Lab 15 - Multivariate Regression & Interpretation

William Bernard

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 selecting age, I feel like varying parts of the life course will yeild varrying interpretations of loss in these individuals.

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 think age might intereact with the likelyhood of these things occuring (like losing a child or death of a spouse etc) I think that will be reflected in the model.

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

this is a continuous variable

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.

```
sat_life_num <- as.numeric(changing_lives_subset_final$sat_life)

child_die_num <- as.numeric(changing_lives_subset_final$child_die)

age <- (changing_lives_subset_final$age_r)

child_sat_lm <- lm(sat_life_num ~ age + factor(child_die_num))

child_sat_lm_2 <- lm(sat_life_num ~ child_die_num + factor(age))

summary(child_sat_lm)
```

```
##
## Call:
## lm(formula = sat_life_num ~ age + factor(child_die_num))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.5128 -0.4649 -0.1464  0.6470  3.0860
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              2.7279277  0.0685163  39.814   <2e-16 ***
## age                     -0.0086069  0.0009065  -9.495   <2e-16 ***
## factor(child_die_num)2 -0.0220669  0.0423911  -0.521    0.603
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.925 on 3586 degrees of freedom
##   (28 observations deleted due to missingness)
## Multiple R-squared:  0.02554,    Adjusted R-squared:  0.02499
## F-statistic: 46.99 on 2 and 3586 DF,  p-value: < 2.2e-16
```

```
summary(child_sat_lm_2)
```

```
##
## Call:
## lm(formula = sat_life_num ~ child_die_num + factor(age))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.6769 -0.5904 -0.1654  0.6596  3.2546
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.38334    0.13432  17.744  < 2e-16 ***
## child_die_num -0.02146    0.04295  -0.500  0.61729
## factor(age)26  0.27442    0.14894   1.842  0.06549 .
## factor(age)27  0.07979    0.14990   0.532  0.59454
## factor(age)28 -0.02892    0.15677  -0.184  0.85366
## factor(age)29  0.25959    0.15961   1.626  0.10395
## factor(age)30  0.11768    0.14123   0.833  0.40478
## factor(age)31 -0.03803    0.15548  -0.245  0.80678
## factor(age)32 -0.12356    0.14893  -0.830  0.40679
## factor(age)33  0.11421    0.15424   0.740  0.45907
## factor(age)34  0.31507    0.15142   2.081  0.03754 *
## factor(age)35  0.08271    0.14801   0.559  0.57631
## factor(age)36  0.05175    0.16729   0.309  0.75709
## factor(age)37 -0.03033    0.14801  -0.205  0.83767
## factor(age)38  0.28267    0.16273   1.737  0.08245 .
## factor(age)39  0.04669    0.15196   0.307  0.75867
## factor(age)40  0.18402    0.16190   1.137  0.25577
## factor(age)41  0.08086    0.17381   0.465  0.64180
## factor(age)42  0.04254    0.16192   0.263  0.79278
## factor(age)43  0.04559    0.16928   0.269  0.78770
## factor(age)44 -0.15198    0.17766  -0.855  0.39235
## factor(age)45 -0.14791    0.17905  -0.826  0.40884
## factor(age)46 -0.03886    0.17262  -0.225  0.82189
## factor(age)47 -0.17934    0.16929  -1.059  0.28951
## factor(age)48  0.08038    0.17380   0.462  0.64376
## factor(age)49 -0.11736    0.18891  -0.621  0.53447
## factor(age)50  0.14376    0.18523   0.776  0.43775
## factor(age)51 -0.16863    0.18052  -0.934  0.35029
## factor(age)52  0.24999    0.19474   1.284  0.19934
## factor(age)53 -0.08512    0.19692  -0.432  0.66560
## factor(age)54 -0.05701    0.19719  -0.289  0.77250
## factor(age)55 -0.18669    0.17519  -1.066  0.28667
## factor(age)56  0.12021    0.17641   0.681  0.49563
## factor(age)57 -0.15895    0.19491  -0.816  0.41484
## factor(age)58 -0.17505    0.16545  -1.058  0.29010
## factor(age)59 -0.02371    0.16542  -0.143  0.88606
```

```
## factor(age)60 -0.22213    0.13864  -1.602  0.10922
## factor(age)61 -0.03993    0.14499  -0.275  0.78300
## factor(age)62 -0.09756    0.14441  -0.676  0.49935
## factor(age)63 -0.20994    0.14396  -1.458  0.14485
## factor(age)64 -0.15026    0.14727  -1.020  0.30765
## factor(age)65 -0.40450    0.14020  -2.885  0.00394 **
## factor(age)66 -0.32211    0.14520  -2.218  0.02659 *
## factor(age)67 -0.23659    0.14631  -1.617  0.10597
## factor(age)68 -0.14585    0.15058  -0.969  0.33281
## factor(age)69 -0.24835    0.14732  -1.686  0.09193 .
## factor(age)70 -0.44081    0.15112  -2.917  0.00356 **
## factor(age)71 -0.41387    0.15990  -2.588  0.00968 **
## factor(age)72 -0.21956    0.15768  -1.392  0.16389
## factor(age)73 -0.41586    0.16206  -2.566  0.01033 *
## factor(age)74 -0.14977    0.14793  -1.012  0.31138
## factor(age)75 -0.15801    0.16447  -0.961  0.33675
## factor(age)76 -0.41986    0.17928  -2.342  0.01924 *
## factor(age)77 -0.51927    0.16649  -3.119  0.00183 **
## factor(age)78 -0.12391    0.17399  -0.712  0.47639
## factor(age)79 -0.11680    0.18233  -0.641  0.52184
## factor(age)80 -0.32209    0.18575  -1.734  0.08302 .
## factor(age)81 -0.59501    0.20435  -2.912  0.00362 **
## factor(age)82 -0.14814    0.21338  -0.694  0.48757
## factor(age)83 -0.21380    0.22428  -0.953  0.34051
## factor(age)84 -0.40807    0.24831  -1.643  0.10039
## factor(age)85 -0.16647    0.24791  -0.671  0.50196
## factor(age)86 -0.35212    0.29875  -1.179  0.23862
## factor(age)87 -0.10651    0.34465  -0.309  0.75731
## factor(age)88 -0.21289    0.36598  -0.582  0.56082
## factor(age)89 -0.34995    0.32610  -1.073  0.28328
## factor(age)90  0.15422    0.47405   0.325  0.74495
## factor(age)91 -0.85114    0.47434  -1.794  0.07284 .
## factor(age)92  0.15422    0.47405   0.325  0.74495
## factor(age)93 -0.34041    0.93006  -0.366  0.71438
## factor(age)94 -0.86188    0.66308  -1.300  0.19375
## factor(age)95 -0.02139    0.54448  -0.039  0.96867
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.924 on 3517 degrees of freedom
##   (28 observations deleted due to missingness)
## Multiple R-squared:  0.04641,    Adjusted R-squared:  0.02716
## F-statistic: 2.411 on 71 and 3517 DF,  p-value: 6.179e-10
```

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

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

after switching age and the loss of a child variable it is very apparent that age is more strongly associated with satisfaction than loss of a child is.

- how the focal association changed when you incorporatedthe control variable

controling for a child death points towards age as a more significant motivating factor.

- the implications of these results for your focal association

child loss is less impactful on life satisfaction than I would have predicted,

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

it helps identify some things that could be wrong with the other model, it could be that as someone gets older they are more or less content and that is simultaiously making them more likely to have lost a child given they have been alive longer. possible explanation.

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.

positive self attitude could be a completely seperate confounding variable.

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 think it is has a stronger connection to satisfaction.

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

it is a catagorical variable,

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.

```
child_three <- as.numeric(changing_lives_subset_final$child_die_three_yr)

pos_self_att_num <- as.numeric(changing_lives_subset_final$pos_self_att)


child_three_sat_lm <- lm(sat_life_num ~ child_three + factor(pos_self_att_num))

summary(child_three_sat_lm)
```

```
##
## Call:
## lm(formula = sat_life_num ~ child_three + factor(pos_self_att_num))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9610 -0.3883 -0.1097  0.6117  2.8903
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 1.74546    0.20320   8.590  < 2e-16 ***
## child_three                 0.18214    0.10233   1.780   0.0752 .
## factor(pos_self_att_num)2   0.27855    0.03293   8.459  < 2e-16 ***
## factor(pos_self_att_num)3   0.85123    0.07552  11.271  < 2e-16 ***
## factor(pos_self_att_num)4   0.64482    0.14527   4.439 9.32e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9101 on 3496 degrees of freedom
##   (116 observations deleted due to missingness)
## Multiple R-squared:  0.05107,    Adjusted R-squared:  0.04998
## F-statistic: 47.04 on 4 and 3496 DF,  p-value: < 2.2e-16
```

```r
factor_test_child_pos <- lm(sat_life_num ~ child_three + pos_self_att_num)

summary(factor_test_child_pos)
```

```
##
## Call:
## lm(formula = sat_life_num ~ child_three + pos_self_att_num)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0471 -0.4183 -0.1038  0.5817  2.8962
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1.43488    0.20633   6.954 4.21e-12 ***
## child_three       0.17727    0.10251   1.729   0.0838 .
## pos_self_att_num  0.31442    0.02417  13.008  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9119 on 3498 degrees of freedom
##   (116 observations deleted due to missingness)
## Multiple R-squared:  0.04691,    Adjusted R-squared:  0.04636
## F-statistic: 86.08 on 2 and 3498 DF,  p-value: < 2.2e-16
```

```r
anova(factor_test_child_pos, test = "Chisq")
```

```
## Analysis of Variance Table
##
## Response: sat_life_num
##                    Df  Sum Sq Mean Sq F value Pr(>F)
## child_three         1    2.46   2.457   2.955 0.0857 .
## pos_self_att_num    1  140.69 140.692 169.201 <2e-16 ***
## Residuals        3498 2908.61   0.832
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

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

The child dying in the last 3 years lost its significance as the positive self attitude was introduced.

- how the focal association changed when you incorporated each control variable

the relationship between satisfaction and children dying in the last three years decreased as the postive self attude was introduced.

- the implications of these results for your focal association

that positive self attitude may be more imortant than a kid dying in the last three years.

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 models seem to be pretty consistant looking at lr and anova so I think this does the job fine. I tested anova considering my reliance on catagorical variables. the r squared value is small but considering the data its not too small (.05).

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.

I am going to stick with the Lr for now, the anova could also be used but considering that they seem to be fairly consistant I see not reason to swap. what we see here is positive self attitude has a strong relationship with life satisfaction (2e-16) and when factoring that in, loss of a child in the last three years has a moderate association (~.08). there is an intercpt at 1.4 - 1.7 but that means less considering the catagorical nature but we can expect to see child dying in the last 3 years increase their life satisfaction by .17 consistantly (on the scale of 1-4) and about .3 on a scale of 1-4. thats really interesting, I wouldnt have expected to see that cause an increase (child death in the last 4 years). but that being said it is a weaker relationship made weaker by positive self attitude. it kind of turns my theory on its head in some ways and boulsters it in others.