Multinomial regression of spouse characteristics for the ChitwanABM

Author: Alex Zvoleff
Email: azvoleff@mail.sdsu.edu
Date: July, 2012

Analyzes the relationship grid data from the CVFS to build a multinomial regression predict the probability of marrying a spouse within each of a set of age groups, together with the probability of marrying someone of a different ethnicity.

The results of this analysis are used to inform the marriage process in the ChitwanABM. Once the determination has been made that a woman is marrying (based on the probability derived from event history analysis), the multinomial regression developed here is used to assign a proability of marriage to each possible spouse, based on the age of that person, so that the age differential between spouses is realistic.

Load the data and setup R

```
library(ggplot2)
library(mlogit)
library(arm)
library(rms)  # Note 'Design' package was renamed to 'rms'
theme_update(theme_bw(base_size = 10))
```

```
load("V:/Nepal/CVFS_R_format/hhrel_with_respIDs.Rdata")
hhrel <- hhrel_with_respIDs

# Drop 'other' ethnicity for consistency with existing work
hhrel <- hhrel[!(hhrel$ethnic == "Other"), ]
hhrel$ethnic <- factor(hhrel$ethnic)
hhrel$CENGENDR <- factor(hhrel$CENGENDR)

hhrel$age_cat <- cut(hhrel$CENAGE, breaks = c(0, 15, 20, 30, 40,
    50, 60, 999), ordered_result = TRUE)
```

Basic statistics

First look at some basic statistics on who is married, and on how many spouses they have, by gender.

```
hhrel$HASSPOUSE1 <- !is.na(hhrel$SPOUSE1)
hhrel$HASSPOUSE2 <- !is.na(hhrel$SPOUSE2)
hhrel$HASSPOUSE3 <- !is.na(hhrel$SPOUSE3)
xtabs(~CENGENDR + HASSPOUSE1, data = hhrel)
```

```
##         HASSPOUSE1
## CENGENDR FALSE TRUE
##   female  2218 1892
##   male    2293 1849
```

```
xtabs(~CENGENDR + HASSPOUSE2, data = hhrel)
```

```
##         HASSPOUSE2
## CENGENDR FALSE TRUE
##   female  4110    0
##   male    4109   33
```

```
xtabs(~CENGENDR + HASSPOUSE3, data = hhrel)
```

```
##         HASSPOUSE3
## CENGENDR FALSE TRUE
##   female  4110    0
##   male    4140    2
```

Now look at who has more than one spouse (only males do) by age group:

```
xtabs(~age_cat + HASSPOUSE2, data = hhrel)
```

```
##           HASSPOUSE2
## age_cat    FALSE TRUE
##   (0,15]    3228    0
##   (15,20]    989    0
##   (20,30]   1393    1
##   (30,40]    970    7
##   (40,50]    711    7
##   (50,60]    470   10
##   (60,999]   458    8
```

```
xtabs(~age_cat + HASSPOUSE3, data = hhrel)
```

```
##           HASSPOUSE3
## age_cat    FALSE TRUE
##   (0,15]    3228    0
##   (15,20]    989    0
##   (20,30]   1394    0
##   (30,40]    977    0
##   (40,50]    718    0
##   (50,60]    479    1
##   (60,999]   465    1
```

Having multiple wives is mostly confied to older men. Now look into the difference in spouse age, by gender.

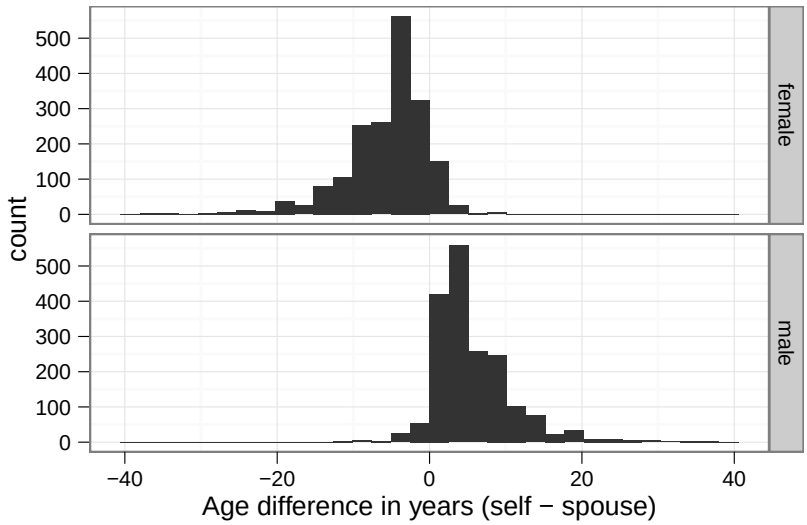
```
spouse1_row <- match(hhrel$SPOUSE1, hhrel$RESPID)
hhrel$sp_age <- hhrel$CENAGE[spouse1_row]
hhrel$sp_age_cat <- hhrel$age_cat[spouse1_row]
hhrel$sp_age_diff <- hhrel$CENAGE - hhrel$sp_age
xtabs(~age_cat + sp_age_cat, data = hhrel)
```

```
##           sp_age_cat
## age_cat    (0,15] (15,20] (20,30] (30,40] (40,50] (50,60] (60,999]
##   (0,15]        0       1       2       0       0       0        0
##   (15,20]       1      88     160       8       0       1        0
##   (20,30]       2     160     689     271      27       6        1
##   (30,40]       0       7     265     425     194      31        8
##   (40,50]       0       0      26     189     268     146       41
##   (50,60]       0       1       4      28     142     133       85
##   (60,999]      0       0       1       7      38      83      188
```

```
qplot(sp_age_diff, facets = CENGENDR ~ ., xlab = "Age difference in years (self - spouse)",
    data = hhrel)
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust
## this.
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust
## this.
```



Comparison of spouse age difference by gender

Setup some age categories for the difference in spouse age.

```
hhrel$sp_age_diff_cat <- cut(hhrel$sp_age_diff, breaks = c(-999,
    -15, -10, -5, -2, -1, 0, 1, 2, 5, 10, 15, 999), ordered_result = TRUE)
xtabs(~sp_age_diff_cat + age_cat + CENGENDR, data = hhrel)
```

```
## , , CENGENDR = female
##
##                 age_cat
## sp_age_diff_cat (0,15] (15,20] (20,30] (30,40] (40,50] (50,60] (60,999]
##      (-999,-15]      0       4      29      34      41      14        2
##      (-15,-10]       1      17      64      68      59      30       12
##      (-10,-5]        1      79     242     146      95      51       30
```

```
##      (-5,-2]         1      78     219     122      77      23      23
##      (-2,-1]         0      25      48      24      14      12       7
##      (-1,0]          0       4      27      30      16      11       8
##      (0,1]           0       3       8       8       5       6       3
##      (1,2]           0       0       5       6       5       2       3
##      (2,5]           0       1       3       5       8       3       5
##      (5,10]          0       0       1       0       4       1       2
##      (10,15]         0       0       0       0       1       0       2
##      (15,999]        0       0       0       0       1       0       0
##
## , , CENGENDR = male
##
##                 age_cat
## sp_age_diff_cat (0,15] (15,20] (20,30] (30,40] (40,50] (50,60] (60,999]
##      (-999,-15]      0       0       1       0       0       0       0
##      (-15,-10]       0       0       1       2       0       0       3
##      (-10,-5]        0       0       1       3       6       1       3
##      (-5,-2]         0       2       9       9      10       4       4
##      (-2,-1]         0       3       9       8       4       6       3
##      (-1,0]          0       6      27      30      14      12       8
##      (0,1]           0      17      54      25      13      14       7
##      (1,2]           0       8      81      42      36      13      14
##      (2,5]           0      11     212     162     100      45      28
##      (5,10]          0       0     106     165      99      75      60
##      (10,15]         0       0       9      33      49      50      37
##      (15,999]        0       0       0       8      13      20      53
##
```

Do people marry outside of their ethnic group?

```
spouse1_row <- match(hhrel$SPOUSE1, hhrel$RESPID)
hhrel$spouse_ethnicity <- hhrel$ethnic[spouse1_row]
xtabs(~ethnic + spouse_ethnicity, data = hhrel)
```

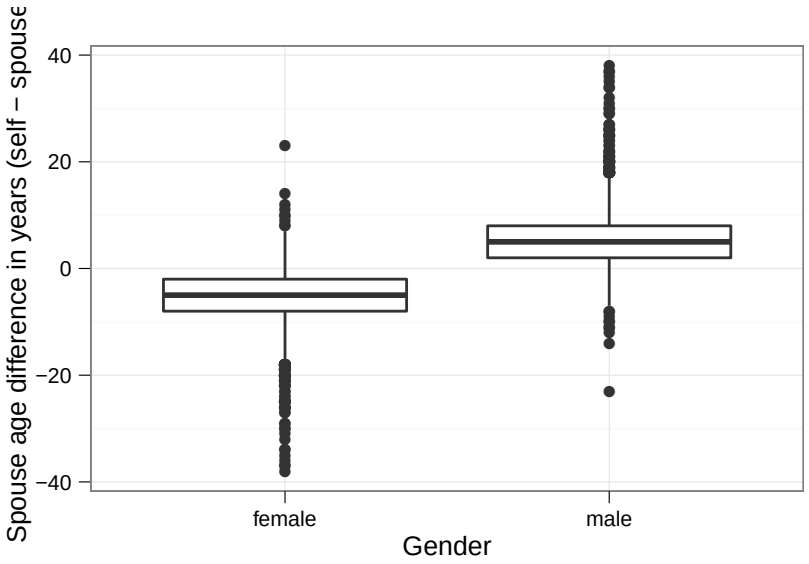
```
##             spouse_ethnicity
## ethnic        UpHindu HillTibeto LowHindu Newar TeraiTibeto
##   UpHindu        1733          2        0     0           0
##   HillTibeto        2        633        0     0           2
##   LowHindu          0          0      413     0           0
##   Newar             0          0        0   244           0
##   TeraiTibeto       0          2        0     0         696
```

Marriages outside of your ethnic group are VERY uncommon. There are only 4 in the data. So we will disallow these marriages in the model - there are not enough of them to develop any kind of predictive model of when they might occur.

Make a few final summary plots: the mean of spouse_age_diff versus gender, and a histogram of spouse_age_diff by gender.

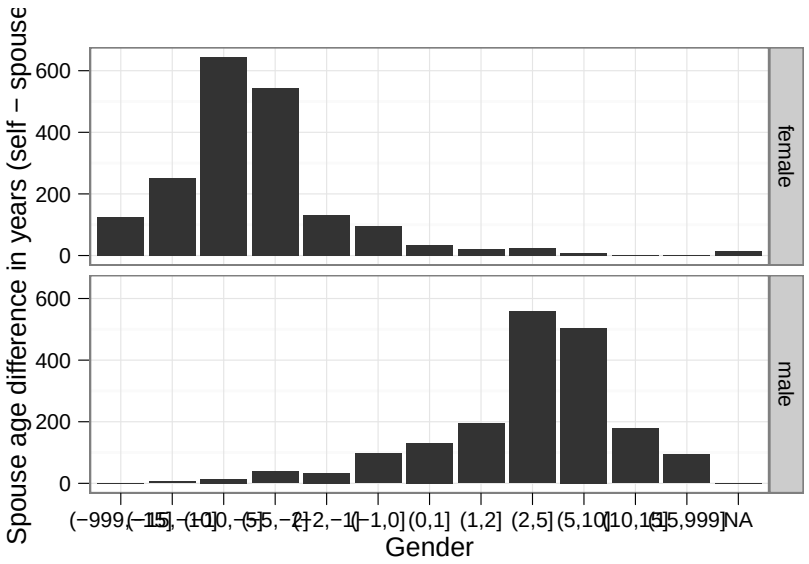
```
qplot(CENGENDR, sp_age_diff, geom = "boxplot", xlab = "Gender", ylab = "Spouse age difference in years (self - spouse)",
    data = hhrel)
```

```
## Warning: Removed 4525 rows containing non-finite values (stat_boxplot).
```



plot of chunk mean-sp-age-diff-versus-age

```
qplot(sp_age_diff_cat, facets = CENGENDR ~ ., geom = "histogram",
    xlab = "Gender", ylab = "Spouse age difference in years (self - spouse)",
    data = hhrel[hhrel$HASSPOUSE1, ])
```



Histogram of marriages by spouse age categories and gender

Multinomial logistic regression predicting spouse age difference based on gender

```
hhrel_mlogit_data <- mlogit.data(hhrel, varying = NULL, choice = "sp_age_diff_cat",
    shape = "wide")
mlogit_spouseage <- mlogit(sp_age_diff_cat ~ 1 | CENGENDR, data = hhrel_mlogit_data)
summary(mlogit_spouseage)
```

```
##
## Call:
## mlogit(formula = sp_age_diff_cat ~ 1 | CENGENDR, data = hhrel_mlogit_data,
##     method = "nr", print.level = 0)
##
## Frequencies of alternatives:
##     (-1,0]   (-10,-5]  (-15,-10]    (-2,-1]    (-5,-2] (-999,-15]
##     0.0518     0.1765     0.0690     0.0437     0.1559     0.0335
##      (0,1]      (1,2]    (10,15]   (15,999]      (2,5]     (5,10]
##     0.0437     0.0577     0.0486     0.0255     0.1564     0.1376
##
## nr method
## 8 iterations, 0h:0m:2s
## g'(-H)^-1g = 3.65E-06
## successive fonction values within tolerance limits
##
## Coefficients :
##                           Estimate Std. Error t-value Pr(>|t|)
## (-10,-5]:(intercept)         1.903      0.109   17.40  < 2e-16 ***
## (-15,-10]:(intercept)        0.961      0.120    8.01  1.1e-15 ***
## (-2,-1]:(intercept)          0.303      0.135    2.25    0.024 *
## (-5,-2]:(intercept)          1.733      0.111   15.65  < 2e-16 ***
## (-999,-15]:(intercept)       0.256      0.136    1.88    0.060 .
## (0,1]:(intercept)           -1.068      0.202   -5.29  1.2e-07 ***
## (1,2]:(intercept)           -1.520      0.241   -6.31  2.8e-10 ***
## (10,15]:(intercept)         -3.466      0.586   -5.91  3.4e-09 ***
## (15,999]:(intercept)        -4.564      1.005   -4.54  5.6e-06 ***
## (2,5]:(intercept)           -1.345      0.225   -5.99  2.1e-09 ***
## (5,10]:(intercept)          -2.485      0.368   -6.75  1.5e-11 ***
## (-10,-5]:CENGENDRmale       -3.839      0.306  -12.54  < 2e-16 ***
## (-15,-10]:CENGENDRmale      -3.744      0.437   -8.56  < 2e-16 ***
## (-2,-1]:CENGENDRmale        -1.381      0.242   -5.70  1.2e-08 ***
## (-5,-2]:CENGENDRmale        -2.670      0.221  -12.08  < 2e-16 ***
## (-999,-15]:CENGENDRmale     -4.831      1.014   -4.76  1.9e-06 ***
## (0,1]:CENGENDRmale           1.361      0.242    5.62  2.0e-08 ***
## (1,2]:CENGENDRmale           2.213      0.271    8.16  2.2e-16 ***
## (10,15]:CENGENDRmale         4.073      0.600    6.79  1.1e-11 ***
## (15,999]:CENGENDRmale        4.533      1.016    4.46  8.1e-06 ***
## (2,5]:CENGENDRmale           3.095      0.250   12.38  < 2e-16 ***
## (5,10]:CENGENDRmale          4.135      0.384   10.76  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log-Likelihood: -6720
## McFadden R^2:  0.212
## Likelihood ratio test : chisq = 3620 (p.value = <2e-16)
```

```
mlogit_spouseage_odds <- cbind(odds_ratio = exp(coef(mlogit_spouseage)),
    exp(confint.default(mlogit_spouseage)))
(mlogit_spouseage_odds <- round(mlogit_spouseage_odds, 4))
```

```
##                           odds_ratio   2.5 %   97.5 %
## (-10,-5]:(intercept)          6.7083  5.4136   8.3127
## (-15,-10]:(intercept)         2.6146  2.0666   3.3079
## (-2,-1]:(intercept)           1.3542  1.0402   1.7629
```

```
## (-5,-2]:(intercept)          5.6562  4.5529   7.0270
## (-999,-15]:(intercept)       1.2917  0.9895   1.6860
## (0,1]:(intercept)            0.3437  0.2315   0.5105
## (1,2]:(intercept)            0.2187  0.1364   0.3508
## (10,15]:(intercept)          0.0312  0.0099   0.0986
## (15,999]:(intercept)         0.0104  0.0015   0.0747
## (2,5]:(intercept)            0.2604  0.1677   0.4044
## (5,10]:(intercept)           0.0833  0.0405   0.1714
## (-10,-5]:CENGENDRmale        0.0215  0.0118   0.0392
## (-15,-10]:CENGENDRmale       0.0237  0.0100   0.0558
## (-2,-1]:CENGENDRmale         0.2512  0.1562   0.4040
## (-5,-2]:CENGENDRmale         0.0693  0.0449   0.1068
## (-999,-15]:CENGENDRmale      0.0080  0.0011   0.0583
## (0,1]:CENGENDRmale           3.8988  2.4247   6.2690
## (1,2]:CENGENDRmale           9.1429  5.3742  15.5543
## (10,15]:CENGENDRmale        58.7216 18.1262 190.2349
## (15,999]:CENGENDRmale       93.0309 12.7110 680.8861
## (2,5]:CENGENDRmale          22.0899 13.5320  36.0599
## (5,10]:CENGENDRmale         62.4742 29.4146 132.6904
```

```
write.csv(coef(mlogit_spouseage), file = "mlogit_spouseage_coefs.csv")
write.csv(mlogit_spouseage_odds, file = "mlogit_spouseage_odds.csv")
```