Adrian Varallyay
October 4, 2020

QMSS-5015-Lab-4

1. Run a simple regression, and interpret your results. (Did the results fit your expectations? Why? Why not?)

I wanted to see how one's degree of risk taking, level of caring for environment, age, and work type affect opinions on prioritizing protecting the environment vs economic growth. I used the subset of respondents from USA and China for this inquiry. I implemented the variables *V81, V76, V78*, and *V230* in the regression, and renamed them *envir, risk, carenvir*, and *worktype,* respectively.

As table 1 indicates, *envir* asked respondents which of two viewpoints with which they identify, 1) Protecting the environment should be given priority, even if it causes slower economic growth and some loss of jobs or 2) Economic growth and creating jobs should be the top priority, even if the environment suffers to some extent. I recoded this variable to be 1 if they identified with protecting the environment, and zero for anything else.

Table 2 shows the responses to the statement, "Adventure and taking risks are important to this person; to have an exciting life", and uses a 6-point scale ranging from "Very much like me" (1) to "Not at all like me" (6). I recoded this variable by reversing the order of responses so that the degree of identification with the statement increases from 1 to 6. Likewise with *carenvir*, as table 3 indicates, it used the same 6-point scale, thus I also reversed the order of responses here so that the degree of identification with the statement, "Looking after the environment is important to this person; to care for nature and save life resources", increases from 1 to 6.

Table 4 shows *worktype* of the respondents, which is comprised of 4 categories, Government or public institution (1), Private business or industry (2), Private non-profit organization (3), MX, EC: Autonomous/Informal sector; PE, CH: Other (4).

After regressing Y on the X variables, the coefficient on the intercept is 0.65; we also see that only 3 factors were very statistically significant ($P<0.001$), one was just statistically significant ($p<0.05$), and one was approaching significance ($p<0.1$). Age is the first of these very statistically significant results, with a coefficient of -0.002, meaning that, on average, net of everything else (the other x variables), as an individual's age increases by one, responses to "Protecting the environment should be given priority, even if it causes slower economic growth and some loss of jobs" decrease by 0.002 scale points. Essentially individuals incrementally discount prioritizing protecting the environment as they age.

As figure 1 shows, of the factors for *carenvir*, net of all the other variables, only the "not like me" category was significant at the $P< 0.001$ level, with a coefficient of 0.46. This means that, on average, compared to its reference group (not at all like me), those who identify with the "not like me" response are 0.46 scale points higher on responses to the statement

Adrian Varallyay
October 4, 2020

"Protecting the environment should be given priority, even if it causes slower economic growth and some loss of jobs". This seems very counter intuitive, considering the question being asked, but I suppose it could make sense when taking into account that this is in relation to the reference group, which in this case is the subset of individuals who do not identify at all with the pro-environmental statement. In addition, though only approaching statistical significance, those who identified with the "somewhat like me" category, had a coefficient of 0.06 ($P<0.1$), meaning, on average, net of the other variables, and compared to its reference group (not at all like me), respondents are 0.06 scale points higher on responses to the statement "Protecting the environment should be given priority, even if it causes slower economic growth and some loss of jobs".

Of the factors for *worktype*, only the fourth category, "MX, EC: Autonomous/Informal sector; PE, CH: Other", was statistically significant ($p<0.05$) with a coefficient of 0.07. This means, on average, net of the other variables, and compared to its work type reference group (Government or public institution) individuals who reported association/identification with the fourth category are 0.07 scale points higher on responses to the statement "Protecting the environment should be given priority, even if it causes slower economic growth and some loss of jobs".

Finally, *country,* in particular the USA, was very statistically significant ($p<0.001$) with a coefficient of about -0.2. This means, on average, and net of the other variables, that respondents in the USA are generally 0.2 scale points lower on responses to the statement "Protecting the environment should be given priority, even if it causes slower economic growth and some loss of jobs", than to respondents from china. From these results, it would seem that Americans place less importance on protecting the environment than the Chinese.

The adjusted R-squared for the first model is 0.1282, meaning about 13% of the variation is explained by this model.

In general, I would say my results were not what I was expecting. I thought *risk* would have been strongly related to the outcomes on views regarding protecting the environment, specifically I would have thought those higher on risk identification would be lower on protecting the environment. It should be noted that this variable may not have been just tapping into risk taking values, but just adventure in general, and this "adventure" could look like many things, to many different people. Though most of the signs on the coefficients for the different levels of *risk* are negative, none of the results were statistically significant. Furthermore, I would have surely thought that responses on *carenvir* would have all been statistically significant, this was not the case. The two most interesting outcomes for me were the results showing views on protecting the environment between USA and China (I would have thought USA respondents would have been more environmentally conscious, jut considering the anecdotes we hear here regarding china and the environment)), and *worktype.* I'm not familiar with what types of jobs fall under the "MX, EC: Autonomous/Informal sector; PE, CH: Other (4)" umbrella, but nonetheless, I would have thought there would have been a stronger relationship between industry sectors and views on protecting the environment.

Table 1:

Adrian Varallyay
October 4, 2020

83) Here are two statements people sometimes make when discussing the environment and economic growth. Which of them comes closer to your own point of view? (V81)

	TOTAL	%
-5) DE, SE: Inapplicable; RU: Inappropriate response; Missing (Inappropriate)	51	0.1
-4) No asked in survey	999	1.3
-2) No answer	955	1.3
-1) Don't know	3540	4.8
1) Protecting the environment should be given priority, even if it causes slower economic growth and some loss of jobs.	34411	46.5
2) Economic growth and creating jobs should be the top priority, even if the environment suffers to some extent.	31477	42.5
3) Other answer	2609	3.5
TOTAL	74042	100.0

Table2:

78) Now I will briefly describe some people. Using this card, would you please indicate for each description whether that person is very much like you, like you, somewhat like you, not like you, or not at all like you? Adventure and taking risks are important to this person; to have an exciting life (V76)

	TOTAL	%
-5) BH: Missing; RU, DE: Inappropriate response (Inappropriate)	393	0.5
-2) No answer	698	0.9
-1) Don't know	1339	1.8
1) Very much like me	7377	10.0

2) Like me	11037	14.9
3) Somewhat like me	12610	17.0
4) A little like me	12558	17.0
5) Not like me	16947	22.9
6) Not at all like me	11083	15.0
TOTAL	74042	100.0

Table 3:

80) Now I will briefly describe some people. Using this card, would you please indicate for each description whether that person is very much like you, like you, somewhat like you, not like you, or not at all like you? Looking after the environment is important to this person; to care for nature and save life resources (V78)

	TOTAL	%
-5) Missing; RU, DE: Inappropriate response (Inappropriate)	81	0.1
-2) No answer	567	0.8
-1) Don't know	1037	1.4
1) Very much like me	17813	24.1
2) Like me	22191	30.0
3) Somewhat like me	16671	22.5
4) A little like me	10217	13.8
5) Not like me	3971	5.4
6) Not at all like me	1494	2.0

TOTAL	74042	100.0

Table 4:

283) Are you working for the government or public institution, for private business or industry, or for a private non-profit organization? If you do not work currently, characterize your major work in the past! Do you or did you work for? (V230)

	TOTAL	%
-5) SE: Inapplicable; SG: Refused; BH: Missing (Inappropriate)	14	0.0
-4) Not asked in survey	1500	2.0
-3) Not applicable (Have never worked; Currently does not work)	11308	15.3
-2) No answer	2047	2.8
-1) Don't know	1623	2.2
1) Government or public institution	17764	24.0
2) Private business or industry	31518	42.6
3) Private non-profit organization	3843	5.2
4) MX, EC: Autonomous/Informal sector; PE, CH: Other	4425	6.0
TOTAL	74042	100.0

Figure 1:

```
## Does one's degree of risk taking, level of caring for environment, age and work type affect o
pinions on prioritizing protecting environment vs  economic growth in subset of respondents fr
om USA and China
lm1 = lm(envir ~ age + lab.r_risk  + lab.r_carenvir + as.factor(worktype) + as.factor(country), dat
a=wvs, subset = (country==156 | country==840))
summary(lm1)

##
## Call:
```

```
## lm(formula = envir ~ age + lab.r_risk + lab.r_carenvir + as.factor(worktype) +
##     as.factor(country), data = wvs, subset = (country == 156 |
##     country == 840))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9773 -0.4072  0.1149  0.4054  0.9349
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          0.6520425  0.0372040  17.526  < 2e-16 ***
## age                 -0.0024028  0.0006533  -3.678  0.00024 ***
## lab.r_risk.L        -0.0529543  0.0366374  -1.445  0.14848
## lab.r_risk.Q        -0.0470500  0.0335201  -1.404  0.16055
## lab.r_risk.C        -0.0005141  0.0277776  -0.019  0.98524
## lab.r_risk^4        -0.0257932  0.0234300  -1.101  0.27106
## lab.r_risk^5         0.0178118  0.0210572   0.846  0.39770
## lab.r_carenvir.L     0.4587134  0.0450120  10.191  < 2e-16 ***
## lab.r_carenvir.Q     0.0232923  0.0408638   0.570  0.56873
## lab.r_carenvir.C     0.0587627  0.0340642   1.725  0.08463 .
## lab.r_carenvir^4     0.0357551  0.0273841   1.306  0.19177
## lab.r_carenvir^5    -0.0315899  0.0210135  -1.503  0.13288
## as.factor(worktype)2 -0.0021086  0.0215941  -0.098  0.92222
## as.factor(worktype)3 -0.0074648  0.0413498  -0.181  0.85675
## as.factor(worktype)4  0.0716506  0.0314912   2.275  0.02297 *
## as.factor(country)840 -0.1950592  0.0217611  -8.964  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4664 on 2626 degrees of freedom
##   (1890 observations deleted due to missingness)
## Multiple R-squared:  0.1332, Adjusted R-squared:  0.1282
## F-statistic: 26.89 on 15 and 2626 DF,  p-value: < 2.2e-16
```

2. Add an interaction term to that model that you think might moderate the original relationship between X1 and X2. Explain why you think an interaction might be present and in what direction it would work. Explain your results. Did it work out? Yes? No?

I didn't know how to preemptively estimate if there would have been an interaction between a couple of the variables I used. Nor did any my variables really justify using a quadratic term (except for maybe age). As such, I retained the age*country interaction professor Eirich used in his example, however, I don't have a framework to make a prediction for the direction of the interaction effect(s).

Adrian Varallyay
October 4, 2020

As figure 1 indicates, compared to the previous model (only difference is age*country interaction replacing age & country), statistical significances change after regressing y on the x variables. The second category for *carenvir* (not like me) is still statistically significant (P<0.001) but with a marginally larger coefficient of 0.46. The interpretation of this result is relatively consistent with the previous regression result. The fourth category of *carenvir* (somewhat like me) remains almost statistically significant (P<0.1) but the coefficient, decreasing slightly, is 0.058. The interpretation of this result is also relatively consistent with the previous result. The fourth category of *worktype* also remains statistically significant (P<0.05) but the coefficient, slightly increasing, is about 0.08.

Finally, age and country are no longer statistically significant, but the age*country interaction is very significant (P<0.001) with a coefficient of -0.004 (difference in slopes between USA and China on *age)*. This means, on average, net of the other variables, as age increases by one for individuals in USA, respondents are 0.004 scale points lower on responses to the statement "Protecting the environment should be given priority, even if it causes slower economic growth and some loss of jobs", but that is not the case for individuals in China. The age gradient is stronger for the USA than for China, such that, the interaction for age and the USA is negative, and the interaction for age and China is marginally positive, though essentially flat within the range of the data. This is depicted in figure 3. This interaction makes sense to me considering that cultural values (e.g. country) would play a role in forming one's values and beliefs, and therefore, *country* would interact with age, as different age groups also tend to have different levels of developed values and beliefs (i.e. culture). As such, it does not imply an individual in China necessarily sees eye-to-eye with an individual from the USA just because of membership to the same age group, or vice versa.

The adjusted R-squared for the second model is 0.13, meaning 13% of the variation is explained by this model. I would say this regression worked out, though not in the sense of validating my prediction, but in the sense that the outcome was an interaction that had different effects to be analyzed. Figure 2 shows the results of the anova of the two regression models. The F statistic of 10.917 is very statistically significant (P=0.00097). Because F is that large and the statistical significance is high, the results support the use of the interaction model over the first model. Intuitively I see this when I just glance at the R-squared of the two models; the first is `0.1282 and the second is` 0.1315. The second model's R-squared explains more of the variation than the first, 12.8% versus 13.2%. However, I wouldn't make an actual claim using this methodology of comparison; but, it's interesting to see that this glanced evaluation is consistent with the results of the anova.

Figure 1:

```
lm2 = lm(envir ~ lab.r_risk  + lab.r_carenvir + as.factor(worktype) + age*as.factor(country), data
=wvs, subset = (country==156 | country==840))
summary(lm2)

##
## Call:
## lm(formula = envir ~ lab.r_risk + lab.r_carenvir + as.factor(worktype) +
```

```
##     age * as.factor(country), data = wvs, subset = (country ==
##     156 | country == 840))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9840 -0.4205  0.1237  0.4167  0.9298
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     0.536401   0.051028  10.512  < 2e-16 ***
## lab.r_risk.L                   -0.053594   0.036569  -1.466 0.142889
## lab.r_risk.Q                    -0.045079   0.033462  -1.347 0.178041
## lab.r_risk.C                   -0.001963   0.027729  -0.071 0.943560
## lab.r_risk^4                    -0.026020   0.023386  -1.113 0.265960
## lab.r_risk^5                     0.013764   0.021053   0.654 0.513322
## lab.r_carenvir.L                0.462292   0.044940  10.287  < 2e-16 ***
## lab.r_carenvir.Q                 0.021707   0.040790   0.532 0.594662
## lab.r_carenvir.C                 0.057529   0.034002   1.692 0.090780 .
## lab.r_carenvir^4                 0.038788   0.027348   1.418 0.156217
## lab.r_carenvir^5                -0.032748   0.020977  -1.561 0.118604
## as.factor(worktype)2             0.006193   0.021699   0.285 0.775345
## as.factor(worktype)3            -0.001024   0.041318  -0.025 0.980233
## as.factor(worktype)4             0.079820   0.031529   2.532 0.011411 *
## age                            0.000216   0.001026   0.210 0.833340
## as.factor(country)840           -0.007465   0.060788  -0.123 0.902278
## age:as.factor(country)840 -0.004193   0.001269  -3.304 0.000965 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4655 on 2625 degrees of freedom
##   (1890 observations deleted due to missingness)
## Multiple R-squared:  0.1367, Adjusted R-squared:  0.1315
## F-statistic: 25.99 on 16 and 2625 DF,  p-value: < 2.2e-16
```

Figure 2:

```
anova(lm1, lm2) ## Did adding the interaction improve my model? ##

## Analysis of Variance Table
##
## Model 1: envir ~ age + lab.r_risk + lab.r_carenvir + as.factor(worktype) +
##     as.factor(country)
## Model 2: envir ~ lab.r_risk + lab.r_carenvir + as.factor(worktype) + age *
```

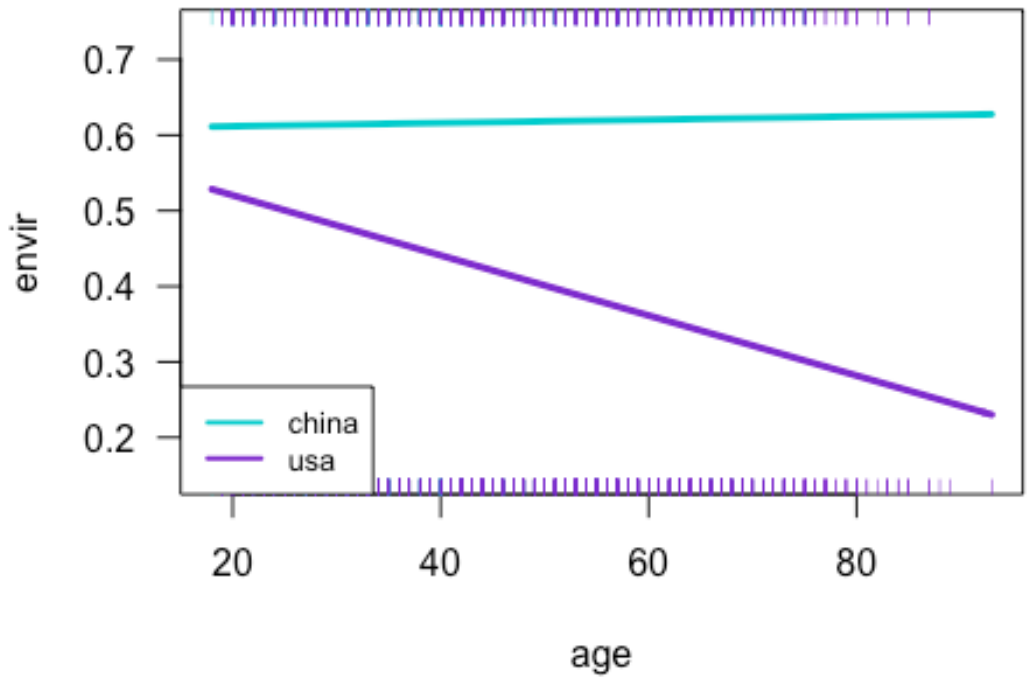
Adrian Varallyay
October 4, 2020

```
##    as.factor(country)
##  Res.Df    RSS Df Sum of Sq     F    Pr(>F)
## 1   2626 571.21
## 2   2625 568.84  1    2.3658 10.917 0.0009655 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 3:

```
wvs2 <-subset(wvs, country==156 | country==840) ## subset the data to use in visreg below ##

##install.packages("visreg")
library(visreg)
visreg(lm(envir ~ age*as.factor(country) + lab.r_risk + lab.r_carenvir + as.factor(worktype), data
= wvs2),
    xvar = "age", by = "country", overlay=T, partial = F, band = F, legend = F,
    line = list(col = c("cyan3", "purple3")))
legend("bottomleft", c("china", "usa"), lwd = 2, col = c("cyan3", "purple3"), cex = 0.8)
```



Adrian Varallyay
October 4, 2020

3. Give me an update on your independent project. What do you plan to investigate? What are your hypotheses? What date are you using? How can we help?

As per your last feedback, I will be using the data I have from my work with Professor Subiaul. I'm currently consolidating the 2 experiments (rather than the entire set of studies) into one excel sheet. I will be exploring the relationships between demographic variables and performances on tasks. I'm actually set to meet with Subiaul to help me give this exploration more of a point and to evaluate potential hypothesis. I may like to bounce some ideas around with you after I meet with Subiaul, if that's okay?

Adrian Varallyay
October 4, 2020

Appendix (codes)

Q1 codes

```
wvs = rename(wvs, c("V76"="risk"))
wvs = rename(wvs, c("V78"="carenvir"))
wvs = rename(wvs, c("V81"="envirVecon"))
wvs = rename(wvs, c("V2"="country"))
wvs = rename(wvs, c("V230"="worktype"))
wvs = rename(wvs, c("V242"="age"))
wvs$envir = ifelse((wvs$envirVecon==1), 1, 0)
## Reverse code a variable and then add labels and make it ordered:
## To reverse code: (highest category + 1) - orginal_variable
wvs$r_risk = 7-wvs$risk
wvs$r_carenvir = 7-wvs$carenvir

## make factor variable into an ORDERED factor, with value labels: "Not at all like me", "Not like me", "A little like me", "Somewhat like me", "Like me", "Very much like me", reversed so that risk identity (Adventure and taking risks are important to this person; to have an exciting life) increases from 1 to 6
wvs$r_risk.fact = as.factor(wvs$r_risk)
wvs$lab.r_risk <- ordered(wvs$r_risk.fact, levels = c(1,2,3,4,5,6), labels = c("Not at all like me", "Not like me", "A little like me", "Somewhat like me", "Like me", "Very much like me" ))

## make factor variable into an ORDERED factor, with value labels: "Not at all like me", "Not like me", "A little like me", "Somewhat like me", "Like me", "Very much like me", reversed so that envir-care identity (Looking after the environment is important to this person; to care for nature and save life resources ) increases from 1 to 6
wvs$r_carenvir.fact = as.factor(wvs$r_carenvir)
wvs$lab.r_carenvir <- ordered(wvs$r_carenvir.fact, levels = c(1,2,3,4,5,6), labels = c("Not at all like me", "Not like me", "A little like me", "Somewhat like me", "Like me", "Very much like me" ))
```