MLE - Lab 5

Andy Ballard

To-Do

- Read .dta files in R
- What is expected value and why do I care?
- Estimate linear model via MLE by hand
- Interaction Effects

Loading Stata files in R

Much of the data you'll find in the wild is not in a format native to R (.Rdata), or even in a format that automatically plays nice with multiple platforms (e.g. .csv). Much of the data you'll work with is in .dta format, which is native to Stata, the most popular statistical software for economists and their ilk. So how do we read Stata files in R?

First, you'll have to load the `foreign` package:

```
library(foreign)
```

Then, use the `read.dta()` command:

```
corruption <- read.dta(paste0(labPath, "corruption.dta"))
```

As we can see, it's about the same syntax as some of the other ways we've read data into R. This is one of the data sets you'll be using for your homework this week, so it's a good idea to get this working.

Expected value

What is expected value?

Let's start with a random variable. The expected value of that variable is the sum of all possible values, each multipled by the probability it occurs.

Say a basketball player, who has a lifetime free throw percentage of 70%, takes 10 free throws. What is the expected number of free throws made?

For this, we take the expected value of a Bernoulli random variable we will call FT, where the probability of making the shot is $p = 0.70$, or $P(FT = 1) = 0.70$. Now, for a Bernoulli random variable, the expected value $E(X)$ is equal to p. Here's why:

For a Bernoulli distributed random variable X with $Pr(X = 1) = p$ and $Pr(X = 0) = (1 - p) = q$, $E[X] = p * 1 + q * 0 = p$. We got this by multiplying the possible values $\{1, 0\}$ by their corresponding probability of occuring $\{p, q\}$.

Back to our friend, the baller. For a single free-throw, the expected number of free throws made is equal to the probability of success, p, which is 0.70. If there are 10 free throw attempts, the expected number of free throws made is $10 * E(FT) = 10 * 0.70 = 7$. Note that the expected value does NOT have to be a value that is possible to happen in the real world. If p were instead 0.65, then the expected number of free throws made out of 10 would be 6.5.

Expected value of the normal distribution

The Bernoulli distribution is discrete. That is, each of its possible values are countable. We can have 0, or we can have 1; that's it. Thus, finding the expected value is as simple as knowing the probability of success.

The Normal distribution, on the other hand, is continuous. There are infinitely many values that a normally distributed random variable can take, so we can't find the expected value simply by directly summing up all the possible values and multiplying each by its probability of occurring. Instead, we need to take an integral. Here is our normal distribution pdf:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\sigma^2\pi}} exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

For this example we're going to use the standard normal distribution, which is a friendlier version of the normal distribution where $\mu = 0$ and $\sigma^2 = 1$. We can rewrite the pdf as:

$$f(x) = \frac{1}{\sqrt{2\pi}} exp(-\frac{1}{2}x^2)$$

Just as the expected value of a Bernoulli random variable is p the probability of success, the expected value of a normal random variable is the mean, μ. So for the case of a standard normal distribution that is a function of x, $E(x) = 0$. How might we prove that? Well, we have to take the integral over the "support" of the function. A function's support is all the possible values it can take. For the Bernoulli distribution, that's just $\{0, 1\}$. For the normal distribution, that's all real numbers (written as $x \in \mathbf{R}$). Thus, we have to take our integral from $-\infty$ to ∞. Just like the Bernoulli, we're taking all the possible values x and multiplying them by their probability to occur, given by the pdf $f(x)$. This gives us a general form for the expected value of a continuous distribution where the support is bounded by {a,b} (for the normal distribution, $a = -\infty$ and $b = \infty$) :

$$E[x] = \int_a^b xf(x)dx$$

Which in this case can be rewritten as:

$$E[x] = \int_{-\infty}^{\infty} x(2\pi)^{-\frac{1}{2}} exp(-\frac{1}{2}x^2)dx$$

And now we only have to do the integral

$$\begin{aligned}
E[x] &= \int_{-\infty}^{\infty} x(2\pi)^{-\frac{1}{2}} exp(-\frac{1}{2}x^2)dx \\
&= (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{0} x * exp(-\frac{1}{2}x^2)dx + 2(\pi)^{-\frac{1}{2}} \int_{0}^{\infty} x * exp(-\frac{1}{2}x^2)dx \\
&= (2\pi)^{-\frac{1}{2}} \Big[-exp(-\frac{1}{2}x^2) \Big]_{-\infty}^{0} + (2\pi)^{-\frac{1}{2}} \Big[-exp(-\frac{1}{2}x^2) \Big]_{0}^{\infty} \\
&= (2\pi)^{-\frac{1}{2}}[-1+0] + (2\pi)^{-\frac{1}{2}}[0+1] \\
&= (2\pi)^{-\frac{1}{2}} - (2\pi)^{-\frac{1}{2}} \\
&= 0
\end{aligned}$$

Ta da!

Variance

The variance is also defined in relation to expected value.

$$Var[x] = E[x^2] - E[x]^2$$

For the standard normal distribution, $\sigma^2 = Var[x] = 1$, which we can show by finding $E[x]^2$ and $E[x^2]$. If $E[x] = 0$, then $E[x]^2 = 0$. $E[x^2]$ is calculated just like $E[x]$, but with x^2 instead of x:

$$E[x^2] = \int_{-\infty}^{\infty} x^2 f(x) dx$$

This integral is even longer than the last one, so I'll spare you computing it here (you can look it up online if you're really interested). Just trust me that $E[x^2] = 1$ so that $Var[x] = E[x^2] - E[x]^2 = 1 - 0 = 1$, since $\sigma^2 = 1$ is the variance of the standard normal distribution.

Simple linear model (OLS analog) in MLE by hand

Usually, we think of MLE being used for binary, ordinal, or categorical outcome variables, but the framework is flexible enough to run many other types of models. For instance, you can run an MLE version of a simple linear model (OLS).

Islam & Authoritarianism

Fish (2002) argues that predominantly Muslim societies are distinctly disadvantaged in democratization. To test this he regresses Freedom House scores (`fhrev`) on an Islamic religious tradition variable (`muslim`) that takes the value of one when a country is predominantly Muslim and zero otherwise. He also goes on to add a number of additional controls for level of development (`income`), sociocultural division (`elf`), economic growth (`growth`), British colonial heritage (`britcol`), Communist heritage (`postcom`), and OPEC membership (`opec`).

We'll run his full model later (and add a twist), but for now let's focus on a simple bivariate relationship: the effect of development levels (`income`) on Freedom House scores (`fhrev`).

First, we can fit the model using least squares regression.

```
summary(fish.m1 <- lm(fhrev ~ income, data=fish))
```

```
##
## Call:
## lm(formula = fhrev ~ income, data = fish)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1946 -0.9149  0.3588  1.0855  2.9324
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.1390     0.6044  -1.885   0.0615 .
## income        1.6576     0.1832   9.046 8.13e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## 
## Residual standard error: 1.48 on 147 degrees of freedom
## Multiple R-squared:  0.3576, Adjusted R-squared:  0.3532
## F-statistic: 81.83 on 1 and 147 DF,  p-value: 8.128e-16
```

```r
plot(fish$income, fish$fhrev)
abline(fish.m1, col="red")
```
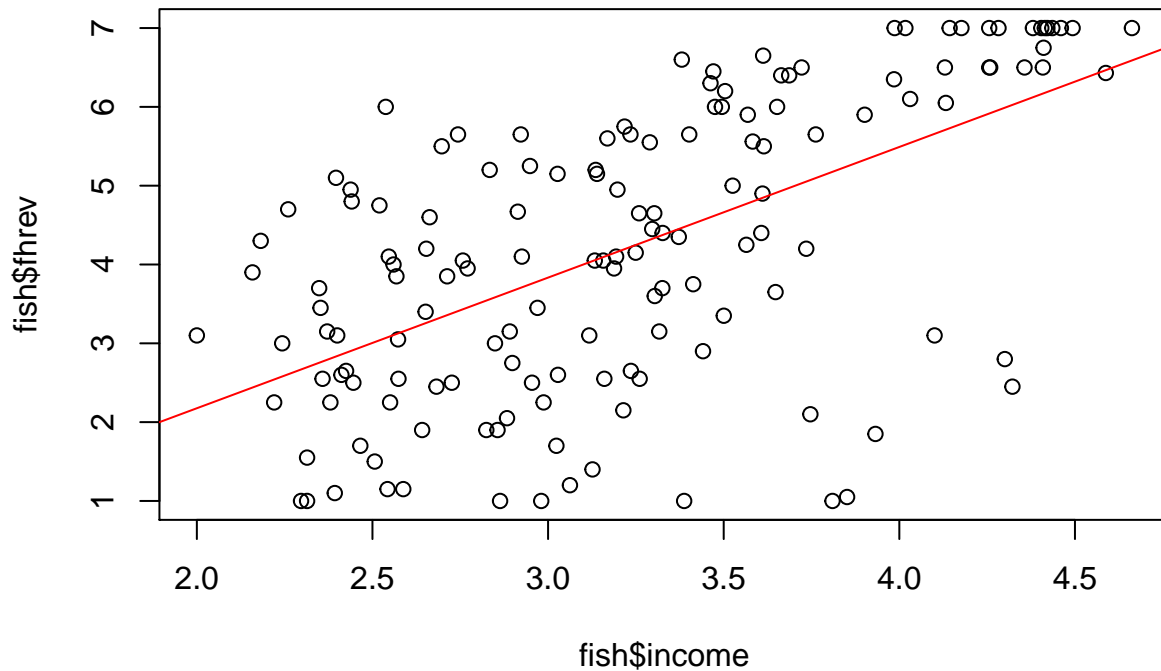

Looks like a pretty solid linear relationship, which the model bears out. We have a highly significant coefficient for `income` and a reasonable R^2 value (0.35).

Now let's see if we can get the same results using MLE.

```r
# First, put the data into matrices for the MLE procedure
x <- cbind(1,as.matrix(fish$income)) #intercepts are fun, we like incercepts
y <- as.matrix(fish$fhrev)
ones <- x[,1]

# Calculate number of parameters to estimate (K) and number of observations (n)
K <- ncol(x)
K1 <- K + 1
n <- nrow(x)

# Define the function to be optimized

llik.regress <- function(par,X,Y) {
Y <- as.vector(y)
X <- as.matrix(x)
```

```r
xbeta <- X%*%par[1:K]
Sig <- par[K1:K1]
  sum(-(1/2)*log(2*pi)-(1/2)*log(Sig^2)-(1/(2*Sig^2))*(y-xbeta)^2) #Normal, y'all
}

# Now let's use the above function to estimate the model.

model <- optim(c(-2,0,2),llik.regress, method = "BFGS", control = list(trace=1,maxit=100,fnscale = -1),
      hessian = TRUE)
```

```
## initial  value 1022.775448
## iter  10 value 495.002645
## iter  20 value 268.837711
## final  value 268.833729
## converged
```

```r
model #Why are there 3 parameter estimates? Sigma!
```

```
## $par
## [1] -1.139155  1.657597 -1.470088
##
## $value
## [1] -268.8337
##
## $counts
## function gradient
##       49       22
##
## $convergence
## [1] 0
##
## $message
## NULL
##
## $hessian
##               [,1]          [,2]          [,3]
## [1,] -6.894454e+01 -222.78658123 -5.281890e-04
## [2,] -2.227866e+02 -750.09688071 -3.570250e-03
## [3,] -5.281890e-04   -0.00357025 -1.378866e+02
```

```r
summary(fish.m1)
```

```
##
## Call:
## lm(formula = fhrev ~ income, data = fish)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1946 -0.9149  0.3588  1.0855  2.9324
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.1390     0.6044  -1.885   0.0615 .
## income        1.6576     0.1832   9.046 8.13e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.48 on 147 degrees of freedom
## Multiple R-squared:  0.3576, Adjusted R-squared:  0.3532
## F-statistic: 81.83 on 1 and 147 DF,  p-value: 8.128e-16
```

Notice that the coefficients from this model are the same as those produced through OLS estimation. Now let's get the variances, standard errors, and Z statistics from our model.

```
# Calculate the variance matrix from the Hessian matrix.

v <- -solve(model$hessian)
v
```

```
##                [,1]          [,2]          [,3]
## [1,]  3.604121e-01 -1.070462e-01  1.391112e-06
## [2,] -1.070462e-01  3.312698e-02 -4.476939e-07
## [3,]  1.391112e-06 -4.476939e-07  7.252335e-03
```

```
# Calculate the standard errors from the variance matrix.

se <- sqrt(diag(v))
se
```

```
## [1] 0.60034334 0.18200819 0.08516064
```

```
summary(fish.m1)$coefficients[,2]
```

```
## (Intercept)      income
##   0.6044091   0.1832408
```

```
# Calculate the z statistics from the coefficients and standard errors

b <- model$par
b
```

```
## [1] -1.139155  1.657597 -1.470088
```

```
zstat <-b/se
zstat
```

```
## [1]  -1.897505   9.107268 -17.262523
```

```
# Calculate p-values for the z statistics
p.vals <- 2 * (1 - pnorm(abs(zstat)))
p.vals
```

```
## [1] 0.05776128 0.00000000 0.00000000
```

```
# Put the results together in a table.
table <- cbind(b, se, zstat, p.vals)
table
```

```
##              b         se      zstat     p.vals
## [1,] -1.139155 0.60034334  -1.897505 0.05776128
## [2,]  1.657597 0.18200819   9.107268 0.00000000
## [3,] -1.470088 0.08516064 -17.262523 0.00000000
```

```
summary(fish.m1)$coefficients
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) -1.139014  0.6044091 -1.884508 6.147053e-02
```

```
## income       1.657552  0.1832408  9.045758 8.127888e-16
```

Why are the p-values slightly different? (normal distribution vs t distribution). This is just another reason p-values by themselves are silly: they're not fixed, and partially based on a distribution you choose!

```r
# Note that the estimate of the error variance is biased by n/n-K
# we can correct the preceding estimate of the standard error of #estimates as follows
Sig <- model$par[3]
Sig
```

```
## [1] -1.470088
```

```r
Sigunb <- sqrt(n/(n-K)*Sig^2)
Sigunb #Same as the residual standard error
```

```
## [1] 1.480054
```

Interaction Effects

A conditional hypothesis is one in which a relationship between two or more variables depends on the value of one or more other variables. Examples:

- An increase in X is associated with an increase in Y when condition Z is met, but not when condition Z is absent.
- X has a positive effect on Y that gets stronger as Z increases.

Back to the Fish data

We will rerun the model from his paper and run one of our own that includes an interaction term between Islamic religious tradition and level of economic development. Our hypothesis is that higher levels of economic development should be associated with higher Freedom House scores for non-Muslim countries, but that same relationship should not hold for Muslim countries.

```r
#Fish Model 4, Dep: fhrev. Indep: muslim, income, opec
lm1 <- lm(fhrev ~ muslim + income + elf + growth + britcol + postcom + opec, data=fish)
round(summary(lm1)$coefficients,3)
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)     0.169      0.676   0.250    0.803
## muslim         -1.245      0.254  -4.892    0.000
## income          1.401      0.178   7.853    0.000
## elf            -0.316      0.389  -0.813    0.418
## growth          0.067      0.045   1.504    0.135
## britcol         0.246      0.267   0.924    0.357
## postcom         0.226      0.314   0.720    0.473
## opec           -1.332      0.455  -2.926    0.004
```

```r
# Add interaction between muslim and income
lm2 <- lm(fhrev ~ muslim*income + elf + growth + britcol + postcom + opec, data=fish)
round(summary(lm2)$coefficients,3)
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1.092      0.661  -1.652    0.101
## muslim          5.074      1.202   4.220    0.000
## income          1.778      0.178  10.009    0.000
## elf            -0.208      0.356  -0.583    0.561
```

```
## growth            0.043     0.041    1.056    0.293
## britcol           0.141     0.245    0.578    0.564
## postcom           0.215     0.287    0.750    0.454
## opec             -0.492     0.445   -1.105    0.271
## muslim:income    -2.143     0.400   -5.357    0.000
```

```r
# Did adding the interaction improve the fit of the model?
# Lets check using the RMSE
rmse <- function(mod){ sqrt(mean(resid(mod)^2)) }
print(paste0('RMSE under Model without the Ineraction Term: ', round(rmse(lm1),3)))
```

```
## [1] "RMSE under Model without the Ineraction Term: 1.202"
```

```r
print(paste0('RMSE under Model with the Ineraction Term: ', round(rmse(lm2),3)))
```

```
## [1] "RMSE under Model with the Ineraction Term: 1.095"
```

```r
# Lets use some of the other diagnostics we have learned as well

# R^2
print(paste0("Adjusted R^2 under model without interaction term: ", round(summary(lm1)$adj.r.square, 3))
```

```
## [1] "Adjusted R^2 under model without interaction term: 0.549"
```

```r
print(paste0("Adjusted R^2 under model with interaction term: ", round(summary(lm2)$adj.r.square, 3)))
```

```
## [1] "Adjusted R^2 under model with interaction term: 0.623"
```

```r
# F test
anova(lm1, lm2)
```

```
## Analysis of Variance Table
##
## Model 1: fhrev ~ muslim + income + elf + growth + britcol + postcom +
##     opec
## Model 2: fhrev ~ muslim * income + elf + growth + britcol + postcom +
##     opec
##   Res.Df    RSS Df Sum of Sq     F    Pr(>F)
## 1    141 215.18
## 2    140 178.57  1    36.608 28.7 3.387e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Likelihood Ratio test
lrtest(lm1, lm2)
```

```
## Likelihood ratio test
##
## Model 1: fhrev ~ muslim + income + elf + growth + britcol + postcom +
##     opec
## Model 2: fhrev ~ muslim * income + elf + growth + britcol + postcom +
##     opec
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1   9 -238.80
## 2  10 -224.91  1 27.785  1.355e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The r-squared, F-test, and the Likelihood Ratio test indicate that we should include the interaction term, but the RMSE suggests it may not improve model fit TOO much.

Interpretation

When interpreting the effects of a covariate we need to keep the conditional relationship in mind. In Fish's original model, to interpret the effect of a one unit change in income on Freedom House scores we would just note that β_{income} equaled 1.4.

After incorporating the interaction effect, however, it is no longer as straightforward because:

$$\frac{\partial fhrev}{\partial income} = \hat{\beta}_{income} + \hat{\beta}_{muslim:income} \times muslim$$

We can see that now the effect of income on Freedom House scores is contingent on the value of $\beta_{muslim:income}$ and `muslim`. If `muslim` equals zero, the effect of income on Freedom House scores can be understood through just β_{income}, but if muslim equals one then the effect of income is measured through $\beta_{income} + \beta_{muslim:income}$.

The necessity of taking the interactive relationship into account is also true for how we estimate the effect of a country having an Islamic religous tradition as well.

$$\frac{\partial fhrev}{\partial muslim} = \hat{\beta}_{muslim} + \hat{\beta}_{muslim:income} \times income$$

The best way to understand an interactive effect is through a visualization, so lets do that now. On the y-axis, we will have predicted FH scores, and on the x-axis we will have the range of income values in the Fish dataset. To incorporate the Muslim variable, we will use two different colored lines to show the conditional effect of a country having an Islamic religious tradition.
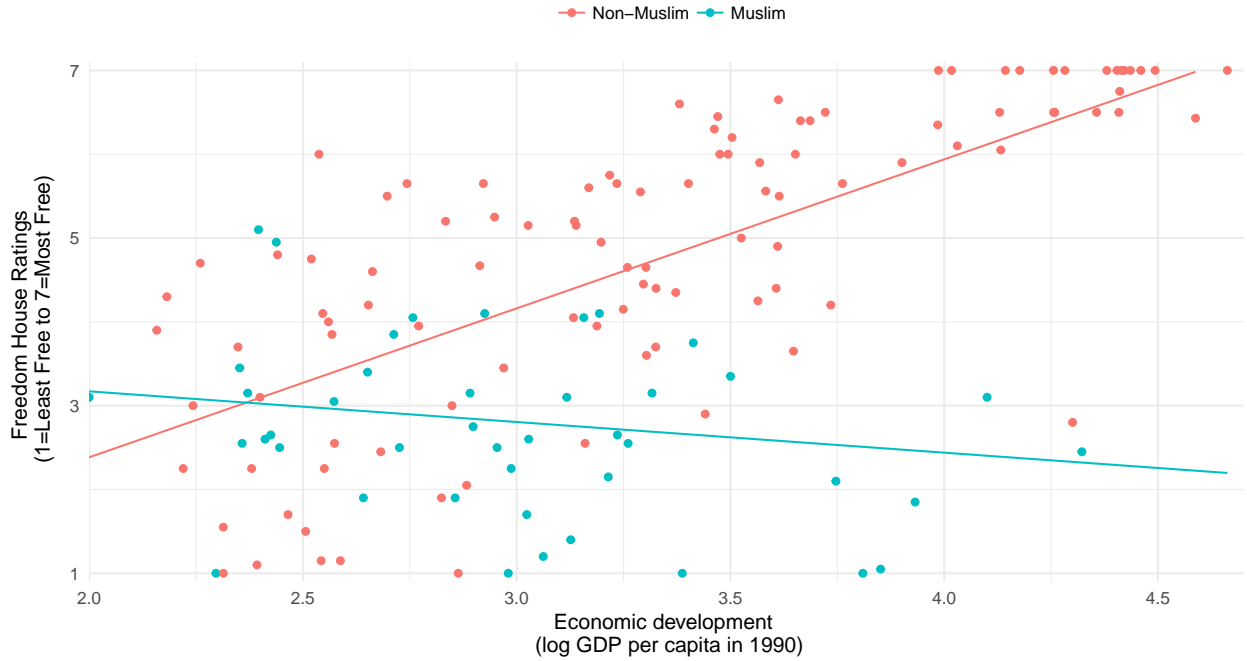

Figure 1: Interactive Effect of Economic Development and Religious Tradition on Democracy

Are these Marginal Effects Significant?

In the figure above, we have done a pretty decent job in showing how our interactive effect works in predicting FH scores across the different values of economic development and religious tradition. However, before we can fully understand the substantive meaning of this interaction effect we also need to take into account the standard errors around the interactive effect. This will help us to answer the question of whether marginal effect of our covariate is significant.

From above we know that:

$$\frac{\partial fhrev}{\partial income} = \hat{\beta}_{income} + \hat{\beta}_{muslim:income} \times muslim$$

Lets denote the `muslim` variable by M and `income` by I, then we can express the variance of $\frac{\partial fhrev}{\partial income}$ as:

$$
\begin{aligned}
Var(\frac{\partial fhrev}{\partial I}) &= Var(\hat{\beta}_I + \beta_{M:I} \times M) \\
&= Var(\hat{\beta}_I) + Var(\hat{\beta}_{M:I} \times M) + 2Cov(\hat{\beta}_I, \beta_{M:I} \times M) \\
&= Var(\hat{\beta}_I) + M^2 Var(\hat{\beta}_{M:I}) + 2 M Cov(\hat{\beta}_I, \beta_{M:I})
\end{aligned}
$$

Now lets do a simple numerical example for how to get a confidence interval around the marginal effect. Before we can start we need to get the necessary parameter values, specifically, we need the coefficient estimates, variances, and covariances. We can find each of these by using the variance-covariance martix from our regression results, we can access this matrix by calling the `vcov` function on our model:

```
round(vcov(lm2)[c('muslim','income','muslim:income'),c('muslim','income','muslim:income')],4)
```

```
##                muslim  income muslim:income
## muslim         1.4455  0.0900       -0.4719
## income         0.0900  0.0315       -0.0281
## muslim:income -0.4719 -0.0281        0.1601
```

The diagnoals of this matrix are filled with the variance estimates of our regressions and the off-diagnoals their covariances. From this matrix we can pull out the necessary information we need to calculate the confidence interval around the marginal effect of income on FH scores.

Marginal effect of Income on FH scores when muslim = 0: $\hat{\beta}_I + \hat{\beta}_{M:I} \times 0 = 1.78 + \text{-2.14} \times 0 = 1.78$

Corresponding standard error: $\sqrt{var(\hat{\beta}_I)} = \sqrt{0.0315} = 0.18$

Marginal effect of Income on FH scores when muslim = 1: $\hat{\beta}_I + \hat{\beta}_{M:I} \times 1 = 1.78 + \text{-2.14} \times 1 = \text{-0.366}$

Corresponding standard error:

$\sqrt{var(\hat{\beta}_I) + M^2 var(\hat{\beta}_{M:I}) + 2M cov(\hat{\beta}_I, \hat{\beta}_{M:I})} = \sqrt{0.0315} + 0.1601 + 2 \times -0.0281 = 0.37$

From this information we can now answer the question that we started with this section. To calculate the 95% confidence interval, we follow the same procedure as always:

Upper 95% CI for marginal effect of income when muslim=0: $1.78 + 1.96 \times 0.18 = 2.13$

Lower 95% CI for marginal effect of income when muslim=0: $1.78 - 1.96 \times 0.18 = 1.43$

Upper 95% CI for marginal effect of income when muslim=1: $\text{-0.37} + 1.96 \times 0.37 = 0.36$

Lower 95% CI for marginal effect of income when muslim=1: $\text{-0.37} - 1.96 \times 0.37 = \text{-1.09}$

Our conclusion from this is that the marginal effect of income on Freedom House scores is significant at a 95% confidence interval when muslim equals zero and not significant at that interval when muslim equals one.

However, tables are boring lets make a plot to illustrate this result, specifically what is known as a marginal effects plot. The interpretation of this plot is quick and easy, the marginal effect of income on democracy is significant for non-Muslim countries but it is not significant for Muslim countries.
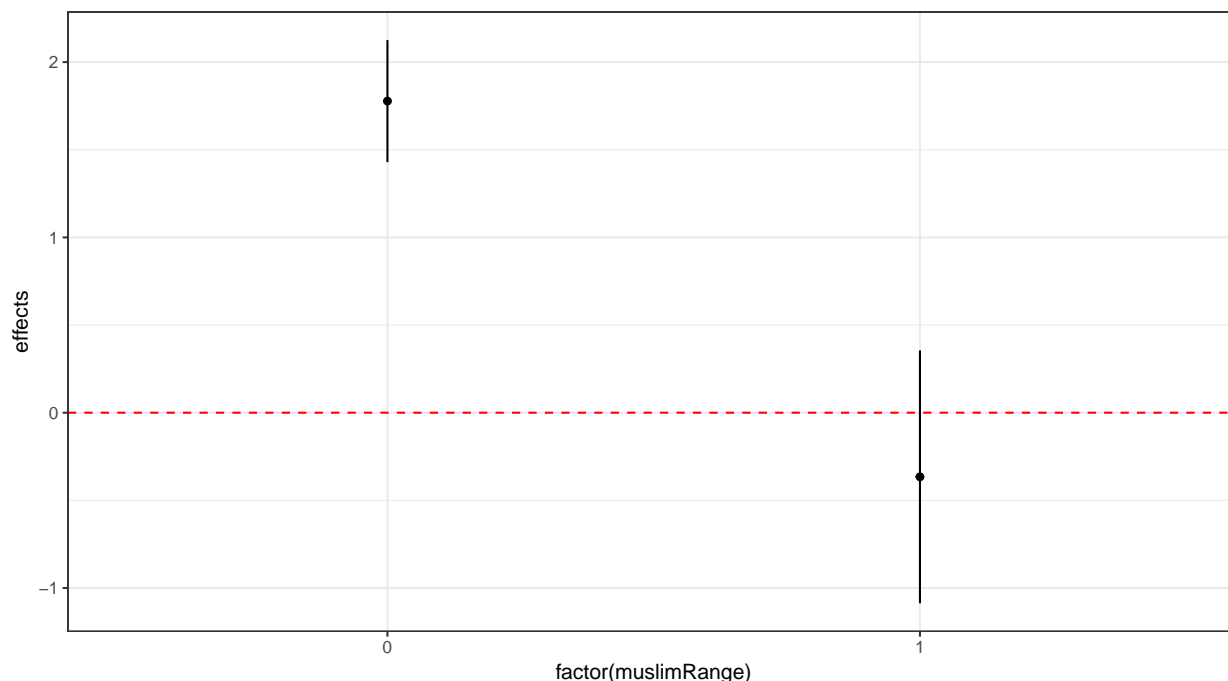

Figure 2: Marginal Effect of Income on Democracy at Levels of Muslim

Now lets do this same analysis to assess the signifiance of the marginal effect of muslim on Freedom House scores at the varying levels of income in the Fish dataset.

Also notice that this analysis would not lend itself to a tabular format as the income variable is continuous. Our conclusion from the analysis shown in this figure is that the marginal effect of the muslim variable on democracy is not significant at low levels of income but it is significant at higher levels.

I find that marginal effect plots are useful to an extent but I prefer thinking of the interactive relationship in a holistic sense that allows us to directly relate the scenarios we are modeling to the dependent variable.

To do this we're going to bring back the plot that we started with and add some confidence intervals to it. What you can see here is that the piecemeal information we were capturing from the two separate marginal effect plots is all captured in one here.

First notice that the predicted democracy scores for Muslim countries across income levels is pretty much flat, indicating that higher levels of income do not have a significant effect on democracy for Muslim countries.

Second, notice that at low levels of income the predicted democracy scores for Muslim and non-Muslim countries overlap, and they only begin to distinguish themselves at higher levels of income.

Thus we have summarized all the information that was in the marginal effect plots into one neater looking plot – or at least it looks neater to me.
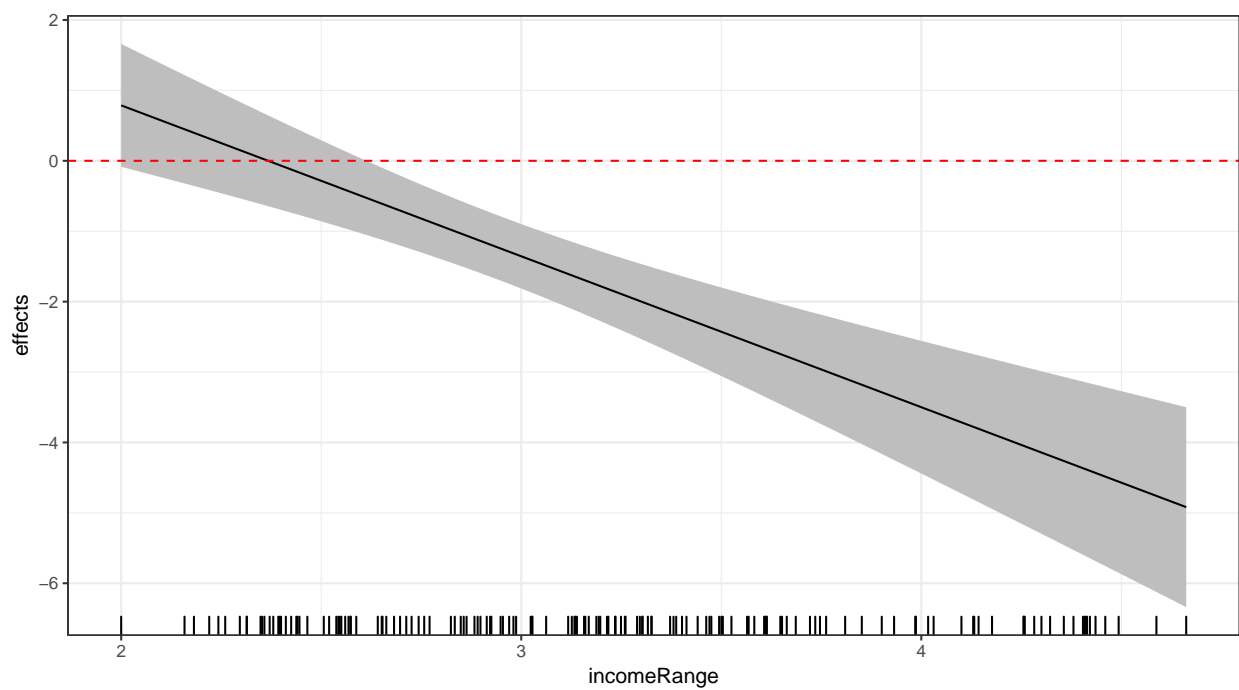
11

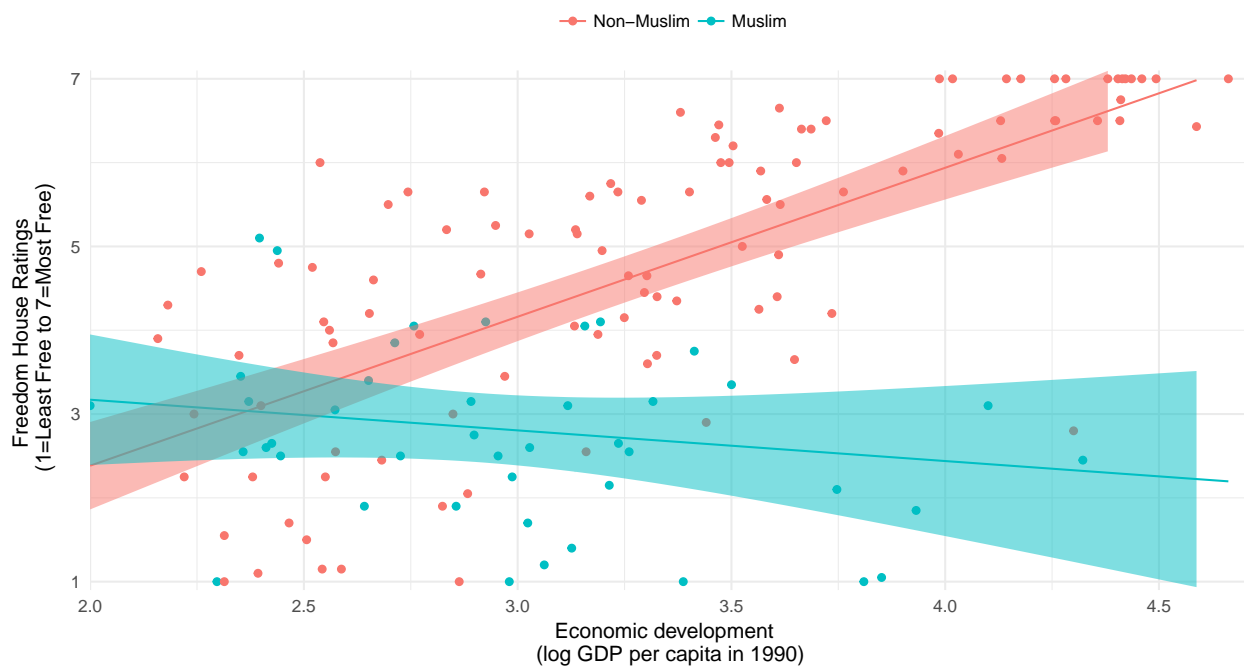Figure 3: Marginal Effect of Muslim on Democracy at Varying Levels of Income

Figure 4: Interactive Effect with Confidence Intervals of Economic Development and Religious Tradition on Democracy