# SOS3003
# Examination question 1

Fall 2009

Erling Berge



# FALL 2009 QUESTIONS

- The questions use data from Malawi collected during field work in 2007. The data come from long interviews and questionnaire forms collected from 270 households plus some additional informers. The data also comprise trust game data from 267 pairs of players. In the present questions we use data from the trust game. More on the sample and variables is presented below.

## **QUESTION 1** (OLS-regression, weight 0,5)

- In this question we explore the propensity to be generous to people within your own community when you do not know the identity of the person you show generosity. It was determined that trust (as measured by answering “yes, most people can be trusted” to the question “Generally speaking, do you think most people can be trusted or that they cannot be trusted?”) did not show any realionship with level of generosity. Instead three other types of factors were considered: general personal characteristics, indicators of wealth, and indicators of culture.



## Question 1: 4 parts of the question

- a) Describe the impact of “Mattress owned” on Generosity as it is estimated by model 4. Find a 95% confidence interval for the impact.
- b) Determine if the interaction between “Sex of respondent” and the variables “Mattress owned” and “Radio owned” contribute significantly to the explanation of variance in the dependent variable. Use a 0.05 level of significance for the test and state explicitly the hypothesis that is being tested.
- c) Present the assumptions that need to be satisfied if the estimates and tests of the 9 models are to be trustworthy. Determine if the tables presented give any reason to doubt that the model assumptions are satisfied
- d) Based on the tables presented what can you say about the factors affecting level of generosity? Discuss in particular the impact of sex and age.

# Question 1 a)

- **a) Describe the impact of "Mattress owned" on Generosity as it is estimated by model 4. Find a 95% confidence interval for the impact.**
- The dependent variable for models 1-9 is Generosity, the amount of Kwacha returned over or below what is defined as a fair share of the profit from the initial investment in a trust game. The variable "Mattress owned" is an indicator of relative wealth in communities where many sleep directly on the floor. It takes the value of 1 if the respondent owns a mattress, zero otherwise.



# Question 1 a)

- Model 4 tells us that a person owning a mattress returns 19.2 Kwacha more than a fair division of the profit controlling for differences between sexes and age groups, as well as ownership of radio. It is a bit puzzling that ownership of radio has a large negative impact, meaning that if you own both mattress and radio you are not nearly as generous as if you own only mattress.

# Question 1 a)

- In OLS regressions estimates of the model parameters, $b_k$ , are known to follow a t-distribution if the estimates come from a simple random samples, and the null hypothesis of zero value of the population parameter is true.
- Then a (1-$\alpha$) confidence interval for the population parameter $\beta_k$ from a model with K parameters estimated on n cases is found as

$$b_k - t_\alpha * SE_{b_k} < \beta_k < b_k + t_\alpha * SE_{b_k}$$

where $t_\alpha$ is the critical value from the t-distribution with n-K degrees of freedom in a two tailed test with $\alpha$ level of significance.



# Question 1 a)

- We find in model 4 that $b_{\text{Mattress owned}}$ = 19.268, $SE_{b(\text{Mattress owned})}$ = 8.398, n = 116, and K = 8. Hence n-K = 108, and since the table of the t-distribution in Hamilton (page 350) for $\alpha$ = 0.05 gives us critical values for 60 (t=2) and 120 degrees of freedom (t=1.98), we see that for df=108 1.98 < $t_\alpha$ < 2.00. Since df=108 is closer to 120 than to 60, one may here interpolate using the conservative value of 1.99. Normally one will choose to use the value of 2, but also 1.98 will be acceptable.

## Question 1 a)

- This means that the 0.95 confidence interval will be
- $19.268-8.398*1.99<\beta_k<19.268+8.398*1.99$
- $19.268 - 16{,}71202< \beta_k < 19.268 + 16{,}71202$

  - **$2{,}55598 < \beta_k < 35{,}98002$**
    - **$2{,}55 < \beta_k < 35{,}98$**



## Question 1 b)

- **b) Determine if the interaction between "Sex of respondent" and the variables "Mattress owned" and "Radio owned" contribute significantly to the explanation of variance in the dependent variable. Use a 0.05 level of significance for the test and state explicitly the hypothesis that is being tested.**

# Question 1 b)

- We want to determine if the two variables (interaction terms) “Sex*Own mattress” and “Sex*Own radio” contribute significantly to the model of Generosity. We want to test
- H0: $b_{Sex}$* Own mattress = 0 and $b_{Sex}$* Own radio = 0 against the alternative
- HA: $b_{Sex}$* Own mattress ≠ 0 and $b_{Sex}$* Own radio ≠ 0
- In testing if interactions between sex and indicators of wealth contribute to the model, we inspect models 4 and 5. This is where they appear for the first time. Due to multicollinearity tests of single coefficients cannot be trusted.



# Question 1 b)

| Model 4 | B | Std. Er. | Beta | t | Sig. | Tolerance | VIF |
|---|---|---|---|---|---|---|---|
| (Constant) | 32.819 | 29.539 | | 1.111 | .269 | | |
| Sex of respondent | -119.871 | 46.696 | -1.749 | -2.567 | .012 | .017 | 59.219 |
| Age of respondent | -2.065 | 1.483 | -1.048 | -1.393 | .167 | .014 | 72.309 |
| Age squared | .020 | .016 | .976 | 1.277 | .204 | .013 | 74.545 |
| Sex * Age | 5.089 | 2.213 | 3.691 | 2.299 | .023 | .003 | 328.752 |
| Sex * Age squared | -.050 | .023 | -2.322 | -2.136 | .035 | .007 | 150.811 |
| Mattress owned | 19.268 | 8.398 | .220 | 2.294 | .024 | .849 | 1.178 |
| Radio owned | -14.420 | 7.194 | -.201 | -2.004 | .048 | .778 | 1.285 |

# Question 1 b)

| Model 5 | B | Std. Er. | Beta | t | Sig. | Tolerance | VIF |
|---|---|---|---|---|---|---|---|
| (Constant) | 31.475 | 30.019 | | 1.049 | .297 | | |
| Sex of respondent | -117.719 | 47.343 | -1.717 | -2.486 | .014 | .017 | 59.854 |
| Age of respondent | -1.994 | 1.533 | -1.012 | -1.300 | .196 | .013 | 75.964 |
| Age squared | .020 | .016 | .946 | 1.197 | .234 | .013 | 78.327 |
| Sex * Age | 4.879 | 2.333 | 3.539 | 2.092 | .039 | .003 | 359.166 |
| Sex * Age squared | -.047 | .025 | -2.206 | -1.907 | .059 | .006 | 167.850 |
| Mattress owned | 22.310 | 11.337 | .255 | 1.968 | .052 | .474 | 2.111 |
| Radio owned | -16.126 | 9.540 | -.225 | -1.690 | .094 | .450 | 2.222 |
| Sex * Own mattress | -6.975 | 17.062 | -.057 | -.409 | .684 | .408 | 2.450 |
| Sex * Own radio | 4.130 | 14.694 | .058 | .281 | .779 | .188 | 5.328 |



# Question 1 b)

- Sex is also involved in all models 1-3 including interactions with age. This results in a very high degree of multicollinearity in the models from 3 on. But the variance inflation factor (VIF=1/tolerance) of Sex does not increase very much from model 4 to 5. Hence the test of the contribution of the interaction terms with the wealth indicators can leave out sex alone.

# Question 1 b)

- In the change statistics of the Model summary table, the value of the F-statistic for the contributions of Sex*Own mattress and Sex*Own radio to the model is 0.098 with 2 and 106 degrees of freedom. The probability of finding this low or lower values of the F-statistic, given that the population values of the model parameters for these two variables are zero, is 0.907 (“Sig.F-change” column of the table). But the tolerances, particularly for Sex*Own radio, are low.



# Question 1 b)

- Inspecting the coefficients of Own Mattress and Own radio in the two models we see that the p-values of the two increases significantly from model 4 to 5.
- Since both interaction terms do not contribute significantly to the model, and since the p-values in the tests of the Own mattress and Own radio variables increase, the F-test of the two interaction terms should be considered to be valid.
- In general a test statistic (in this case F) is constructed assuming the null hypothesis of no impact of the tested variables is true. The hypothesis we want to test here is then:
- H0: In model 5 $b_{Sex}$*Own mattress and $b_{Sex}$*Own radio are both equal to 0

## Question 1 b)

Then we have to compare model 4 and 5. The F-statistic:

$$F_{n-K}^{H} = \frac{\dfrac{RSS_{[K-H]} - RSS_{[K]}}{H}}{\dfrac{RSS_{[K]}}{n-K}}$$

- follows a F-distribution with H and n-K degrees of freedom if it is true that the H extra variables included in the big model have no effect (if H0 "No impact of the new variables" is true) and the assumptions of OLS regression are met.



## Question 1 b)

- In this formula the RSS[K] is the sum of squares of the residuals of the big model with K parameters (or K-1 variables) and RSS[K-H] is the sum of squared residuals in the small model where the H new variables are not included. We reject the null-hypothesis that the H new variables do not have an impact with level of significance $\alpha$ if $F^{H}_{n-K}$ is larger than the critical value for level of significance $\alpha$ in the table of the F-distribution with H and n-K degrees of freedom.

# Question 1 b)

- In this case we have that H=2, n=116, K=10, and from table A4.2 in Hamilton we see that the critical value for $F^2_{106}$ with level of significance 0.05 is approximately 3.07. We conclude that the two interaction terms do not contribute to the model specification if we find that the computed value $F^2_{106}$ is less than the critical value 3.07 (table value of $F^2_{120}$ for $\alpha$=0.05) of assuring a test level of 0.05. This F-value has already been computed in the Model summary table and is there given as 0.098, far below the critical value.



# Question 1 b)

- *Alternatively:*
- An alternative avenue for finding the F-value, is to look up the residual sums of squares in the ANOVA table and compute the value according to the formula given above.
- The F-value we compute this way is of course exactly the same as the one reported by SPSS in the Model summary table for the test of changes in model. The conclusion is that the two interaction terms do not contribute to the model specification. They are irrelevant variables and should be removed.

# Question 1 c)

- **c) Present the assumptions that need to be satisfied if the estimates and tests of the 9 models are to be trustworthy. Determine if the tables presented give any reason to doubt that the model assumptions are satisfied**



# Question 1 c)

- All models in question 1 are regression models of the form
- $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + ... + \beta_{19} x_{i19} + \varepsilon_i$
- where "i" runs over the household population of 18 Malawian villages. If we let k=0, 1, 2, 3, … ,19, $\beta_k$ will be the unknown parameters showing how many measurement units of y will be added to y per unit increase in $X_k$
- "$\varepsilon_i$" is the error term, a variable that comprises all relevant factors not observed as well as random noise in the measurement of y. The 19 x-variables are defined in the model 9 table and the section of variable definitions

## Question 1 c)

- An OLS (ordinary least squares) estimate of the model parameters defined above can be found as the b-values of
- $\hat{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + ... + b_{19} x_{i19} + e_i$
- that minimizes the sum of squared residuals,

$$RSS = \sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n} e_i^2$$

(For "$\hat{y}_i$" read "estimated" or "predicted" value of $y_i$ or just "y-hat")



## Question 1 c)

- OLS estimates will be unbiased and efficient with a known sampling distribution if the following assumptions are true:
- I: The model is correct, that is
  - All relevant variables are included
  - No irrelevant variables are included
  - The model is linear in the parameters

# Question 1 c)

- II: The Gauss-Markov requirements for “Best Linear Unbiased Estimates” (BLUE)
  - Fixed x-values (no random component in their measurement)
  - The error terms have an expected value of 0 for all cases “i”
    - $E(e_i) = 0$ for all “i”
  - The error terms have constant variance for all cases “i” (homoscedasticity) for all “i”
    - $var(e_i) = \sigma^2$ for all “i”
  - The error terms do not correlate with each other across cases (no autocorrelation) for all “i” ≠ “j”
  - $cov(e_i, e_j) = 0$ for all “i” ≠ “j”

Fall 2009 © Erling Berge 25

# Question 1 c)

- III: The error terms are normally distributed
  - The error terms are normally distributed (and with the same variance) for all cases for all “i”
    - $\varepsilon_i \sim N(0, \sigma^2)$ for all “i”

- Inferences from a sample to a population can be obtained with a known confidence if the estimates come from a simple random sample from the population of interest.

Fall 2009 © Erling Berge 26

# Question 1 c)

Some of the stated assumptions cannot be tested. In particular we cannot test if

- All relevant variables are included
- Variables are without measurement error
- The error term in reality has mean 0 and variance 1

We can test if

- irrelevant variables have been included in the model
- the model is curvilinear in the included variables
- there is heteroscedasticity and/ or autocorrelation
- the error term is normally distributed



# Question 1 c)

*Discussion of the assumptions in relation to model 9.*

- As concluded in point b) above there are irrelevant variables in the model. The consequence of irrelevant variables is that variances are larger than they otherwise would be, making confidence intervals wider and estimates less precise.

# Question 1 c)

- All variables except Age are binary coded. Hence only age can be curvilinearly related to the dependent variable Generosity. And already in model 3 it is established that age is curvilinearly related to Generosity in interaction with sex. In the models 6-9 this relationship seems to be disappearing. But fluctuations in the p-values for sex and age (Sig.) may be assumed to be related to the introduction of several interaction terms with sex that perhaps might be irrelevant. A model without the interaction terms between sex and Own mattress, Own radio, and Region, possibly also marriage system, should be estimated before judgment is passed on sex and age in a larger model than model 3.



# Question 1 c)

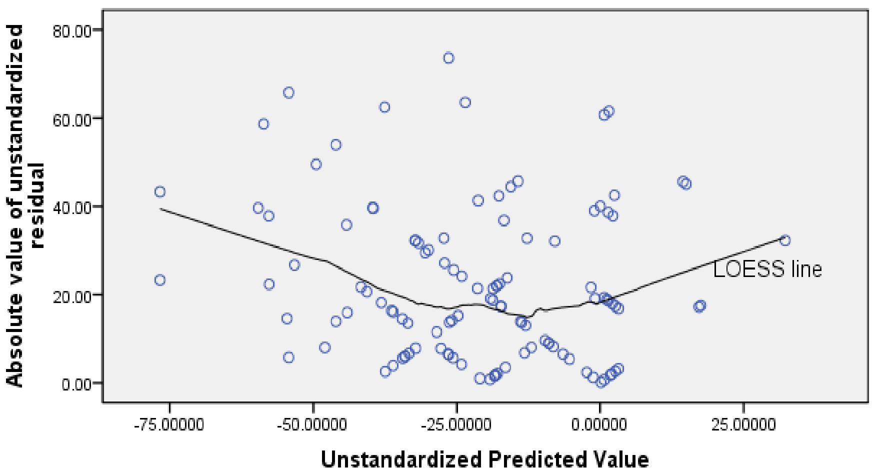
- In judging the degree of heteroscedasticity we look at the plot of predicted values against the residuals. The LOESS line is for central parts of the scatter plot fairly level. This suggests that the degree of heteroscedasticity is low and probably introduced through limited variation both on the dependent and the independent variables. The fluctuations of the LOESS line are mirrored in the deviations from the diagonal of the Normal Probability plot. Another possible reason for the curved LOESS line may be influential cases at the extremes of the predicted values scale. Taking a closer look at the LOESS line in the scatter plot of the absolute value of the unstandardized residual we see 2 cases with ca -75 as predicted values and one with ca +35.

# Question 1 c)

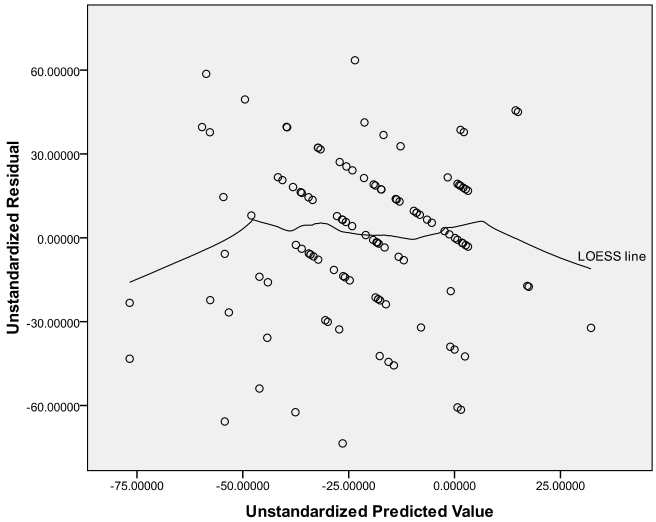
- In the box plot of the standardized predicted value 2 cases, 8037 and 10006, appear as extreme outliers. These are both women, 48 years old, and have the same values also on all other variables included in the model except the dependent. Together they have high influence. The conclusion here is that the sample probably is too small. One should probably also consider to report results both with these two and without them.



# Question 1 c)



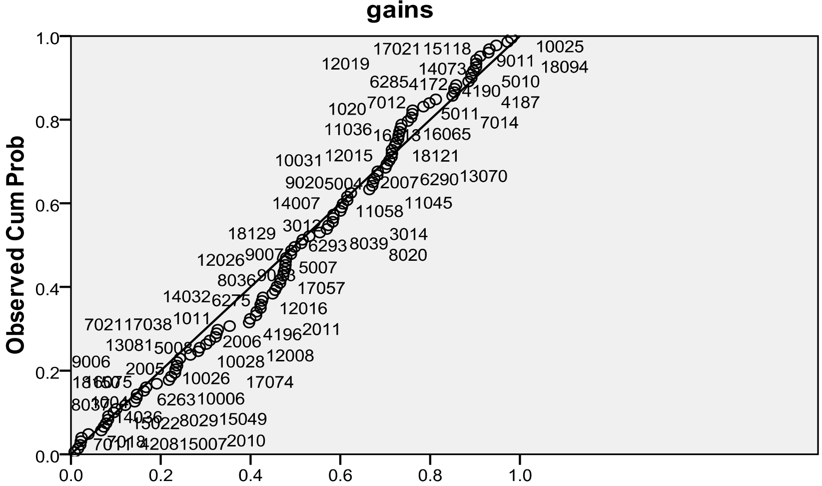
# Question 1 c)






# Question 1 c)

**Dependent Variable: Returned more or less than 50% of capital gains**



# Question 1 c)

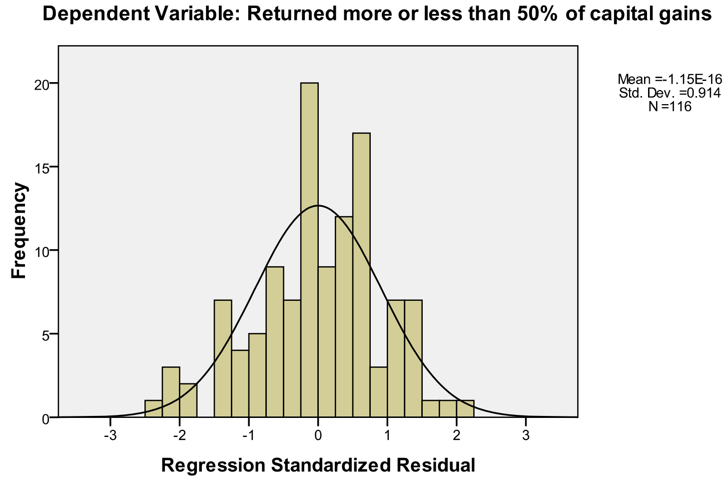
- To judge autocorrelation we need to think about the possible causes of such correlations. Regions and districts are purposively selected. Sorting data according to geographical proximity might reveal any autocorrelation due to this. Assuming this has been done before the computation of the Durbin-Watson statistic of 1.92 (see the model summary table) one may conclude that there probably is not any autocorrelation in the data. Hamilton’s table A4.4 gives for samples of 100 and models of 5 variables an upper limit of 1.78. Model 9 has 19 variables and is estimated on 116 cases. So at most the test would be inconclusive, but probably we could reject the hypothesis of autocorrelation.



# Question 1 c)

- To evaluate the requirement of normally distributed residuals we inspect the distribution of the residual in the histogram of the residual. There are some deviations but they do not seem to be systematic in relation to the distribution. However, in a sample of only 116 cases also this distribution ought to alert us to the possibility of influential observations. “No influential case” is not a requirement per se, but their presence may destroy the normal distribution of the error term.

# Question 1 c)






# Question 1 c)

- There are several statistics we can inspect to evaluate the possible presence of influential cases. One basic statistic is the leverage, h. SPSS reports the centered leverage, that is the leverage minus the mean. The sample mean of the leverage is K/n, or in this case 20/116 = 0.172. The maximum of the centered leverage is 0.57. Thus the absolute value of the maximum is 0.742. This is above the 0.5 where Hamilton advices us to avoid the case. Looking at the box plot of the h statistic we find 4 cases outside the 1.5IQR distance from the median. They are 2006, 6285, 7014, and 17038.

# Question 1 c)

- In the box plot of the standardized residual and the standardized predicted value we find large values for the cases 4208, 8037, and 10006.
- Another general indicator is Cook's D statistic. Inspecting the box plot of Cook's D statistic we see that there are 8 cases with values more than 1.5*IQR from the median. Three of these are among those with large h. Looking similarly at the box plots of the standardized residual and standardized predicted value we find the cases 4208, 6285, 7014, 7018, 8029, 8037, 10025, and 17038.



# Question 1 c)

- Looking also at the box plots for the DFBETAS, we see numerous cases with values exceeding 1.5IQR from the median. Looking for gaps in the distribution we find the cases: 2006, 4208, 6285, 7014, 7021, 8037, 10025, 15118, 17038, and 18094. They are distributed across the variables as follows:

# Question 1 c)

| Variable group | Cases with highest values DFBETAS (single case or groups) | | | | |
|---|---|---|---|---|---|
| Sex | 6285 | 7014 | | | |
| Age | - | | | | |
| Wealth indicators | 4208 | 15118 | 18094 | | |
| Region dummies | 7014 | 7021 | 8037 | 10025 | 17038 |
| Marriage system dummies | 2006 | 7014 | 17038 | | |



# Question 1 c)

- Five cases are found as potentially influential by only one statistic, the cases 7018, 8029, 10006, 15118, and 18094. Case 10006 has the largest predicted value. We shall look at this together with the cases 2006, 4208, 6285, 7014, 8037, 10025, and 17038.

# Question 1 c)

| | Variables | Case 2006 | Case 4208 | Case 6285 | Case 7014 | Case 8037 | Case 10006 | Case 10025 | Case 17038 |
|---|---|---|---|---|---|---|---|---|---|
| 2 | MatriMatri | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | MatriPatri | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 4 | PatriPatri | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | OtherMarri | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |
| 6 | Generosity | -40 | -100 | 0 | 0 | -120 | -100 | 0 | 0 |
| 7 | OwnMattr | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 8 | OwnRadio | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 |
| 17 | Sex | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 18 | Age | 52 | 27 | 85 | 25 | 48 | 48 | 28 | 47 |
| 20 | North | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 21 | Central | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 |
| 22 | South | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |



# Question 1 c)

- It is not obvious why these are influential cases. We see that 3 of the 6 most non-generous players are included here, and none of the generous. Two of these are middle aged women living in the central districts.
- The large DFBETAS for Sex on cases 6285 and 7014 must be due to the particular combination of values for these cases. 6285 is an old man living in the north and with 0 Generosity. 7014 is a young woman living in the Central region also with 0 Generosity.
- The influence we see is probably a consequence of few cases rather than any other kind of problems.

# Question 1 d)

- **d) Based on the tables presented what can you say about the factors affecting level of generosity? Discuss in particular the impact of sex, age, and wealth in the models 1-4.**



# Question 1 d)

- There are estimates of 9 models all nested hierarchically so that all previous models are contained in the last. The dependent variable is “Returned more or less than 50% of capital gains”.
- In the first 3 models the only variables involved were sex and age:

## Question 1 d)

| Model | | Unstandardized Coefficients B | Std. Error | Standardized Coefficients Beta | t | Sig | Collinearity Statistics Tolerance | VIF |
|---|---|---|---|---|---|---|---|---|
| 1 | (Constant) | -17.432 | 8.513 | | -2.048 | .043 | | |
| | Sex of respondent | -12.328 | 6.384 | -.180 | -1.931 | .056 | .988 | 1.013 |
| | Age of respondent | .055 | .184 | .028 | .300 | .765 | .988 | 1.013 |
| 2 | (Constant) | -.841 | 23.650 | | -.036 | .972 | | |
| | Sex of respondent | -11.923 | 6.419 | -.174 | -1.857 | .066 | .981 | 1.020 |
| | Age of respondent | -.788 | 1.136 | -.400 | -.694 | .489 | .026 | 38.636 |
| | Age squared | .009 | .012 | .433 | .752 | .454 | .026 | 38.515 |
| 3 | (Constant) | 43.248 | 29.905 | | 1.446 | .151 | | |
| | Sex of respondent | -126.694 | 47.557 | -1.848 | -2.664 | .009 | .017 | 58.641 |
| | Age of respondent | -2.722 | 1.465 | -1.382 | -1.858 | .066 | .015 | 67.360 |
| | Age squared | .026 | .016 | 1.271 | 1.683 | .095 | .014 | 69.473 |
| | Sex * Age | 4.979 | 2.257 | 3.611 | 2.206 | .029 | .003 | 326.372 |
| | Sex * Age squared | -.046 | .024 | -2.128 | -1.925 | .057 | .007 | 148.803 |



## Question 1 d)

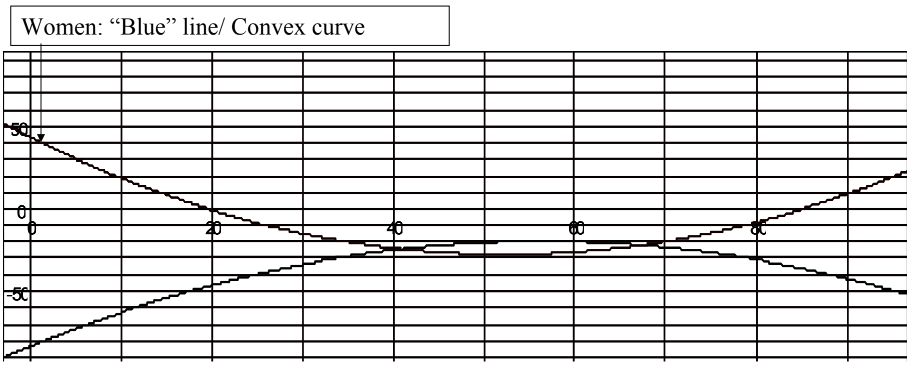
- It is remarkable that Sex alone is not quite significant at 5% level, and Age is far from being significant alone. Age as curvilinear variable seems to do better than as a linear variable, but is still far from being significant. But introducing the interaction between Sex and Age as a curvilinear variable makes the group clearly significant at 5% level and even more remarkable the least significant single element, Age squared has a p-value of 0.095 despite a very high degree of multicollinearity.

# Question 1 d)

- A conditional effect plot of age and sex might be interesting to inspect. Plotting the relationship as determined in model 3 we find that
- Y=43.248 -126.694Sex -2.722Age +0.026Age*Age +4.979Sex*Age -0.046Sex*Age*Age
- will provide 2 curves showing how generosity varies by age for men and women.
- Such a curve is presented below:



# Question 1 d)




$y=43.248 - 126.694\times 1 - 2.722\times x + 0.026\times x^2 + 4.979\times x\times 1 - 0.046\times x^2\times 1$

$y=43.248 - 126.694\times 0 - 2.722\times x + 0.026\times x^2 + 4.979\times x\times 0 - 0.046\times x^2\times 0$

Black equation= men; Blue equation = women

# Question 1 d)

- From this we see that both young and old women are more generous than men, while men and women between about 30 and 80 years of age are about equal in generosity. However, we should also note that there minimum observed age is 15 and maximum is 85, and that there probably are very few cases below 20 and above 80. Hence a figure like this will exaggerate the differences between the sexes. Extrapolation from the observed range of a variable is not advisable.
- In the models 4-9 the other explanatory factors are added:



# Question 1 d)

| Explanatory factors | Variables | Results |
|---|---|---|
| In models 4 and 5 the indicators of wealth and their interactions with Sex are introduced | Mattress owned<br>Radio owned<br>Sex * Own mattress<br>Sex * Own radio | In model 4 the wealth indicators are significant. In model 6 they become marginally insignificant while the interaction terms clearly are irrelevant variables |
| | | |
| In models 6 and 7 the indicators of regional cultures and their interaction with Sex are introduced. The reference category here is the Central region | North region<br>South region<br>Sex*North<br>Sex*South | In model 6 we see that the region variable is significant while model 7 shows that the interaction terms are irrelevant variables |
| | | |
| Then finally in models 8 and 9 the indicators of marriage system and their interactions with Sex are introduced | Matrilineal and matrilocal<br>Patrilineal and paralegal<br>Other marriage patterns<br>Sex * Matrilineal and matrilocal<br>Sex * Patrilineal and patrilocal<br>Sex * Other marriage patterns | In model 8 we find that the marriage system variable do not contribute to the model by themselves, however, model 9 shows that their interaction with Sex contributes significantly |

# Question 1 d)

- In model 4 the wealth indicators “Own mattress” and Own radio” are added.
- The simple reasoning behind wealth as and explanation for generosity is that the relatively wealthier would be more generous towards their fellows. The reasoning may be too simple. The two wealth indicators work in different directions in relation to generosity, and they do so consistently across all models. Why this should be so is not obvious. One might want to rethink the reasoning behind their interpretation as wealth indicators.
- The last model estimated is model 9. To facilitate the discussion we drop the irrelevant interaction terms without re-estimating the model.



# Question 1 d)

| Model | | B | Std. Error | t | Sig. | | VIF |
|---|---|---|---|---|---|---|---|
| 9 | (Constant) | -16.644 | 38.993 | -.427 | .670 | | |
| | Sex of respondent | -98.703 | 54.634 | -1.807 | .074 | | 92.589 |
| | Age of respondent | .235 | 1.552 | .151 | .880 | | 90.380 |
| | Age squared | -.002 | .016 | -.129 | .897 | | 90.816 |
| | Sex * Age | 3.228 | 2.323 | 1.389 | .168 | | 413.649 |
| | Sex * Age squared | -.031 | .024 | -1.257 | .212 | | 189.957 |
| | | | | | | | |
| | Mattress owned | 32.334 | 11.899 | 2.717 | .008 | | 2.701 |
| | Radio owned | -19.555 | 9.052 | -2.160 | .033 | | 2.324 |
| | | | | | | | |
| | North region | 3.847 | 16.893 | .228 | .820 | | 7.802 |
| | South region | 27.801 | 17.082 | 1.627 | .107 | | 8.276 |
| | | | | | | | |
| | Matrilineal and matrilocal | -14.310 | 21.658 | -.661 | .510 | | 13.707 |
| | Patrilineal and patrilocal | -6.344 | 22.706 | -.279 | .781 | | 14.095 |
| | Other marriage patterns | -46.917 | 21.813 | -2.151 | .034 | | 5.876 |
| | Sex * Matrilineal and matrilocal | 35.404 | 26.953 | 1.314 | .192 | | 6.453 |
| | Sex * Patrilineal and patrilocal | 29.770 | 26.353 | 1.130 | .261 | | 14.573 |
| | Sex * Other marriage patterns | 75.449 | 27.447 | 2.749 | .007 | | 3.856 |

# Question 1 d)

- The regional variables may indicate differences in culture as well as correlate with differences in the research teams collecting data in the 3 regions. The north and central region are not so different. But living in the southern region clearly increases generosity compared to living in the central region. Now, it is also the case that the southern region basically is matrilineal and matrilocal while the north basically is patrilineal and patrilocal and the central region mixed but perhaps leaning towards the patrilineal values. This means that there may be inter-correlations between region and marriage system further complicating the interpretation of the coefficients.



# Question 1 d)

- The fact that the interaction terms are significant where the marriage system alone is not, speaks to the reasonable suspicion that being man in a matrilineal culture is very different from being man in a patrilineal culture. But further interpretation depends on re-estimating the model with fewer variables, and more attention to the limited number of cases.