Generalized Linear Models Project
Statistical Methods Reveal What Makes University of Chicago MBA Students Happy

JUN XU
Boston University
February 12, 2015

1 Introduction

The data for our analysis were collected from 39 students in a University of Chicago MBA class. Five variables were recorded: *happy*, *money*, *sex*, *love* and *work*. *Happy* was measured on a 10-point scale, with 1 representing a suicidal state, 5 representing a feeling of "just muddling along", and 10 representing a euphoric state. *Money* was measured by annual family income in thousands of dollars. *Sex* was measured by a dummy variable taking the values 0 or 1, with 1 indicating a satisfactory level of sexual activity. *Love* was measured on a 3-point scale, with 1 representing loneliness and isolation, 2 representing a set of secure relationships, and 3 representing a deep feeling of belonging and caring in the context of some family or community. *Work* was measured on a 5-point scale, with 1 indicating that an individual is seeking other employment, 3 indicating the job is okay and 5 indicating that the job is enjoyable.

There are two reasons why I choose this dataset. For one thing, being happy is always good and everybody pursues happiness. With this dataset, we are able to reveal which factors have an influence on happiness. In particular, we are able to answer questions such as can money buy happiness, which has sparked a heated debate for a long time. Furthermore, Abraham Maslow pointed out a long time ago in his hierarchy of needs, that when a person is physically comfortable and no longer driven by fear of starving or freezing to death, emotional pleasure becomes a primary pursuit. Hence we cannot help but wonder is that the case for these MBA students? For the other, I prefer dataset with shorter name in the sense that I will type less while coding.

2 Methods

We first construct a scatterplot to check the relationship between the response and all predictors as shown in Figure 1. In particular, we plot boxplots of the *happy* by *sex*, *love* and *work* to see how each predictor affect the mean or variance of the response. We see clearly that *money*, *love* and *work* are linearly related to *happy* in the scatterplot. Those with more family income are more likely to be happy among the 39 University of Chicago MBA students. Likely, those with better love status, or greater job also tend to have higher happiness index. However, it remains uncertain if *sex* is related to the response because the mean happiness indices in both *sex* group are almost the same. Meanwhile, we think that *money*, *sex* and *love* have an effect on the variance of the response. These two observations are confirmed by the respective boxplots and imply that we may construct a Joint Mean and Variance Model to capture the mean and variance at the same time. Also it seems unnecessary to transform any variable.

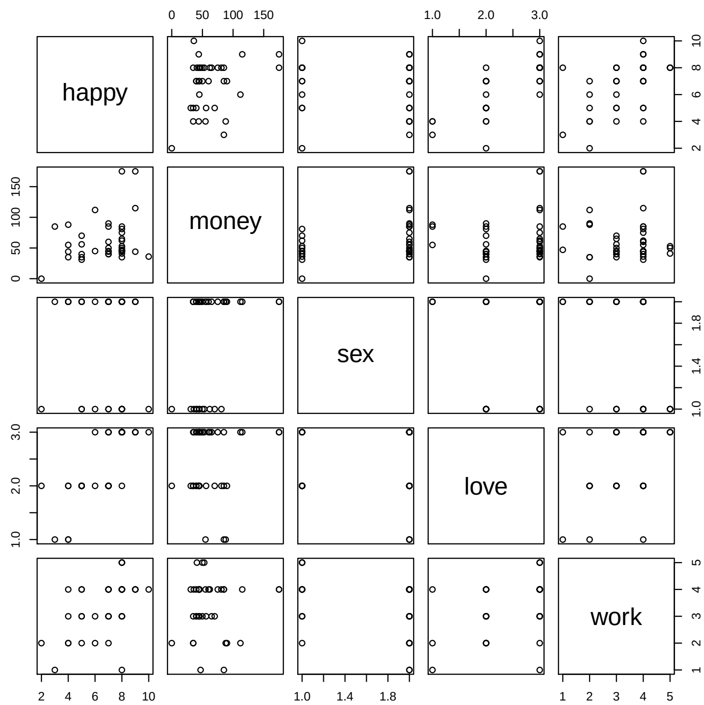
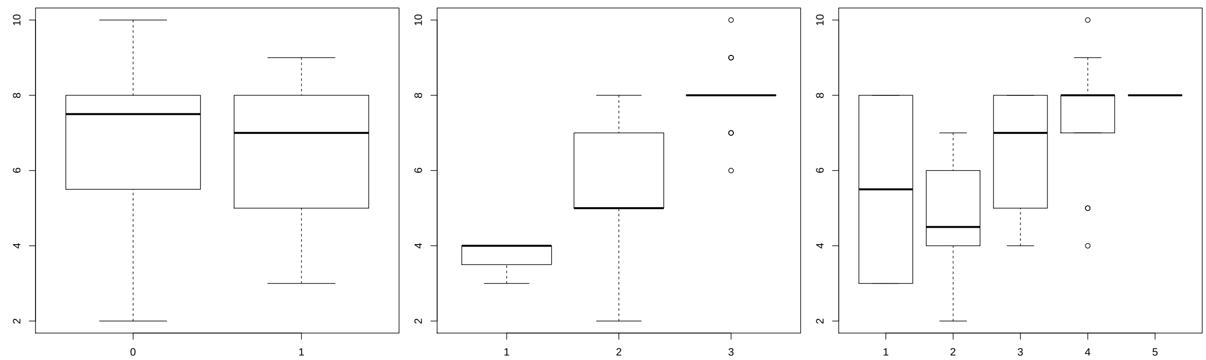


Figure 1: Scatterplot Matrix of Response and Predictors

Second, we construct ordinary linear models to select variables by AIC in a stepwise algorithm. It shows that *money*, *love* and *work* should be included in the model, while *sex* should be discarded. This agrees with our observations at first glance. Hence, we constructed three models including a Polynomial Model, a Quasi-Gaussian Model and a Joint Mean and Variance Model using these selected predictors.

If we treat *happy* as an ordinal variable, which is supposed to be, then a Polynomial Model is appropriate. This is based on the assumption that we have proportional odds. That is, we define the cumulative probabilities $\gamma_j = \sum_{i=1}^{j} \pi_i$, where π_i denotes the probability that one's happiness



(a) happy vs sex (b) happy vs love (c) happy vs work

Figure 2: Boxplots of Response by Each Discrete Predictor

index is i. By that,

$$\begin{aligned} P(y \leq j) = \gamma_j &= P(Z \leq \theta_j) \\ &= P(Z - X\beta \leq \theta_j - X\beta) \\ &= \text{logit}^{-1}(\theta_j - X\beta), \end{aligned}$$

where y denotes the happiness index and Z is a latent variable such that $Z - X\beta \sim \textit{logistic}(0,1)$. And then

$$\begin{aligned} \frac{\gamma_j(x_1)/(1-\gamma_j(x_1))}{\gamma_j(x_2)/(1-\gamma_j(x_2))} &= \frac{e^{(\theta_j - x_1^\top \beta)}}{e^{(\theta_j - x_2^\top \beta)}} \\ &= e^{(x_2 - x_1)^\top \beta}, \end{aligned}$$

which is unrelated to θ_j.

However, if we assume that *happy* is a continuous variable, then we can run a Quasi-Gaussian Model or a Joint Mean and Variance Model. This assumption is reasonable because given such a 10 point scale with 1 stands for the least happy and 10 the most, it is natural for the surveyed to think of the ten scores as a decile. In other words, one may naturally assume that the difference between the consecutive happiness scores are constant when he or she is investigated. For the former, we are assuming that all we know is the Quasi-Score, defined as

$$U = \frac{y - \mu}{\sigma^2 V(\mu)},$$

where y denotes the response, μ is the mean of the response, σ^2 is the dispersion and $V(\mu)$ is a function of μ. From that, we can derive

$$(D^\top V^{-1} D)\beta^{(t+1)} = D^\top V^{-1}(D\beta^{(t)} + y - \mu(\beta^{(t)}))$$

to estimate the coefficients β iteratively, where $D = \partial\mu/\partial\beta$. Without doubt, the estimated coefficients of the Ordinary Linear Model and the Quasi-Gaussian Model are the same.

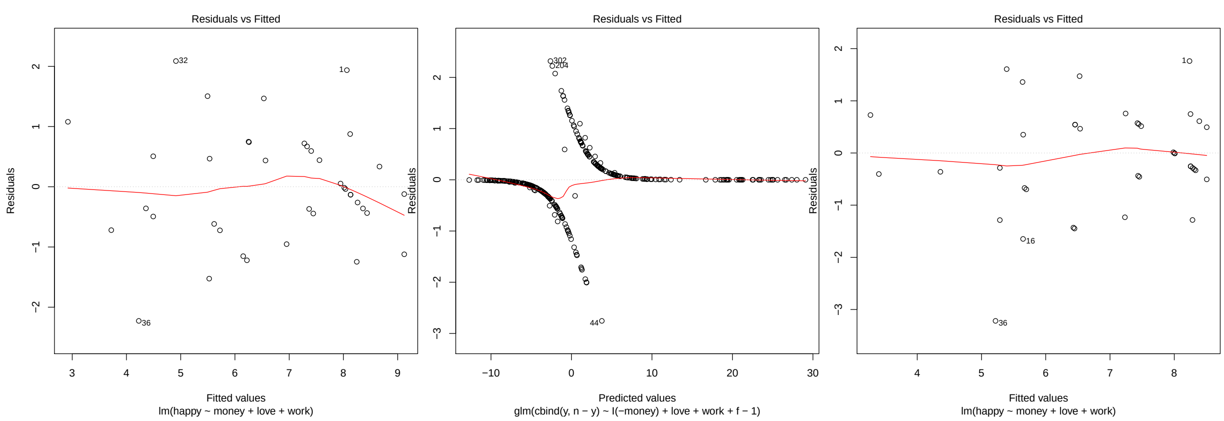
For the latter, we are assuming

$$E[y_i] = \mu_i,\ \eta_i = x_i^\top \beta = g(\mu_i),\ Var(y_i) = \phi_i V(\mu_i),$$

$$E[d_i] = \phi_i,\ \xi_i = u_i^\top \gamma = h(\phi_i),\ Var(d_i) = \tau^2 V_D(\phi_i),$$

where y_i denotes the i'th response, μ_i is the mean of the i'th response, ϕ_i is the i'th dispersion, $V(\mu_i)$ is a function of μ_i, d_i is some measure of dispersion and often equals the square of i'th pearson residual, τ^2 is the dispersion for d and $V_D(\phi_i)$ is a function of ϕ_i. We end up with iteration between two steps: (1) fit a GLM $y \sim X$ with weights $\hat{\phi}_i^{-1}(t)$; and (2) fit a Gamma GLM $d \sim U$.

We compare the results from all the models, with respect to residual against fitted plots in particular. Figure 3 shows the residual against fitted plots of the four models, respectively. Notice here we transform the response into cumulative probabilities in order to draw the plot as shown in Figure 3.b. Figure 3.a and 3.c are slightly alike, since the main difference between the two models are the weights. Actually we see here the Ordinary Linear Model fits the data well already. However, the trend line resulting from the LOWESS smoother in Figure 3.c turns out to be closer to the x-axis, compared to the other two. What is more, except observation 36, a possible outlier, the points in Figure 3 are more uniformly distributed in both direction. Given these, we think the Joint Mean and Variance Model is better.



(a) Linear/Quasi-Gaussian Model (b) Multinomial Model (c) Joint Mean and Variance Model

Figure 3: Residuals vs Fitted Plots of Four Models

3 Results

3.1 Estimates

We regress *happy* on *money*, *love* and *work* and report the results from the Joint Mean and Variance Model.

- The estimated *intercept* is reported as **3.239**. It means that for a University of Chicago MBA student, if he or she doesn't have a job or an income and feel lonely, then his or her mean happiness index is 3.239.
- The estimated coefficient of *money* is reported as **0.002**. It means that $1,000 increase in family income leads to 0.002 unit increase in the mean happiness index for the 39 students in a University of Chicago MBA class, controlling for all other covariates.
- The estimated coefficient of *love2* is reported as **2.116**. It means that the mean happiness index of the University of Chicago MBA students who are in secure relationships is 2.116 higher than that of the lonely ones, controlling for all other covariates.
- The estimated coefficient of *love3* is reported as **3.913**. It means that the mean happiness index of the University of Chicago MBA students who have deep feeling of belonging and caring is 3.913 higher than that of the lonely ones, controlling for all other covariates.
- The estimated coefficient of *work2* is reported as $-\mathbf{0.136}$. It means that the mean happiness index of the University of Chicago MBA students whose work index is 2 is 0.136 lower than that of the unemployed ones, controlling for all other covariates.
- The estimated coefficient of *work3* is reported as **0.207**. It means that the mean happiness index of the University of Chicago MBA students whose have okay jobs is 0.207 higher than that of the unemployed ones, controlling for all other covariates.
- The estimated coefficient of *work4* is reported as **1.016**. It means that the mean happiness index of the University of Chicago MBA students whose work index is 4 is 1.016 higher than that of the unemployed ones, controlling for all other covariates.
- The estimated coefficient of *work5* is reported as **0.755**. It means that the mean happiness index of the University of Chicago MBA students whose have great jobs is 0.755 lower than that of the unemployed ones, controlling for all other covariates.

Table 1: Analysis of Happiness using the Joint Mean and Variance Model

	Estimate	Std.Error	95% C.I.	t value	Pr(>|t|)	
Intercept	3.239	0.557	(2.103, 4.374)	5.817	2.07e-6	$***$
money	0.002	0.004	(-0.006, 0.010)	0.502	0.6190	
love2	2.116	0.529	(1.037, 3.195)	3.999	0.0004	$***$
love3	3.913	0.449	(2.997, 4.830)	8.713	7.73e-10	$***$
work2	-0.136	0.617	(-1.395, 1.123)	-0.221	0.8267	
work3	0.207	0.636	(-1.089, 1.504)	0.326	0.7464	
work4	1.016	0.562	(-0.130, 2.162)	1.808	0.0803	
work5	0.755	0.743	(-0.760, 2.270)	1.017	0.3172	

3.2 Confidence Intervals

- The 95% confidence interval is computed as $(\mathbf{2.103}, \mathbf{4.374})$, meaning we are 95% confident that the *intercept* lies between 2.103 and 4.374.
- The 95% confidence interval of the coefficient for *money* is computed as $(-\mathbf{0.006}, \mathbf{0.010})$, meaning that we are 95% confident the coefficient lies between -0.006 and 0.010.

- The 95% confidence interval of the coefficient for *love2* is computed as $(\mathbf{1.037}, \mathbf{3.195})$, meaning that we are 95% confident the coefficient lies between 1.037 and 3.195.
- The 95% confidence interval of the coefficient for *love3* is computed as $(\mathbf{2.997}, \mathbf{4.830})$, meaning that we are 95% confident the coefficient lies between 2.997 and 4.830.
- The 95% confidence interval of the coefficient for *work2* is computed as $(-\mathbf{1.395}, \mathbf{1.123})$, meaning that we are 95% confident the coefficient lies between -1.395 and 1.123.
- The 95% confidence interval of the coefficient for *work3* is computed as $(-\mathbf{1.089}, \mathbf{1.504})$, meaning that we are 95% confident the coefficient lies between -1.089 and 1.504.
- The 95% confidence interval of the coefficient for *work4* is computed as $(-\mathbf{0.130}, \mathbf{2.162})$, meaning that we are 95% confident the coefficient lies between -0.130 and 2.162.
- The 95% confidence interval of the coefficient for *work5* is computed as $(-\mathbf{0.760}, \mathbf{2.270})$, meaning that we are 95% confident the coefficient lies between -0.760 and 2.270.

3.3 Tests

- Dividing the estimated coefficient of *money* 0.002 by its standard error 0.004 obtains the t-value $\mathbf{0.502}$. The p-value is computed as 0.6190. Thus we do not reject the null hypothesis; there is no significant evidence of correlation between happiness and the family income among the 39 University of Chicago MBA students, controlling for all other covariates.
- Dividing the estimated coefficient of *love2* 2.116 by its standard error 0.529 obtains the t-value $\mathbf{3.999}$. The p-value is computed as 0.0004. Thus we reject the null hypothesis; there is a significant difference in happiness between the University of Chicago MBA students who are in secure relationships and the lonely ones, controlling for all other covariates.
- Dividing the estimated coefficient of *love3* 3.913 by its standard error 0.449 obtains the t-value $\mathbf{8.713}$. The p-value is computed as $7.73e-10$. Thus we reject the null hypothesis; there is a significant difference in happiness between the University of Chicago MBA students who have deep feeling of belonging and caring and the lonely ones, controlling for all other covariates.
- Dividing the estimated coefficient of *work2* -0.136 by its standard error 0.617 obtains the t-value $-\mathbf{0.221}$. The p-value is computed as 0.8267. Thus we do not reject the null hypothesis; there is no significant difference in happiness between the University of Chicago MBA students whose work index is 2 and the unemployed ones, controlling for all other covariates.
- Dividing the estimated coefficient of *work3* 0.207 by its standard error 0.636 obtains the t-value $\mathbf{0.326}$. The p-value is computed as 0.7464. Thus we do not reject the null hypothesis; there is no significant difference in happiness between the University of Chicago MBA students who have okay jobs and the unemployed ones, controlling for all other covariates.
- Dividing the estimated coefficient of *work4* 1.016 by its standard error 0.562 obtains the t-value $\mathbf{1.808}$. The p-value is computed as 0.0803. Thus we do not reject the null hypothesis; there is no significant difference in happiness between the University of Chicago MBA students whose work index is 4 and the unemployed ones, controlling for all other covariates.
- Dividing the estimated coefficient of *work5* 0.755 by its standard error 0.743 obtains the t-value $\mathbf{1.017}$. The p-value is computed as 0.3172. Thus we do not reject the null hypothesis; there is

no significant difference in happiness between the University of Chicago MBA students who have great jobs and the unemployed ones, controlling for all other covariates.

3.4 Comparison of Estimates and Predictions

Although we report the results from the Joint Mean and Variance Model only, we do compute the estimated coefficients of all models as listed below. Notice here the response of the Multinomial Model is on a log scale, different from the others, we therefore do not directly compare the estimates of the Multinomial Model with the others. As mentioned before, the estimated coefficients of the Linear Model is exactly the same as those of the Quasi-Gaussian Model. However, due to the fact that we have captured the variance of the response using the Quasi-Gaussian Model, the standard errors become smaller, leading to narrower confidence intervals. As to the Joint Mean and Variance Model, the estimated coefficients for *love3* and *work5* are nearly the same to those of the first and third models. The estimated coefficients for the rest predictors slightly deviate. Anyway, the standard errors are much smaller. So the 95% confidence intervals are the narrowest among the three models. In particular, smaller standard errors means that we are more likely to detect significant predictors.

Table 2: Point and Interval Estimates from Four Models

Models	money	love2	love3	work2	work3	work4	work5
Linear	0.008	1.976	3.849	-0.822	0.141	0.867	0.713
	(-0.003, 0.018)	(0.463, 3.489)	(2.386, 5.311)	(-2.684, 1.040)	(-1.679, 1.961)	(-0.862, 2.596)	(-1.384, 2.809)
Multi-	0.017	3.729	7.614	-1.352	0.172	1.928	1.658
nomial	(-0.004, 0.039)	(0.849, 7.298)	(4.308, 11.639)	(-4.869, 1.913)	(-3.091, 3.413)	(-1.195, 5.099)	(-2.164, 5.628)
Quasi-	0.008	1.976	3.849	-0.822	0.141	0.867	0.713
Gaussian	(-0.003, 0.018)	(0.522, 3.430)	(2.443, 5.254)	(-2.612, 0.968)	(-1.608, 1.891)	(-0.795, 2.529)	(-1.302, 2.727)
Joint Mean	0.002	2.116	3.913	-0.136	0.207	1.016	0.755
& Variance	(-0.006, 0.010)	(1.037, 3.195)	(2.998, 4.830)	(-1.395, 1.123)	(-1.089, 1.504)	(-0.130, 2.162)	(-0.760, 2.270)

Consider that there are two University of Chicago MBA students Adam and Robert. Assume that Adam has no job (i.e., *work*=1) and his family income is only $40,000. But he is in a secure relationship and satisfied with sexual activity. While Robert has a great job (i.e., *work*=5) and his family income is up to $150,000. But he is not satisfied with sexual activity and feel lonely often. Guess who is happier?

- Both the Linear Model and the Quasi-Gaussian Model indicate that Adam's expected happiness index is $3.073 + 0.008 \times 40 + 1.976 = \mathbf{5.354}$. While Robert's expected happiness index is $3.073 + 0.008 \times 150 + 0.713 = \mathbf{4.927}$.
- The Multinomial Model indicates that the probabilities that Adam's happiness index is 2 through 10 are 0.013, 0.017, 0.238, 0.442, 0.104, 0.164, 0.022, 0.000 and 0.000, respectively. Hence Adam's expected happiness index is **5.189**. While the probabilities that Robert's happiness index is 2 through 10 are 0.016, 0.022, 0.281, 0.439, 0.091, 0.133, 0.017, 0.000 and 0.000, respectively. And Robert's expected happiness index is **5.037**.

- The Joint Mean and Variance Model indicates that Adam's expected happiness index is $3.239 + 0.002 \times 40 + 2.116 = \mathbf{5.432}$. While Robert's expected happiness index is $3.239 + 0.002 \times 150 + 0.755 = \mathbf{4.283}$.

In closing, regardless of the model we build, Adam appears happier than Robert. It seems that an excellent job does not offer Robert too much in terms of happiness. However, this is after all a special case. And we are not comparing materials such as money with emotional pleasure here in a general sense.

3.5 Interaction

Up to now, our analysis suggests that money do not buy happiness, at least for these 39 University of Chicago MBA students. So is that true? Is it possible that some of them become happier when they are wealthier? Put it another way, is there an interaction between two of the predictors? We, therefore construct a model with an interaction item, for example, $money * love$ and compare it with the one without this interaction. To be explicit,

$$M_1 : happy \sim money + love + work + money * love, \text{ and}$$
$$M_2 : happy \sim money + love + work.$$

Here we use Quasi-Gaussian Models instead of Joint Mean and Variance Models. Because on one hand, it is comparatively simple to construct a Quasi-Gaussian Model in R. On the other, the alternative Polynomial Model will generate an error when we focus on the subset who have deep feeling of belonging and caring. After all, the results from these models are not so different as discussed above. Besides, we still focus on the selected predictors *money*, *love* and *work*. The hypotheses are

H_0: the estimated coefficient of the interaction between *money* and *love* is equal to zero;
H_A: the estimated coefficient of the interaction between *money* and *love* is not equal to zero.

The deviance is reported as **11.375** with degree of freedom 2. The p-value is computed as 0.0007. We therefore reject the null hypothesis; the estimated coefficient of the interaction between *money* and *love* is not zero. It means that for people in different love status groups, money, or family income has a different weights in terms of happiness.

Afterwards, we run a Quasi-Gaussian Model in each love status stratification. To be clear, we run the model $happy \sim money + work$ in each stratification. It turns out that *money* is significantly related to *happy* (p-value=0.072) only among those who have deep feeling of belonging and caring, i.e., *love*=3. The estimated coefficient is reported as **0.001**, meaning that $1,000 increase in family income leads to 0.001 increase in the happiness index for those who have deep feeling of belonging and caring. But it is not the case for those who are lonely or in secure relationships (p-value=NA and 0.8887, respectively).

Furthermore, neither $money * work$ nor $love * work$ is significantly related to *happy* (p-value=0.2262 and 0.4812, respectively) when we follow the same procedure as above.

4 Conclusion

Adjusting for all other covariates, love status is significantly related to the happiness index among the 39 University of Chicago MBA students. The mean happiness index of those who are in secure relationships and those who have deep feeling of belonging and caring is 2.116 and 3.913 higher than that of the lonely ones, respectively. While neither family income nor sexual activity is significant related to the happiness index for them. In addition, those who get whatever jobs do not feel significantly happier than the unemployed ones among the 39 MBA students.

However, if checking more carefully the data, we will discover that there is an significant interaction between family income and love status among these 39 University of Chicago MBA students. For those who have deep feeling of belong and caring, family income are significantly related the happiness index. $1,000 increase in family income leads to 0.001 increase in the happiness index for those who have deep feeling of belonging and caring, adjusting for all other covariates.

5 Discussion

Our method of variable selection is quite naive. After all, using AIC as a criterion in a Linear Model does not guarantee that the chosen are suitable predictors in a Generalized Linear Model. However, stochastic search variable selection developed by Edward George and Robert McCulloch confirms that one of the most probable models is *money*, *love* and *work*[1].

As to the problem we ask at the beginning, money does not buy happiness in general for the 39 University of Chicago MBA students. And those who get whatever jobs do not feel significantly happier than the unemployed ones. If we think of family income and employment as some form of economic safety, the second layer of Maslow's hierarchy of needs[2], then it looks reasonable to assume that these student are relatively satisfied with safety needs. And not surprisingly, love and belonging, the third layer, becomes dominant.

However, what do surprise us is the result that family income are significantly related the happiness index for those who have deep feeling of belong and caring. Reasonable explanations may include that those in intimate relationships are willing to do more for their partners, such as purchasing anniversary gifts. Besides, maintaining a family do costs also.

References

1. George, E., & McCulloch, R. Variable Selection Via Gibbs Sampling. *Journal of the American Statistical Association*, *88*, 881-889.

2. Maslow's hierarchy of needs. (2014, April 25). *Wikipedia*. Retrieved May 2, 2014, from http://en.wikipedia.org/wiki/Maslow's_hierarchy_of_needs

Appendices – R Scripts

```r
#
# Filename: proj.r
# Author: Jun Xu
# Email: junx@bu.edu
# Created Time: Tue 29 Apr 2014 03:25:47 PM EDT
#

# 1. load and save data
library(faraway)
str(happy)
happy.arch <- happy
happy.new <- happy[order(happy[,2],happy[,3],happy[,4],happy[,5],happy[,1]),]
write.table(happy.new, file='happy.csv')

# 2. data preprocessing
happy <- within(happy,{
  sex <- factor(sex)
  love <- factor(love)
  work <- factor(work)
})
str(happy)

# 3. simple relationship exploration
# scatterplot matrix
pdf(file='31.pdf')
plot(happy)
dev.off()
# boxplots
pdf(file='32.pdf')
boxplot(happy ~ sex,happy)
dev.off()
pdf(file='33.pdf')
boxplot(happy ~ love,happy)
dev.off()
pdf(file='34.pdf')
boxplot(happy ~ work,happy)
dev.off()

# 4. OLM
lm <- lm(happy ~ .,happy)
lm <- step(lm)
m1 <- lm
summary(m1)
# c.i.
confint(m1)
# residuals vs fitted
pdf(file='41.pdf')
plot(m1,1)
dev.off()

# 5. multinomial model
library(nnet)
m2 <- multinom(happy ~ money+love+work,happy)
summary(m2)
# c.i.
confint(m2)

library(MASS)
m3 <- polr(as.factor(happy) ~ money+love+work,happy)
summary(m3)
# c.i.
confint(m3)

# residuals vs fitted
happy.new <- read.csv('happy2.csv',sep='\t',comm='#',header=T)
happy.new <- within(happy.new,{
```

```r
  f <- factor(f)
  sex <- factor(sex)
  love <- factor(love)
  work <- factor(work)
})
pm <- glm(cbind(y,n-y)~I(-money)+love+work+f-1,happy.new,family='binomial')
summary(pm)
confint(pm)
pdf(file='51.pdf')
plot(pm,1)
dev.off()

# 6. quasi-likelihood model
m4 <- glm(happy ~ money+love+work,happy,family=quasi(link='identity',variance='constant'))
summary(m4)
# c.i.
confint(m4)
# residuals vs fitted
pdf(file='61.pdf')
plot(m4,1)
dev.off()

# 7. joint mean & variance model
beta <- coef(lm)
repeat{
  d <- residuals(lm)^2
  gm <- glm(d ~ money+sex+love,happy,family=Gamma)
  lm <- lm(happy ~ money+love+work,happy,weights=1/fitted(gm))
  beta.new <- coef(lm)
  if(sqrt(sum((beta.new-beta)^2))<1e-6) break
  beta <- beta.new
}
m5 <- lm
summary(m5)
# c.i.
confint(m5)
# residuals vs fitted
pdf(file='71.pdf')
plot(m5,1)
dev.off()

# 8. prediction
# Adam
predict(m1,newdata=data.frame(money=40,sex='1',love='2',work='1'))
p1 <- predict(m3,newdata=data.frame(money=40,sex='1',love='2',work='1'),type='prob')
p1[1]*2+p1[2]*3+p1[3]*4+p1[4]*5+p1[5]*6+p1[6]*7+p1[7]*8+p1[8]*9+p1[9]*10
predict(m4,newdata=data.frame(money=40,sex='1',love='2',work='1'),type='response')
predict(m5,newdata=data.frame(money=40,sex='1',love='2',work='1'))

# Robert
predict(m1,newdata=data.frame(money=150,sex='0',love='1',work='5'))
p2 <- predict(m3,newdata=data.frame(money=150,sex='0',love='1',work='5'),type='prob')
p2[1]*2+p2[2]*3+p2[3]*4+p2[4]*5+p2[5]*6+p2[6]*7+p2[7]*8+p2[8]*9+p2[9]*10
predict(m4,newdata=data.frame(money=150,sex='0',love='1',work='5'),type='response')
predict(m5,newdata=data.frame(money=150,sex='0',love='1',work='5'))

# 9. interaction
pdf(file='91.pdf')
plot(happy ~ money,happy,col=sex,pch=c(1,19))
plot(happy ~ money,happy,col=love,pch=19)
plot(happy ~ money,happy,col=work,pch=19)
dev.off()

# money*love
m.noint <- glm(happy ~ money+love+work,happy,family=quasi(link='identity',variance='constant'))
m.int <- glm(happy ~ money+love+work+money*love,happy,family=quasi(link='identity',variance='constant'))
```

```
anova(m.noint,m.int,test='Chisq')

# money*work
m.int <- glm(happy ~ money+love+work+money*work,happy,family=quasi(link='identity',variance='constant'))
anova(m.noint,m.int,test='Chisq')

# love*work
m.int <- glm(happy ~ money+love+work+love*work,happy,family=quasi(link='identity',variance='constant'))
anova(m.noint,m.int,test='Chisq')

# check if money is significant in each stratification
summary(glm(happy ~ money+work,happy[happy$love==1,],family=quasi(link='identity',variance='constant')))
summary(glm(happy ~ money+work,happy[happy$love==2,],family=quasi(link='identity',variance='constant')))
summary(glm(happy ~ money+work,happy[happy$love==3,],family=quasi(link='identity',variance='constant')))
```