

# SEM\_Exercise

Ashok Sekar

08/11/2016

## Introduction

This document is about testing and performing Structural equation modeling (SEM) with the Sense of Place (SOP) dataset. The SEM will relate SOP dimentions of a neighborhood with demographic variables and travel responses.

To develop the model we first perform a confirmatory factor analysis. The factor analysis helps in identifying the sense of place (SOP) dimensions that exists in a neighborhood. For example, in the survey, several likert scale based questions that aimed at capturing the responder's attitude towards the place were asked. Those responses are used to identify 6 latent factors also called SOP dimensions. The hypothesis is that neighborhoods could have differing SOP dimensions.

A factor analysis was performed, not described here, and the resulting factor strucutre forms the basis for this analysis. For simplicity, we look at the final factor structure for each neighborhood seperately. Factor structure for RPM is given in Figure 1. Based on the figure we know of the 6 dimensions dependence and identity does not explain the SOP of RPM. Using the factor struture we add on travel and demographic variables to identify the final structure.

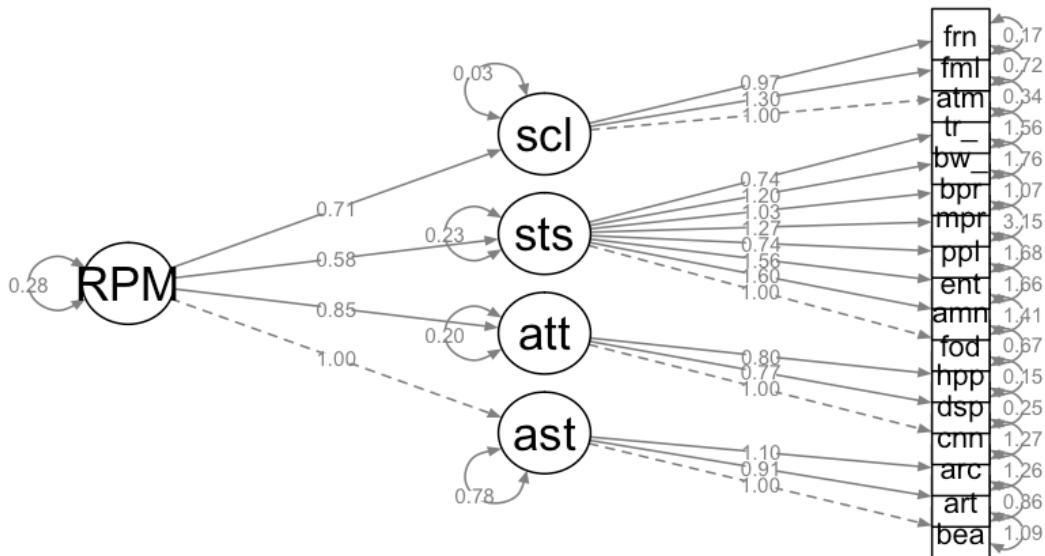


Figure 1: Factor Structure for Rochester Public Market

## Datasets

For this analysis we depend on a dataset named *mainFile*. The file contains both the attitudinal responses, travel and socio-demographic characteristics. Note that for factor analysis the attitudinal responses for dimensions were used. For SEM we would be using attitudinal responses, travel and socio-demographic variables.

The *mainFile* was used to create two different files depending on how the missing values were treated. Numerous responses to likert scale questions contained a *NA* category, therefore we decided to convert the *NA* values to “4” being neutral. Note that *NA* means responders did not have any attitude towards that question. For the other file no changes were made. However, it is important to note that for our analysis we use the dataset that has *NA* values converted to 4. The code for the data preparation is not shown here. In addition all the categorical data were converted into binary variables. E.g., many demographic variables such as age, income were collected as categories. They were converted to binary variables.

## Experimental Model:

This is my first experimental model using *SEM* package. I used *SEM* package because it seemed easy to set arrows with this package. This is followed by using Lavaan based *semPlot* package for visualizing the path diagram. In this simple model, a regression model is set up between number of vehicles in a household and the number of drivers with valid driver’s license. The estimate is saved as the variable *pi* and the variance are recorded as *var1* and *var2*. Because we are using the *SEM* package the factor structure for RPM is again evaluated using *SEM* package. The same result was identified. To keep the document simple I have not added the results below.

## Methodology

We add travel and demographic variables to each dimension and slowly build the model based on significance test. Other Notes: After completing the modeling of Satisfaction with demographics and travel variables I tried to build a full model that contains all the factor structures to start with and experiment with adding the demographic variables. However it is too difficult to proceed in that way because the reason for error was difficult to gauge. Therefore I will do whatever I did for satisfaction and then try to summarize a full model.

## Rochester Public Market - Satisfaction

The following section documents the experiments to find the structural model for satisfaction dimension of the RPM. We test for significance of each demographic variable with one sense of place dimension before moving to the other dimensions. Based on the results from the combination of tests a final model for RPPM-satisfaction will be presented.

List of dimensions and the factor structure are already discussed earlier. Check Figure 1.

## Demographic Variables

### Satisfaction - AGE

The figure below (Figure 3) shows the only significant model for age in satisfaction structure of RPM. Trial and error method was used to identify the significant variable.

### **Satisfaction - Gender**

Follow the same procedure as above but for gender. However gender was not a significant variable therefore they will not be included.

### **Satisfaction - How long did the visitor live in rochester**

In this experiment all the variables were found insignificant **unfortunately**.

### **Satisfaction - Income**

None of the income variables were significant. Although some of the income groups were close to significance with *zvalue* = 1.4.

### **Satisfaction - number of household members**

Notes: there seems to be an interaction between *hhmem\_1* and *hhmem\_2*. When running the model separately with *hhmem\_1* we get a significant value for *hhmem\_1*. When I run *hhmem\_1* and *hhmem\_2* together *hhmem\_1* becomes insignificant, while *hhmem\_2* is significant. When I run *hhmem\_2* separately, *hhmem\_2* is significant. In addition to these two values *hhmem\_5ormore* was also significant. When I ran a model with *hhmem\_2* and *hhmem\_5* together *hhmem\_5ormore* became insignificant. Therefore from this point forward variable *hhmem\_2* will be included while the other two that were significant independently will be given further inspection later.

### **Satisfaction - number of household members employed full time employed**

In this experiment people with 1 fulltime employed household member seemed to be a significant indicator of satisfaction.

### **Satisfaction - number of children in household**

In this experiment no significant variables were found.

### **Satisfaction - relationship with household members**

In this experiment people who were living immediately with family were significant in explaining satisfaction.

## **Travel Variables**

Now I am experimenting with the travel variables. Compared to demographics there are only few travel related variables viz.,

*vist* - how often people visited a location.

*mode* - what mode did they use? *biwalk* - have they ever biked or walked to the location *no\_vehicles* - number of vehicles at their home.

*no\_drivers* - number of drivers at their home. *no\_bicycles* - number of bicycles at home \* several variables indicating with whom they traveled to the location on day of interview

Of all the *travel* related variables discussed above, few of them could be informed by SOP dimension. For e.g., *mode*, *vist*, *biwalk*. While they also can be informed by other variables such as distance from home, number

of vehicles and drivers using a simple regression. To not complicate the modeling exercise I will currently perform test the significance of the the variables *mode*, *visit*, *biwalk*. Because satisfaction, as a latent variable is already being informed by numerous satisfaction responses these travel variables will not affect the existing satisfaction factor structure.

### **Satisfaction - Visit**

Responders who visited RPM all the three days a week are satisfied. For all other visitors no significance was found.

### **Satisfaction - Mode**

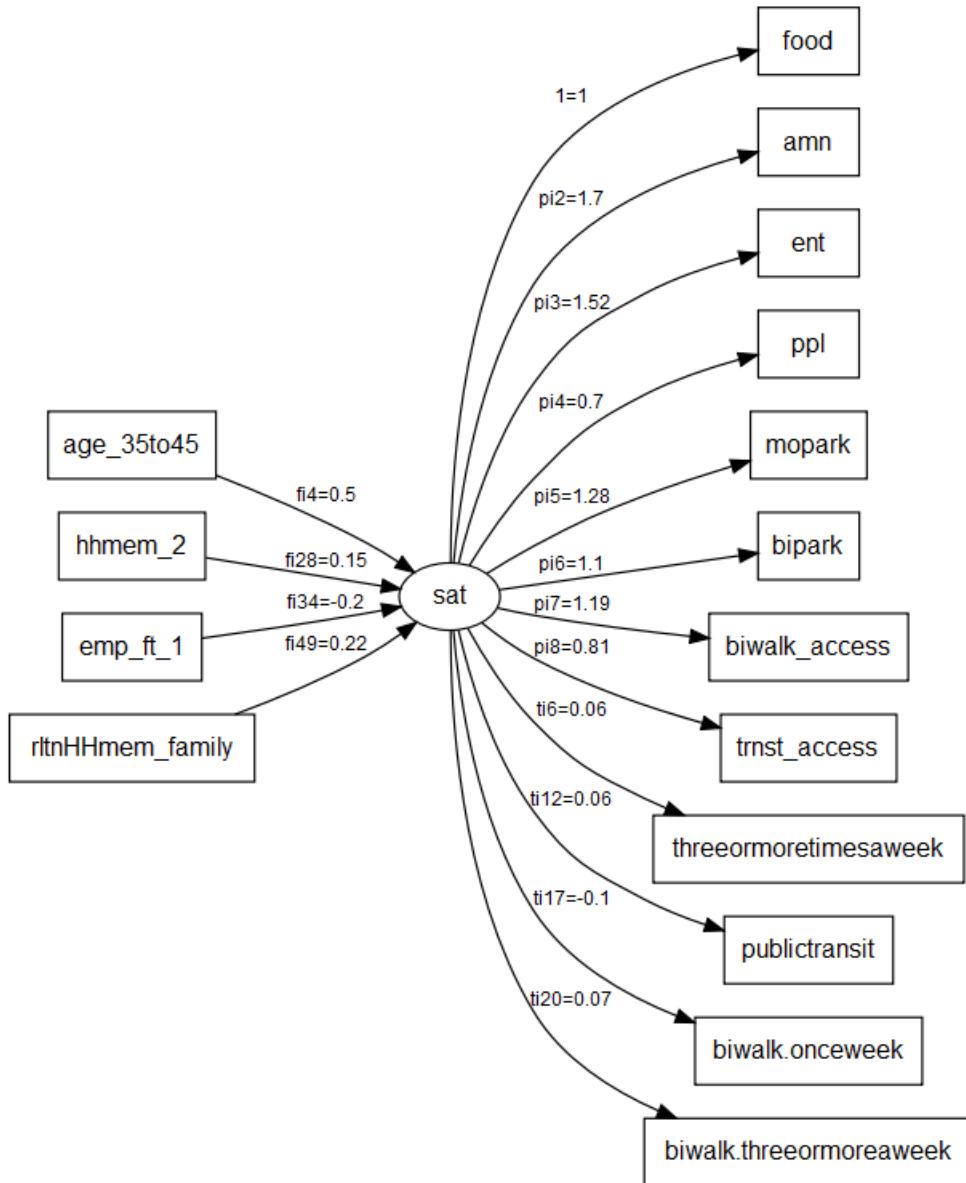
Public transit was the only significant variable.

### **Satisfaction - Cycle or Walk in Summer**

Notes: The variables *biwalk.oncein3weeks* and *biwalk.threecormoretimesaweek* seems to be interacting. When modeled individually both are significant however when modeled together *biwalk.oncein3weeks* is significant.

## Final Model for RPM. Relationship between Satisfaction, travel and demographic variables

The final model structure is shown below as Figure below. The summary statistics also provided.



```
## NOTE: it is generally simpler to use specifyEquations() or cfa()
##       see ?specifyEquations
```

```
##
##  Model Chisquare = 151.6167   Df = 104 Pr(>Chisq) = 0.001614224
##  AIC = 215.6167
##  BIC = -357.7586
##  Normalized Residuals
```

```

##      Min. 1st Qu. Median Mean 3rd Qu. Max.
## -2.14700 -0.59830 -0.04336 -0.02394 0.38840 3.21200
##
## R-square for Endogenous Variables
##          sat          food          amn
##          0.1435       0.3159       0.4049
##          ent          ppl          mopark
##          0.2994       0.0835       0.1403
##          bipark       biwalk_access   trnst_access
##          0.2701       0.2000       0.1188
## threeormoretimesaweek publictransit biwalk.onceweek
##          0.0771       0.0352       0.0451
## biwalk.threeormoreaweek
##          0.0480
##
## Parameter Estimates
##   Estimate Std. Error z value Pr(>|z|)
## pi2    1.69719708 0.348443475 4.870796 1.111495e-06
## pi3    1.52156010 0.340918087 4.463125 8.077283e-06
## pi4    0.70064202 0.258752612 2.707768 6.773738e-03
## pi5    1.27747275 0.378455124 3.375493 7.368347e-04
## pi6    1.10005785 0.254939303 4.314979 1.596178e-05
## pi7    1.18848733 0.306514672 3.877424 1.055683e-04
## pi8    0.80726815 0.256141923 3.151644 1.623541e-03
## var1   0.27248410 0.091082756 2.991610 2.775104e-03
## err1   0.68898527 0.100326012 6.867464 6.535325e-12
## err2   1.34680464 0.215282318 6.255993 3.949949e-10
## err3   1.72386630 0.247520281 6.964546 3.294653e-12
## err4   1.71395881 0.216863827 7.903387 2.714246e-15
## err5   3.18145600 0.412983721 7.703587 1.322990e-14
## err6   1.04020989 0.145989038 7.125260 1.038842e-12
## err7   1.79714590 0.240851676 7.461629 8.545911e-14
## err8   1.53758950 0.197575401 7.782292 7.122212e-15
## fi4    0.50327849 0.201190433 2.501503 1.236674e-02
## err22  0.08214566 0.010073347 8.154753 3.498936e-16
## fi28   0.15296502 0.109853655 1.392444 1.637881e-01
## err44  0.25047694 0.030715453 8.154753 3.498936e-16
## fi34   -0.19820514 0.121509003 -1.631197 1.028487e-01
## err50  0.20788913 0.025493000 8.154753 3.498936e-16
## fi49   0.22203708 0.119266352 1.861691 6.264668e-02
## err65  0.21956009 0.026924186 8.154753 3.498936e-16
## ti6    0.05994830 0.022932757 2.614090 8.946534e-03
## err72  0.01368030 0.001726385 7.924251 2.295253e-15
## ti12   0.06327785 0.034755124 1.820677 6.865597e-02
## err78  0.03492904 0.004336691 8.054308 7.992998e-16
## ti17   -0.09931993 0.048548913 -2.045770 4.077897e-02
## ti20   0.06634388 0.031506775 2.105702 3.523024e-02
## err83  0.06646742 0.008282883 8.024672 1.017975e-15
## err86  0.02778997 0.003466849 8.015916 1.093188e-15
##
## pi2    amn <--- sat
## pi3    ent <--- sat
## pi4    ppl <--- sat
## pi5    mopark <--- sat

```

```

## pi6 bipark <--- sat
## pi7 biwalk_access <--- sat
## pi8 trnst_access <--- sat
## var1 sat <--> sat
## err1 food <--> food
## err2 amn <--> amn
## err3 ent <--> ent
## err4 ppl <--> ppl
## err5 mopark <--> mopark
## err6 bipark <--> bipark
## err7 biwalk_access <--> biwalk_access
## err8 trnst_access <--> trnst_access
## fi4 sat <--- age_35to45
## err22 age_35to45 <--> age_35to45
## fi28 sat <--- hhmem_2
## err44 hhmem_2 <--> hhmem_2
## fi34 sat <--- emp_ft_1
## err50 emp_ft_1 <--> emp_ft_1
## fi49 sat <--- rltHHmem_family
## err65 rltHHmem_family <--> rltHHmem_family
## ti6 threeormoretimesaweek <--- sat
## err72 threeormoretimesaweek <--> threeormoretimesaweek
## ti12 publictransit <--- sat
## err78 publictransit <--> publictransit
## ti17 biwalk.onceweek <--- sat
## ti20 biwalk.threeormoreaweek <--- sat
## err83 biwalk.onceweek <--> biwalk.onceweek
## err86 biwalk.threeormoreaweek <--> biwalk.threeormoreaweek
##
## Iterations = 30

```

## Rochester Public Market - Attachment

This section shows the final model for RPM attachment. Coding Notes: Always change the *sat* to *att* in the *model\_RPM\_att\_XXX* file/

### Demographic Variables

#### Attachment - Age

Significant variables - *age\_55to64*, *age\_45to55*

#### Attachment - Gender

Both variables were *close* to 90% significance therefore gender is ignored.

#### Attachment - How long did the visitor live in rochester

variables significant were - *notfromROC*,

### **Attachment - Income**

*Inc25kto50k* very close to significance. No other variables were significant.

### **Attachment - Number of household members**

*hhmem\_5ormore* was the only significant variable.

### **Attachment - Number of household members employed full time employed**

*emp\_ft\_3* is the only significant variable.

### **Attachment - Number of children in household**

*child\_3* was significant. However when it was added *hhmem\_5ormore* became insignificant. When a cross tabulation was performed between them it was identified that the two variables were correlated. Therefore I won't be adding *child\_3* in the equation.

### **Attachment - Relationship with household members**

No variables were significant. Although I did get an unusual error when running variables for *acquaintances* and *alone*

## **Travel Variables**

Now I am experimenting with the travel variables. Compared to demographics there are only few travel related variables viz.,

*vist* - how often people visited a location.

*mode* - what mode did they use? *biwalk* - have they ever biked or walked to the location *no\_vehicles* - number of vehicles at their home.

*no\_drivers* - number of drivers at their home. *no\_bicycles* - number of bicycles at home \* several variables indicating with whom they traveled to the location on day of interview

Of all the *travel* related variables discussed above, few of them could be informed by SOP dimension. For e.g., *mode*, *visit*, *biwalk*. While they also can be informed by other variables such as distance from home, number of vehicles and drivers using a simple regression. To not complicate the modeling exercise I will currently perform test the significance of the the variables *mode*, *visit*, *biwalk*. Because satisfaction, as a latent variable is already being informed by numerous satisfaction responses these travel variables will not affect the existing satisfaction factor structure.

### **Attachment - Visit**

Significant variables: *once a week* and *less than once a month*

### **Attachment - Mode**

*bike* mode was a significant indicator for attachment.

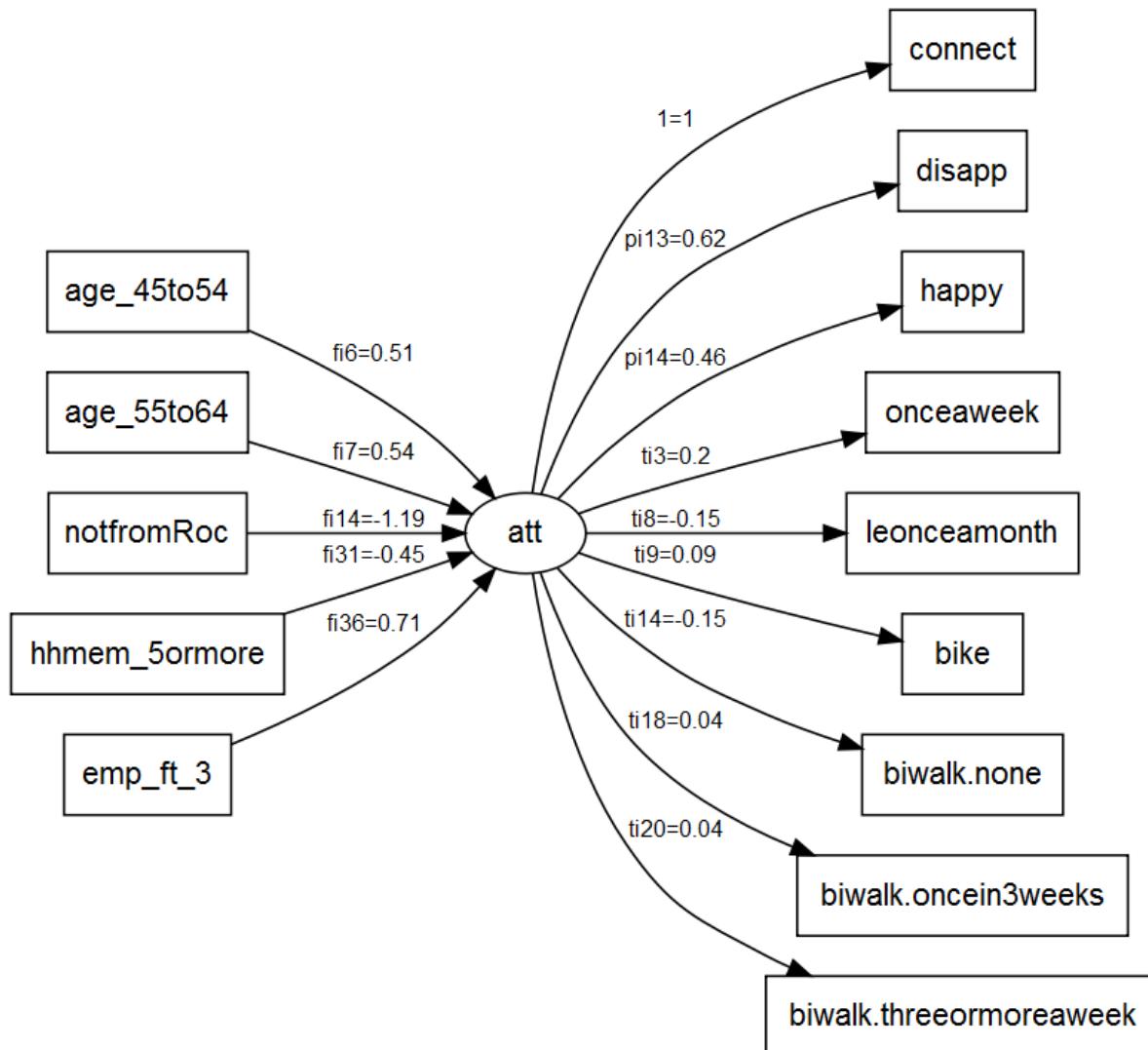
## Attachment - Cycle or Walk in Summer

Notes: Some interesting results here. *biwalk.none*, *threeormoreawek*, *oncein3weeks* were significant while *biwalk.onceawek*, *lemonth* and *oncemonth* showed the following error. >Error in w\_mat %>% p\_deriv\_mat %>% invMat : requires numeric/complex matrix/vector arguments.

At this time I couldn't find the reason for the error, however as we consolidate the variables these can be looked again.

## Final Model for RPM. Relationship between Attachment, travel and demographic variables

The final model structure is shown below as Figure below. The summary statistics also provided.



```

## NOTE: it is generally simpler to use specifyEquations() or cfa()
##       see ?specifyEquations
  
```

```

##
```

```

## Model Chisquare = 169.7979 Df = 77 Pr(>Chisq) = 6.052065e-09
## AIC = 225.7979
## BIC = -207.3357
##
## Normalized Residuals
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## -4.2680 -0.7511 -0.1489 -0.1060 0.6922 3.2870
##
## R-square for Endogenous Variables
## att connect disapp
## 0.3131 0.3899 0.5007
## happy onceaweeek leonceamonth
## 0.3470 0.1043 0.1004
## bike biwalk.none biwalk.oncein3weeks
## 0.0814 0.0757 0.0355
## biwalk.threeormoreaweek
## 0.0356
##
## Parameter Estimates
## Estimate Std Error z value Pr(>|z|)
## pi13 0.61545873 0.110464012 5.571577 2.524440e-08
## pi14 0.46246772 0.091684253 5.044135 4.555784e-07
## var4 0.46018069 0.140316485 3.279591 1.039576e-03
## err15 1.04844675 0.163443992 6.414716 1.410859e-10
## err16 0.25302693 0.046566178 5.433706 5.519540e-08
## err17 0.26969821 0.040156765 6.716134 1.866094e-11
## fi6 0.51173109 0.191756511 2.668650 7.615668e-03
## fi7 0.54301296 0.207224058 2.620415 8.782290e-03
## err23 0.16210302 0.019878348 8.154753 3.498936e-16
## err24 0.13825609 0.016954049 8.154753 3.498936e-16
## fi14 -1.19410163 0.357173082 -3.343202 8.281770e-04
## err30 0.04988217 0.006116944 8.154753 3.498936e-16
## fi31 -0.45081314 0.237513515 -1.898053 5.768916e-02
## err47 0.10015711 0.012282053 8.154753 3.498936e-16
## fi36 0.71018580 0.290143799 2.447703 1.437702e-02
## err52 0.06957693 0.008532070 8.154753 3.498936e-16
## ti3 0.19709493 0.063922158 3.083358 2.046784e-03
## ti8 -0.15173901 0.050068298 -3.030640 2.440356e-03
## err69 0.22347842 0.028480559 7.846700 4.271258e-15
## err74 0.13819560 0.017583109 7.859565 3.854688e-15
## ti9 0.09209122 0.033464094 2.751941 5.924317e-03
## err75 0.06416046 0.008100200 7.920849 2.358950e-15
## ti14 -0.15233041 0.057239354 -2.661288 7.784233e-03
## ti18 0.03937903 0.021186553 1.858680 6.307248e-02
## ti20 0.03940687 0.021187084 1.859948 6.289295e-02
## err80 0.18987067 0.023917508 7.938564 2.045348e-15
## err84 0.02818702 0.003498165 8.057660 7.776883e-16
## err86 0.02818562 0.003498053 8.057517 7.785947e-16
##
## pi13 disapp <--- att
## pi14 happy <--- att
## var4 att <--> att
## err15 connect <--> connect
## err16 disapp <--> disapp

```

```

## err17 happy <--> happy
## fi6   att <--- age_45to54
## fi7   att <--- age_55to64
## err23 age_45to54 <--> age_45to54
## err24 age_55to64 <--> age_55to64
## fi14  att <--- notfromRoc
## err30 notfromRoc <--> notfromRoc
## fi31  att <--- hhmem_5ormore
## err47 hhmem_5ormore <--> hhmem_5ormore
## fi36  att <--- emp_ft_3
## err52 emp_ft_3 <--> emp_ft_3
## ti3   onceaweeek <--- att
## ti8   leonceamonth <--- att
## err69 onceaweeek <--> onceaweeek
## err74 leonceamonth <--> leonceamonth
## ti9   bike <--- att
## err75 bike <--> bike
## ti14  biwalk.none <--- att
## ti18  biwalk.oncein3weeks <--- att
## ti20  biwalk.threeormoreaweek <--- att
## err80 biwalk.none <--> biwalk.none
## err84 biwalk.oncein3weeks <--> biwalk.oncein3weeks
## err86 biwalk.threeormoreaweek <--> biwalk.threeormoreaweek
##
## Iterations = 164

```

## Rochester Public Market - Aesthetics

This section shows the final model for RPM aesthetics Coding Notes: Always change the *att* to *aes* in the *model\_RPM\_att\_XXX* file.

### Demographic Variables

#### Aesthetics - Age

Significant variables - *age\_45to55*

#### Aesthetics - Gender

*age\_45to55* and *female* seems interacting however their correlation was  $-0.13$ . This issue needs to be solved. same issue with indicator *male*. Their correlation is  $0.13$ . Confused whether to add gender variable. for now I am not adding it. I think in this situation chi-squared values could be helpful. Chisquared value when using only age variable ( $0.38$ ) is smaller than using gender variable ( $3.71$ ). Therefore I reject the use of age for aesthetics.

#### Aesthetics - How long did the visitor live in rochester

None were significant.

### **Aesthetics - Income**

Significant variables *inc\_100kto125k, inc\_125kto150k*

### **Aesthetics - Number of household members**

No variables were significant. Although there were some errors of hhmem\_5ormore which i couldn't solve it.

### **Aesthetics - Number of household members employed full time employed**

No variables were significant

### **Aesthetics - Number of children in household**

No variables were significant.

### **Aesthetics - Relationship with household members**

Significant variable - rltnHHmem\_acqaintances. Although only 5 people responded in this category. Need to question the relationship.

## **Travel Variables**

Now I am experimenting with the travel variables. Compared to demographics there are only few travel related variables viz.,

*vist* - how often people visited a location.

*mode* - what mode did they use? *biwalk* - have they ever biked or walked to the location *no\_vehicles* - number of vehicles at their home.

*no\_drivers* - number of drivers at their home. *no\_bicylces* - number of bicylces at home \* several variables indicating with whom they traveled to the location on day of interview

Of all the *travel* related variables discussed above, few of them could be informed by SOP dimension. For e.g., *mode*, *visit*, *biwalk*. While they also can be informed by other variables such as distance from home, number of vehicles and drivers using a simple regression. To not complicate the modeling exercise I will currently perform test the significance of the the variables *mode*, *visit*, *biwalk*. Because satisfaction, as a latent variable is already being informed by numerous satisfaction responses these travel variables will not affect the existing satisfaction factor structure.

### **Aesthetics - Visit**

Significant variables: *onceintwo weeks*

### **Aesthetics - Mode**

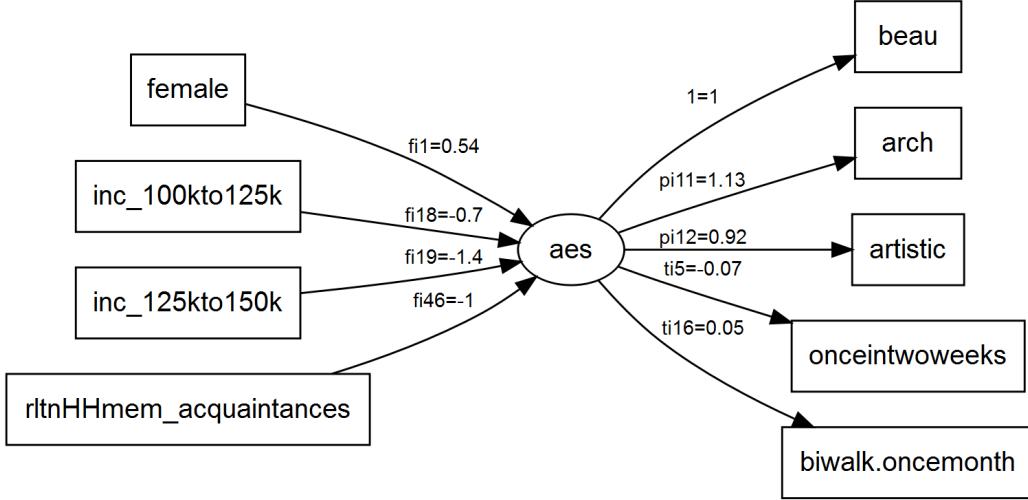
No variables were significant.

### **Aesthetics - Cycle or Walk in Summer**

Notes: Some interesting results here.

## Final Model for RPM. Relationship between Aesthetics, travel and demographic variables

The final model structure is shown below as Figure below. The summary statistics also provided.



```

## NOTE: it is generally simpler to use specifyEquations() or cfa()
##       see ?specifyEquations

##
##  Model Chisquare =  29.37993   Df =  27 Pr(>Chisq) = 0.3427214
##  AIC =  65.37993
##  BIC = -102.8617
##
##  Normalized Residuals
##      Min. 1st Qu. Median     Mean 3rd Qu.     Max.
## -0.98750 -0.11460  0.02705  0.16240  0.39140  2.70100
##
##  R-square for Endogenous Variables
##          aes           beau         arch        artistic
##          0.2009        0.4899        0.5235        0.5118
##          onceintwoweeks biwalk.oncemonth
##          0.0441        0.0494
##
##  Parameter Estimates
##      Estimate Std. Error z value Pr(>|z|)
##  pi11  1.12647050 0.178665316 6.304920 2.883431e-10
##  pi12  0.92044221 0.146563567 6.280157 3.382312e-10
##  var3  0.85667540 0.220835427 3.879248 1.047799e-04
##  err12 1.11630221 0.194521665 5.738704 9.540393e-09
##  err13 1.23810603 0.229998568 5.383103 7.321254e-08
##  err14 0.86653288 0.157247084 5.510645 3.575208e-08
##  fi1   0.54304262 0.200360992 2.710321 6.721810e-03
##  err19 0.23925485 0.029339313 8.154753 3.498936e-16
##  fi18 -0.69949775 0.298395875 -2.344194 1.906826e-02
##  fi19 -1.39907678 0.570041874 -2.454340 1.411433e-02
##  err34 0.10593648 0.012990765 8.154753 3.498936e-16
##  err35 0.02917742 0.003577965 8.154753 3.498936e-16
  
```

```

## fi46 -0.99619881 0.506544117 -1.966658 4.922270e-02
## err62 0.03619122 0.004438053 8.154753 3.498936e-16
## ti5 -0.06781266 0.032082259 -2.113712 3.453986e-02
## err71 0.10674389 0.013267524 8.045502 8.589288e-16
## ti16 0.05105343 0.022865855 2.232737 2.556632e-02
## err82 0.05380481 0.006698924 8.031860 9.600623e-16
##
## pi11 arch <--- aes
## pi12 artistic <--- aes
## var3 aes <--> aes
## err12 beau <--> beau
## err13 arch <--> arch
## err14 artistic <--> artistic
## fi1 aes <--- female
## err19 female <--> female
## fi18 aes <--- inc_100kto125k
## fi19 aes <--- inc_125kto150k
## err34 inc_100kto125k <--> inc_100kto125k
## err35 inc_125kto150k <--> inc_125kto150k
## fi46 aes <--- rlttnHHmem_acquaintances
## err62 rlttnHHmem_acquaintances <--> rlttnHHmem_acquaintances
## ti5 onceintwoweeks <--- aes
## err71 onceintwoweeks <--> onceintwoweeks
## ti16 biwalk.oncemonth <--- aes
## err82 biwalk.oncemonth <--> biwalk.oncemonth
##
## Iterations = 110

```