828	-0.9198		-0.039757	-0.5511	
TDUMMY:Shared	-0.659177	-3.3392	***	-0.518963	-2.1646	*
TDUMMY:Train	0.273643	1.2163		0.174004	0.7690	
iv	—	—	—	0.562125	2.2496	*
N – Sample Size		1125			1125	
LL – Convergence		-612.28			-610.18	

Table 39: Model Estimation Results for Model m7 and Model n2

We can see in this case that the value of the scale parameter is 0.562125, which suggests that our nesting structure is valid at first glance. However, we must keep in mind that the t-test is testing whether the value of a parameter is better to be equal to zero or not. On the other hand, here we want to check whether the value of the scale parameter is equal to or greater than 1 or not. So we need to do the t-test manually as well.

$$t = \frac{\hat{\beta} - 1}{\mathrm{SE}(\hat{\beta})}$$

```
> (coef(n2)['iv'] - 1) / sqrt(vcov(n2)['iv', 'iv'])
         iv
-1.752384
```

We can see that the t-value is in the acceptance region (between -1.96 and 1.96), so we accept the null hypothesis that the value of the scale parameter coefficient is equal to 1. So the nested logit is not valid, and MNL is sufficient in this case. Moreover, we double-check our deduction by performing an LRT between these two models. In this case, the value of Chi-Squared for 95% confidence and $df = 1$ is 3.841459. So if the value of LR is greater than this, then we are at the Reject Area.

Model	#Df	LogLik	Df	Chisq	Pr(> Chisq)
m7	11	-612.28	—	—	—
n2	12	-610.18	1	4.2068	0.04026*

Table 40: Likelihood Ratio Test Comparing Model m7 and Model n2

Summary: Based on our model comparison results for Nested Logit Structure 2 (n2) vs. the Multinomial Logit Model (m7), there is moderate support for the nested logit model.

- **LRT:** Since $p = 0.04026 < 0.05$, the improvement in fit is statistically significant. Consequently, there is evidence to prefer the nested logit model over the simpler multinomial logit.
- **t-test:** Estimate: 0.562 and $p = 0.0244$ (statistically significant). Test for difference from 1: $t = -1.75$. This t-value tells us that we fail to reject the hypothesis that $IV = 1$. This implies the nested structure (Not_shared vs. Shared) is distinct from the multinomial structure, but not by a wide margin. The LRT is significant so n2 fits the data significantly better than m7. The IV is significantly different from 0, supporting the validity of the nested logit. The IV is marginally different from 1, suggesting some departure from the independence of irrelevant alternatives (IIA) assumption. In conclusion, it seems that the data do not support this nesting structure.

Structure 3 (n3) Model vs MNL (m7) Model

First we take a look at the summary of model m7 and model n3.

Variable	Model m7			Model n3		
	Coefficient	t-statistic	Significance	Coefficient	t-statistic	Significance
(Intercept):Shared	-1.331716	-3.8672	***	-2.993515	-1.7894	
(Intercept):Train	0.540643	1.5646		0.753358	1.2106	
IVTT	-0.038364	-3.5809	***	-0.032515	-1.7093	***
OVTT	-0.053697	-4.9962	***	-0.052406	-3.3497	***
TOTCOST	-0.788570	-13.5329	***	-0.903396	-11.7652	
VEHWORK:Shared	-0.755700	-3.3792	***	-1.378493	-1.8253	**
VEHWORK:Train	-0.987458	-4.2903	***	-1.133767	-2.7225	
HHSIZE:Shared	0.099726	1.4010		0.222961	1.1023	
HHSIZE:Train	-0.070828	-0.9198		-0.057428	-0.4796	
TDUMMY:Shared	-0.659177	-3.3392	***	-1.217836	-1.9579	
TDUMMY:Train	0.273643	1.2163		0.172626	0.4494	
iv	—	—	—	2.151776	2.2265	*
N – Sample Size		1125			1125	
LL – Convergence		-612.28			-607.76	

Table 41: Model Estimation Results for Model m7 and Model n3

We can see in this case that the value of the scale parameter is 2.151776, which suggests that our nesting structure is not valid at first glance. However, we must keep in mind that the t-test is testing whether the value of a parameter is better to be equal to zero or not. On the other hand, here we want to check whether the value of the scale parameter is equal to or greater than 1 or not. So we need to do the t-test manually as well.

$$t = \frac{\hat{\beta} - 1}{\mathrm{SE}(\hat{\beta})}$$

```
> (coef(n3)['iv'] - 1) / sqrt(vcov(n3)['iv', 'iv'])
       iv
1.19178
```

We can see that the t-value is in the acceptance region (between -1.96 and 1.96), so we accept the null hypothesis that the value of the scale parameter coefficient is equal to 1. So the nested logit is not valid, and MNL is sufficient in this case. Moreover, we double-check our deduction by performing an LRT between these two models. In this case, the value of Chi-Squared for 95% confidence and $df = 1$ is 3.841459. So if the value of LR is greater than this, then we are at the Reject Area.

Model	**#Df**	**LogLik**	**Df**	**Chisq**	**Pr(> Chisq)**
m7	11	-612.28	—	—	—
n2	12	-607.76	1	9.0337	0.00265**

Table 42: Likelihood Ratio Test Comparing Model m7 and Model n3

Summary: Based on our model comparison results for Nested Logit Structure 2 (n2) vs. the Multinomial Logit Model (m7), there is moderate support for the nested logit model.

- **LRT:** Since $p = 0.04026 < 0.05$, the improvement in fit is statistically significant. Consequently, there is evidence to prefer the nested logit model over the simpler multinomial logit.
- **t-test:** Estimate: 2.151776 which is greater than 1 and indicates that the nested structure is not valid. Test for difference from 1: $t = 1.19$. This t-value tells us that we fail to reject the hypothesis that $IV = 1$. This implies the nested structure (Auto vs. Transit) is not valid. The LRT is significant, so n3 fits the data significantly better than m7. In conclusion, it seems that the data do not support this nesting structure.

Nested Logit Estimation (Estimation and Interpretation of other Model Specifications)

Now we add other Individual-Specific Variables one after another and create the following models:

- **m8:** CHOSEN IVTT + OVTT + TOTCOST—VEHWORK + HHSIZE + TDUMMY + DIST
- **m9:** CHOSEN IVTT + OVTT + TOTCOST—VEHWORK + HHSIZE + TDUMMY + DIST + MALE
- **m10:** CHOSEN IVTT + OVTT + TOTCOST—VEHWORK + DIST + TDUMMY + DIST + MALE + NEWINC
- **m11:** CHOSEN IVTT + OVTT + TOTCOST—VEHWORK + DIST + TDUMMY + DIST + MALE + NEWINC + AEMPDENS
- **m12:** CHOSEN IVTT + OVTT + TOTCOST—VEHWORK + DIST + TDUMMY + DIST + MALE + NEWINC + AEMPDENS + POPDENS
- **m13:** IVTT + OVTT + TOTCOST—VEHWORK + DIST + TDUMMY + DIST + MALE + NEWINC + AEMPDENS + POPDENS + AGE

We only show the result in Table 43. Our analysis and explanation on how we have chosen the final model in each comparison are the same as the previous section (Note: We already have reached to the best model of **m6_7_1_2** in the previous sections, and it is sufficient to only do this part on this model, however, since it is asked to add all other variables, we have decided to create all possible models).

Model	IV Value	t-test Result	LRT Result (Chi-Squared)	Decision
n8_1	0.8066303	-1.22078	1.1536	Not Valid
n8_2	0.1926810	-2.815734	7.1896	Valid
n8_3	30.869895	1.260363	19.768	Not Valid
n9_1	0.8002296	-1.268368	1.234	Not Valid
n9_2	0.1998839	-2.739245	7.0573	Valid
n9_3	36.150246	1.289469	19.273	Not Valid
n10_1	0.79028	-1.341108	1.38	Not Valid
n10_2	0.16964	-2.90444	7.6582	Valid
n10_3	27.265	1.253706	18.587	Not Valid
n11_1	0.83385	-0.888781	0.5592	Not Valid
n11_2	0.096987	-2.84058	7.5097	Valid
n11_3	64.455	1.695757	27.789	Not Valid
n12_1	0.80629	-1.048986	0.7746	Not Valid
n12_2	0.099325	-2.818453	7.3268	Valid
n12_3	86.123	1.558815	29.734	Not Valid
n13_1	0.81041	-1.009099	0.698	Not Valid
n13_2	0.096961	-2.78379	7.6139	Valid
n13_3	78.188	1.589008	28.34	Not Valid

Table 43: Model comparison results including inclusive value (IV), t-tests, and likelihood ratio tests (LRT)

It can be observed that for all the models, nesting the mode choice based on Not-shared and Shared is a valid nesting model.