Lab 15 - Multivariate Regression & Interpretation

Jackson Nahpi

November 30, 2017

Complete the following exercises below and include all code used to find the answers. Knit together the PDF document and commit both the Lab 15 RMD file and the PDF document to Git. Push the changes to GitHub so both documents are visible in your public GitHub repository.

1. Select a second explanatory variable from your dataset that you think has implications for the theoretical association of your focal relationship.

a. Describe the theoretical reasoning for selecting this variable.

I am using the percieved probability of finishing a graduate education as my response variable, my first explanatory variable will be the level of rewards the respondent is given for grades, and my second explanatory variable will be the level of the respondent having their homework checked in their household. I believe that these are all good actions that can be done towards children to ensure them to pursue higher education by teaching good work ethic. Therefore, I think both explanatory variables will have a positive relationship with the response variable I have chosen.

b. What type of relationship do you think this variable has with your focal variables? Given that, what do you expect to happen to your focal relationship when it is added to the model?

I believe there will be positive relationships between the response variable and both of the explanatory variables. However, I think the original significance between my response variable and my first explanatory variable will be weakened by the addition of a new variable.

c. Is it a continuous or categorical variable? What implications does this have for a multivariate regression equation?

The variable I am adding in is a categorical variable which means we have the potential of having simpsons paradox occur which means that the groups can act as a confounding variable. A confounding variable's effect can cause results given to be irrelevant if they are not handled well.

d. Conduct a multivariate linear regression with this additional explanatory variable and save the model object. Print out the full results by calling `summary()` on your model object.

```
##
## Call:
## lm(formula = probfingraddeg ~ rwrdfrgrdslyr + checkhwlyr, data = freshman_dataset_v4)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.190 -1.179  0.298  1.951  2.439
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    7.35498    0.09737  75.533  < 2e-16 ***
## rwrdfrgrdslyr  0.14090    0.02926   4.816 1.52e-06 ***
## checkhwlyr     0.06520    0.03278   1.989   0.0468 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.187 on 3646 degrees of freedom
## Multiple R-squared:  0.009764,   Adjusted R-squared:  0.009221
```

```
## F-statistic: 17.98 on 2 and 3646 DF,  p-value: 1.705e-08
```

```
##
## Call:
## lm(formula = probfingraddeg ~ rwrdfrgrdslyr + factor(checkhwlyr),
##     data = freshman_dataset_v4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.2277 -1.1498  0.3436  1.9930  2.4873
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          7.51360    0.09114  82.436  < 2e-16 ***
## rwrdfrgrdslyr        0.14283    0.02926   4.882  1.1e-06 ***
## factor(checkhwlyr)2 -0.14376    0.08933  -1.609  0.10762
## factor(checkhwlyr)3 -0.07791    0.10515  -0.741  0.45877
## factor(checkhwlyr)4  0.08684    0.13713   0.633  0.52661
## factor(checkhwlyr)5  0.54649    0.17784   3.073  0.00214 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.184 on 3643 degrees of freedom
## Multiple R-squared:  0.01326,    Adjusted R-squared:  0.0119
## F-statistic:  9.79 on 5 and 3643 DF,  p-value: 2.622e-09
```

```
##
## Call:
## lm(formula = probfingraddeg ~ factor(rwrdfrgrdslyr) + factor(checkhwlyr),
##     data = freshman_dataset_v4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.4454 -1.3131  0.2946  2.0877  2.4498
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)             7.68252    0.08317  92.368  < 2e-16 ***
## factor(rwrdfrgrdslyr)2  0.15528    0.11280   1.377  0.16871
## factor(rwrdfrgrdslyr)3  0.19852    0.10933   1.816  0.06950 .
## factor(rwrdfrgrdslyr)4  0.22979    0.11977   1.919  0.05512 .
## factor(rwrdfrgrdslyr)5  0.76290    0.13603   5.608  2.2e-08 ***
## factor(checkhwlyr)2    -0.13237    0.09077  -1.458  0.14485
## factor(checkhwlyr)3    -0.06152    0.10642  -0.578  0.56322
## factor(checkhwlyr)4     0.11447    0.13781   0.831  0.40622
## factor(checkhwlyr)5     0.51762    0.17800   2.908  0.00366 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.182 on 3640 degrees of freedom
## Multiple R-squared:  0.01581,    Adjusted R-squared:  0.01365
## F-statistic: 7.311 on 8 and 3640 DF,  p-value: 1.114e-09
```

e. Describe the results of the multivariate analysis, highlighting:

- the apparent association between the control variable and the focal response variable

- how the focal association changed when you incorporated the control variable
- the implications of these results for your focal association

I ran two seperate multivariate analyses in order to see what happens when you account for a factor of a categorical value. What I found was that the model where I factored the second explanatory variable resulted in more significant results. The relationship is significant to the 0.001 level with the factored control variable. The association overall decreased the significance of the rwrdfrgrdslyr's significance but the model gained anothe significant variable of checkhwlyr at the 5th level. These results imply that checkhwlyr belongs in our model. The third table was done out of curiosity just to see what would happen, however I need to consult my professor before reporting on it's validity.

f. How well does this model fit the data? Is it an improvement over the bivariate model? Why or why not?

The first and second model in this question both have better fit than in the bivariate case. The second model has a better fit than the first model with a fit of 0.01326 (and an adjusted R-squared of 0.0119!), even thought this is still a pretty weak fit it is an improvement. By adding more variables we may be able to further improve our multivariate regression.

2. Select any additional variables you want to incorporate into your final model. For each additional variable added to the model answer the following questions:

a. Describe the theoretical reasoning for selecting this variable.

I want to add in 4 new variables. First is percieved level of punishment for bad grades (pnshfrgrdslyr) towards the respondent in their last year of secondary school. I feel this variable is significant and different than rewards for grades. Punishment can be used as a deterent to bad grades and the reward the respondent would recieve is not being punishment. Even though using punishment for bad grades as a deterent is a cruel method of teaching children good work ethic, it might be useful in seeing whether it has any effect of percieved possibility of achieving a graduates degree. I have a feeling this will show a positive relationship. The second variable I want to include is the percieved level of punishment for disobediance (pnshdeviancelyr) in the last year of their secondary education. I believe a positive relationship will be present since punishment for disobediance may cause respondents to behave better in school from a perspective of not wanting to have bad feedback given to their parents from teachers. My fourth variable is sex (sex), I do not know what relationship to expect I mainly want to see if it is significant in any way. Lastly, I want to include ethnicity (ethnicity) into the model. This study was primarily focused on seeing why their was a lack of diversity in higher education, so it would be interesting to see whether it does have a significant effect in the percieved possibility of achieving a graduate degree.

b. What type of relationship do you think this variable has with your focal variables? Given that, what do you expect to happen to your focal relationship when it is added to the model?

I believe the type of relationships will occur with the following variables: pnshfrgrdslyr will be positive, pnshdeviancelyr will be positive, sex I do not have any intuition as to which way it will go but I am certain it will be significant, and last is ethnicity which I think will be positively associated with white (when ethnicity=2). I think the relationship between the focal variables will decrease but the model will increase in fit.

c. Is it a continuous or categorical variable? What implications does this have for a multivariate regression equation?

All of these variables are categorical but we can account for this by factoring them which allows us to see what specific levels are actually significant to the model. By doing this we also rule out potential confounding variables.

d. Conduct a multivariate linear regression by adding one explanatory variable at a time and save the model objects. Print out the full results by calling `summary()` on each model object.

```
##
## Call:
```

```
## lm(formula = probfingraddeg ~ rwrdfrgrdslyr + factor(checkhwlyr) +
##     factor(pnshfrgrdslyr) + factor(pnshdeviancelyr) + factor(sex) +
##     factor(ethnicity), data = freshman_dataset_v4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.6403 -1.2302  0.4275  1.7782  3.2321
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 7.662187   0.127134  60.269  < 2e-16 ***
## rwrdfrgrdslyr               0.115441   0.029802   3.874 0.000109 ***
## factor(checkhwlyr)2        -0.085940   0.090794  -0.947 0.343937
## factor(checkhwlyr)3        -0.020264   0.107143  -0.189 0.849999
## factor(checkhwlyr)4         0.210922   0.138768   1.520 0.128608
## factor(checkhwlyr)5         0.514127   0.180019   2.856 0.004315 **
## factor(pnshfrgrdslyr)2     -0.091066   0.095098  -0.958 0.338330
## factor(pnshfrgrdslyr)3      0.008149   0.133505   0.061 0.951329
## factor(pnshfrgrdslyr)4     -0.161049   0.173663  -0.927 0.353800
## factor(pnshfrgrdslyr)5     -0.186250   0.225916  -0.824 0.409756
## factor(pnshdeviancelyr)2 -0.143871   0.094358  -1.525 0.127412
## factor(pnshdeviancelyr)3 -0.203888   0.112283  -1.816 0.069477 .
## factor(pnshdeviancelyr)4  0.019329   0.148949   0.130 0.896756
## factor(pnshdeviancelyr)5  0.203907   0.193335   1.055 0.291641
## factor(sex)1                0.472644   0.073318   6.446 1.30e-10 ***
## factor(ethnicity)2         -0.704814   0.101659  -6.933 4.85e-12 ***
## factor(ethnicity)3         -0.063830   0.101114  -0.631 0.527901
## factor(ethnicity)4         -0.337942   0.100692  -3.356 0.000798 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.152 on 3631 degrees of freedom
## Multiple R-squared:  0.04516,    Adjusted R-squared:  0.04069
## F-statistic:  10.1 on 17 and 3631 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = probfingraddeg ~ rwrdfrgrdslyr + factor(checkhwlyr) +
##     factor(pnshdeviancelyr) + factor(sex) + factor(ethnicity),
##     data = freshman_dataset_v4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.6676 -1.2544  0.4279  1.7825  3.2433
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 7.66758    0.12692  60.411  < 2e-16 ***
## rwrdfrgrdslyr               0.11119    0.02946   3.774 0.000163 ***
## factor(checkhwlyr)2        -0.09646    0.08964  -1.076 0.281991
## factor(checkhwlyr)3        -0.03792    0.10533  -0.360 0.718903
## factor(checkhwlyr)4         0.18317    0.13658   1.341 0.179972
## factor(checkhwlyr)5         0.48217    0.17778   2.712 0.006715 **
## factor(pnshdeviancelyr)2 -0.16346    0.09162  -1.784 0.074498 .
```

```
## factor(pnshdeviancelyr)3 -0.22536    0.10675  -2.111 0.034831 *
## factor(pnshdeviancelyr)4 -0.03468    0.13695  -0.253 0.800096
## factor(pnshdeviancelyr)5  0.11359    0.16864   0.674 0.500618
## factor(sex)1              0.47872    0.07294   6.563 6.03e-11 ***
## factor(ethnicity)2       -0.70029    0.10148  -6.900 6.09e-12 ***
## factor(ethnicity)3       -0.07208    0.10083  -0.715 0.474746
## factor(ethnicity)4       -0.33891    0.10062  -3.368 0.000764 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.152 on 3635 degrees of freedom
## Multiple R-squared:  0.04463,    Adjusted R-squared:  0.04121
## F-statistic: 13.06 on 13 and 3635 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = probfingraddeg ~ rwrdfrgrdslyr + factor(checkhwlyr) +
##     factor(pnshfrgrdslyr) + factor(pnshdeviancelyr) + factor(sex) +
##     factor(ethnicity) + factor(musmlyr), data = freshman_dataset_v4)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.560 -1.228  0.401  1.786  3.306
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                7.67435    0.12693  60.460  < 2e-16 ***
## rwrdfrgrdslyr              0.10008    0.02993   3.344 0.000834 ***
## factor(checkhwlyr)2       -0.11954    0.09126  -1.310 0.190323
## factor(checkhwlyr)3       -0.07675    0.10817  -0.710 0.478013
## factor(checkhwlyr)4        0.12960    0.13971   0.928 0.353665
## factor(checkhwlyr)5        0.43719    0.18000   2.429 0.015194 *
## factor(pnshfrgrdslyr)2    -0.09117    0.09480  -0.962 0.336282
## factor(pnshfrgrdslyr)3     0.02359    0.13323   0.177 0.859497
## factor(pnshfrgrdslyr)4    -0.15263    0.17305  -0.882 0.377851
## factor(pnshfrgrdslyr)5    -0.20829    0.22556  -0.923 0.355856
## factor(pnshdeviancelyr)2 -0.13766    0.09412  -1.463 0.143663
## factor(pnshdeviancelyr)3 -0.19088    0.11191  -1.706 0.088164 .
## factor(pnshdeviancelyr)4  0.05207    0.14855   0.351 0.725947
## factor(pnshdeviancelyr)5  0.23736    0.19278   1.231 0.218300
## factor(sex)1               0.42399    0.07376   5.748 9.77e-09 ***
## factor(ethnicity)2        -0.76989    0.10303  -7.472 9.81e-14 ***
## factor(ethnicity)3        -0.09744    0.10114  -0.963 0.335390
## factor(ethnicity)4        -0.39934    0.10141  -3.938 8.37e-05 ***
## factor(musmlyr)2           0.10047    0.08762   1.147 0.251581
## factor(musmlyr)3           0.38472    0.10750   3.579 0.000350 ***
## factor(musmlyr)4           0.32171    0.17834   1.804 0.071332 .
## factor(musmlyr)5           1.26085    0.28922   4.359 1.34e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.145 on 3627 degrees of freedom
## Multiple R-squared:  0.05301,    Adjusted R-squared:  0.04753
## F-statistic: 9.669 on 21 and 3627 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = probfingraddeg ~ factor(rwrdfrgrdslyr) + factor(checkhwlyr) +
##     factor(pnshdeviancelyr) + factor(sex) + factor(ethnicity) +
##     factor(musmlyr) + factor(college), data = freshman_dataset_v4)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.845 -1.258  0.405  1.762  3.427
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 7.76396    0.16457  47.176  < 2e-16 ***
## factor(rwrdfrgrdslyr)2      0.13094    0.11172   1.172 0.241246
## factor(rwrdfrgrdslyr)3      0.14231    0.10883   1.308 0.191071
## factor(rwrdfrgrdslyr)4      0.15707    0.12010   1.308 0.191010
## factor(rwrdfrgrdslyr)5      0.54867    0.13734   3.995 6.60e-05 ***
## factor(checkhwlyr)2        -0.11788    0.09135  -1.290 0.197021
## factor(checkhwlyr)3        -0.07551    0.10754  -0.702 0.482600
## factor(checkhwlyr)4         0.13386    0.13828   0.968 0.333061
## factor(checkhwlyr)5         0.40310    0.17802   2.264 0.023610 *
## factor(pnshdeviancelyr)2 -0.13930    0.09192  -1.516 0.129729
## factor(pnshdeviancelyr)3 -0.18925    0.10686  -1.771 0.076644 .
## factor(pnshdeviancelyr)4  0.03664    0.13727   0.267 0.789532
## factor(pnshdeviancelyr)5  0.12678    0.16894   0.750 0.453019
## factor(sex)1                0.43874    0.07358   5.963 2.71e-09 ***
## factor(ethnicity)2         -0.75785    0.10284  -7.369 2.12e-13 ***
## factor(ethnicity)3         -0.09924    0.10089  -0.984 0.325352
## factor(ethnicity)4         -0.39195    0.10141  -3.865 0.000113 ***
## factor(musmlyr)2            0.09752    0.08774   1.111 0.266432
## factor(musmlyr)3            0.37124    0.10783   3.443 0.000582 ***
## factor(musmlyr)4            0.30531    0.17846   1.711 0.087195 .
## factor(musmlyr)5            1.19243    0.28902   4.126 3.78e-05 ***
## factor(college)2            0.05654    0.12616   0.448 0.654086
## factor(college)3           -0.12578    0.13368  -0.941 0.346806
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.142 on 3626 degrees of freedom
## Multiple R-squared:  0.05527,    Adjusted R-squared:  0.04953
## F-statistic: 9.642 on 22 and 3626 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = probfingraddeg ~ factor(rwrdfrgrdslyr) + factor(checkhwlyr) +
##     factor(pnshdeviancelyr) + factor(sex) + factor(ethnicity) +
##     factor(musmlyr) + factor(college) + probextraeduc, data = freshman_dataset_v4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.8480 -0.0659  0.2233  0.3886  9.4689
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)                0.252572   0.110535   2.285 0.022371 *
## factor(rwrdfrgrdslyr)2     0.039259   0.056392   0.696 0.486367
## factor(rwrdfrgrdslyr)3     0.019852   0.054942   0.361 0.717871
## factor(rwrdfrgrdslyr)4     0.045908   0.060627   0.757 0.448960
## factor(rwrdfrgrdslyr)5     0.179481   0.069410   2.586 0.009754 **
## factor(checkhwlyr)2        0.032606   0.046131   0.707 0.479724
## factor(checkhwlyr)3        0.031728   0.054288   0.584 0.558965
## factor(checkhwlyr)4        0.043005   0.069797   0.616 0.537836
## factor(checkhwlyr)5       -0.058043   0.089961  -0.645 0.518838
## factor(pnshdeviancelyr)2 -0.021913   0.046407  -0.472 0.636813
## factor(pnshdeviancelyr)3 -0.054935   0.053952  -1.018 0.308638
## factor(pnshdeviancelyr)4 -0.012282   0.069284  -0.177 0.859311
## factor(pnshdeviancelyr)5  0.068721   0.085268   0.806 0.420329
## factor(sex)1               0.127131   0.037259   3.412 0.000652 ***
## factor(ethnicity)2        -0.161199   0.052228  -3.086 0.002041 **
## factor(ethnicity)3        -0.124071   0.050923  -2.436 0.014880 *
## factor(ethnicity)4        -0.126101   0.051249  -2.461 0.013918 *
## factor(musmlyr)2          -0.028393   0.044300  -0.641 0.521607
## factor(musmlyr)3           0.032909   0.054523   0.604 0.546156
## factor(musmlyr)4          -0.033809   0.090131  -0.375 0.707600
## factor(musmlyr)5           0.215031   0.146183   1.471 0.141387
## factor(college)2           0.145918   0.063682   2.291 0.021999 *
## factor(college)3           0.105473   0.067506   1.562 0.118279
## probextraeduc              0.920526   0.008937 103.000  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.081 on 3625 degrees of freedom
## Multiple R-squared:  0.7594, Adjusted R-squared:  0.7579
## F-statistic: 497.5 on 23 and 3625 DF,  p-value: < 2.2e-16
```

e. Describe the results of the multivariate analysis, highlighting:

- the apparent association between each additional control variable and the focal response variable
- how the focal association changed when you incorporated each control variable
- the implications of these results for your focal association

I conducted multiple regressions in the first I saw that the pnshfrgrdslyr was insignificant to the model. So I removed it and conducted another regression with the rest of the variables and saw that a new variable in punishment for deviance became significant.

Out of curiosity I added another variable into the third regression but used the first model I had for this question. I added the variable frequency of museum visits in the last year of secondary school that the adults of your household took you to. To my surprise it came out as a more significant model and even had multiple significant values. So in the final model I removed pnshdeviancelyr from the model and got a better model than the previous one since the adjusted R-squared is higher.

Overall I think the final model for this question gives the best model we can get with the dataset I have created. In this ending model, almost every variable is significant to a minimum of the 5% level. The only variable included that has no statistical significance is the college selection, I kept college in simply to control for the variable. I also made sure to factor for each of the variables in my model since almost all of my data is categorical, with the exception of my dependent variable. In these following paragraphs I'll talk about the factored variables, however I will discuss the variables sex and ethnicity last.

The first independent variable (rwrdfrgrdslyr) has a positive association with the dependent variable, the only significant factor form from the regression is 5th level which is significant to the 0 level. The theoretical argument that can come from this finding is that high levels of rewards for grades in school may inspire

higher pursuits in education.

The second independent variable (checkhwlyr) has a positive association with the dependent variable at the 0.01 level. The significant factored level is the 5th level. Implications for the result recieved is that individuals who's parents check their homework frequently causes their children to have higher educational pursuits.

The third independent variable (pnshdeviancelyr) has a negative association with the dependent variable at the 0.05 level. The only factored version that has significance is at the third level of the factored variable. Implications for research from this result can be that parents who sometimes punish deviant behaviour tend to lower the educational pursuits of their child.

The next independent variable (musmlyr) has a positive association with the dependent variable at multiple factored levels. Its is significant at the 3 and 5 level at a 0 level significance; the 4th level of the variable is significant at the 0.05 level. Implications these results have to research is that childrens parents who bring their child to a museum (at the minimum) sometimes will have children who believe they can pursue higher levels of education.

The independent variable sex is significantly associated with the dependent variable. My only issue is since I factored the variable I got only an output for females and not males. For females they have a positive association with wanting to pursue high levels of education and it is significant at the 0 level. My only worry about using this variable is that I can't compare to what males statistics are so I may change the final model to an unfactored version of this variable.

The last independent variable for this model is ethnicity. The variable is negatively associated with our dependent variable, and is significant negative associations (at the 0 level) for individuals who are white and hispanic. Implications for this would be that maybe students who are white or hispanic have lower views of their educationahl pursuit capabilities. However, this factored version doesn't include african americans. Also, I would change the format of the variable in my recording by trying to create a binary version that way it would be easier to see if a certain race variable would be associated with our dependent variable.

f. How well does the full (all explanatory variables included) model fit? Are any of the other models you ran a better fit? Explain how you came to the conclusion you did.

The model fits not as well as one would like but with an adjusted R-squared near 5 percent with the variables that I am using is pretty good. I believe the regressions I am working with are more along the lines of exploratory models in order to help give me some ideas as to what I should pursue for my research question. This model gives the best fit out of all the models I ran and gave the most significant variables.

g. Select the model that you think best fits the data. Provide a brief synopsis of the analysis of your data using this model and describe the implications for the theoretical arguments you set out to test.

The answer to this question is in question E. above. Also as a side note, the final model produced above is a test to see what adding the students' percieved chance of completing extra education after colleg and seeing if it influences their percieved chance of completing a graduate degree. However, I do not feel this model is valid since many individuals at that age may not know the difference between those two levels of educaiton.