STAT3011 Project 2

Group 3

CHAN, Wai Kit 1155050291

HO, Yui Man 1155078203

LEE, Cheung Kiu 1155079458

TSE, Kwan Nok 1155078905

UNDYONO, Jeremy J 1155083539

YUEN, Chun Wing 1155077960

Table of Contents

Introduction

Our project is to analyze the Dataset 1 from the given questionnaire. the dataset contains questionnaire responses data from China and India which are quantitative data with rating scales from 1 to 10 except for question V91, V146, V353 and V355. The first three questions are closed questions asking for status and the last question is an open question asking for the respondent's age.

Objective

The objective would compare the median of moral value and (minor) crime rate between China and India, finding the relationship between Problem, sex and age, the relationship between moral behavior and satisfaction and the relationship between sex and homosexuality.

Methodology

The dataset is being selected and divided into two main categories – moral value and (minor) crime. Moral value contains question V301 – V306, V312, V313, V316. Crime contains question V296 – V300, V314, V318, V319. We used R programming language to analyze the problems. The techniques used are Chi-square test of independence, Wilcoxon rank sum test, Bayesian ordered probit regression, concordant and discordant pairs and Pearson's correlation are carried out to find the objectives.

Results

Missing Data

Missing data is a big issue in this questionnaire. It bothers our analysis. The first method that quickly comes in mind is list-wise deletion, which deletes all the respondents with missing values. This, however, is not plausible in our case since many respondents have missing values. Simply deleting them will cause a huge data reduction. Besides, those respondents still contain part of the information and we want to utilize it.

```
> pMiss <- function(x) {sum(is.na(x))/length(x)*100}
> apply(china,2,pMiss)
  V91   V96  V116  V132  V146  V296  V297  V298  V299  V300  V301  V302
  2.0   0.4   7.6   0.4   0.3   1.2   0.2   0.4   0.2   0.6   1.1   0.7
 V303  V304  V305  V306  V307  V308  V309  V310  V311  V312  V313  V314
  2.1   1.2   0.8   0.5   2.7   1.0   4.5   3.3   3.4   4.5   1.6   1.2
 V315  V316  V317  V318  V319  V353  V355
  5.3   2.0 100.0   0.4   0.4   0.2   0.0
> apply(india,2,pMiss)
  V91   V96  V116  V132  V146  V296  V297  V298  V299  V300  V301  V302
 1.20  1.56 50.72  1.32  0.00  2.60  0.96  1.96  0.92  1.80  1.36  0.60
 V303  V304  V305  V306  V307  V308  V309  V310  V311  V312  V313  V314
 0.80  1.48  1.52  0.96  5.24  1.36  1.04  0.92  1.72  3.76  0.68  2.56
 V315  V316  V317  V318  V319  V353  V355
 3.00  1.60  2.72  0.80  0.84  0.00  0.00
```

The above table shows the percentage missing data for each question. Apparently, V317 in China and V116 in India contain so many missing data. (> 50%) Therefore, we delete these two questions from the rest of analysis.

```
> which(rmischina>=20)
 [1]  10 188 300 465 483 488 497 533 563 570 605 613 628 638 951 983
> which(rmisindia>=20)
 [1]    1    2    3    4  276  290  317  319  321  356  385  438  525  574
[15]  701  702  705  711  715  790  865  885  891  927  928  935 1165 1213
[29] 1312 1314 1317 1318 1319 1335 1336 1364 1376 1563 1605 1735 1861 1862
[43] 1916 1986 1989 1990 1995 1997 2179
> china<-china[-which(rmischina>=20),]
> india<-india[-which(rmisindia>=20),]
```

We also delete those respondents who answer less than 80% of the questions.

The MICE package in R comes in the place for missing data. The function MICE can impute missing values based on other existing values using different models. The model we used is "Proportional odds model", which is suitable for ordered data. We have run the MICE once, and continue our analysis using the imputed dataset.

Limitation:

- Very time-consuming: around half-hours for 2 iterations.[1]

[1] The imputed dataset is also sent to avoid long running time.

Chi-square Test of Independence

In this part, there is an additional category for V311 and V315, which is called security. Here, we are dealing with two-ways table and we tried to know the relationship between the results in both countries. There are several steps shown in this part. First, we made the social behavior frequency table for both China and India. Second, the proportion table of both countries is made.

```
> m1
          Social Behaviour in China
             1    2    3   4    5   6   7   8   9  10 TOTAL
Moral     4969  493  453 320  597 508 182 325 208 801  8856
Crime     5968  607  520 221  244 167  47  30  17  51  7872
Security  1101  201  180  94  186 124  24  25  11  22  1968
TOTAL    12038 1301 1153 635 1027 799 253 380 236 874 18696
>
```

```
> m2
          Social Behaviour in India
             1    2    3   4    5   6   7   8   9  10 TOTAL
Moral    15681 1075  995 794 1490 461 258 351 219 735 22059
Crime    16060  964  813 585  689 187  85  61  43 121 19608
Security  3406  262  315 244  413  85  42  45  30  60  4902
TOTAL    35147 2301 2123 1623 2592 733 385 457 292 916 46569
```

```
> mpropchi
       Proportion Table of China
          1       2       3       4       5       6       7       8       9       10      TOTAL
Moral     "56.11%" "5.57%"  "5.12%" "3.61%" "6.74%" "5.74%" "2.06%" "3.67%" "2.35%" "9.04%" "100%"
Crime     "75.81%" "7.71%"  "6.61%" "2.81%" "3.1%"  "2.12%" "0.6%"  "0.38%" "0.22%" "0.65%" "100%"
Security "55.95%" "10.21%" "9.15%" "4.78%" "9.45%" "6.3%"  "1.22%" "1.27%" "0.56%" "1.12%" "100%"
```

```
> mpropind
       Proportion Table of India
          1       2       3       4       5       6       7       8       9       10      TOTAL
Moral     "71.09%" "4.87%"  "4.51%" "3.6%"  "6.75%" "2.09%" "1.17%" "1.59%" "0.99%" "3.33%" "100%"
Crime     "81.91%" "4.92%"  "4.15%" "2.98%" "3.51%" "0.95%" "0.43%" "0.31%" "0.22%" "0.62%" "100%"
Security "69.48%" "5.34%"  "6.43%" "4.98%" "8.43%" "1.73%" "0.86%" "0.92%" "0.61%" "1.22%" "100%"
```

H_0: Moral values, crime, and security are independent variables.

H_a: They are not independent variables.

H_0 and H_a are the same for both China and India.

By the tables, the Chi-square test of independence can be conducted.

$$\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected}$$

$$Expected\ score = \frac{(\sum row)(\sum column)}{\sum overall}$$

$$D.F = (\#rows - 1)(\#columns - 1)$$

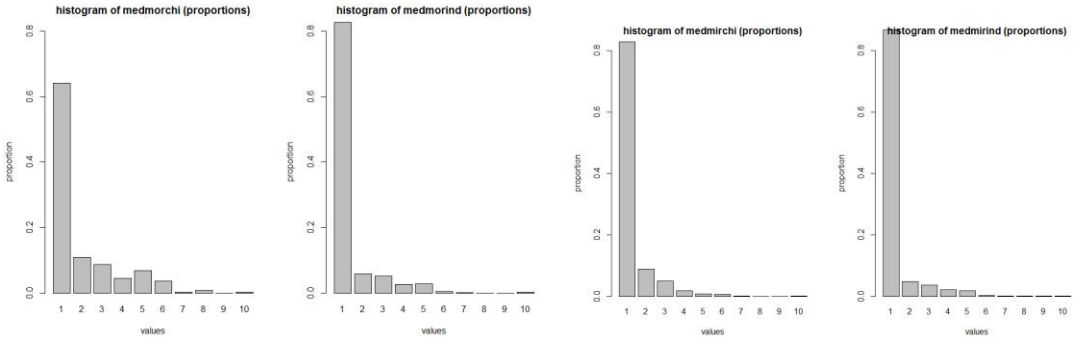
The result of the computation for China, χ^2=1856.2, d.f = 18, p-value < 2.2e-16. Since the p-value is extremely small, we reject the null hypothesis because there are highly significant evidences that morality, crime, and security are not independent variables.

The result of the computation for India, χ^2 =1387.6, df = 18, p-value < 2.2e-16. The former conclusion is also true for India because they have the same p-values. Therefore, we reject the null hypothesis because there are highly significant evidences that morality, crime, and security are not independent variables.

Limitation: the chi-square test of independence is not really suitable for ordinal data because we assume each category is interchangeable to the other, which is definitely not the case. Therefore, another test, which is called Wilcoxon Rank Sum Test, will be conducted to address the issue.

Wilcoxon Rank Sum Test

Histograms are used to identify the distribution of the median of each country. For both countries, the distribution of the median of the two categories are both negatively skewed. The two groups are independent to each other and are negatively skewed, Wilcoxon rank sum test would be more appropriate than using two-sample t-test.



The left ones are the histograms of median of moral values whereas the right ones are regarding median of crime for both China and India.

Wilcoxon rank sum test is a non-parametric test and it does not deal with the parameters in the mathematical formulae of the distributions as the exact distribution of both groups are unknown. Hence, the parameter is unknown.

From the histograms, most of the respondents answered 1 which means never for both categories in both countries.

The setting of the test is as follows:

For moral value,

H_0: $There\ is\ no\ difference\ in\ the\ median\ of\ moral\ value\ between\ China\ and\ India.$
H_1: $There\ is\ difference\ in\ the\ median\ of\ moral\ value\ between\ China\ and\ India.$

For crime,

H_0: $There\ is\ no\ difference\ in\ the\ median\ of\ crime\ between\ China\ and\ India.$
H_1: $There\ is\ difference\ in\ the\ median\ of\ crime\ between\ China\ and\ India.$

Questions are classified into three categories and each category contains China data set and India data set. The number of option "1" to option "10" are counted. Then the t_i (multiplicity) is counted, i.e. the total number of each option which contains China data set and India data set are calculated. $t_i(t_i^2 - 1)$are also calculated. After that, the rank of each option are calculated. Then, the rank sum can be also obtained which is the product of the rank and the number of option. Moreover, W is obtained by adding 10 "Rank Sum" numbers. Then E(W) and Var(W) are also calculated. The formulae are shown as follow.

$$E(W) = \frac{1}{2} n(N+1)$$

$$Var(W) = \frac{1}{12} mn(N+1) - \frac{mn}{12N(N-1)} \sum_{1}^{g} t_j(t_j^2 - 1)$$

Finally, Z - test is adopted, the value of Z is shown below.

$$\frac{W - E(W)}{\sqrt{Var(W)}} \sim N(0, 1) \text{ in large sample.}$$

For both categories, the p-values are smaller than 0.05.[2] That means the null hypotheses are rejected, there are difference in the median of moral value and crime between China and India.

Bayesian Ordered Probit Regression

Here, we want to analyze how moral values, crime, fighting police relate with sex and age. The common method is the ordinal logistic regression. It, however, has some limitation, mainly the proportional odds assumption. But the moral values are relative description and the others are frequency. It is hard to explain the number in a proportional way.

Instead, we use Bayesian approach with follow model.

$Z_i = \beta^T X_i + \epsilon_i$
$Y_i = g(Z_i)$
with some non-decreasing function g and thresholds $\gamma_1, \dots, \gamma_{n-1}$ determine Y.

[2] For the result, please refer to the excel files attached.

We assign prior $N(0,(X^TX)^{-1})$ to β.

The full conditional distribution is (Cowle,1996)

$P(\beta,\gamma,y_i)\sim N(x_i^T\beta,1)$
$P(Y,Z_i)\sim N((X^TX)^{-1}X^TZ,(X^TX)^{-1})$

The following models will be performed.

- Median of moral problem
- Median of minor crime problem → Sex + Age + Sex × Age
- Fighting Police

(for both countries)

Below is one of the examples.

Median of moral problem ~ Sex + Age + Sex × Age for China

We run the Gibbs sampler for 10,000 iterations with 1000 burns-in.

```
1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                  Mean       SD  Naive SE Time-series SE
(Intercept)    0.39550 0.150236 1.502e-03      5.689e-03
sexchi        -0.04226 0.230094 2.301e-03      3.143e-03
agechi        -0.01745 0.003607 3.607e-05      7.319e-05
sexchi:agechi -0.00429 0.005821 5.821e-05      8.560e-05
gamma2         0.32692 0.033454 3.345e-04      4.519e-03
gamma3         0.64552 0.051468 5.147e-04      1.081e-02
gamma4         0.85484 0.060163 6.016e-04      1.411e-02
gamma5         1.35007 0.079006 7.901e-04      2.306e-02
gamma6         1.98498 0.163594 1.636e-03      9.295e-02
gamma7         2.15067 0.192595 1.926e-03      1.037e-01
gamma8         3.08289 0.422246 4.222e-03      4.161e-01

2. Quantiles for each variable:

                  2.5%       25%       50%        75%     97.5%
(Intercept)    0.09858  0.296523  0.394688  0.4971861  0.690449
sexchi        -0.49095 -0.199552 -0.043223  0.1119549  0.417110
agechi        -0.02449 -0.019879 -0.017451 -0.0150485 -0.010370
sexchi:agechi -0.01591 -0.008186 -0.004232 -0.0003251  0.007026
gamma2         0.26442  0.304689  0.324613  0.3475807  0.397357
gamma3         0.54578  0.612641  0.643099  0.6753770  0.753699
gamma4         0.73446  0.815037  0.853259  0.8924541  0.972196
gamma5         1.20901  1.281958  1.357141  1.4162090  1.486493
gamma6         1.75422  1.834742  1.981591  2.1361541  2.257856
gamma7         1.84971  1.977532  2.129123  2.3105094  2.531065
gamma8         2.51784  2.675339  3.054885  3.4124303  3.816554
```

If 95% posterior interval includes 0, we conclude that variable is not significant to explain Y. In this case, only age is significant to explain moral problem of China data. Negative slope indicates older the person, moral sense is better.

Here is the table summarizing all the coefficients in the models.[3] (cross means insignificant.)

	Moral (China)	Moral (India)	Crime (China)	Crime (India)	Police (China)	Police (India)
Sex	×	-ve	×	-ve	×	-ve
Age	-ve	-ve	×	-ve	×	-ve

[3] The table can be obtained in the appendix.

Sex × Age	×	+ve	×	+ve	×	×

From the table, we have some general findings.

- Older the person, the sense of moral, crime, fighting police are better.
- In China, sex and age are not a significant factor to determine those senses.
- In India, female usually have a better sense than male.

Limitation:

- The coefficients are not comparable for each regression.
- The prior may not be accurate for the model.

Correlation: Cramer's V, Goodman and Kruskal's Gamma and Pearson's Correlation

Since the sample size of the two datasets differ, chi-square statistic cannot provide a reliable guide to distinguish the strength of the association underneath. So, the below measurements are adjusted to take account of differences in sample size (Gingrich, 2004). Here we mainly discuss Cramer's V here because it can compare the strength of association between two two-way tables with different dimension. Cramer's V is defined as square root of chi-sq statistics divided by sample size and minimum of row or column number. $V = \sqrt{\frac{\chi^2}{nt}}\ where\ t = Min\ (r-1, c-1)$. Cramer's V represents the strength of relationship by the range from 0 to 1. It generally has a maximum value of 1 when there is a very strong relationship between two variables (Gingrich, 2004).

The second measure we use is Goodman and Kruskal's Gamma which is derived from concordant and discordant pairs. A pair is concordant if the subject ranked higher on X also rank higher on Y, and the discordant pairs is the subject ranking higher on X and lower on Y (Agrest,2013). Gamma is defined as the proportion of the difference between the number of concordant pairs and discordant pairs and the summation of the number of concordant pairs and discordant pairs. $Gamma = (\Pi_c - \Pi_d)/(\Pi_c + \Pi_d)$ The sample version of gamma is (C-D)/(C+D) (Agrest,2013). Gamma is ranging from -1 to 1, it means that of the untied pairs, the proportion of concordant pairs is the value of gamma times higher than the proportion of discordant pairs and the extent of the tendency of y when x increases. The test statistic of association hypothesis test is $z = G\sqrt{(n_c + n_d)/[N(1-G^2)]}. Reject\ H_0 if\ |z| < 1.96.$

The third measure we used is Pearson correlation coefficient. The coefficient describes the strength of a linear trend in the population for interval data which is also available for the 10-point Likert scale response here. We will find the sample correlation $r = \frac{\sum_{i,j} (u_i - \underline{u})(v_j - \underline{v})p_{ij}}{\sqrt{[\sum_{i,j} (u_i - \underline{u})^2 p_{i+}][\sum (v_j - \underline{v})^2 p_{+j}}}$ and conduct a hypothesis test for $\rho = 0$.

Independence between X and Y $\rho = 0$. The larger $|\rho|$ is, the farther the data fall from independence in the linear dimension.

Independence between X and Y $\rho = 0$. The larger $|\rho|$ is, the farther the data fall from independence in the linear dimension.

Having a glance on the sample data, all the testing relationship show weak tendency between the variables. And for Item 1 to 3, the absolute values of Cramer V revealed that the relationships in China are stronger than India do.

Investigating the population independence of data, all the data in India except the correlation test of threatening and satisfaction are consistent not to reject the null hypothesis. Independence is then concluded to the relationships. Yet, all the null hypothesis in the China dataset are rejected to show there is association between the testing relationship.

The analysis shows that moral is independent of satisfaction in India, there is a weak tendency that higher satisfaction, less moral behavior problem in China. The investigation in correlation between crime and satisfaction shows a similar result that it is independent of satisfaction in India, and a weak correlation in China. Our final tests shows that sex is independent of homosexuality.

Overall, the results obtained in Cramer's V, gamma and correlation are almost consistent.

Threatening and Satisfaction	**Cramer V's**	**Gamma**	**P- value for gamma Reject H_0?**
China	0.098	0.098	0.004461026 Rejected
India	-0.199	-0.199	0.08767119 Not rejected
	Sample Correlation	**95% C.I. of Correlation**	
China	-0.1291621	-0.19012271	-0.06720934
India	-0.05290921	-0.08913102	-0.01014099

Limitation: Gamma may overestimate the strength of association if there are too much tied pairs which is the rankings are same on both variables (Odek, 2009). The problem is not serious in our dataset, gamma is also acceptable in our measure. Pearson's correlation coefficient is easier to be affected by the outliers (Abdullah,1990). It is a potential reason that the results of hypothesis test differ between gamma and correlation coefficient.

Moral Behavior and Satisfaction	**Cramer V's**	**Gamma**	**P- value for gamma Reject H_0?**
China	0.164	-0.316	1.866011e-06 Rejected
India	0.073	0.002	0.5143379 Not reject
	Sample Correlation	**95% C.I. of Correlation**	
China	-0.2506032	-0.3082705	-0.1911008
India	-0.01042733	-0.04999940	0.02917743

Fighting with Police and Satisfaction	**Cramer V's**	**Gamma**	**P- value for gamma Reject H_0?**
China	0.132	0.132	0.0004211362 Reject
India	-0.192	-0.192	0.1605667 Not Reject
	Sample Correlation	**95% C.I. of Correlation**	
China	-0.1870044	-0.2466175	-0.1259815
India	-0.03964255	-7.911110e-02	0.001903292

Crime and Satisfaction	**Cramer V's**	**Gamma**	**P- value for gamma Reject H_0?**
China	0.118	-0.223	0.003922538 Rejected
India	0.082	-0.036	0.2659388 Not reject
	Sample Correlation	**95% C.I. of Correlation**	
China	-0.1054489	-0.16684462	-0.04323852
India	-0.01166723	-0.05123628	0.02793839

Sex and Homosexuality	**Cramer V's**	**Gamma**	**P- value for gamma Reject H_0?**
	0.059	-0.086	0.1715651 Not reject
	Sample Correlation	**95% C.I. of Correlation**	
	-0.023	-0.05662490	0.01022617

Conclusion

- Chi-squared Test of independence & Rank Sum test:

The data set variables in China are not independent variables

The data set variables in Inida are not independent variables

There is difference in the median of China's moral value between India's moral value

There is difference in the median of China's crime between India's crime

- Bayesian ordered probit regression

Older the person, the sense of moral, crime, fighting police are better.

In China, sex and age are not a significant factor to determine those senses.

In India, female usually have a better sense than male.

- Goodman and Kruskal's Gamma & Pearson's Correlation

The analysis shows that moral is independent of satisfaction in India, there is a weak tendency that higher satisfaction, less moral behavior problem in China.

The investigation in correlation between crime and satisfaction shows a similar result that it is independent of satisfaction in India, and a weak correlation in China.

Our final test shows that sex is independent of homosexuality.

Overall, the results obtained in Cramer's V and correlation are consistent.

Reference

1. Abdullah, M. B. (1990). On a robust correlation coefficient. The Statistician, 455-460.

2. Acock, A., & Stavig, G. (1979). A measure of association for nonparametric statistics. Social Forces, 57(4), 1381-1386.

3. Agresti, A., ProQuest, & ProQuest CSA. (2013). Categorical data analysis (3rd ed., Wiley series in probability and statistics ; 792). Hoboken, NJ: Wiley.

4. Andrew D. Martin, Kevin M. Quinn, Jong Hee Park (2011). MCMCpack: Markov Chain Monte Carlo in R. Journal of Statistical Software. 42(9): 1-21. URL http://www.jstatsoft.org/v42/i09/.

5. Falissard, B. (2012). Analysis of questionnaire data with R. Boca Raton, FL: CRC Press.

6. Gingrich, P. (n.d.). Social Studies 201 Text Introductory Statistics for the Social Sciences. Retrieved April 15, 2018, from http://uregina.ca/~gingrich/text.htm

7. M. K. Cowles. (1996.) "Accelerating Monte Carlo Markov Chain Convergence for Cumulative-link Generalized Linear Models." Statistics and Computing. 6: 101-110.

8. Presnell, B. (2000, March 28). An Introduction to Categorical Data Analysis Using R. Retrieved April 15, 2018, from http://users.stat.ufl.edu/~presnell/Courses/sta4504-2000sp/R/R-CDA.pdf

9. R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

10. Stef van Buuren, Karin Groothuis-Oudshoorn (2011). mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3), 1-67. URL http://www.jstatsoft.org/v45/i03/

11. Odek, W. (2009). SPSS for Social Scientists. Sociological Research Online, 14(4), 1-2.

Appendix

The result of Bayesian regression

```
> resultlind<-MCMCoprobit(medmorind~sexind+ageind+sexind*ageind)
> summary(resultlind)

Iterations = 1001:11000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                  Mean       SD  Naive SE Time-series SE
(Intercept)   -0.197772 0.248548 2.485e-03      4.365e-03
sexind        -0.460583 0.167654 1.677e-03      2.957e-03
ageind        -0.015818 0.006582 6.582e-05      1.173e-04
sexind:ageind  0.009537 0.004408 4.408e-05      7.728e-05
gamma2         0.261760 0.018846 1.885e-04      1.457e-03
gamma3         0.595794 0.032133 3.213e-04      3.839e-03
gamma4         0.869983 0.040080 4.008e-04      5.801e-03
gamma5         1.562842 0.074388 7.439e-04      1.906e-02
gamma6         2.134728 0.215771 2.158e-03      1.334e-01
gamma7         2.414239 0.224987 2.250e-03      1.464e-01

2. Quantiles for each variable:

                  2.5%       25%       50%       75%     97.5%
(Intercept)   -0.680932 -0.366026 -0.198436 -0.03103  0.287162
sexind        -0.793687 -0.573711 -0.459150 -0.34865 -0.134102
ageind        -0.028754 -0.020269 -0.015764 -0.01136 -0.002952
sexind:ageind  0.001004  0.006579  0.009481  0.01249  0.018187
gamma2         0.224870  0.249322  0.261713  0.27435  0.299081
gamma3         0.534036  0.573641  0.595123  0.61779  0.660259
gamma4         0.784639  0.844183  0.870255  0.89567  0.949948
gamma5         1.438191  1.510776  1.552491  1.61137  1.724217
gamma6         1.727036  1.982565  2.099382  2.30013  2.533198
gamma7         1.937651  2.348969  2.470359  2.56152  2.745589
```

```
> summary(result2chi)

Iterations = 1001:11000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                  Mean       SD  Naive SE Time-series SE
(Intercept)   -0.202023 0.405758 4.058e-03      0.0068589
sexchi        -0.226326 0.279103 2.791e-03      0.0047091
agechi        -0.015185 0.010213 1.021e-04      0.0001857
sexchi:agechi  0.002993 0.007051 7.051e-05      0.0001325
gamma2         0.444458 0.041668 4.167e-04      0.0059843
gamma3         0.912829 0.063906 6.391e-04      0.0137259
gamma4         1.260384 0.075739 7.574e-04      0.0191769
gamma5         1.615887 0.082969 8.297e-04      0.0226494
gamma6         2.528603 0.165792 1.658e-03      0.0773797
gamma7         2.911692 0.122225 1.222e-03      0.0491535

2. Quantiles for each variable:

                  2.5%       25%       50%       75%  97.5%
(Intercept)   -0.98532 -0.475601 -0.201735  0.071853 0.5945
sexchi        -0.77843 -0.412352 -0.228662 -0.037405 0.3158
agechi        -0.03528 -0.022040 -0.015286 -0.008307 0.0045
sexchi:agechi -0.01055 -0.001805  0.003031  0.007667 0.0168
gamma2         0.36259  0.413650  0.445118  0.474485 0.5269
gamma3         0.78898  0.868403  0.918863  0.959321 1.0343
gamma4         1.09035  1.215411  1.263513  1.312847 1.3939
gamma5         1.44882  1.564260  1.619329  1.668024 1.7703
gamma6         2.21192  2.400571  2.539480  2.667395 2.7910
gamma7         2.69514  2.823041  2.887798  3.012266 3.1336
```

```
> summary(result3chi)

Iterations = 1001:11000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                  Mean       SD  Naive SE Time-series SE
(Intercept)    0.536296 0.312867 3.129e-03      3.648e-03
sexchi         0.104827 0.211204 2.112e-03      2.467e-03
agechi        -0.003701 0.007599 7.599e-05      9.033e-05
sexchi:agechi -0.006478 0.005202 5.202e-05      6.278e-05
gamma2         0.292115 0.026261 2.626e-04      3.379e-03
gamma3         0.581999 0.035191 3.519e-04      6.630e-03
gamma4         0.740186 0.037901 3.790e-04      6.948e-03
gamma5         1.172060 0.045195 4.519e-04      8.985e-03
gamma6         1.755590 0.062443 6.244e-04      1.710e-02
gamma7         1.961782 0.070852 7.085e-04      2.187e-02
gamma8         2.275820 0.077251 7.725e-04      2.672e-02
gamma9         2.483335 0.098625 9.862e-04      4.352e-02

2. Quantiles for each variable:

                  2.5%       25%       50%       75%     97.5%
(Intercept)   -0.07050  0.329683  0.530244  0.746912 1.150775
sexchi        -0.30465 -0.038620  0.105576  0.247774 0.513179
agechi        -0.01853 -0.008790 -0.003633  0.001404 0.011316
sexchi:agechi -0.01677 -0.009984 -0.006524 -0.002935 0.003575
gamma2         0.24368  0.273122  0.291484  0.309588 0.348271
gamma3         0.51773  0.555358  0.583315  0.606306 0.650223
gamma4         0.67184  0.712188  0.738361  0.766425 0.813969
gamma5         1.09366  1.138281  1.169963  1.203321 1.260992
gamma6         1.65110  1.709634  1.749690  1.799541 1.885097
gamma7         1.84398  1.908545  1.952270  2.011683 2.101044
gamma8         2.13575  2.212794  2.277847  2.336389 2.407998
gamma9         2.29541  2.397902  2.500241  2.563975 2.642472
```

```
> summary(result2ind)

Iterations = 1001:11000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                  Mean       SD  Naive SE Time-series SE
(Intercept)   -0.10256 0.246989 2.470e-03      0.0043501
sexind        -0.58367 0.165824 1.658e-03      0.0029505
ageind        -0.02024 0.006586 6.586e-05      0.0001190
sexind:ageind  0.01323 0.004367 4.367e-05      0.0000773
gamma2         0.32081 0.021036 2.104e-04      0.0019890
gamma3         0.59354 0.029661 2.966e-04      0.0037992
gamma4         0.92101 0.045038 4.504e-04      0.0088018
gamma5         1.46640 0.102401 1.024e-03      0.0410594
gamma6         1.77894 0.156367 1.564e-03      0.0864554
gamma7         2.06003 0.196791 1.968e-03      0.1398967
gamma8         2.19300 0.221043 2.210e-03      0.1648015
gamma9         2.30906 0.226504 2.265e-03      0.1506097

2. Quantiles for each variable:

                  2.5%       25%       50%       75%      97.5%
(Intercept)   -0.581596 -0.27077 -0.10420  0.06409  0.386499
sexind        -0.906803 -0.69706 -0.57998 -0.47129 -0.258600
ageind        -0.033254 -0.02468 -0.02012 -0.01578 -0.007384
sexind:ageind  0.004696  0.01026  0.01319  0.01624  0.021778
gamma2         0.281250  0.30592  0.32067  0.33577  0.361582
gamma3         0.532062  0.57555  0.59470  0.61246  0.653937
gamma4         0.842916  0.88923  0.91831  0.94878  1.024219
gamma5         1.314485  1.40048  1.44521  1.50656  1.715369
gamma6         1.559295  1.66784  1.74175  1.86160  2.113665
gamma7         1.792638  1.88892  1.99529  2.24086  2.421902
gamma8         1.904727  1.97109  2.19153  2.36258  2.570075
gamma9         1.972339  2.07602  2.34279  2.50381  2.662345
```

```
> result3ind<-MCMCoprobit(polind~sexind+ageind+sexind*ageind)
> summary(result3ind)

Iterations = 1001:11000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                  Mean       SD  Naive SE Time-series SE
(Intercept)    0.642895 0.200986 2.010e-03      2.661e-03
sexind        -0.361470 0.134671 1.347e-03      1.773e-03
ageind        -0.01426  0.005349 5.349e-05      7.241e-05
sexind:ageind  0.003213 0.003616 3.616e-05      4.931e-05
gamma2         0.169228 0.012955 1.296e-04      9.414e-04
gamma3         0.434771 0.021090 2.109e-04      2.321e-03
gamma4         0.660328 0.024289 2.429e-04      2.960e-03
gamma5         1.276059 0.031389 3.139e-04      4.848e-03
gamma6         1.497643 0.033455 3.346e-04      5.852e-03
gamma7         1.639079 0.036889 3.689e-04      7.911e-03
gamma8         1.806822 0.045237 4.524e-04      9.647e-03
gamma9         1.978525 0.055866 5.587e-04      1.525e-02

2. Quantiles for each variable:

                  2.5%       25%       50%       75%     97.5%
(Intercept)    0.257832  0.5026193  0.642194  0.779211  1.038840
sexind        -0.625809 -0.4514298 -0.361447 -0.268851 -0.101632
ageind        -0.024926 -0.0180715 -0.014450 -0.010757 -0.004158
sexind:ageind -0.003819  0.0007941  0.003243  0.005627  0.010397
gamma2         0.143564  0.1605015  0.169213  0.177576  0.195851
gamma3         0.393093  0.4212685  0.435486  0.448430  0.476915
gamma4         0.610689  0.6443903  0.662971  0.676860  0.703337
gamma5         1.206200  1.2579499  1.278542  1.297214  1.331648
gamma6         1.414373  1.4802074  1.502621  1.520891  1.550821
gamma7         1.552265  1.6198958  1.643022  1.663933  1.701722
gamma8         1.708547  1.7772097  1.809497  1.842345  1.883539
gamma9         1.865970  1.9374732  1.980797  2.020509  2.077203
```

The graphs represent (from left to right, up to down) moral values(India), crime(China and India), police(China and India)