What Contributes to Your Happiness?

Sara Bey, Jayanth Sivakumar, Wei Yang

Department of Mathematical Sciences
Binghamton University

December 2, 2018

Introduction

- What factors contribute most to our happiness?
- Is it money, sex, love, or work? Or, is it a combination of these factors?
- How do each of these factors, and combinations of them, contribute to our happiness?

Data Set: *happy*

- Observation of 39 students in the MBA class at University of Chicago
  - Data about student's happiness, money, sexual activity, love and work are provided
  - *happy*: rated on a scale from 1 to 10
  - *money*: measured by looking at family income of the student in thousands of dollars
  - *sex*: is measured by responding "1" for "satisfactory sexual activity" and "0" for "not" satisfactory sexual activity
  - *love*: is measured by responding "1" for "lonely", "2" for "secure relationships" or "3" for "deep feeling of belonging and caring"
  - *work*: is measured on a 5 point scale with "1" representing "no job", "3" representing "OK job" and "5" representing "great job"

Hypotheses

- We hypothesize that people are happier when they have more money, satisfactory sexual activity, deeper feelings of belonging and caring and a great job.
  - Can we find evidence to support the claim that “money can’t buy happiness”?
- We believe that there will be a strong correlation between *money* and *work*, as those who have a great job tend to be earning more.
- We conjecture that there is a strong relationship between *love* and *sex* since someone in a sexually active relationship tends to have deeper connections and have a stronger relationship.

Levels of Happiness

- We first wanted to get an idea of how many people reported happiness at each level 1 - 10.

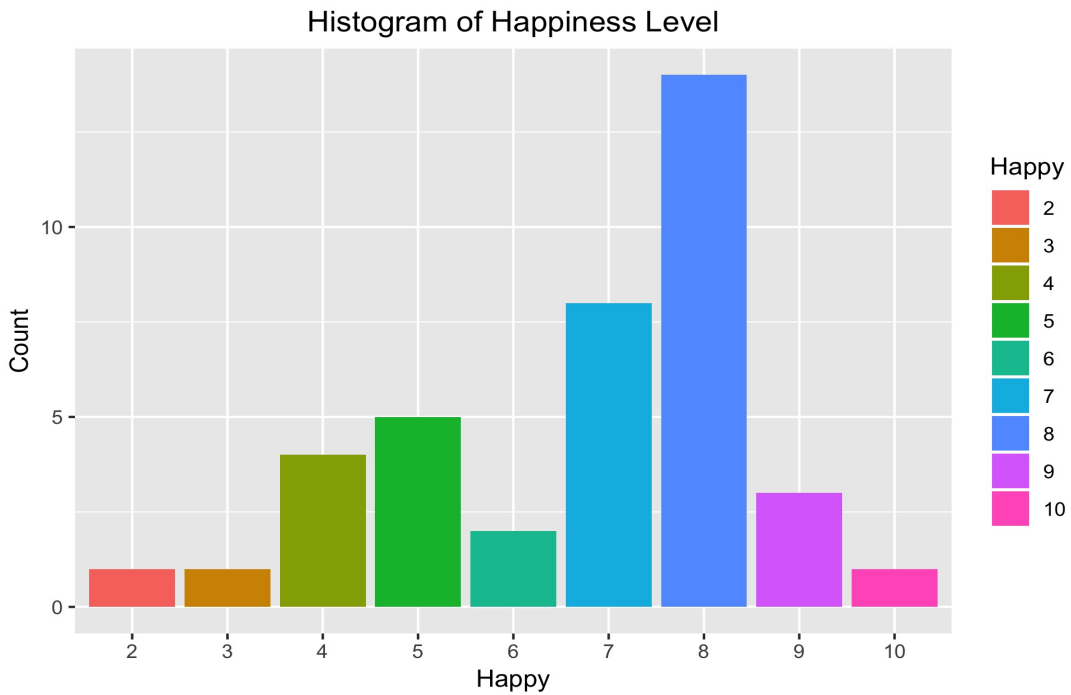


Figure: *Overall Happiness Rating*

- We see that most people have an *happy* level greater than or equal to 7, i.e. people are generally happy.

Variables Relationship to Happiness

- Then, we wanted to look at the breakdown of how many people gave each rating for the remaining four variables provided in the data set; *money, sex, love*, and *work*.

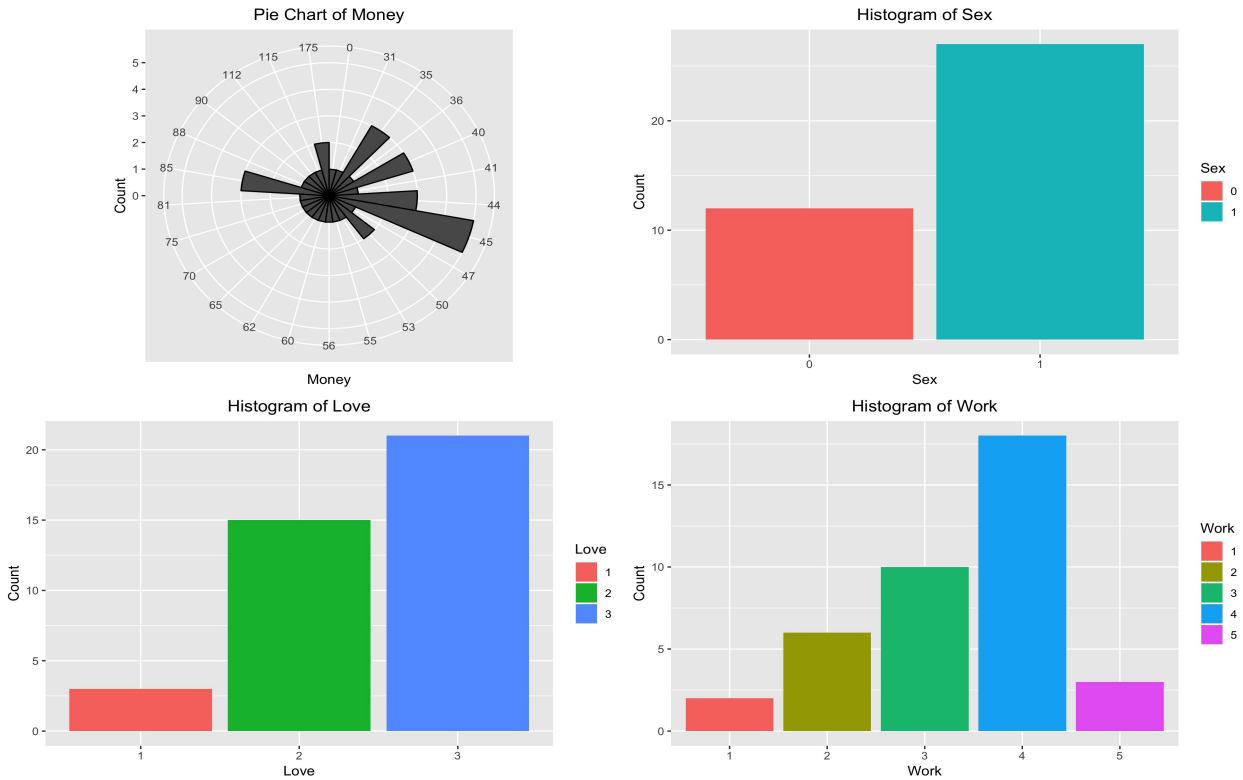


Figure: *Histograms*

Variables Relationship to Happiness continued

- In the upper left histogram we can see most people have an income around $\$45k$.
- In the upper right histogram, we can see that there are more people have satisfactory sexual activity, denoted by “1”, than not, denoted by “0”.
- In the lower left, we can see very few people consider themselves lonely, denote by “1”, most people said that they have secure relationships, denoted by “2”, or have a deep feeling of belonging and caring, denoted by “3”.
- In the lower right we can see that very few people have no job, denoted by “1”, most people are in the range of having an Ok job to having a great job, denoted by “3” and “5”, respectively.

Relationships Between Variables

- We then wanted to gain further insight into the relationship between the predictors themselves.

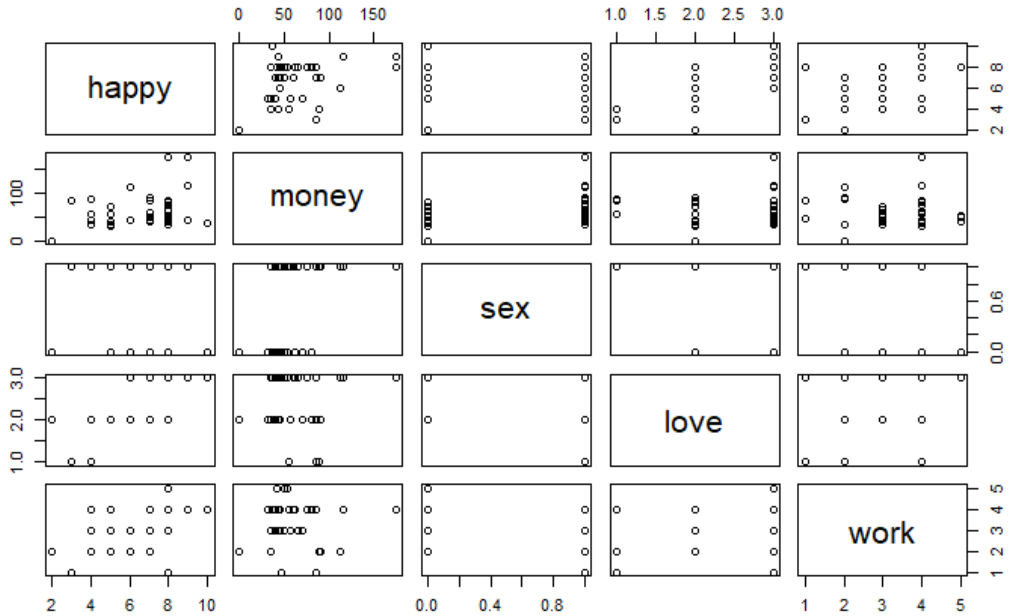


Figure: *Pairwise Scatter plots*

- From the scatter plot, we see that there is no clear pairwise linear trend among the variables.

Relationships Between Variables Continued

- Correlation between variables:

	happy	money	sex	love	work
happy	1.000	0.271	-0.033	0.784	0.539
money	0.271	1.000	0.307	0.126	0.068
sex	-0.033	0.307	1.000	0.047	-0.316
love	0.784	0.126	0.047	1.000	0.386
work	0.539	0.068	-0.316	0.386	1.000

- Variation Inflation Factors:

happy	money	sex	love	work
3.450936	1.257565	1.303891	2.740516	1.615127

- The small covariance among the explanatory variables and the fact that all VIF values are less than 3 further implies that there is no multicollinearity among the variables.

Full Model Analysis

- We started our investigation with the full model that has *happy* as the response variable and *money, sex, love* and *work* as the explanatory variables.

```
lm(formula = happy ~ ., data = happy)

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.072081   0.852543  -0.085   0.9331
money        0.009578   0.005213   1.837   0.0749 .
sex         -0.149008   0.418525  -0.356   0.7240
love         1.919279   0.295451   6.496 1.97e-07 ***
work         0.476079   0.199389   2.388   0.0227 *

Multiple R-squared:  0.7102,Adjusted R-squared:  0.6761
```

Figure: *Regression Analysis on Full Model*

- We see that there is a small adjusted R^2 of 0.6761, and only 2 variables, *love* and *work*, are significant at a significance level of $\alpha = 0.05$.
- Thus, we have to modify our model.

Analysis of the Predictor *sex*

- The variable *sex* appears to have a very small correlation, close to 0 (-0.033), with the variable *happy*.
- We wanted to investigate the relationship between the two.
- Using an F-test, we tested the hypothesis that $\beta_{sex} = 0$.
- Our full model predicts happiness based on the remaining four variables, where as our reduced model predicts happiness without including *sex* in the model.

F-test: $\beta_{sex} = 0$ vs $\beta_{sex} \neq 0$

```
Model 1: happy ~ money + love + work
Model 2: happy ~ money + sex + love + work
Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     35 38.229
2     34 38.087  1     0.142 0.1268  0.724
```

Figure: $\beta_{sex} = 0$ *vs* $\beta_{sex} \neq 0$

- Based on the F-test, we got a *P*-value of 0.724.
- At an α level of 0.05, we fail to reject the null hypothesis that $\beta_{sex} = 0$.
- There is evidence to suggest that our rating for *sex* may not have an effect on our happiness.
- Further investigation can be done to determine whether or not *sex* is insignificant in our predictions for happiness.

Subset Selection, Akaike Information Criterion (AIC), and Mallow's CP(C_p) (Dang 2018, slide 6-8)

- We select the best model for each size from 1 to 4, and observed that for size $n = 3$, the model without *sex* is the best model.

```
regsubsets.formula(happy ~ ., data = happy)
1 subsets of each size up to 4
Selection Algorithm: exhaustive
money sex love work
1  ( 1 ) " "   " " "*"  " "
2  ( 1 ) " "   " " "*"  "*"
3  ( 1 ) "*"   " " "*"  "*"
4  ( 1 ) "*"   "*" "*"  "*"
```

Figure: *Subset Selection*

Subset Selection, Akaike Information Criterion (AIC), and Mallow's CP(C_p) continued

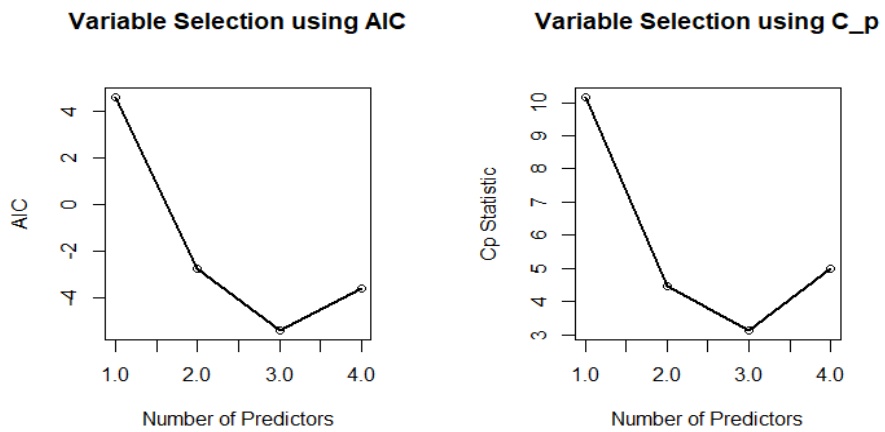


Figure: *Variable Selection using AIC and CP*

- It turns out that the model without *sex* minimizes both AIC and C_p (Dang 2018, slide 15).
- Hence we should remove *sex* from our model

Predicting *happy* without *sex* in the Model

- Using *money, love and work* as our predictors for happiness we get the following model:

```
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.185936   0.780372  -0.238   0.8131
money        0.008959   0.004852   1.846   0.0733 .
love         1.901709   0.287644   6.611 1.22e-07 ***
work         0.503602   0.181486   2.775   0.0088 **

Multiple R-squared:  0.7091,Adjusted R-squared:  0.6842
```

Figure: *Predicting Happiness Without Sex*

$$Happy = 0.008959 * money + 1.901709 * love + 0.503602 * work - 0.185936 + \epsilon$$

Analysis of Model that Predicts *happy* without *sex*

- The intercept term is -0.185936, which is the estimated value for *happy* when *money, love*, and *work* are all 0.
- But, *happy* takes value between 0 and 10, hence it does not make sense for us to have a negative intercept in our model.
- If *money, love* and *work* are all 0, assuming we have included all significant predictors, *happy* is expected to be 0.
- In other words, a model without the intercept term is more reasonable, but further investigation is needed.

F-test: $\beta_0 = 0$ vs $\beta_0 \neq 0$

```
Analysis of Variance Table
Model 1: happy ~ money + love + work - 1
Model 2: happy ~ money + love + work
Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     36 38.291
2     35 38.229  1  0.062009 0.0568 0.8131
```

Figure: $\beta_0 = 0$ *vs* $\beta_0 \neq 0$

- With a *p*-value of 0.8131 $\geq$ 0.05, we fail to reject H_0, at a significance level of $\alpha = 0.05$.
- There is evidence supporting the alternative hypothesis for a reduced model where $\beta_0 = 0$.

Predicting *happy* without *sex* or an intercept term

- Now we fit a model for *happy* based on *money, love* and *work*, without an intercept term:

```
lm(formula = happy ~ money + love + work - 1, data = happy)
Coefficients:
Estimate Std. Error t value Pr(>|t|)
money 0.008644   0.004608   1.876  0.06879 .
love  1.862795   0.233659   7.972 1.82e-09 ***
work  0.485124   0.161919   2.996  0.00493 **
---
Multiple R-squared:  0.9799,Adjusted R-squared:  0.9782
```

Figure: *Predicting Happiness Without sex or Intercept*

- We can see from above that at a significance level of $\alpha = 0.05$, *money* is not a significant predictor.
- Further investigation can be done to determine whether or not *money* is insignificant in our predictions for happiness.

Predicting *happy* without *sex*, *money*, or an intercept

- We proceed by removing *money* from the model in which we previously removed the intercept and *sex* as a predictor.

```
lm(formula = happy ~ love + work - 1, data = happy)

Coefficients:

Estimate Std. Error t value Pr(>|t|)

love    2.0081     0.2278    8.814 1.27e-10 ***

work    0.5330     0.1652    3.225  0.00263 **

---

Multiple R-squared:  0.9779,Adjusted R-squared:  0.9767
```

Figure: *Predicting Happiness Without sex, money, or Intercept*

- Removing *money* gives us a simpler model, but only leads to a small reduction in R^2.

Analysis of Model that Predicts *happy* without *sex*, *money*, or Intercept

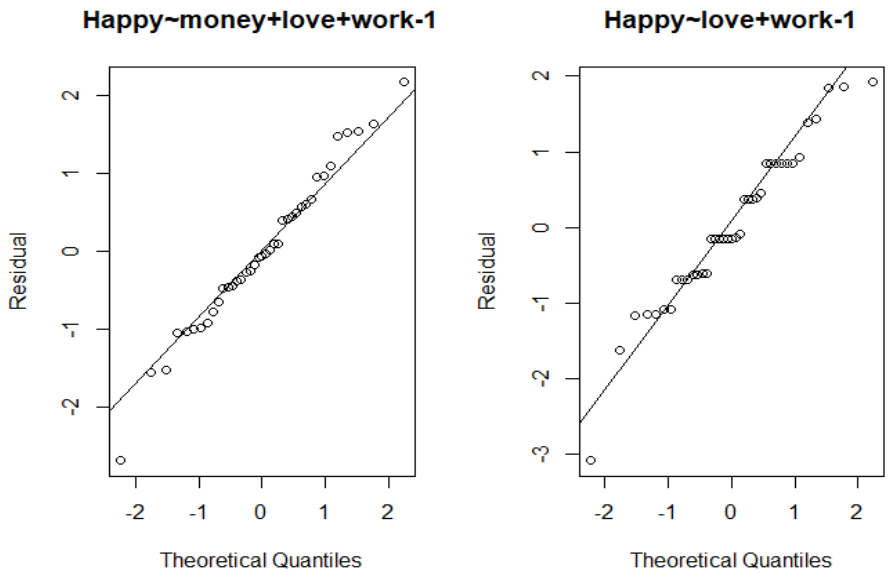


Figure: *QQ plot for Happy* $\sim$ *love+work-1*

- THe QQ plots indicate that the normality assumption for the error terms, $\epsilon_i's$, no longer holds if *money* is removed from the model.

Role of *money* in Predicting *happy*

- We should keep *money* in our model, but a possible transformation is needed since *money* was found to be insignificant at the significance level $\alpha = 0.05$.
- *Money* needed to be included in our model as a predictor, but could be altered to create a model that better fit our data.
- It is a well known fact from Economics that money has diminishing returns (Lane 2000).
- The function $f(x) = \sqrt{x}$ is increasing and concave down on $[0, \infty)$.
- We may incorporate the idea of diminishing returns of money into our model by transforming the predictor *money* with the square root function.

Model that Predicts *happy* with $\sqrt{money}$

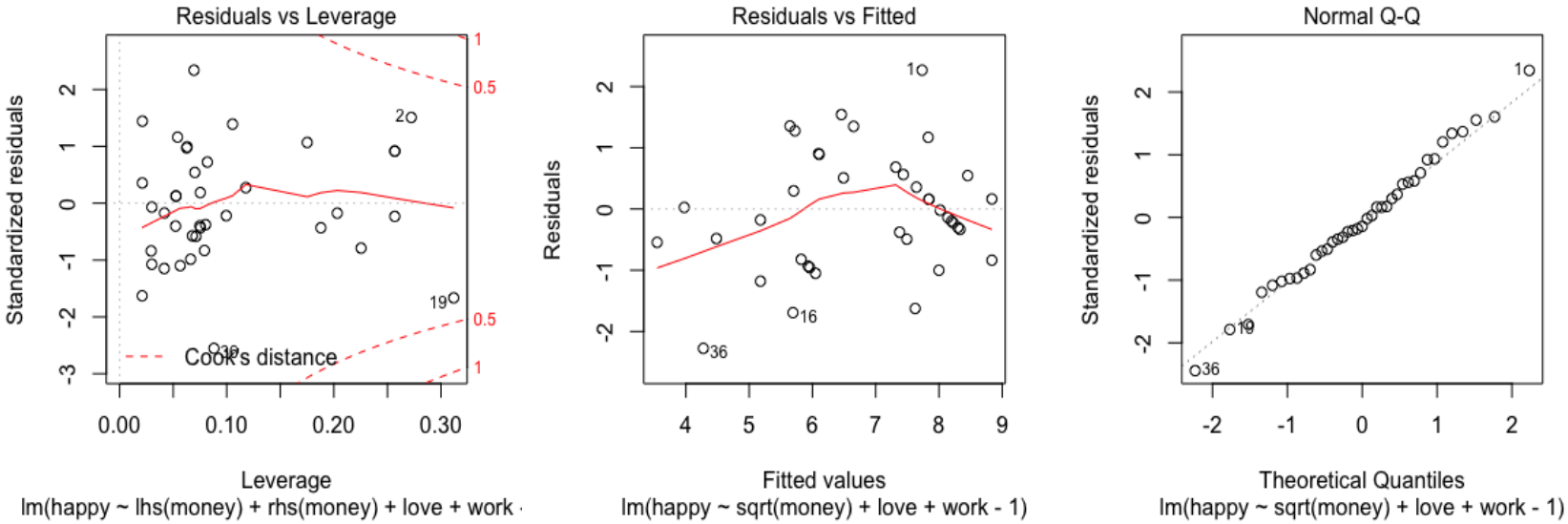
```
lm(formula = happy ~ sqrt(money) + love + work - 1, data = happy)
Coefficients:
Estimate Std. Error t value Pr(>|t|)
sqrt(money)  0.15263    0.06218   2.455   0.0191 *
love         1.73346    0.24128   7.184 1.88e-08 ***
work         0.40422    0.16369   2.469   0.0184 *
Residual standard error: 1 on 36 degrees of freedom
Multiple R-squared:  0.9811,Adjusted R-squared:  0.9795
F-statistic: 622.9 on 3 and 36 DF,  p-value: < 2.2e-16
```

```
AIC(all,nosex,nointsex,nointsexmoney,rootmoney)
             df      AIC
all           6 121.7534
nosex         5 119.8985
nointsex      4 117.9617
nointsexmoney 3 119.5990
rootmoney     4 115.5633
```

Figure: *Model with Square Root of money*

- In taking the square root of *money* we were able to create a simple model which has the smallest AIC (115.5) among all the models that we have tried.
- All the variables contained in this model are significant and the model gives an adjusted R^2 value of 0.9795.

Analysis of Model that Predicts *happy* with $\sqrt{money}$



- The residual vs. fitted values are randomly scattered, indicating that the constant variance assumption holds.
- The Normal QQ plot is linear, indicating that the normality assumption of the residuals holds.
- Thus $happy = 0.1526254 * \sqrt{money} + 1.7334612 * love + 0.4042174 * work + \epsilon$ can be consider to be a good model.

Implications of the Model

- In our model, *money* has a diminishing effect on happiness, and plays the least important role out of the included predictors.
- According to our model, earning $\$170k$ a year only contributes about 2 points to our overall rating of happiness.
- On the other hand, love plays the most important role in a person's happiness.
- Suppose a person has no job and makes no money. As long as that person has deep feeling of belonging and caring, our model indicates that this person can expect to be generally satisfied, with a rating for *happy* of 5.

Making Predictions Using Our Model

- Consider a person with median *money, love* and *work*.
- That is to say a person with an annual salary of $\$77.97k$, a person who has a deep feeling of belonging and caring (love=3) and a person having a more than ok job (work=4).
- Our model predicts that this person is going to have a happiness level of 7.89647 and we are 68% confident that his/her rating of happiness is in the interval $(6.947661, 8.845294)$ and 95% confident that his/her happiness is in the interval $(5.821091, 9.971864)$.

Making Predictions Using Our Model Continued

- We can do the same for a person whose statistics fall in the 0^{th}, 25^{th}, 75^{th} or 100^{th} quantiles.

quantile	love	work	money
0%	1	1	0.00000
25%	2	3	42.48676
50%	3	4	50.00000
75%	3	4	77.97114
100%	3	5	175.00000

```
predict(rootmoney,x0,interval="prediction",level=0.95)
      fit       lwr       upr
0%  2.137679 0.07579235  4.199565
25% 5.674415 3.62628012  7.722550
50% 7.896478 5.82109083  9.971864
75% 8.164955 6.10069013 10.229219
100% 9.240514 7.10116518 11.3798641
```

- We are 95% confident that a person with stats in the 25^{th} percentile (with secure relationships, an ok job, and a salary of \$42.48 k) will have a happiness rating in $(3.62628012, 7.722550)$.

- This implies that even if a person makes only $\$42k$, there is still a chance that he/she is feeling quiet happy, ($happy = 7$), given that he/she is in a secure relationship and has an ok job.

Conclusion

- Based on the regression analysis that we have carried out, we have determined that our happiness can be predicted based on our *money, love* and *work*, but that our rating of *sex* in insignificant in making this prediction.
- The model that removes the intercept and the predictor *sex*, but takes the square root of *money* provides the best fit for predictions of *happy* for the observations given in the dataset.

Key References

- Dang, S 2018, *Variable Selection*, lecture notes, Regression I MA531, Binghamton University, delivered 15, Oct 2018.

- Lane, E 2000, 'Diminishing returns to income, companionship–and happiness', *Journal of Happiness Studies*, vol. 1, no.1, pp. 103-119.

Semester's Over! YAY!