ON UNMITIGATED COMMUNION AND PROBLEM SHARING

Steven W. Knox

Department of Statistics, Carnegie Mellon University

January 5, 2002

Abstract. This report summarizes analysis of data collected by Prof. Vicki Helgeson in a study of the effect of unmitigated communion on a person's daily interactions. Prof. Helgeson posed five questions, each asking whether a person's level of unmitigated communion affects a particular aspect of their daily interactions. This report concludes with a negative answer to each of the five questions.

0. Introduction

A person with high *unmitigated communion* (UC) is one whose need to help others is so great that his or her physical and mental health is degraded by it. The effect of UC on a person's daily interactions is being studied by Professor Vicki Helgeson of the Carnegie Mellon University Department of Psychology. This report summarizes an analysis of Professor Helgeson's data, focussed on the following five questions:

(1) If a problem is shared with a subject in a personal interaction, does a high UC subject feel better about the interaction that a normal UC subject?

(2) If a subject shares a problem in a personal interaction, does a high UC subject perceive the share-ee as less helpful than does a normal UC subject?

(3) Is a subject's UC related to the proportion of personal interactions in which a problem is shared with him or her?

(4) Are high UC subjects different from normal UC subjects in how they think other people feel after interacting with them?

(5) Are high UC subjects different from normal UC subjects in how they feel about themselves after a personal interaction?

Section 1 of this paper describes the data and Section 2 provides detail on UC and identification of interactions in which a problem is shared with a subject. Sections 3 through 7 treat Questions 1 through 5, one section per question. Conclusions are stated very briefly in Section 8.

1. The Data

A group of first-year students at Carnegie Mellon University were given a test which rated their UC on a continuous scale from 0 to 5. Twenty students from among the top third in UC score and twenty-one from among the middle third participated in a 7-day study of their personal interactions.

At the end of each study-day, each participant completed a telephone survey form for every personal interaction of at least ten minutes' length he or she had that day. Except for the subjects' UC scores,

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

all data used in this report came from these interaction surveys. Additionally, any personal interaction of less than ten minutes' length which involved conflict was also reported in a survey form and each participant completed one survey form per day which assessed their physical and mental health. Copies of the survey forms which include the variable names used in this paper are attached.

The "simple" variables of interest in this study are UC (a single number) and responses to the interaction survey questions 6, 8, 9 and 10. Indicating the day by d (1 through 7) and the interaction number by n (1 through 16), the responses to these questions are coded in the seven variables listed in Table 1.1. From these simple variables, "compound" variables, such as the total number of interactions per day, will be constructed as needed.

Variable	**Question**	**Answers**
QdnSHR	6. Did anyone share a problem during the interaction?	1 — Self & Other 2 — Self 3 — Other 4 — Neither
QdnHLP	6. Did you feel that your response was helpful?	1 — Very Unhelpful ⋮ 5 — Very Helpful
QdnHPU	6. Was the other person(s) response helpful?	1 — Very Unhelpful ⋮ 5 — Very Helpful
QdnHAP	8. How did you feel after the interaction?	1 — Very Unhappy ⋮ 5 — Very Happy
QdnCLM	8. How did you feel after the interaction?	1 — Very Anxious ⋮ 5 — Very Calm
QdnOHA	9. How do you *think* the other person(s) felt after the interaction?	1 — Very Unhappy ⋮ 5 — Very Happy
QdnSLF	10. How did you feel **about yourself** after the interaction?	1 — Very Bad About Self ⋮ 5 — Very Good About Self

TABLE 1.1. The seven variables of interest and their associated questions and possible responses.

2. Two Common Variables

Unmitigated Communion (UC).

A histogram of UC is shown in Figure 2.1, plotted over the interval of possible values, [0,5]. The visible skew is consistent with the selection of subjects for the study (half from the highest third of UC scores, half from the middle third and none from the lowest third).

Interaction Types (QdnSHR).

For the n^{th} interaction on the d^{th} day, QdnSHR takes the value 1 if both the subject him- or herself and at least one other person shared a problem during the interaction, 2 if only the subject shared

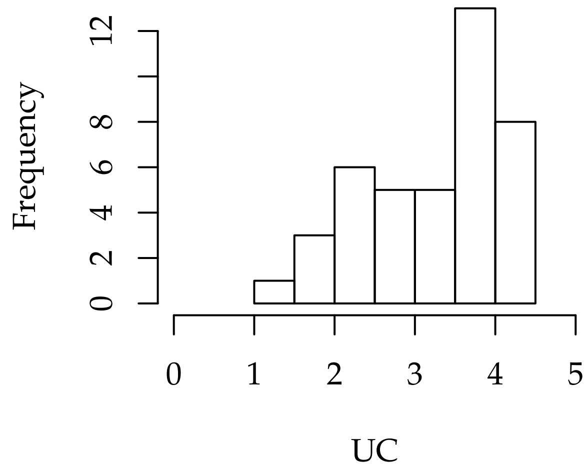

Figure 2.1. A histogram of UC plotted on the interval [0,5].

a problem, 3 if only another person shared a problem and 4 if no one shared a problem. Histograms illustrating the frequencies of the four types of problem-sharing are shown in Figure 2.2. For comparison, a histogram of the total number of interactions is shown in Figure 2.3.

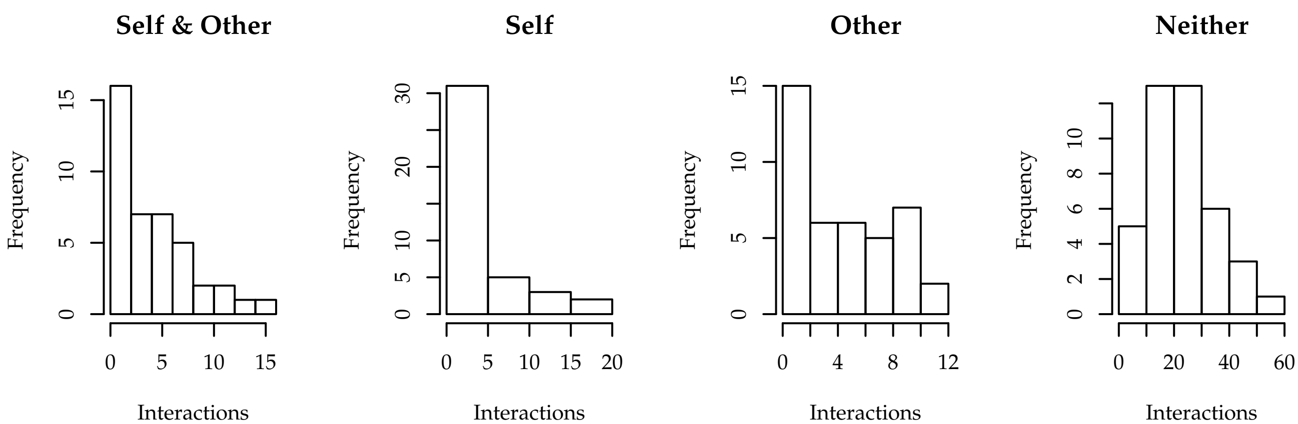

Figure 2.2. Four histograms showing, from left to right, frequencies of interactions in which a study subject or another person shared a problem (QdnSHR = 1, 2, 3 and 4, respectively).

One may ask "in how many interactions did another person share a problem with a subject?" There are two reasonable interpretations of this question: either one counts the number of interactions for which QdnSHR is *either* 1 or 3, or one counts *only* the number of interactions for which QdnSHR is 3. We shall investigate both interpretations in Sections 3 and 5.

3. Question 1

Question 1 asks whether UC is significantly related to the difference in how a subject feels about

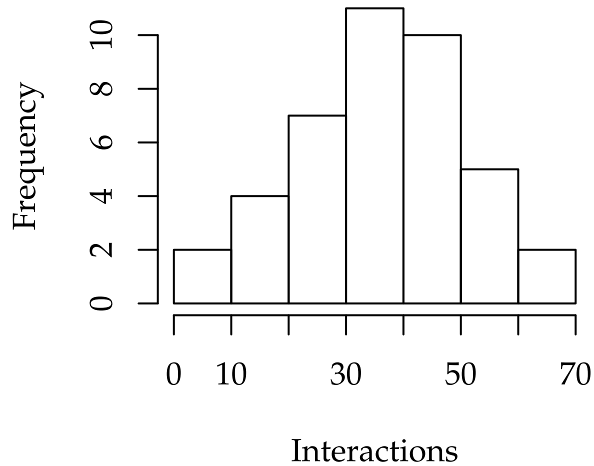

FIGURE 2.3. A histogram of the total number of interactions in the study.

interactions, depending on whether or not a problem was shared with the subject. The results in this section are inconclusive.

There are two variables which measure how a subject felt immediately after an interaction: QdnHAP and QdnCLM. We will consider each of these variables separately. Also, as discussed in Section 2, there are two ways to decide whether a problem was shared in an interaction: QdnSHR is exactly 3 (the "exclusive" case) or QdnSHR is either 1 or 3 (the "inclusive" case). We will consider each of these interpretations separately, so there are a total of four cases to consider in this section.

Exclusive Case.

For each subject, QdnHAP and QdnCLM are averaged over interactions in which problems are shared and over interactions in which problems are not shared, dichotomized according to whether QdnSHR is exactly 3. Four new variables are thus constructed,

$$\begin{aligned}
\text{SHRHAP} &= \text{average Q}dn\text{HAP when a problem is shared} \\
\text{NOSHRHAP} &= \text{average Q}dn\text{HAP when no problem is shared} \\
\text{SHRCLM} &= \text{average Q}dn\text{CLM when a problem is shared} \\
\text{NOSHRCLM} &= \text{average Q}dn\text{CLM when no problem is shared} \; .
\end{aligned}$$

The "happiness effect" HE and "calmness effect" CE are the differences of these,

$$\text{HE} = \text{SHRHAP} - \text{NOSHRHAP} \qquad \text{and} \qquad \text{CE} = \text{SHRCLM} - \text{NOSHRCLM} \; .$$

These are the predictor variables of interest because they measure, in an average sense, how sharing a problem in an interaction affects how a subject feels about the interaction. In terms of these variables, Question 1 asks whether there is a significant relationship between UC and either HE or CE.

Two linear models were fit. In the first model, UC was regressed on the subjects' HE (36 observations — five subjects reported no interactions in which a problem was shared) and in the second, UC was regressed on the subjects' CE (36 observations).

Using all available data, the HE model had a marginally significant slope coefficient and the CE model had a significant slope coefficient at the 5% level. On further investigation, it appeared that subject number 2 was a high-leverage data point (subject 2 is the point at the lower left of the plots in Figure 3.1). Removing subject 2 and refitting the models, neither slope coefficient was significant at the 5% level, though the slope coefficient of the CE model was marginally significant (p-value 0.0995). Figure 3.1 shows scatterplots of the data for the two models together with fitted least-squares lines.

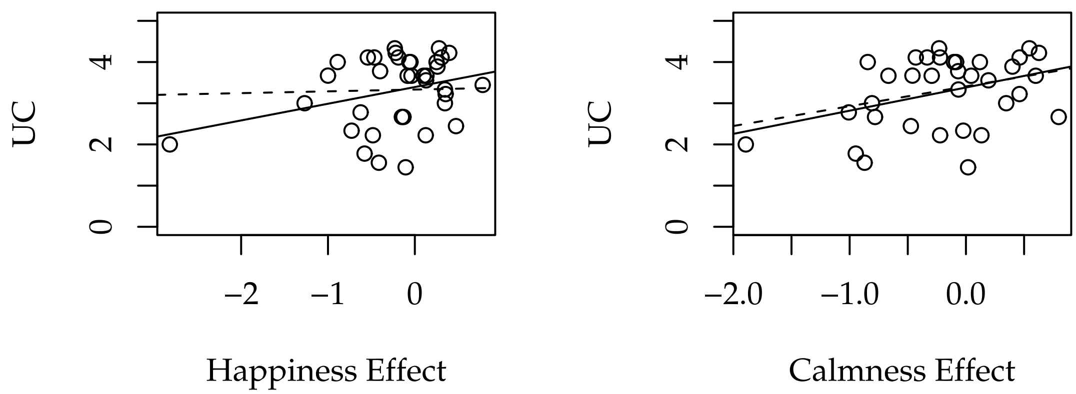

FIGURE 3.1. Scatterplots showing the data and least-squares lines of the two models for Question 1 with an exclusive interpretation of sharing. The solid lines were computed using all data points and the dashed lines were computed without subject number 2.

Inclusive Case.

For each subject, QdnHAP and QdnCLM are averaged over interactions in which problems are shared and over interactions in which problems are not shared, dichotomized according to whether QdnSHR is either 1 or 3 or is either 2 or 4.

As in the exclusive case, SHRHAP, NOSHRHAP, SHRCLM, NOSHRCLM, HE and CE are constructed. In terms of these variables, Question 1 asks whether there is a significant relationship between UC and either HE or CE.

Two linear models were fit. In the first model, UC was regressed on the subjects' HE (39 observations — two subjects reported no interactions in which a problem was shared) and in the second, UC was regressed on the subjects' CE (39 observations).

Using all available data, the HE model did not have a significant slope coefficient and the CE model had a significant slope coefficient at the 5% level. On further investigation, it appeared that subject number 2 was a high-leverage data point (subject 2 is the point at the lower left of the plots in Figure 3.2). Removing subject 2 and refitting the models, neither slope coefficient was significant at the 5% level, though the slope coefficient of the CE model was marginally significant (p-value 0.071). Figure 3.2 shows scatterplots of the data for the two models together with fitted least-squares lines.

4. QUESTION 2

Question 2 asks whether there is a significant relationship between UC and the variable QdnHPU. The answer is probably no. Figure 4.1 shows frequency histograms of QdnHPU for each of the 41 subjects in the study, in decreasing order of UC. No pattern is evident to the eye.

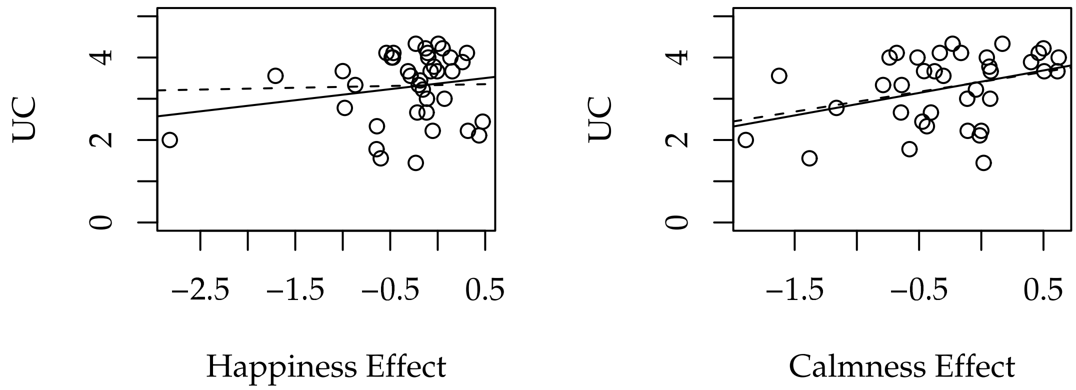

FIGURE 3.2. Scatterplots showing the data and least-squares lines of the two models for Question 1 with an inclusive interpretation of sharing. The solid lines were computed using all data points and the dashed lines were computed without subject number 2.

Three linear models were fit. In the first model, UC was regressed on the subjects' mean QdnHPU (40 observations — subject number 16 reported no relevant interactions) and in the second, UC was regressed on the subjects' QdnHPU standard error (40 observations). In the third model, every interaction of every subject was treated independently (350 observations) and UC was regressed on QdnHPU.

None of the three models produced slope coefficients which were significantly different from zero at the 5% level. Figure 4.2 shows scatterplots of the data for the three models together with fitted least-squares lines.

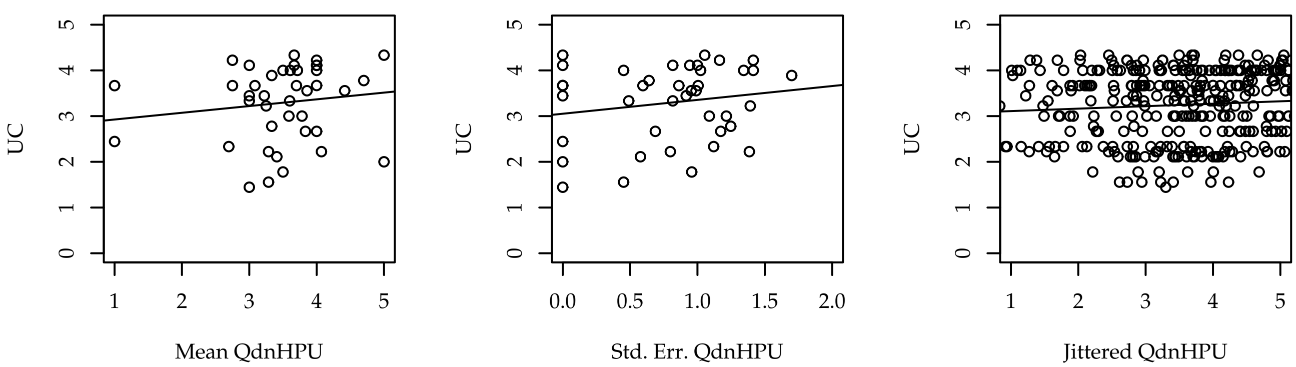

FIGURE 4.2. Scatterplots showing the data and least-squares lines of the three models for Question 2. The data points of the third model have been jittered for visual clarity.

5. QUESTION 3

Question 3 asks whether a subject's UC is related to the proportion of personal interactions in which a problem is shared with him or her. The answer is probably no. This may be concluded from the scatterplots shown in Figure 5.1, in which UC is plotted against the proportion of personal interactions

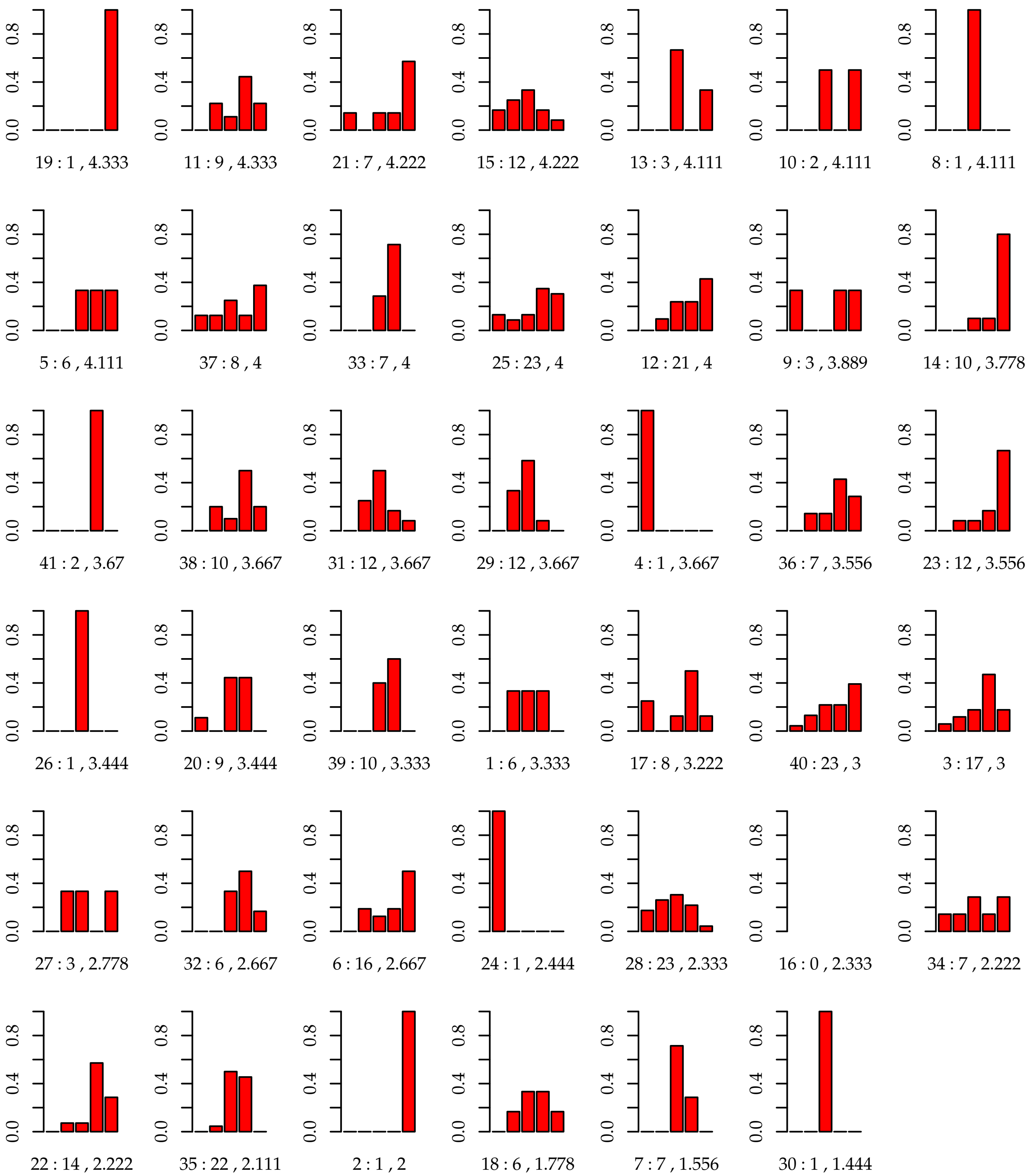

FIGURE 4.1. Frequency histograms of QdnHPU plotted on the domain $\{1, 2, 3, 4, 5\}$, in decreasing order of UC. The numbers below each histogram are (subject number: total interactions reported, UC).

in which another person shares a problem with a subject. The slopes of the least-squares regression lines shown in the scatterplots are not significantly different from zero at the 5% level.

One might also ask whether UC is related to the total number of personal interactions. Again the

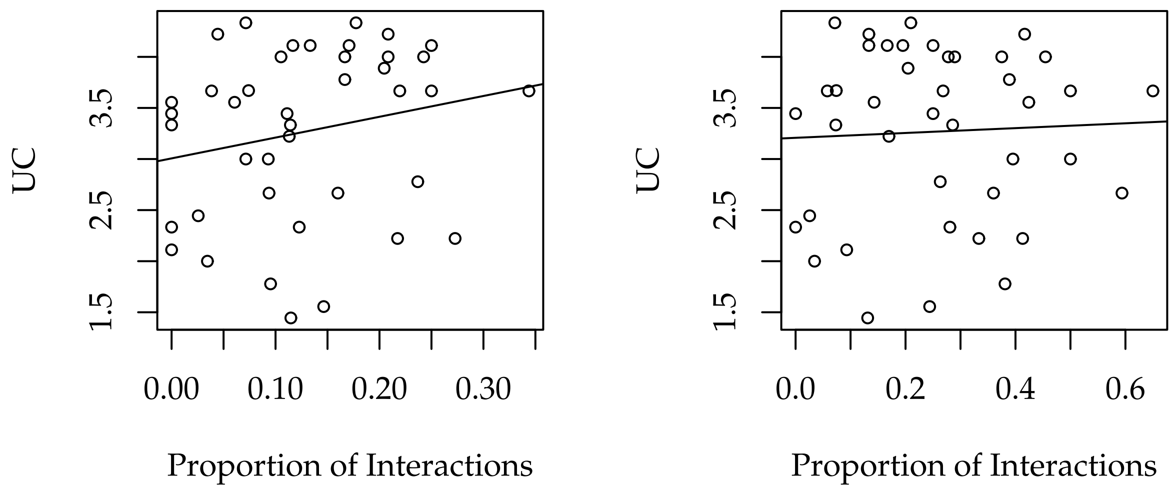

FIGURE 5.1. Scatterplots of UC against the proportion of interactions in which a problem is shared with the subject: only QdnSHR = 3 is counted as sharing in the left-hand plot, while QdnSHR = 1 or 3 is counted as sharing in the right-hand plot. Least-squares lines are shown.

answer is no, as indicated by Figure 5.2. The slope of the least-squares regression line is not significantly different from zero at the 5% level. Categorizing the subjects by sex led to no improvement in significance.

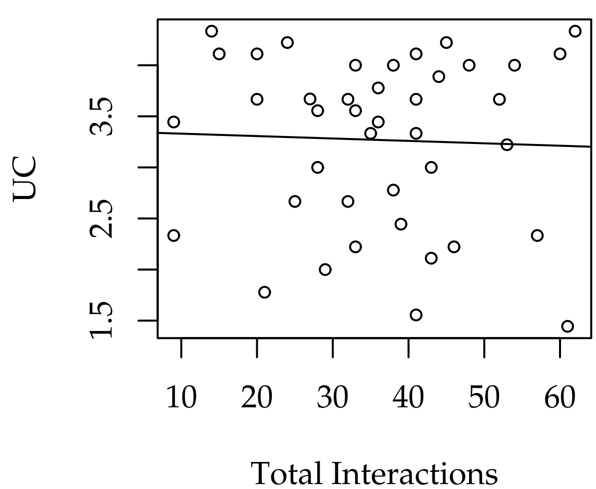

FIGURE 5.2. Scatterplot UC against the total number of interactions had by a subject. The least-squares line is shown.

6. QUESTION 4

Question 4 asks whether there is a significant relationship between UC and the variable QdnOHA.

The answer is probably no.

Figure 6.1 shows frequency histograms of QdnOHA for each of the 41 subjects in the study, in decreasing order of UC. No pattern is evident to the eye.

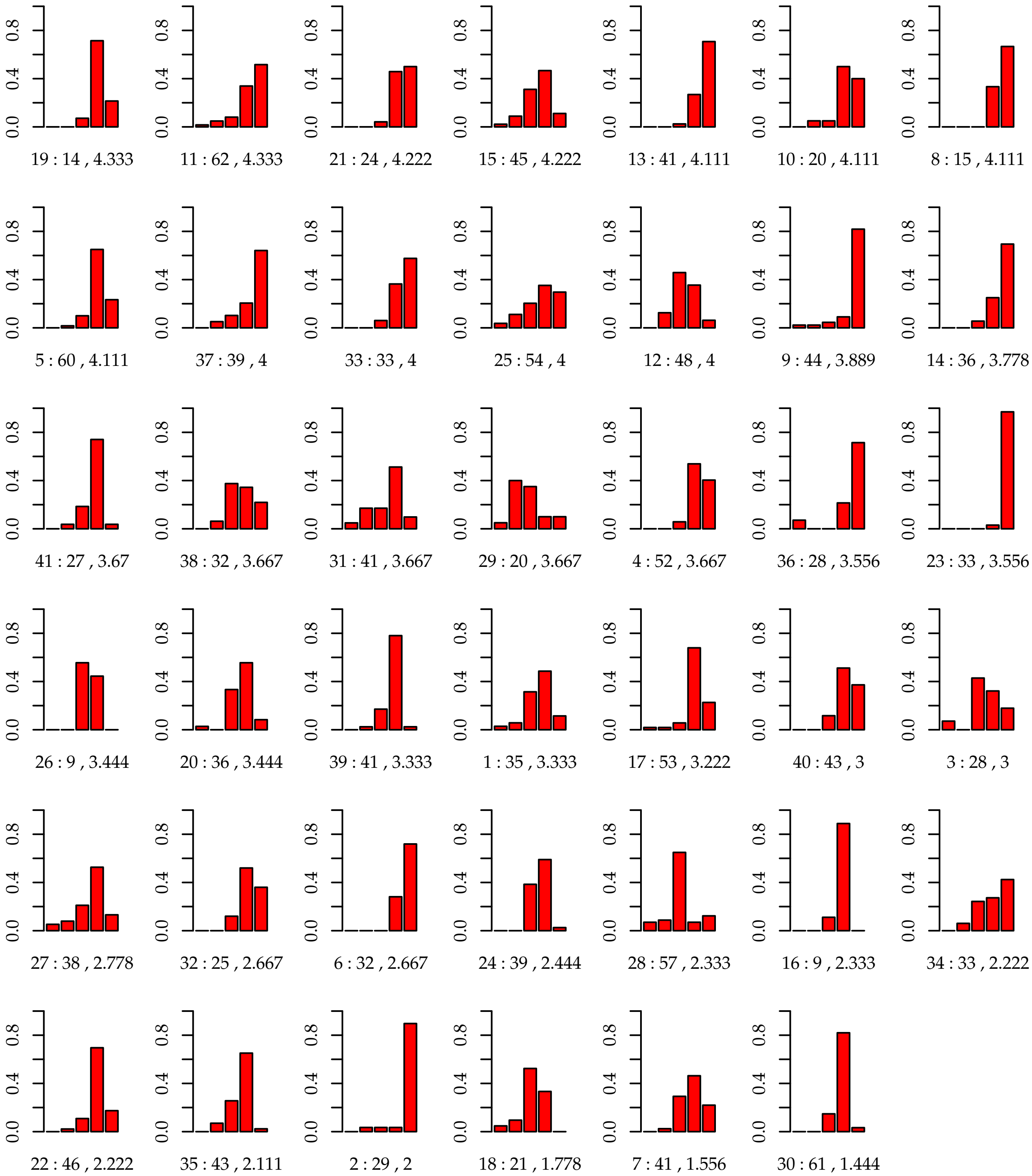

FIGURE 6.1. Frequency histograms of QdnOHA plotted on the domain $\{1, 2, 3, 4, 5\}$, in decreasing order of UC. The numbers below each histogram are (subject number: total interactions reported, UC).

Three linear models were fit. In the first model, UC was regressed on the subjects' mean QdnOHA (41 observations) and in the second, UC was regressed on the subjects' QdnOHA standard error (41 observations). In the third model, every interaction of every subject was treated independently (1487 observations) and UC was regressed on QdnOHA.

Neither of the first two models produced slope coefficients which were significantly different from zero at the 5% level. The third model produced a slope coefficient which was significantly different from zero at the 1% level,

$$\text{UC} \;=\; 0.127 \; \text{Q}dn\text{OHA} \; + 2.749 \; .$$

The adjusted R^2 of the third model is very small (0.0174) and no linearity is visually apparent in the data: we see no explanatory value in this model. Figure 6.2 shows scatterplots of the data for the three models together with fitted least-squares lines.

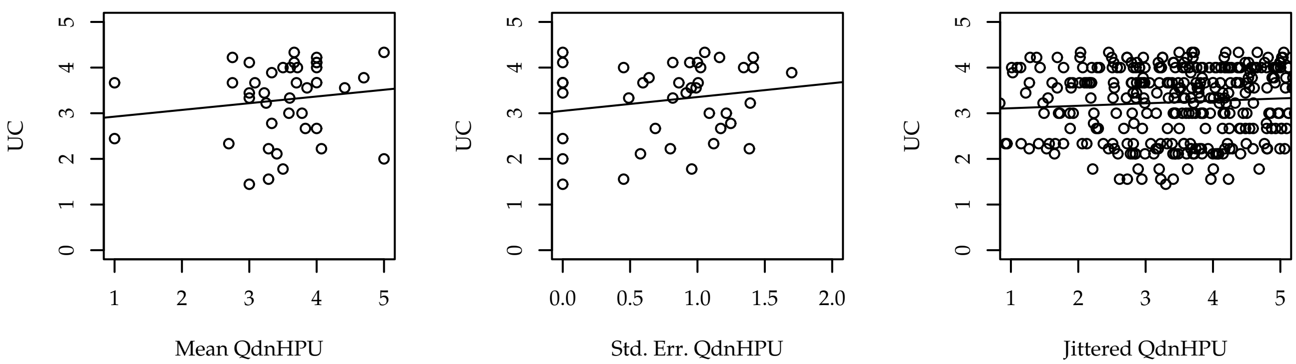

FIGURE 6.2. Scatterplots showing the data and least-squares lines of the three models for Question 2. The data points of the third model have been jittered for visual clarity.

7. Question 5

Question 5 asks whether there is a significant relationship between UC and the variable QdnSLF. The answer is probably no. Figure 7.1 shows frequency histograms of QdnSLF for each of the 41 subjects in the study, in decreasing order of UC. No pattern is evident to the eye.

Three linear models were fit. In the first model, UC was regressed on the subjects' mean QdnSLF (41 observations) and in the second, UC was regressed on the subjects' QdnSLF standard error (41 observations). In the third model, every interaction of every subject was treated independently (1486 observations) and UC was regressed on QdnSLF.

None of the three models produced slope coefficients which were significantly different from zero at the 5% level. Figure 7.2 shows scatterplots of the data for the three models together with fitted least-squares lines.

8. Conclusion

Subjects with high UC do not appear to differ significantly from other subjects in any way addressed by the five questions in the introduction.

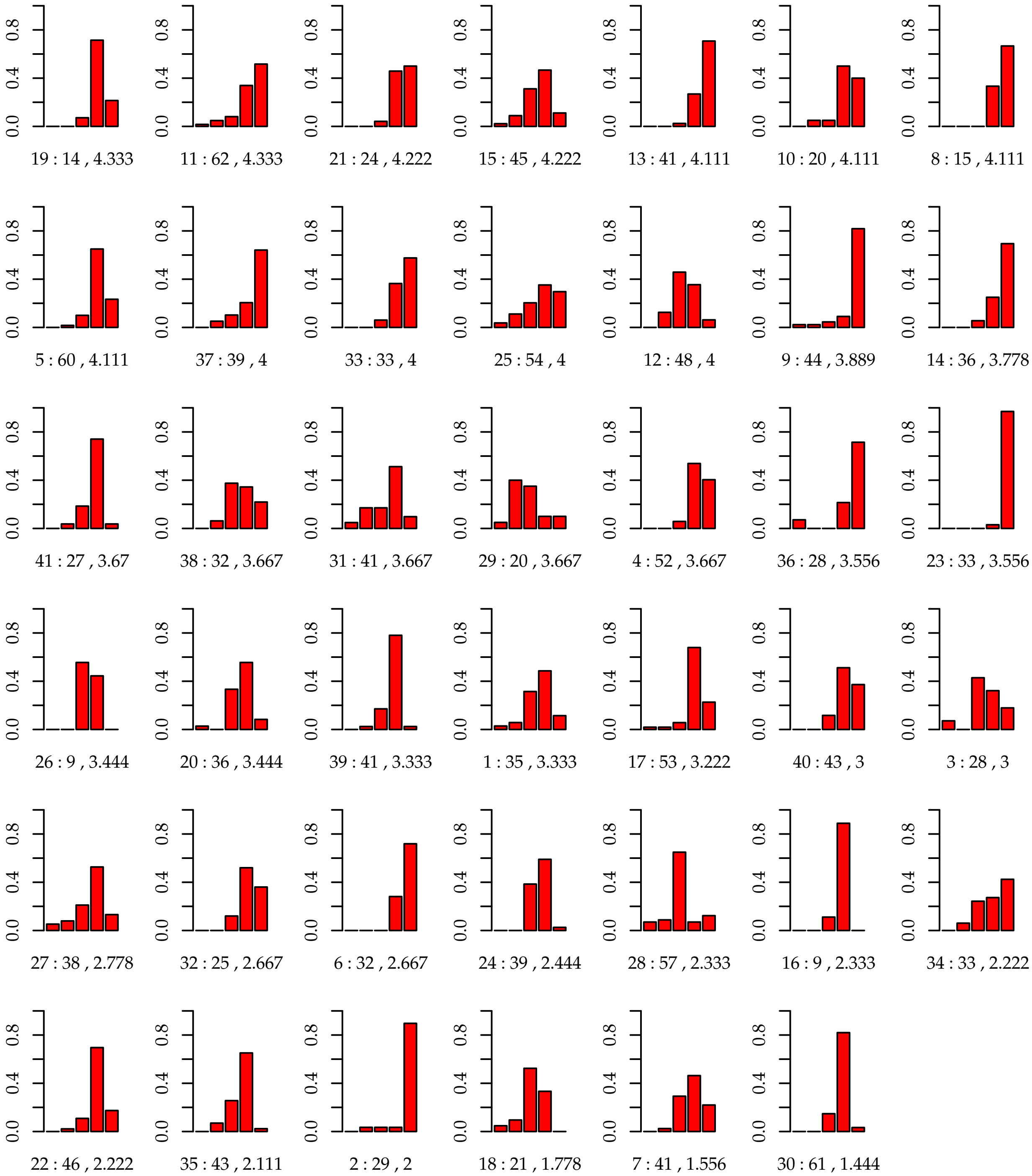

FIGURE 7.1. Frequency histograms of QdnSLF plotted on the domain $\{1, 2, 3, 4, 5\}$, in decreasing order of UC. The numbers below each histogram are (subject number: total interactions reported, UC).

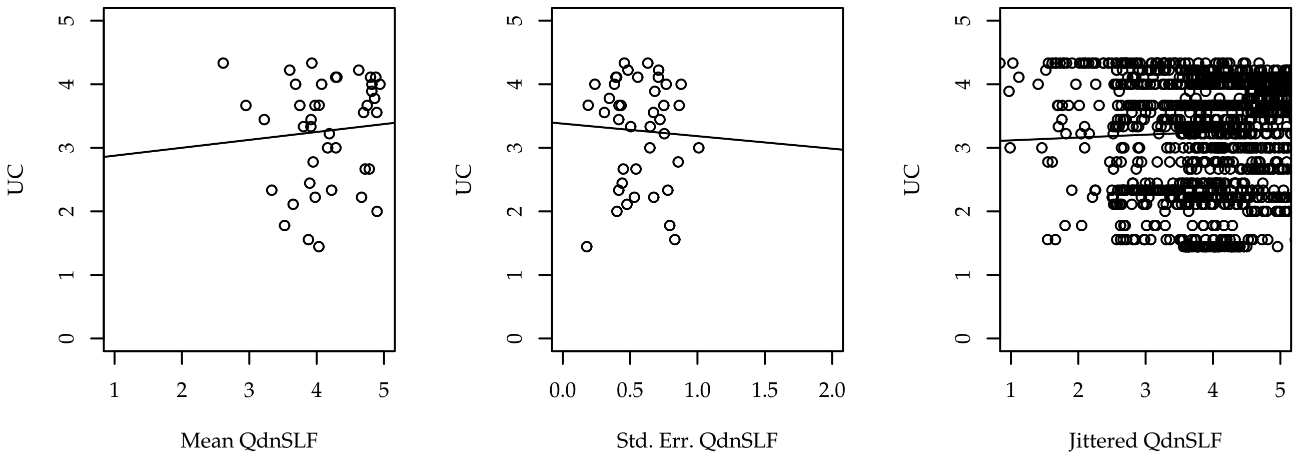

FIGURE 7.2. Scatterplots showing the data and least-squares lines of the three models for Question 2. The data points of the third model have been jittered for visual clarity.

Data Processing.

The data is read in,

```
data_read.table("Yggdrasill:Desktop Folder:Laboratory:36707:Project2:DATA:
                        interaction4b.datSKmac",header=T,na.strings="*");
attach(data);
```

Each day is presented as $480 = 16 \times (16 + 14)$ interaction variables, $30 = 5 \times 6$ conflict variables and 21 end-of-day variables: a total of 531 variables per day.

We create a variable which indexes the days,

```
days_rep(NA,7);
days[1]_3;
for(i in 2:7) days[i]_days[i-1]+531;
days[6]_days[6]+1;
days[7]_days[7]+1;
names(data)[days];
days_c(3,534,1065,1596,2127,2659,3190);
```

We did not check why the two corrections are needed for days 6 and 7. We creat an offset value to get the specific interactions,

```
intoff_30;
allint2_0;
for(i in 0:6) {for(j in 0:15) allint2[16*i+j+1]_days[i+1]+(j*intoff);};
names(data)[allint2];
```

so that now all of the "...SHR", "...HLP", "...HPU" variables can be accessed by

```
shroff_21;
hlpoff_22;
hpuoff_23;
hapoff_26;
clmoff_27;
ohaoff_28;
slfoff_29;
names(data)[allint2+hlpoff];
```

All of the variables of interest are stored in matrices

```
shrmat_data[,allint2+shroff];
hlpmat_data[,allint2+hlpoff];
hpumat_data[,allint2+hpuoff];
hapmat_data[,allint2+hapoff];
clmmat_data[,allint2+clmoff];
ohamat_data[,allint2+ohaoff];
slfmat_data[,allint2+slfoff];
```

We make up variables IT1, IT2, IT3, IT4 and TOTINT for each person, so that ITi is the number of interactions of type Q..SHR $= i$ for $1 \leq i \leq 4$ and TOTINT is the total number of interactions,

```
shrcntmat_matrix(rep(0,5*41),41);
for(s in 1:41){
for(i in 1:112) {
tmp_shrmat[s,i];
     if(tmp != "NA"){
     shrcntmat[s,tmp]_shrcntmat[s,tmp]+1;
shrcntmat[s,5]_shrcntmat[s,5]+1;
}
}
}
IT1_shrcntmat[,1];
IT2_shrcntmat[,2];
IT3_shrcntmat[,3];
IT4_shrcntmat[,4];
TOTINT_shrcntmat[,5];
```

R Code for the Linear Models.

The following R code was used to set up and estimate coefficients for the linear models described in the body of this paper. The details of the fit (e.g. coefficients and p-values) are given in the R output below.

Question 1.

First, consider that "share" means QdnSHR = 3,

```
shrhmeanv_rep(0,41);
shrcntv_rep(0,41);
noshrhmeanv_rep(0,41);
noshrcntv_rep(0,41);
shrcmeanv_rep(0,41);
noshrcmeanv_rep(0,41);
for(j in 1:41){
for(i in 1:112){
tmp_shrmat[j,i];
if(tmp != "NA") {
if(tmp==3) {
shrhmeanv[j]_shrhmeanv[j]+hapmat[j,i];
shrcntv[j]_shrcntv[j]+1;
shrcmeanv[j]_shrcmeanv[j]+clmmat[j,i];
}
if(tmp==1 [OR] tmp==2 [OR] tmp==4) {
noshrhmeanv[j]_noshrhmeanv[j]+hapmat[j,i];
noshrcntv[j]_noshrcntv[j]+1;
noshrcmeanv[j]_noshrcmeanv[j]+clmmat[j,i];
}
}
}
}
diffhv_shrhmeanv/shrcntv-noshrhmeanv/noshrcntv
diffcv_shrcmeanv/shrcntv-noshrcmeanv/noshrcntv
```

Modelling the "happiness effect",

```
fit1_lm(UC~diffhv);
summary(fit1);
            Estimate Std. Error t value Pr(>abs(t))
(Intercept)    3.3928     0.1444  23.492   <2e-16 ***
diffhv         0.4051     0.2193   1.847   0.0734 .
Residual standard error: 0.8256 on 34 degrees of freedom
Multiple R-Squared: 0.09123, Adjusted R-squared: 0.0645
F-statistic: 3.413 on 1 and 34 DF,  p-value: 0.0734
```

It looks as though subject number 2 has high leverage, so we remove it,

```
tUC_UC[-2]
tdiffhv_diffhv[-2]
fit2_lm(tUC~tdiffhv);
summary(fit2);
            Estimate Std. Error t value Pr(>abs(t))
(Intercept)    3.3881     0.1466   23.11   <2e-16 ***
tdiffhv        0.3103     0.3135    0.99    0.329
Residual standard error: 0.8357 on 33 degrees of freedom
Multiple R-Squared: 0.02885, Adjusted R-squared: -0.0005815
F-statistic: 0.9802 on 1 and 33 DF,  p-value: 0.3293
```

Modelling the "calmness effect",

```
fit1_lm(UC~diffcv);
summary(fit1);
            Estimate Std. Error t value Pr(>abs(t))
(Intercept)   3.3818     0.1460  23.158   <2e-16 ***
diffcv        0.5628     0.2444   2.303   0.0279 *
Residual standard error: 0.8125 on 32 degrees of freedom
Multiple R-Squared: 0.1422, Adjusted R-squared: 0.1154
F-statistic: 5.303 on 1 and 32 DF,  p-value: 0.02794
```

It looks as though subject number 2 has high leverage, so we remove it,

```
tUC_UC[-2]
tdiffcv_diffcv[-2]
fit2_lm(tUC~tdiffcv);
summary(fit2);
            Estimate Std. Error t value Pr(>abs(t))
(Intercept)   3.3825     0.1479  22.874   <2e-16 ***
tdiffcv       0.4931     0.2904   1.698   0.0995 .
Residual standard error: 0.8227 on 31 degrees of freedom
Multiple R-Squared: 0.0851, Adjusted R-squared: 0.05559
F-statistic: 2.884 on 1 and 31 DF,  p-value: 0.0995
```

Now consider that "share" means QdnSHR = 1 or 3,

```
shrhmeanv_rep(0,41);
shrcntv_rep(0,41);
noshrhmeanv_rep(0,41);
noshrcntv_rep(0,41);
shrcmeanv_rep(0,41);
noshrcmeanv_rep(0,41);
for(j in 1:41){
for(i in 1:112){
tmp_shrmat[j,i];
if(tmp != "NA") {
if(tmp==1 [OR] tmp==3) {
shrhmeanv[j]_shrhmeanv[j]+hapmat[j,i];
shrcntv[j]_shrcntv[j]+1;
shrcmeanv[j]_shrcmeanv[j]+clmmat[j,i];
}
if(tmp==2 [OR] tmp==4) {
noshrhmeanv[j]_noshrhmeanv[j]+hapmat[j,i];
noshrcntv[j]_noshrcntv[j]+1;
noshrcmeanv[j]_noshrcmeanv[j]+clmmat[j,i];
}
}
}
}
diffhv_shrhmeanv/shrcntv-noshrhmeanv/noshrcntv
diffcv_shrcmeanv/shrcntv-noshrcmeanv/noshrcntv
```

Modelling the "happiness effect",

```
fit1_lm(UC~diffhv);
summary(fit1);
            Estimate Std. Error t value Pr(>abs(t))
(Intercept)   3.3676     0.1503   22.41   <2e-16 ***
diffhv        0.2688     0.2278    1.18    0.245
Residual standard error: 0.8383 on 37 degrees of freedom
Multiple R-Squared: 0.03628, Adjusted R-squared: 0.01023
F-statistic: 1.393 on 1 and 37 DF,  p-value: 0.2455
```

It looks as though subject number 2 has high leverage, so we remove it,

```
fit2_lm(tUC~tdiffhv);
summary(fit2);
            Estimate Std. Error t value Pr(>abs(t))
(Intercept)  3.33172    0.15421  21.605   <2e-16 ***
tdiffhv      0.04344    0.31661   0.137    0.892
Residual standard error: 0.8378 on 36 degrees of freedom
Multiple R-Squared: 0.0005227, Adjusted R-squared: -0.02724
F-statistic: 0.01883 on 1 and 36 DF,  p-value: 0.8916
```

Modelling the "calmness effect",

```
fit1_lm(UC~diffcv);
summary(fit1);
             Estimate Std. Error t value Pr(>abs(t))
(Intercept)   3.4139     0.1464  23.316   <2e-16 ***
diffcv        0.5451     0.2265   2.407   0.0215 *
Residual standard error: 0.7992 on 35 degrees of freedom
Multiple R-Squared: 0.142, Adjusted R-squared: 0.1175
F-statistic: 5.794 on 1 and 35 DF,  p-value: 0.02149
```

It looks as though subject number 2 has high leverage, so we remove it,

```
tUC_UC[-2]
tdiffcv_diffcv[-2]
fit2_lm(tUC~tdiffcv);
summary(fit2);
             Estimate Std. Error t value Pr(>abs(t))
(Intercept)   3.4090     0.1482  23.003   <2e-16 ***
tdiffcv       0.4807     0.2579   1.864    0.071 .
Residual standard error: 0.8074 on 34 degrees of freedom
Multiple R-Squared: 0.09268, Adjusted R-squared: 0.066
F-statistic: 3.473 on 1 and 34 DF,  p-value: 0.07103
```

Question 2.

```
idv_NULL;
UCv_NULL;
hpuv_NULL;
cnts_matrix(rep(0,41*5),41);
k_1;
for(j in 1:41){
for(i in 1:112){
    tmp_hpumat[j,i];
if(tmp != "NA"){
idv[k]_j;
hpuv[k]_tmp;
UCv[k]_UC[j];
cnts[j,tmp]_cnts[j,tmp]+1;
k_k+1;
}
}
}
```

Model 1,

```
meanv_NULL;
for(i in 1:41){
meanv[i]_sum(cnts[i,] * c(1,2,3,4,5))/sum(cnts[i,])
}
```

```
fit_lm(UC~meanv);
summary(fit);
            Estimate Std. Error t value Pr(>abs(t))
(Intercept)   2.7784     0.5992   4.637  4.1e-05 ***
meanv         0.1471     0.1675   0.878    0.385
Residual standard error: 0.8346 on 38 degrees of freedom
Multiple R-Squared: 0.01989, Adjusted R-squared: -0.005898
F-statistic: 0.7713 on 1 and 38 DF,  p-value: 0.3853
```

Model 2,

```
stderrv_NULL;
for(i in 1:41) {
stderrv[i]_0;
for(j in 1:5) stderrv[i]_stderrv[i]+cnts[i,j]*(j-meanv[i])^2;
stderrv[i]_sqrt(stderrv[i]/sum(cnts[i,]));
}
```

```
fit_lm(UC~stderrv);
summary(fit);
            Estimate Std. Error t value Pr(>abs(t))
(Intercept)   3.0516     0.2524  12.090 1.35e-14 ***
stderrv       0.3032     0.2723   1.114    0.272
Residual standard error: 0.8296 on 38 degrees of freedom
Multiple R-Squared: 0.03161, Adjusted R-squared: 0.006122
F-statistic:  1.24 on 1 and 38 DF,  p-value: 0.2724
```

Model 3,

```
fit_lm(UCv~hpuv);
summary(fit);
            Estimate Std. Error t value Pr(>abs(t))
(Intercept)  3.05597    0.13297  22.982   <2e-16 ***
hpuv         0.05357    0.03555   1.507    0.133
Residual standard error: 0.7627 on 348 degrees of freedom
Multiple R-Squared: 0.006483, Adjusted R-squared: 0.003628
F-statistic: 2.271 on 1 and 348 DF,  p-value: 0.1327
```

Question 3.

```
tmp_IT3/TOTINT;
f1_lm(UC~tmp);
summary(f1);
            Estimate Std. Error t value Pr(>abs(t))
(Intercept)    3.006      0.233  12.905 1.11e-15 ***
tmp            2.036      1.507   1.351    0.184
Residual standard error: 0.8267 on 39 degrees of freedom
Multiple R-Squared: 0.04471, Adjusted R-squared: 0.02021
F-statistic: 1.825 on 1 and 39 DF,  p-value: 0.1845
```

```
tmp_IT3/TOTINT;
f1_lm(UC~tmp*dsex);
summary(f1);
            Estimate Std. Error t value Pr(>abs(t))
(Intercept)   3.1413     0.4566   6.880  4.1e-08 ***
tmp          -0.5217     3.2871  -0.159    0.875
dsex         -0.1519     0.5335  -0.285    0.777
tmp:dsex      3.1293     3.7122   0.843    0.405
Residual standard error: 0.8345 on 37 degrees of freedom
Multiple R-Squared: 0.07662, Adjusted R-squared: 0.001755
F-statistic: 1.023 on 3 and 37 DF,  p-value: 0.3934
```

```
tmp_(IT1+IT3)/TOTINT;
f1_lm(UC~tmp);
summary(f1);
            Estimate Std. Error t value Pr(>abs(t))
(Intercept)   3.2076     0.2470  12.985 8.88e-16 ***
tmp           0.2370     0.8146   0.291    0.773
Residual standard error: 0.8449 on 39 degrees of freedom
Multiple R-Squared: 0.002166, Adjusted R-squared: -0.02342
F-statistic: 0.08464 on 1 and 39 DF,  p-value: 0.7726
```

```
tmp_(IT1+IT3)/TOTINT;
f1_lm(UC~tmp*dsex);
summary(f1);
            Estimate Std. Error t value Pr(>abs(t))
(Intercept)   2.8717     0.4859   5.910 8.32e-07 ***
tmp           0.9571     1.8836   0.508    0.614
dsex          0.4814     0.5692   0.846    0.403
tmp:dsex     -1.0161     2.1012  -0.484    0.632
Residual standard error: 0.8571 on 37 degrees of freedom
Multiple R-Squared: 0.02586, Adjusted R-squared: -0.05313
F-statistic: 0.3274 on 3 and 37 DF,  p-value: 0.8056
```

```
f1_lm(UC~TOTINT);
summary(f1);
             Estimate Std. Error t value Pr(>abs(t))
(Intercept)  3.354072   0.377745   8.879 6.63e-11 ***
TOTINT      -0.002364   0.009765  -0.242     0.81
Residual standard error: 0.8452 on 39 degrees of freedom
Multiple R-Squared: 0.001501, Adjusted R-squared: -0.0241
F-statistic: 0.05863 on 1 and 39 DF,  p-value: 0.8099
```

Question 4.

```
idv_NULL;
UCv_NULL;
ohav_NULL;
cnts_matrix(rep(0,41*5),41);
k_1;
for(j in 1:41){
for(i in 1:112){
    tmp_ohamat[j,i];
if(tmp != "NA"){
idv[k]_j;
ohav[k]_tmp;
UCv[k]_UC[j];
cnts[j,tmp]_cnts[j,tmp]+1;
k_k+1;
}
}
}
```

Model 1,

```
meanv_NULL;
for(i in 1:41){
meanv[i]_sum(cnts[i,] * c(1,2,3,4,5))/sum(cnts[i,])
}
fit_lm(UC~meanv);
summary(fit);
            Estimate Std. Error t value Pr(>abs(t))
(Intercept)   1.7189     1.0173   1.690    0.099 .
meanv         0.3873     0.2522   1.535    0.133
Residual standard error: 0.8214 on 39 degrees of freedom
Multiple R-Squared: 0.057, Adjusted R-squared: 0.03282
F-statistic: 2.357 on 1 and 39 DF,  p-value: 0.1328
```

Model 2,

```
stderrv_NULL;
for(i in 1:41) {
stderrv[i]_0;
for(j in 1:5) stderrv[i]_stderrv[i]+cnts[i,j]*(j-meanv[i])^2;
stderrv[i]_sqrt(stderrv[i]/sum(cnts[i,]));
}
fit_lm(UC~stderrv);
summary(fit);
            Estimate Std. Error t value Pr(>abs(t))
(Intercept)   2.9728     0.4489   6.622  7.1e-08 ***
stderrv       0.4164     0.6046   0.689    0.495
Residual standard error: 0.8408 on 39 degrees of freedom
Multiple R-Squared: 0.01201, Adjusted R-squared: -0.01332
F-statistic: 0.4742 on 1 and 39 DF,  p-value: 0.4951
```

Model 3,

```
fit_lm(UCv~ohav);
summary(fit);
            Estimate Std. Error t value Pr(>abs(t))
(Intercept)  2.74900    0.09962  27.596  < 2e-16 ***
ohav         0.12712    0.02432   5.227 1.97e-07 ***
Residual standard error: 0.8445 on 1485 degrees of freedom
Multiple R-Squared: 0.01807, Adjusted R-squared: 0.0174
F-statistic: 27.32 on 1 and 1485 DF,  p-value: 1.968e-07
```

Question 5.

```
idv_NULL;
UCv_NULL;
slfv_NULL;
cnts_matrix(rep(0,41*5),41);
k_1;
for(j in 1:41){
for(i in 1:112){
    tmp_slfmat[j,i];
if(tmp != "NA"){
idv[k]_j;
slfv[k]_tmp;
UCv[k]_UC[j];
cnts[j,tmp]_cnts[j,tmp]+1;
k_k+1;
}
}
}
```

Model 1,

```
meanv_NULL;
for(i in 1:41){
meanv[i]_sum(cnts[i,] * c(1,2,3,4,5))/sum(cnts[i,])
}

fit_lm(UC~meanv);
summary(fit);
            Estimate Std. Error t value Pr(>abs(t))
(Intercept)   2.7522     0.9716   2.833  0.00727 **
meanv         0.1243     0.2318   0.536  0.59483
Residual standard error: 0.8428 on 39 degrees of freedom
Multiple R-Squared: 0.007319, Adjusted R-squared: -0.01813
F-statistic: 0.2876 on 1 and 39 DF,  p-value: 0.5948
```

Model 2,

```
stderrv_NULL;
for(i in 1:41) {
stderrv[i]_0;
for(j in 1:5) stderrv[i]_stderrv[i]+cnts[i,j]*(j-meanv[i])^2;
stderrv[i]_sqrt(stderrv[i]/sum(cnts[i,]));
}

fit_lm(UC~stderrv);
summary(fit);
            Estimate Std. Error t value Pr(>abs(t))
(Intercept)   3.3807     0.3978   8.498 2.08e-10 ***
stderrv      -0.1971     0.6588  -0.299    0.766
Residual standard error: 0.8449 on 39 degrees of freedom
Multiple R-Squared: 0.002291, Adjusted R-squared: -0.02329
F-statistic: 0.08956 on 1 and 39 DF,  p-value: 0.7663
```

Model 3,

```
fit_lm(UCv~slfv);
summary(fit);
            Estimate Std. Error t value Pr(>abs(t))
(Intercept)  3.07238    0.10969  28.009   <2e-16 ***
slfv         0.04471    0.02606   1.715   0.0865 .
Residual standard error: 0.8516 on 1484 degrees of freedom
Multiple R-Squared: 0.001979, Adjusted R-squared: 0.001306
F-statistic: 2.942 on 1 and 1484 DF,  p-value: 0.08651
```