Quantitative Methods III
Homework 1

Alexander Herzog

October 1, 2007

(I) Huber et al. (2005) data

Note: I named the Huber et al. data set `data.R` which is a really bad name as `.R` is usually reserved for syntax files. The convention is to use the extension `.Rdata`

Load Gelman's "arm" package.

```
> library(arm)
```

Change working directory

```
> setwd("~/NYU/QuantIII/hmwk01/")
```

a) Get the data for the huber et al study into R

```
> load("huber_et_al_2005.Rdata")
```

b) Run some individual country logits

b.1) Individual country regressions for Germany, Ireland, and Slovenia

```
> pid.germany <- glm(simplepid ~ education + age1 + age2 + male,
+                         subset = country == "Germany",
+                         family=binomial(link="logit"))

> display(pid.germany)
glm(formula = simplepid ~ education + age1 + age2 + male, family = binomial(link = "logit"),
    subset = country == "Germany")
            coef.est coef.se
(Intercept) -0.76     0.08
education    0.23     0.03
age1         0.25     0.05
age2        -0.02     0.02
male         0.40     0.10
---
  n = 1906, k = 5
  residual deviance = 2362.6, null deviance = 2481.3 (difference = 118.7)


> pid.ireland <- glm(simplepid ~ education + age1 + age2 + male,
```

```
+                         subset = country == "Ireland",
+                         family=binomial(link="logit"))

> display(pid.ireland)
glm(formula = simplepid ~ education + age1 + age2 + male, family = binomial(link = "logit"),
    subset = country == "Ireland")
            coef.est coef.se
(Intercept) -1.22     0.08
education    0.03     0.03
age1         0.25     0.04
age2        -0.01     0.01
male         0.28     0.09
---
  n = 2291, k = 5
  residual deviance = 2658.5, null deviance = 2727.3 (difference = 68.8)


> pid.slovenia <- glm(simplepid ~ education + age1 + age2 + male,
+                         subset = country == "Slovenia",
+                         family=binomial(link="logit"))

> display(pid.slovenia)
glm(formula = simplepid ~ education + age1 + age2 + male, family = binomial(link = "logit"),
    subset = country == "Slovenia")
            coef.est coef.se
(Intercept) -1.62     0.10
education    0.17     0.03
age1         0.22     0.05
age2        -0.01     0.02
male         0.46     0.12
---
  n = 1869, k = 5
  residual deviance = 1833.9, null deviance = 1898.4 (difference = 64.6)
```

b.2) All countries with regression of country b's for education on a single group level predictor (= "secret weapon")

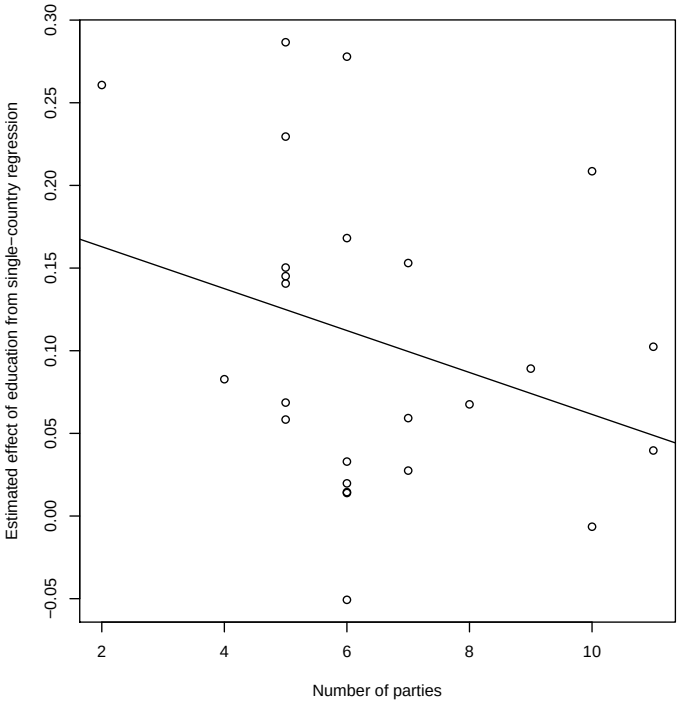
```
> j <- 1
> coefs <- NA

> for(i in unique(country)) {
+   reg.i <- glm(simplepid ~ education + age1 + age2 + male,
+                         subset = country == i,
+                         family=binomial(link="logit"))
+   coefs[j] <- reg.i$coef[2]
+   j <- j + 1
+ }

> no.parties.aggregated <- tapply(no.parties.gt.1, country, mean)

> reg.coef <- lm(coefs ~ no.parties.aggregated)

> plot(no.parties.aggregated,coefs,
+      xlab="Number of parties", ylab="Estimated effect of education
+      from single-country regression")
> abline(reg.coef)
```


Estimated effect of education from single−country regression
−0.05
0.00
0.05
0.10
0.15
0.20
0.25
0.30
2
4
6
8
10
Number of parties

c) Take the full data set and run with interactions

I specify three different models: a complete-pooling model, a no-pooling model, and a no-pooling model with interactions between country dummies and education.

c.1) Complete-pooling model

```
> pooling <- glm(simplepid ~  education + age1 + age2 + male , family=binomial(link="logit"))

> display(pooling)
glm(formula = simplepid ~ education + age1 + age2 + male, family = binomial(link = "logit"))
            coef.est coef.se
(Intercept) -0.23     0.01
education    0.12     0.01
age1         0.19     0.01
age2        -0.02     0.00
male         0.16     0.02
---
  n = 56073, k = 5
  residual deviance = 76210.1, null deviance = 77629.6 (difference = 1419.6)
```

c.2) No-pooling model: Varying-intercept model without interactions

```
> no.pooling <- glm(simplepid ~  education + age1 + age2 + male +
+                   as.factor(country),
+                   family=binomial(link="logit"))

> display(no.pooling)
glm(formula = simplepid ~ education + age1 + age2 + male + as.factor(country),
    family = binomial(link = "logit"))
                             coef.est coef.se
(Intercept)                   1.45     0.07
education                     0.10     0.01
age1                          0.20     0.01
age2                         -0.02     0.00
male                          0.17     0.02
as.factor(country)Belgium     1.31     0.12
as.factor(country)Bulgaria   -1.92     0.09
as.factor(country)Canada     -1.65     0.08
as.factor(country)Czech Rep  -1.51     0.08
as.factor(country)Denmark    -1.68     0.08
as.factor(country)France     -1.69     0.09
as.factor(country)Germany    -2.15     0.08
as.factor(country)Hungary    -1.90     0.08
as.factor(country)Iceland    -1.57     0.08
as.factor(country)Ireland    -2.57     0.08
as.factor(country)Israel     -1.12     0.08
as.factor(country)Japan      -2.16     0.09
as.factor(country)Lithuania  -2.29     0.10
as.factor(country)Netherlands -2.58    0.08
as.factor(country)New Zealand -1.39    0.07
as.factor(country)Norway     -1.82     0.07
as.factor(country)Poland     -1.65     0.08
as.factor(country)Portugal   -1.56     0.08
as.factor(country)Slovenia   -2.91     0.09
as.factor(country)Spain      -1.80     0.08
```

```
as.factor(country)Sweden      -1.62     0.08
as.factor(country)Switzerland -2.06     0.08
as.factor(country)UK          -1.74     0.08
as.factor(country)USA         -1.46     0.09
---
  n = 56073, k = 29
  residual deviance = 71308.5, null deviance = 77629.6 (difference = 6321.1)
```

c.3) No-pooling model with interactions for 'education'

```
> no.pooling.int <- glm(simplepid ~  education + age1 + age2 + male +
+                       as.factor(country) + as.factor(country):education,
+                       family=binomial(link="logit"))

> display(no.pooling.int)
glm(formula = simplepid ~ education + age1 + age2 + male + as.factor(country) +
    as.factor(country):education, family = binomial(link = "logit"))
                                       coef.est coef.se
(Intercept)                             1.49     0.07
education                              -0.04     0.04
age1                                    0.20     0.01
age2                                   -0.01     0.00
male                                    0.17     0.02
as.factor(country)Belgium               1.27     0.12
as.factor(country)Bulgaria             -1.94     0.09
as.factor(country)Canada               -1.76     0.09
as.factor(country)Czech Rep            -1.56     0.08
as.factor(country)Denmark              -1.64     0.09
as.factor(country)France               -1.71     0.10
as.factor(country)Germany              -2.11     0.09
as.factor(country)Hungary              -1.89     0.08
as.factor(country)Iceland              -1.59     0.09
as.factor(country)Ireland              -2.61     0.08
as.factor(country)Israel               -1.11     0.08
as.factor(country)Japan                -2.19     0.09
as.factor(country)Lithuania            -2.45     0.11
as.factor(country)Netherlands          -2.60     0.09
as.factor(country)New Zealand          -1.41     0.08
as.factor(country)Norway               -1.84     0.08
as.factor(country)Poland               -1.59     0.08
as.factor(country)Portugal             -1.63     0.09
as.factor(country)Slovenia             -2.93     0.09
as.factor(country)Spain                -1.84     0.08
as.factor(country)Sweden               -1.65     0.08
as.factor(country)Switzerland          -2.06     0.08
as.factor(country)UK                   -1.78     0.08
as.factor(country)USA                  -1.70     0.10
education:as.factor(country)Belgium     0.22     0.07
education:as.factor(country)Bulgaria    0.19     0.05
education:as.factor(country)Canada      0.21     0.05
education:as.factor(country)Czech Rep   0.06     0.05
education:as.factor(country)Denmark     0.03     0.05
education:as.factor(country)France      0.07     0.05
education:as.factor(country)Germany     0.27     0.05
education:as.factor(country)Hungary     0.21     0.05
education:as.factor(country)Iceland     0.21     0.05
```

```
education:as.factor(country)Ireland      0.06     0.05
education:as.factor(country)Israel       0.07     0.05
education:as.factor(country)Japan       -0.03     0.06
education:as.factor(country)Lithuania    0.28     0.06
education:as.factor(country)Netherlands  0.07     0.05
education:as.factor(country)New Zealand  0.05     0.04
education:as.factor(country)Norway       0.07     0.04
education:as.factor(country)Poland       0.32     0.05
education:as.factor(country)Portugal     0.12     0.05
education:as.factor(country)Slovenia     0.21     0.05
education:as.factor(country)Spain        0.14     0.05
education:as.factor(country)Sweden       0.10     0.05
education:as.factor(country)Switzerland  0.26     0.05
education:as.factor(country)UK           0.13     0.05
education:as.factor(country)USA          0.31     0.05
---
  n = 56073, k = 53
  residual deviance = 71005.8, null deviance = 77629.6 (difference = 6623.9)
```

The different model predictions for the effect of education on the probability of having a party ID are shown in Figure 1.

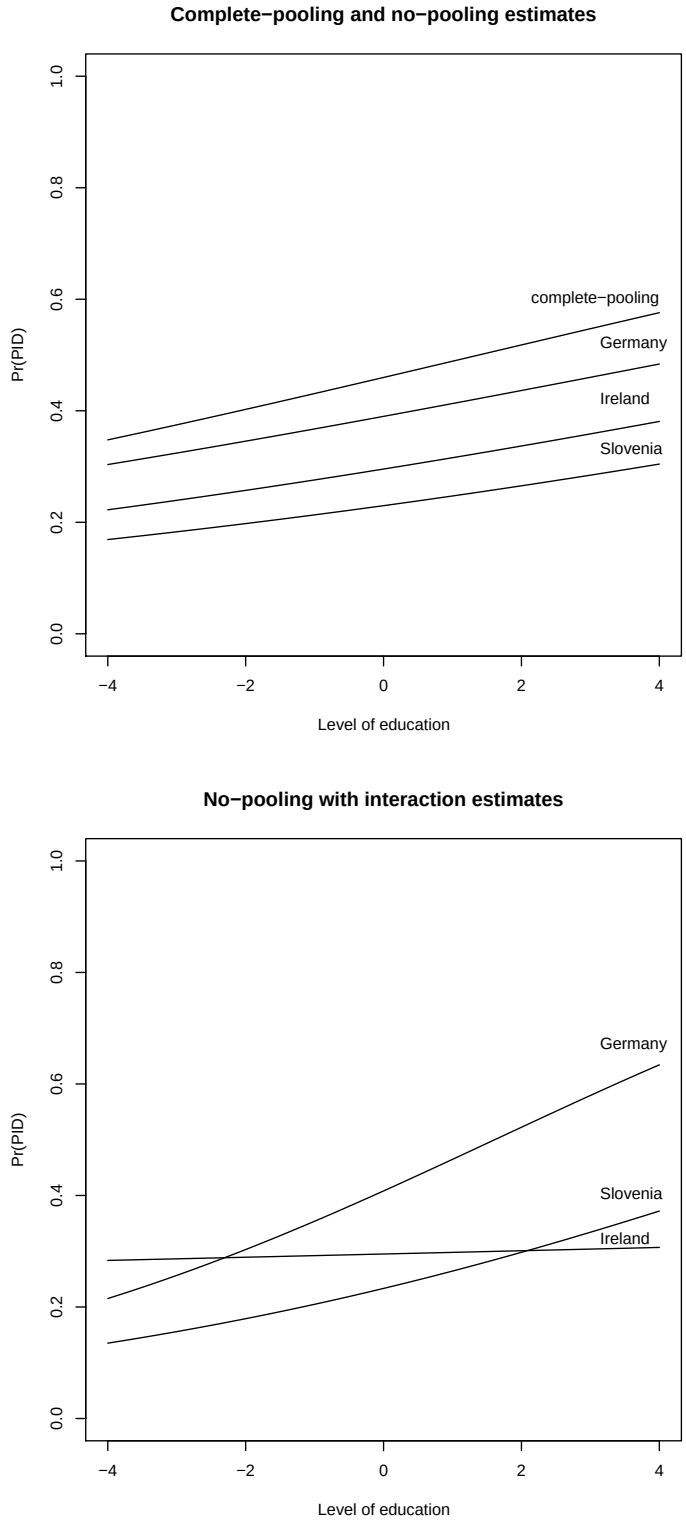


Figure 1: The effect of education on the predicted probability of identifying with a party, derived from a complete-pooling model, a no-pooling model, and a no-pooling model with interactions between the country dummy variables and education. Age variables are held at their mean values and male == 1.

d) Now repeat c with lmer

d.1) Varying intercepts model without group-level predictors

```
> lmer.d1 <- lmer(simplepid ~  education + age1 + age2 + male + (1 | country),
+                     family=binomial(link="logit"))
```

Get coefficient estimates and standard errors for fixed effects.

Two ways: Either use `display()`:

```
> display(lmer.d1)
lmer(formula = simplepid ~ education + age1 + age2 + male + (1 |
    country), family = binomial(link = "logit"))
            coef.est coef.se
(Intercept) -0.21     0.16
education    0.10     0.01
age1         0.20     0.01
age2        -0.02     0.00
male         0.17     0.02

Error terms:
 Groups   Name Std.Dev.
 country       0.81
 Residual      1.00
---
number of obs: 56073, groups: country, 25
AIC = 71487.1, DIC = 71475.1
deviance = 71475.1
```

or use `fixef()` and `se.fixef()`:

```
> fixef(lmer.d1)
(Intercept)   education        age1        age2        male
     -0.210       0.096       0.201      -0.016       0.172

> se.fixef(lmer.d1)
[1] 0.1623 0.0056 0.0076 0.0030 0.0179
```

Get coefficient estimates and standard errors for random effects.

Here we have to use `ranef()` and `se.ranef()` as `display()` doesn't print random effects.

```
> ranef(lmer.d1)
An object of class "ranef.lmer"
[[1]]
            (Intercept)
Australia         1.651
Belgium           2.922
Bulgaria         -0.255
Canada            0.013
Czech Rep         0.156
Denmark          -0.021
```

```
France           -0.032
Germany          -0.483
Hungary          -0.236
Iceland           0.093
Ireland          -0.903
Israel            0.544
Japan            -0.500
Lithuania        -0.622
Netherlands      -0.913
New Zealand       0.268
Norway           -0.159
Poland            0.009
Portugal          0.100
Slovenia         -1.240
Spain            -0.133
Sweden            0.045
Switzerland      -0.396
UK               -0.078
USA               0.199
```

```
> se.ranef(lmer.d1)
An object of class "ranef.lmer"
[[1]]
      [,1]
 [1,] 0.17
 [2,] 0.19
 [3,] 0.17
 [4,] 0.17
 [5,] 0.17
 [6,] 0.17
 [7,] 0.17
 [8,] 0.17
 [9,] 0.17
[10,] 0.17
[11,] 0.17
[12,] 0.17
[13,] 0.17
[14,] 0.18
[15,] 0.17
[16,] 0.16
[17,] 0.16
[18,] 0.17
[19,] 0.17
[20,] 0.17
[21,] 0.17
[22,] 0.17
[23,] 0.17
[24,] 0.17
[25,] 0.17
```

d.2) Varying intercepts and slopes model without group-level predictors

```
> lmer.d2 <- lmer(simplepid ~  education + age1 + age2 + male + (1 + education | country),
+                     family=binomial(link="logit"))
```

Get coefficient estimates and standard errors for fixed effects:

```
> fixef(lmer.d2)
(Intercept)   education        age1        age2        male
     -0.200       0.101       0.199      -0.014       0.168

> se.fixef(lmer.d2)
[1] 0.1620 0.0197 0.0077 0.0031 0.0180
```

Get coefficient estimates and standard errors for random effects:

```
> ranef(lmer.d2)
An object of class "ranef.lmer"
[[1]]
            (Intercept) education
Australia        1.6729   -0.1226
Belgium          2.9090    0.0423
Bulgaria        -0.2448    0.0455
Canada          -0.0617    0.0620
Czech Rep        0.1358   -0.0800
Denmark          0.0504   -0.1023
France          -0.0226   -0.0619
Germany         -0.4250    0.1168
Hungary         -0.1974    0.0575
Iceland          0.0992    0.0625
Ireland         -0.9149   -0.0791
Israel           0.5797   -0.0670
Japan           -0.4935   -0.1511
Lithuania       -0.7277    0.1016
Netherlands     -0.9103   -0.0678
New Zealand      0.2815   -0.0932
Norway          -0.1500   -0.0730
Poland           0.0952    0.1654
Portugal         0.0623   -0.0262
Slovenia        -1.2324    0.0599
Spain           -0.1433   -0.0032
Sweden           0.0434   -0.0378
Switzerland     -0.3662    0.1092
UK              -0.0915   -0.0104
USA              0.0091    0.1478

> se.ranef(lmer.d2)
An object of class "ranef.lmer"
[[1]]
      [,1]  [,2]
 [1,] 0.17 0.041
 [2,] 0.19 0.052
 [3,] 0.17 0.035
 [4,] 0.17 0.035
 [5,] 0.17 0.033
 [6,] 0.17 0.031
 [7,] 0.17 0.037
 [8,] 0.17 0.034
 [9,] 0.17 0.029
[10,] 0.17 0.032
[11,] 0.17 0.032
[12,] 0.17 0.030
[13,] 0.17 0.038
```

```
[14,] 0.18 0.046
[15,] 0.17 0.037
[16,] 0.16 0.025
[17,] 0.16 0.026
[18,] 0.17 0.029
[19,] 0.17 0.031
[20,] 0.17 0.036
[21,] 0.17 0.029
[22,] 0.17 0.031
[23,] 0.17 0.028
[24,] 0.17 0.033
[25,] 0.17 0.036
```