THE ROLE OF ATHLETIC ACTIVITY IN SUBSTANCE USE

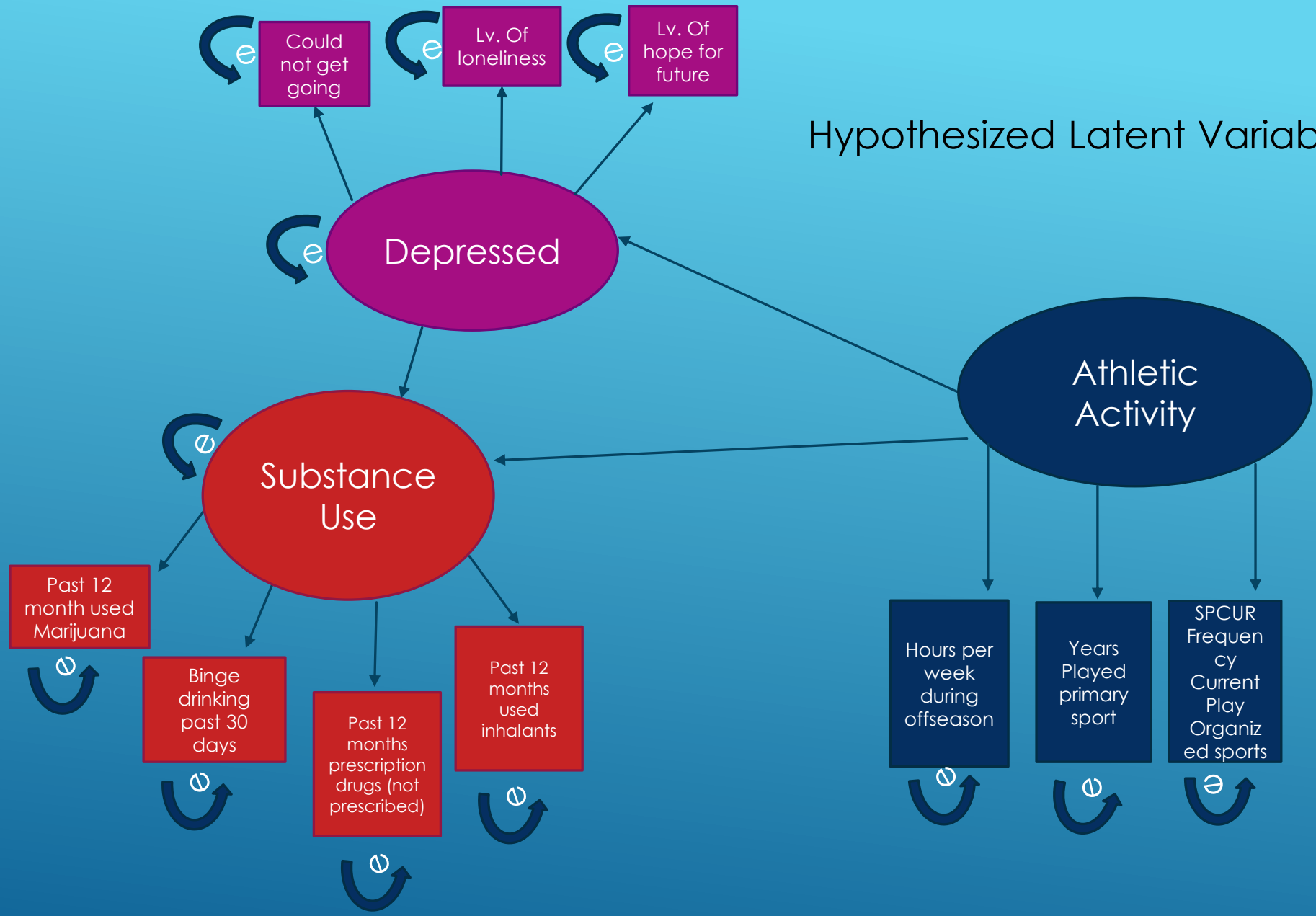
QAC 313

By: Coleton Reitan

RESEARCH STATEMENT

- I was curious to know if people with different levels of athletic activity showed different levels of depression, and if this was related to their substance use levels. This led me to the following research statement:
- The purpose of the research is to have a better understanding of the association between someone's level of athletic activity and substance use while using their level of depression as a mediator variable.
- I hypothesize that as physical activity decreases, depression increases, and thus I would expect one to have higher levels of substance use.
- I will be using R software to complete this research.

Hypothesized Latent Variable Model Diagram



METHODS

METHODS: SAMPLE

The Dataset I worked with for my research was the “Athletic Involvement Study” survey data.

- The survey involved 795 completed observations from students at a large college in northeastern US.
- There is a variety of questions asked in the survey ranging from involvement in athletics to mental health questions.
- I dropped all observations with missing data, leaving n=367 total observations.
- Keep in mind, the sample size is relatively small and may have an affect on test statistics.

METHODS: MEASURES

To have the best model I could, I created three latent variables in which I decided to base my focus of study around.

1. The first latent variable is Athletic Activity, which is comprised of three observed variables:
    - Hours per week during the offseason dedicated towards primary sport (PSHRO). Measured quantitatively, with missing observations dropped.
    - Years played in primary sport (PSYEARS). Measured quantitatively.
    - Frequency Played in Organized Sports (SPCUR). Measured Quantitatively on a scale of 1-5.

METHODS: MEASURES

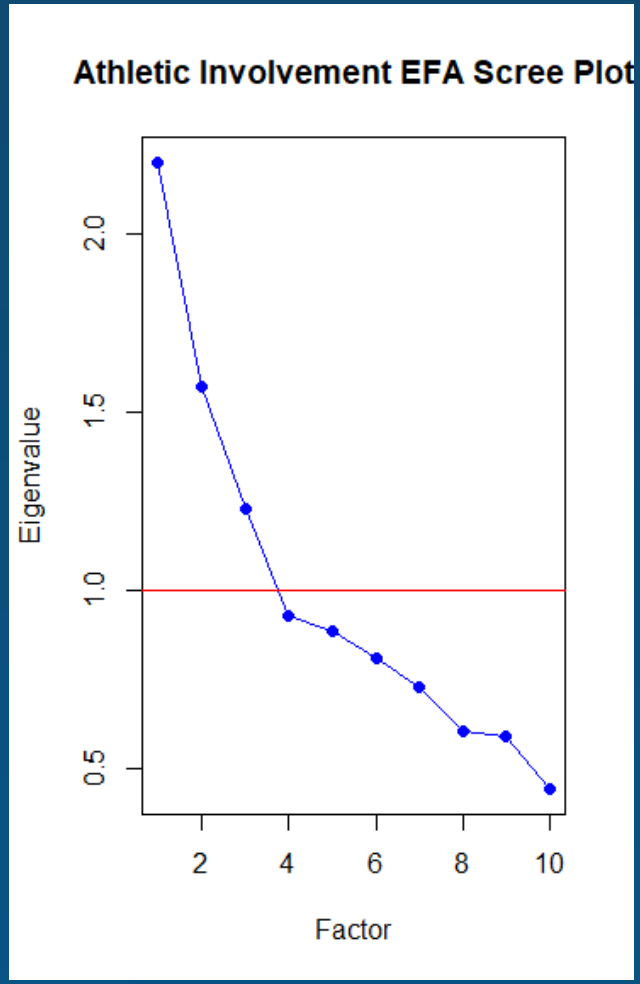
2. The second latent variable is Depression, which is also comprised of three observed variables:
    - Could not get going (CESD10). Measured quantitatively on a scale from 1-4.
    - How lonely one feels (CESD09). Measured quantitatively on a scale from 1-4.
    - Hopeful for the future (CESD05). Measured quantitatively on a scale from 1-4.

3. The third and final latent variable is Substance Use, which is comprised of four observed variables:
    - Past 12 months marijuana use (DAI01). Measured quantitatively on a scale from 0-4.
    - Past 12 months taken prescription drugs without prescription (DAI04). Measured quantitatively on a scale from 0-4.
    - Past 30 days binge drink (SAI06). Measured quantitatively on a scale from 0-4.
    - Past 12 months used inhalants (DAI02). Measured quantitatively on a scale from 0-4.

METHODS: ANALYSIS

In my analysis of the data, I went through four phases that helped me gather a better understanding of my model (after data collection).

1. The first was Exploratory Factor Analysis, in which I was able to learn how many (latent) factors I should retain.
   - Through this process, I saw that my model showed 3 significant eigenvalues (with a value above 1), telling me that I should retain 3 latent factors for my research, just as I had anticipated.
   - We can see there are 3 blue dots on the graph that are above the red line with significant drop off between each of them, signifying there should be 3 factors retained.
   - Although there is a steep drop off between factors 3 and 4 signifying there may be a possible 4th factor involved, but with the value below 1, the 4th factor will not be retained.



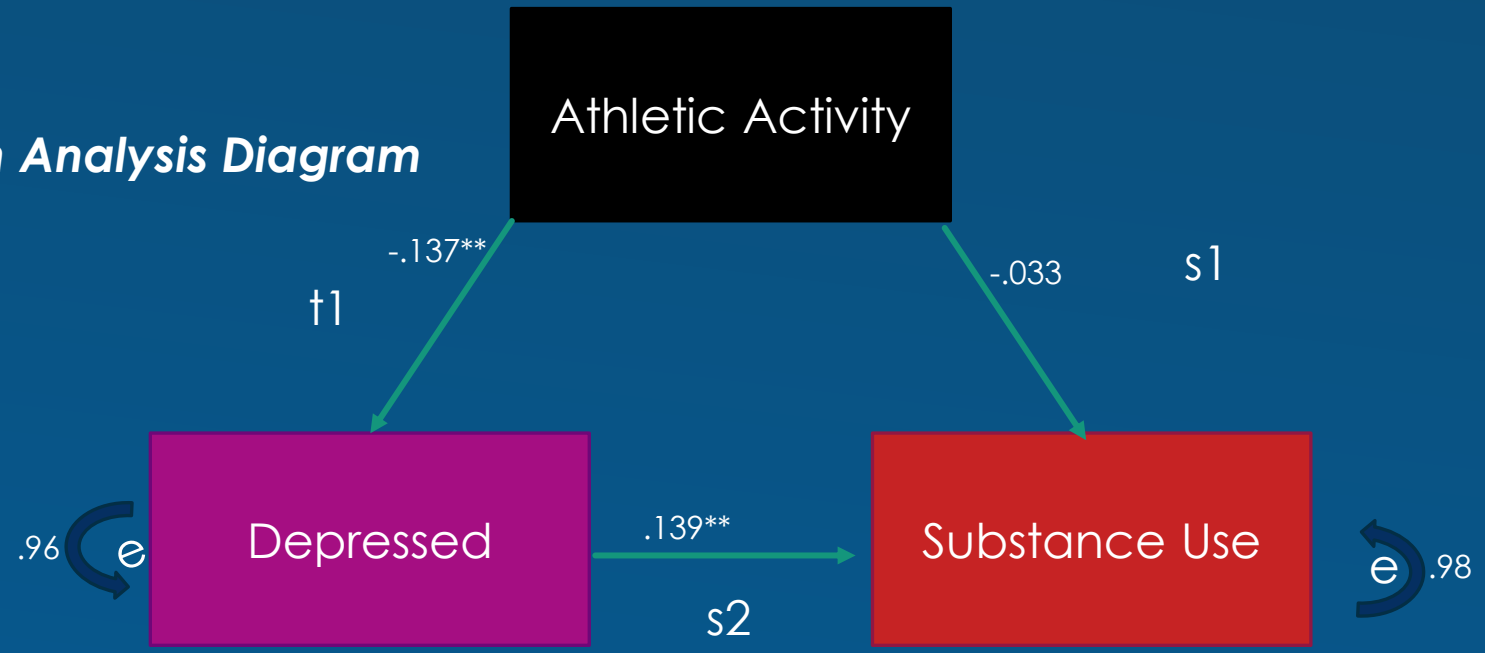
METHODS: ANALYSIS

2. The next phase was performing a confirmatory factor analysis, in which I specify my latent variable model based off the prior explanatory factor analysis.
    - Through this process, I was able to see the association between my latent factors and their respective observed variables.
    - I was also able to get an understanding as to how good the fit of my model is, and whether I could consider it reliable.
    - CFA gives me my final latent factor construct I will be using through the rest of the research.
    - More detail on my confirmatory factor analysis to come.

METHODS: ANALYSIS

3. The following phase is creating and testing a path analysis model, which helps me in understanding the relationship between factors.
   - In order to do this, I must turn all latent variables into observed variables.
   - This is done by creating composite variables from my latent variables using their respective observed variables, then standardizing the values.
   - From this model, I can also see if the fit of my model is good by looking at test statistics given as output.

Path Analysis Diagram



METHODS: ANALYSIS

4. The last phase in my research is to create a final latent variable model.
   - With this model, I can finally have some closure to my research.
   - I can see:
     - the association between latent factors.
     - the association between latent factors and their respective observed variables.
     - Error values
   - I will be looking at the test statistics as well to understand if my model was a good fit.
   - More details to come on my final latent model.

CONFIRMATORY FACTOR ANALYSIS

CFA: FIT STATISTICS

The CFA model has many test statistics and taking a closer look at these statistics could give us a better understanding of whether the model shows a good fit or not.

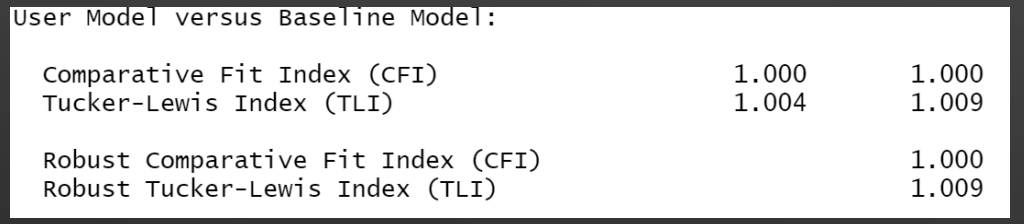
- To start, we'll take a look at the chi-squared value and its respective statistics. It is seen that the chi-square value is 29.854 with 32 degrees of freedom. However, it is also apparent that the p-value is well above .05 with a value of .576, leading me to believe that this model may be misspecified.

Model Test User Model:

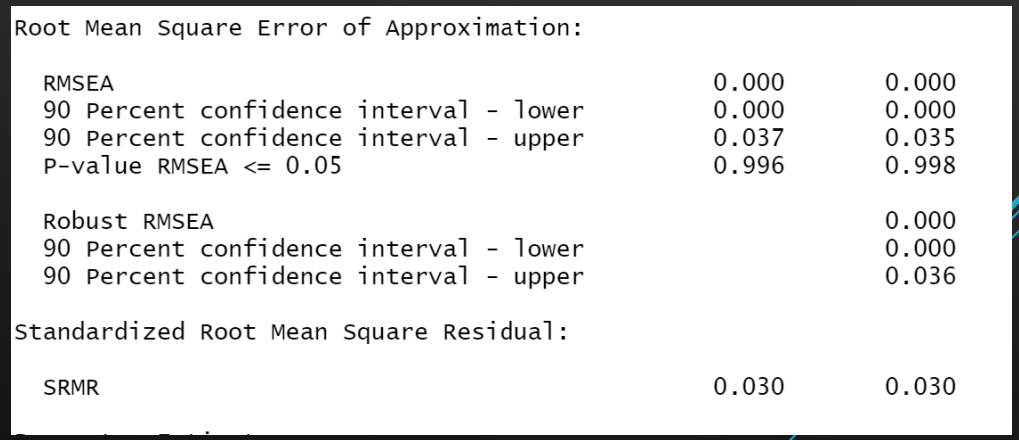
	Standard	Robust
Test Statistic	31.095	29.854
Degrees of freedom	32	32
P-value (Chi-square)	0.512	0.576

CFA: FIT STATISTICS

- The next fit statistics are the incremental indexes, which are the comparative fit index and Tucker Lewis fit index. For both of these values, we would anticipate a value above .90, which typically signals a good fitting model. However, we can see that the CFI and TLI are 1.00, and 1.009, respectively. This also tells me that my model fit may be misspecified.
- The last fit statistics we will be looking at is the RMSEA and SRMR. For the robust RMSEA, it is clear that the total and lower bound is 0.00 while the upper is .036. Although all values are below .08, the 0.00 values may show some more misspecification in my model. My SRMR is below .05, which could shed some light that my model isn't as poorly fit as it may seem.



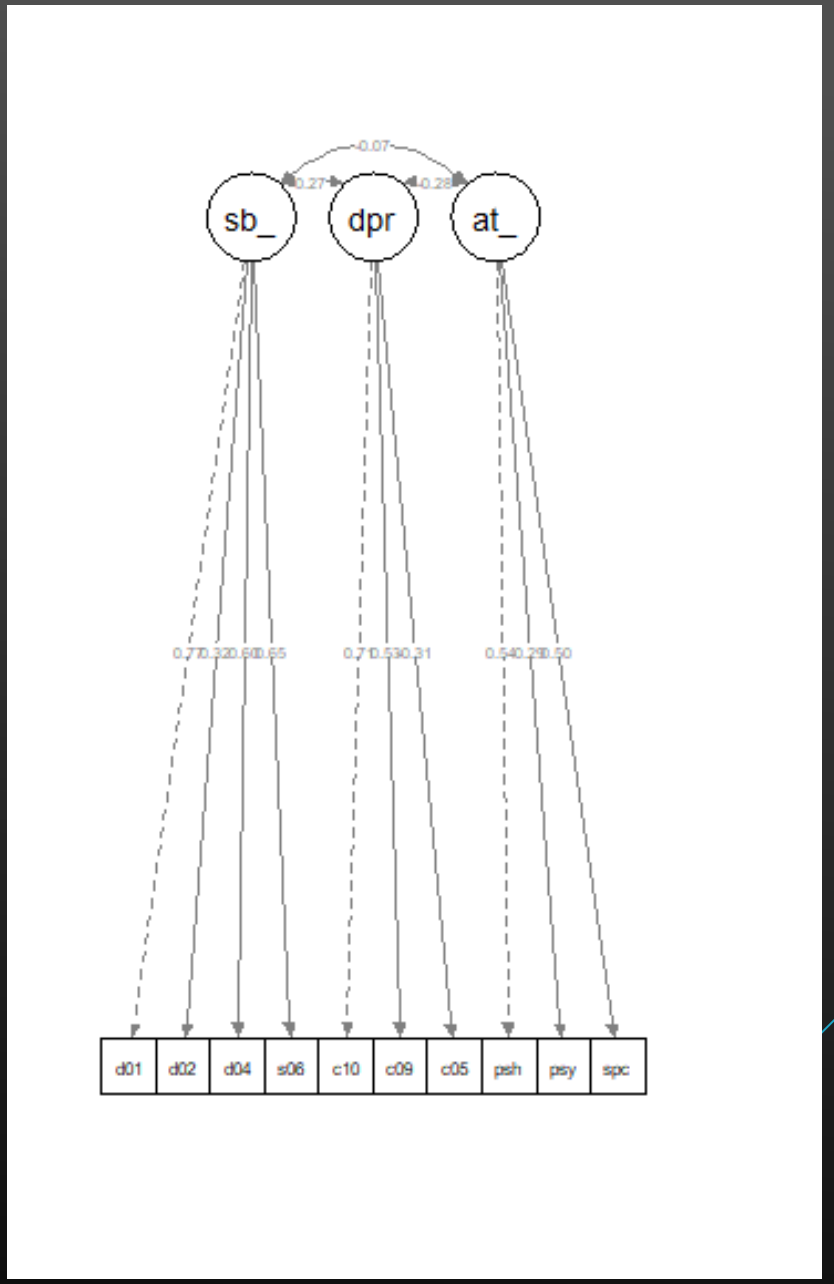
```
User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    1.000       1.000
  Tucker-Lewis Index (TLI)                       1.004       1.009

  Robust Comparative Fit Index (CFI)                         1.000
  Robust Tucker-Lewis Index (TLI)                            1.009
```



```
Root Mean Square Error of Approximation:

  RMSEA                                          0.000       0.000
  90 Percent confidence interval - lower         0.000       0.000
  90 Percent confidence interval - upper         0.037       0.035
  P-value RMSEA <= 0.05                          0.996       0.998

  Robust RMSEA                                               0.000
  90 Percent confidence interval - lower                     0.000
  90 Percent confidence interval - upper                     0.036

Standardized Root Mean Square Residual:

  SRMR                                           0.030       0.030
```

CFA: MODEL DIAGRAM AND OUTPUT

```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  substance_use =~
    dai01             1.000                              1.174    0.770
    dai02             0.093    0.035    2.688    0.007    0.109    0.324
    dai04             0.396    0.056    7.089    0.000    0.464    0.599
    sai06             0.858    0.091    9.430    0.000    1.007    0.649
  depressed =~
    cesd10            1.000                              0.590    0.705
    cesd09            0.803    0.248    3.241    0.001    0.474    0.527
    cesd05           -0.487    0.128   -3.789    0.000   -0.287   -0.309
  athletic_activity =~
    pshro             1.000                              4.856    0.543
    psyears           0.283    0.099    2.856    0.004    1.372    0.295
    spcur             0.113    0.046    2.475    0.013    0.547    0.503

Covariances:
                 Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  substance_use ~~
    athletic_ctvty  -0.404    0.529   -0.764    0.445   -0.071   -0.071
  depressed ~~
    athletic_ctvty  -0.806    0.366   -2.199    0.028   -0.281   -0.281
  substance_use ~~
    depressed        0.185    0.067    2.757    0.006    0.266    0.266
```



CFA: RESULTS SUMMARY

- Looking at the standardized factor loadings, we can see that the p-value for all of the variables is less than .05, which means that they loaded significantly for their own factor, allowing me to assume the correlation given between the factor and variable is true. The std.all values (standardized factor loadings) shows how correlated the variables are with the latent factor, the closer to 1 (or -1), the stronger the relationship. Also for latent observed variables, there are three factor loadings (absolute valued) not greater than .35, indicating that most of the correlations are worth noting.
  - It is seen that the association between latent factors depressed and athletic activity is significant with a p-value of .028 and a value of -.281.
  - The association between latent factors substance use and depressed is significant with a p-value of .006 and a value of .266.
  - The association between latent factors athletic activity and substance use is insignificant, with a p-value of .445.

  - Depressed factor highest/lowest loadings: Could not get going (CESD10) .705 / Hopeful for the future(CESD05) -.309
  - Athletic Activity factor highest/lowest loadings: Primary sport hours (PSHRO) .543 / Primary sport years (PSYEAR) .295
  - Substance Use Factor highest/lowest loadings: Past 12 months marijuana use (DAI01) .770 / Past 12 months inhalants (DAI02) .324

FINAL LATENT VARIABLE MODEL

FINAL LATENT VARIABLE MODEL: FIT STATISTICS

The final latent variable model also has many test statistics and taking a closer look at these statistics could give us a better understanding of whether the model shows a good fit or not.

- Looking at the Model Test User Model, we can see the chi square test statistic of 29.854, as well as the degrees of freedom and p-value. The p-value is above .05, showing that the chi square value is not significant, suggesting that my model may not have a good fit. We can also see that the test statistic is larger than the degrees of freedom, indicating the model fit may be off.

Model Test User Model:		
	Standard	Robust
Test Statistic	31.095	29.854
Degrees of freedom	32	32
P-value (Chi-square)	0.512	0.576

FINAL LATENT VARIABLE MODEL: FIT STATISTICS

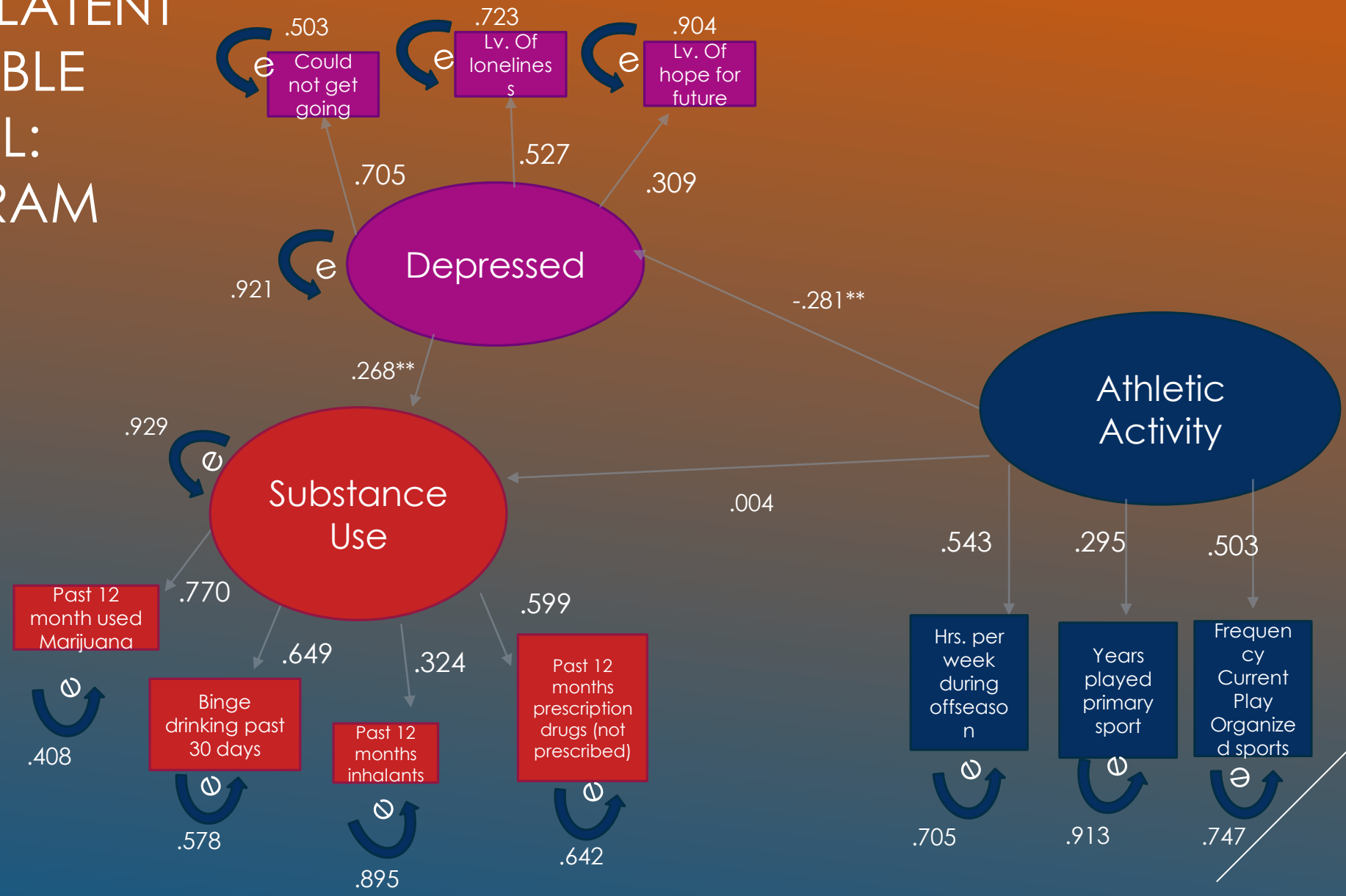
- By looking at the CFI and TLI, it is obvious that both indexes are above .90, possibly indicating the fit of the model actually isn't terrible, but still equal to or above 1.00, telling me there may have been some misspecification.
- Also looking at the robust RMSEA, we can see the total and the confidence intervals are all under .08, showing some display that the model is a good fit. Finally, the SRMR is .032, below .05, showing some upside to my model.

Robust RMSEA	0.000
90 Percent confidence interval - lower	0.000
90 Percent confidence interval - upper	0.036

Standardized Root Mean Square Residual:

SRMR	0.032	0.032

FINAL LATENT VARIABLE MODEL: DIAGRAM



FINAL LATENT. VARIABLE MODEL: RESULTS SUMMARY

Looking at the results of our model, we can see the correlation between athletic activity and depression is significant, with a value of -.281, and a p-value of .034.

- For the correlation between substance use and depression, we can see there is a significant relationship, with a standardized coefficient of .268 and a p-value of .021.
- What these tell me is
  - as someone's athletic activity increases, their depression levels tend to decrease
  - as one's depression levels decrease, their substance use levels tend to decrease.
- However, we see for our last latent regression analysis of athletic activity and substance use, there is a lack of significance, as well as a very small standardized regression coefficient.
- With this information, we cannot say that there is an association between one's athletic activity and substance use with depression as a mediator. We can also see this in the final line of output, as the standardized coefficient of our path is tot_t1s2 = -.001 (very small), and the p-value shows large insignificance , proving my hypothesis somewhat right.
- The Rsquared value for my response variables depression and substance use was .079 and .071, respectively. What this really means is that 7.9% of the response variable, depression, can be explained by athletic activity (7.1% of substance use can be explained by athletic activity.

FINAL LATENT VARIABLE MODEL: RESULTS SUMMARY LIMITATIONS

- Since the sample size was so small, I would be curious to see a model with a sample size at least 2x larger than mine (n>800).
  - The small sample size could create under identification of the latent factors and could create an untrue model.
- I would also be curious to add more observed variables to my latent factors. Because of time and resource constraints, latent factors were possibly misspecified.

THANK YOU!