summary_statistics

Zhenlin Kang

2022-10-21

```
library(base)
library(kim)
```

```
## Package attached: kim v0.5.55; Most recent version available on GitHub: v0.5.131
##
## You have an option to update the package with the function `update_kim()`. If you do so, make sure t
```

The data was collected in 2012, measuring preferences for a nationally representative sample for 76 countries. I will use this dataset to explore the impacts of income level and cultural variables on country-level time preference, risk preference, altruism, trust, positive reciprocity, and negative reciprocity.

The survey data has already been cleaned when I retrieved it. I linked the preference data with world development indicators and hofstede cultural measures (see create_dataset.R)

We first look at the distribution of country-level preference variables, measured as standard deviation from the global individual mean.

As shown in the histogram and summary statistics, the median country is socially impatient. Most impatient countries have their patience level within one standard deviation of the global average individual level. In contrast, there is substantial heterogeneity in patience level among those socially patient countries. Similar pattern has been seen in risk taking.

```
hist(country$patience, xlab = "patience", ylab = "frequency"
     , main = "distribution of patience level")
```

distribution of patience level
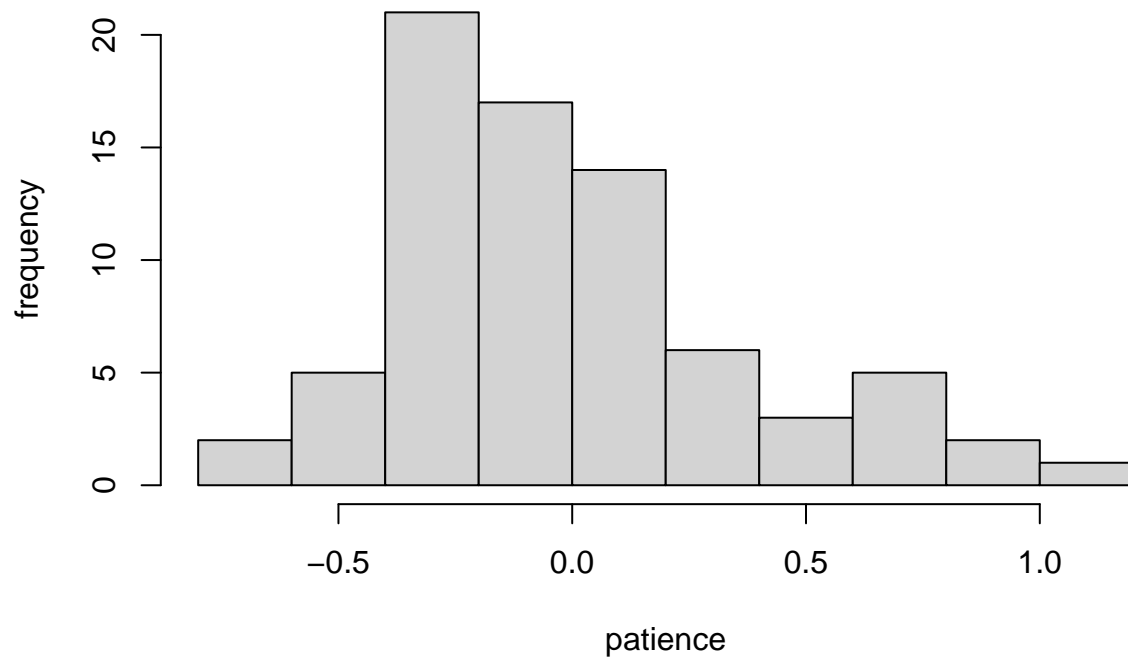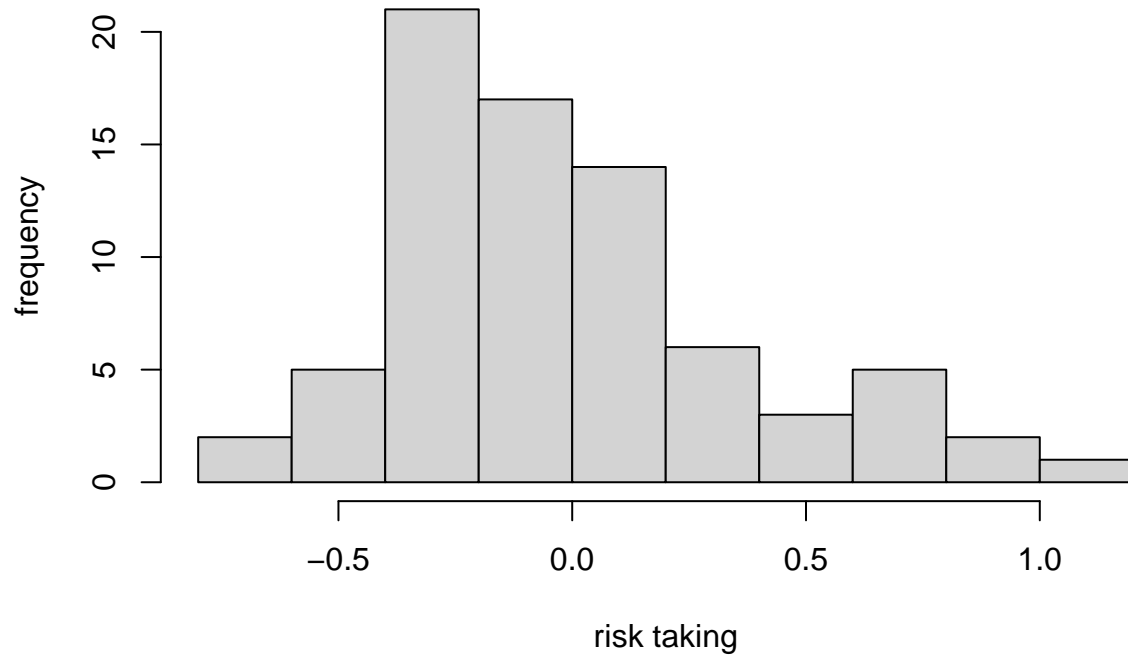


```
summary(country$patience)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -0.612520 -0.257605 -0.092516 -0.003421  0.131628  1.071452
```

```
hist(country$patience, xlab = "risk taking", ylab = "frequency"
    , main = "distribution of risk taking level")
```

distribution of risk taking level
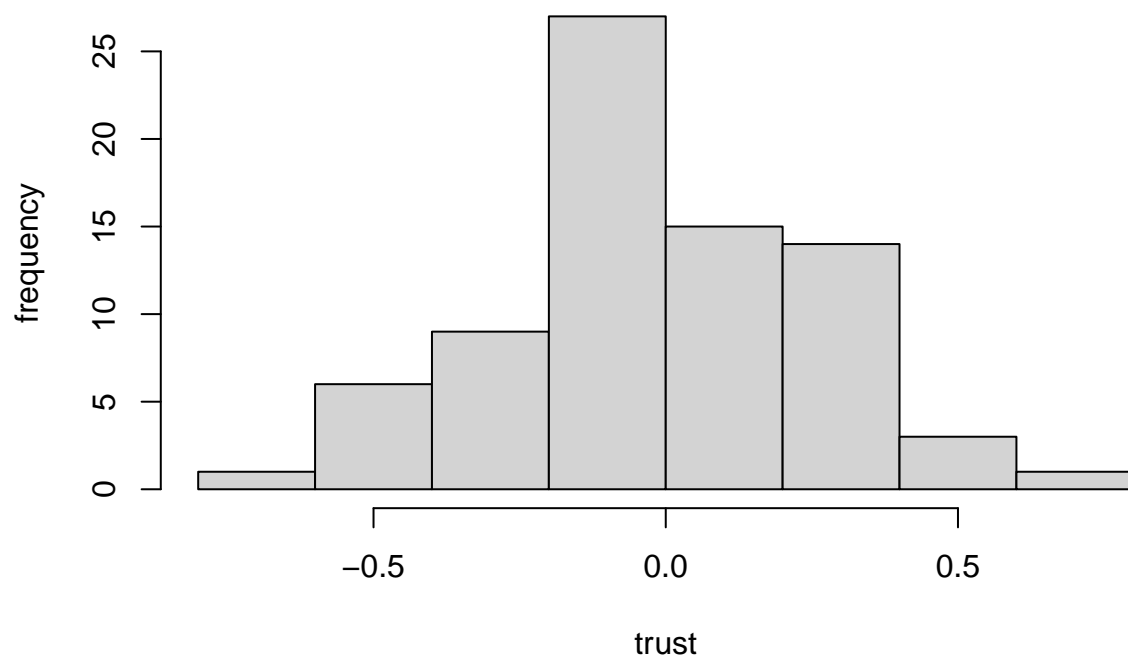

risk taking

```
summary(country$risktaking)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -0.79243 -0.15741 -0.01958  0.01266  0.16339  0.97060
```

Trust, altruism and negative reciprocity level are more symmetrically distributed across countries.

```
hist(country$trust, xlab = "trust", ylab = "frequency"
     , main = "distribution of trust level")
```
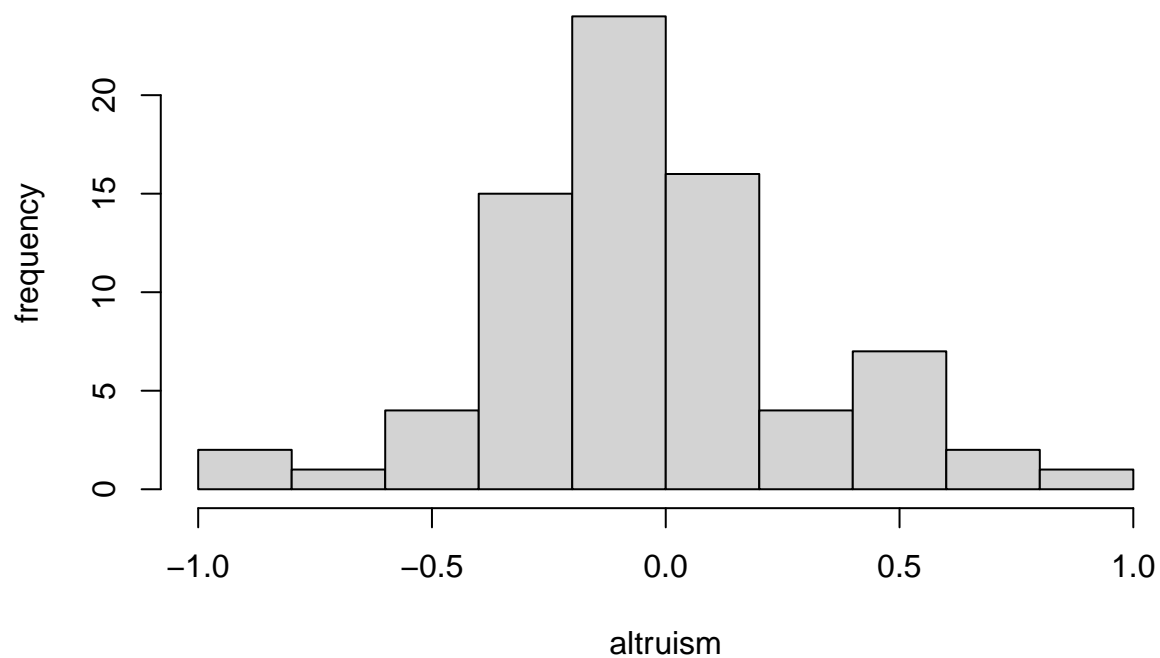
distribution of trust level


```
summary(country$trust)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -0.70644 -0.17740 -0.08117 -0.02215  0.15336  0.60902
```

```
hist(country$altruism, xlab = "altruism", ylab = "frequency"
    , main = "distribution of altruism level")
```

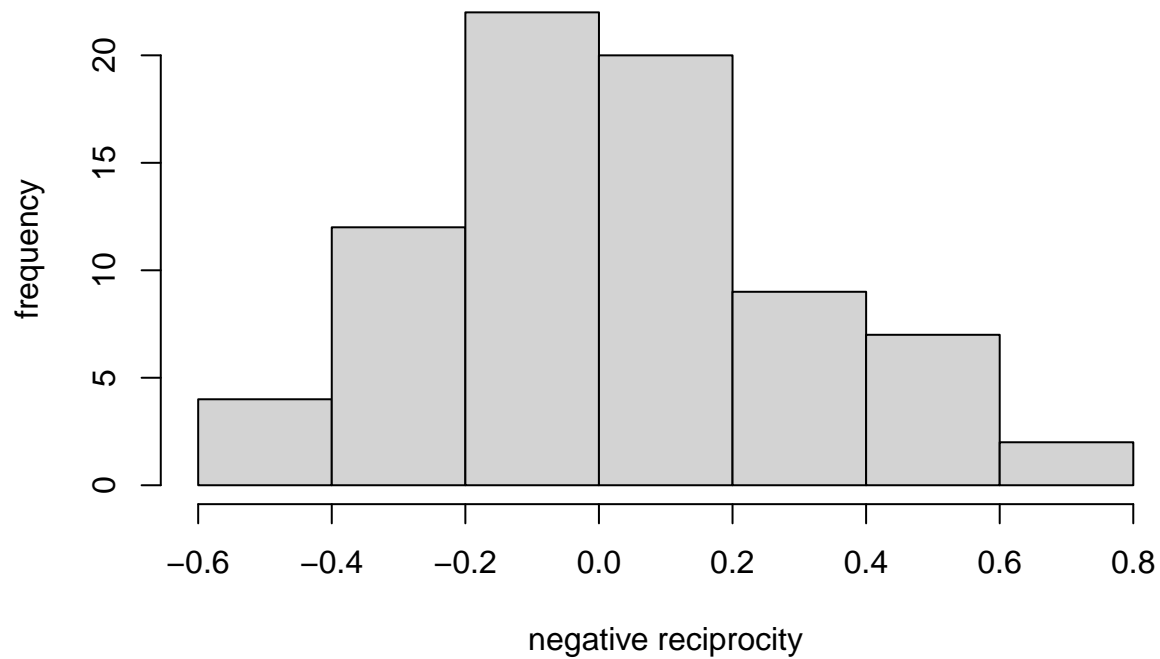**distribution of altruism level**


```
summary(country$altruism)
```

```
##      Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -0.93962 -0.23960 -0.09715 -0.03757  0.15447  0.90646
```

```
hist(country$negrecip, xlab = "negative reciprocity", ylab = "frequency"
     , main = "distribution of negative reciprocity level")
```

distribution of negative reciprocity level


```
summary(country$negrecip)
```
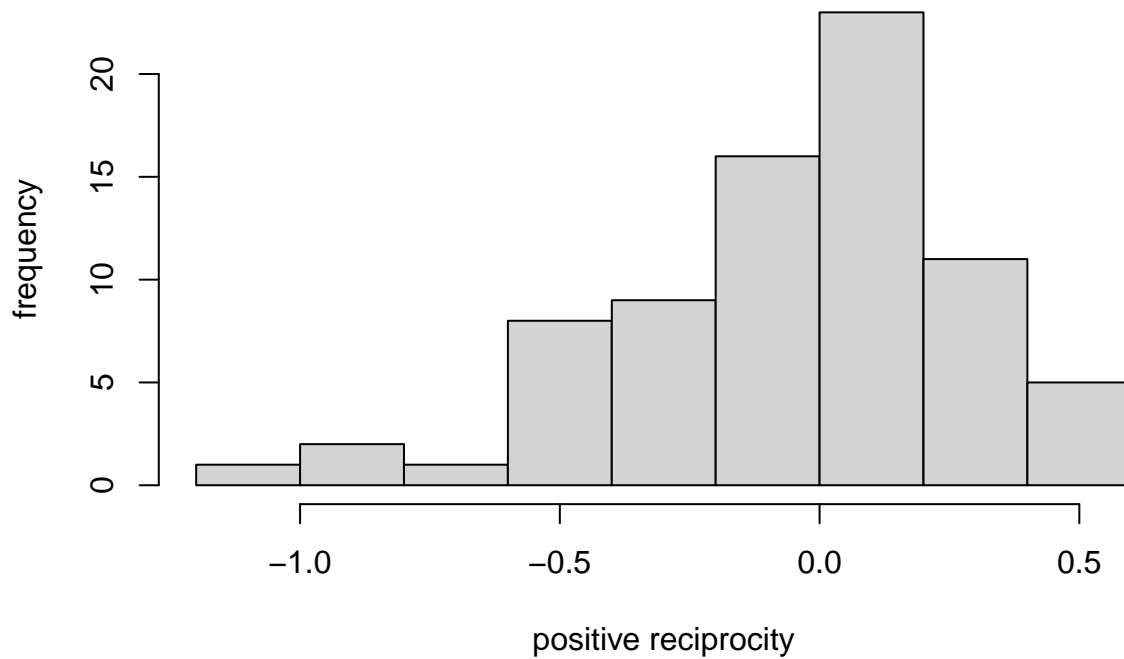
```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -0.489312 -0.167865  0.003909  0.013159  0.182523  0.738908
```

The distribution of positive reciprocity is skewed to the left and more spread out than negative reciprocity.

```
hist(country$posrecip, xlab = "positive reciprocity", ylab = "frequency"
     , main = "distribution of positive reciprocity level")
```

distribution of positive reciprocity level


```
summary(country$posrecip)
```

```
##      Min.  1st Qu.   Median     Mean  3rd Qu.      Max.
## -1.03798 -0.24243  0.02442 -0.03361  0.18695  0.56998
```

Next, we look at the correlation of preferences. The direction of correlation is consistent with expectation

```
preference <- data.frame(matrix(ncol = 0, nrow = 76))
preference$patience <- country$patience
preference$risktaking <- country$risktaking
preference$altruism <- country$altruism
preference$trust <- country$trust
preference$posrecip <- country$posrecip
preference$negrecip <- country$negrecip
correlation <- correlation_matrix(data = preference
, var_names = c("patience", "risktaking", "altruism", "trust", "posrecip", "negrecip")
, output_type = "rp")
correlation
```

```
##       variable patience risktaking altruism  trust posrecip  negrecip
## 1:    patience        1      0.23*    -0.01   0.19     0.02     0.26*
## 2: risktaking    0.23*          1    -0.02  -0.06   -0.26* 0.19 m.s.
## 3:    altruism    -0.01      -0.02        1  0.27*  0.71***    -0.13
## 4:       trust     0.19      -0.06     0.27*      1   0.36**     0.16
## 5:    posrecip     0.02     -0.26*   0.71*** 0.36**        1    -0.15
## 6:    negrecip    0.26*  0.19 m.s.    -0.13   0.16    -0.15         1
```

Finally, we check the correlation between income and preferences

Poorer countries are on average more impatient, consistent with previous findings.

```
# create binary variables for income category
country$low <- ifelse(country$income == "Low income", 1, 0)
country$lowmiddle <- ifelse(country$income == "Lower middle income", 1, 0)
country$uppermiddle <- ifelse(country$income == "Upper middle income", 1, 0)
country$high <- ifelse(country$income == "High income", 1, 0)
reg1 <- lm(patience ~ low + lowmiddle + uppermiddle, data = country)
summary(reg1)
```

```
##
## Call:
## lm(formula = patience ~ low + lowmiddle + uppermiddle, data = country)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -0.68600 -0.20588 -0.00984  0.23156  0.81637
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.25508    0.05901   4.322 4.94e-05 ***
## low         -0.53214    0.16692  -3.188  0.00213 **
## lowmiddle   -0.42690    0.08897  -4.798 8.57e-06 ***
## uppermiddle -0.39309    0.09015  -4.361 4.30e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3123 on 71 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.3143, Adjusted R-squared:  0.2853
## F-statistic: 10.85 on 3 and 71 DF,  p-value: 5.984e-06
```

Risk taking is positively associated with income level. This contradicts expected utility theory (which typically assumes DARA).

```
reg2 <- lm(risktaking ~ low + lowmiddle + uppermiddle, data = country)
summary(reg2)
```

```
##
## Call:
## lm(formula = risktaking ~ low + lowmiddle + uppermiddle, data = country)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -0.75737 -0.17534 -0.00945  0.15604  0.92953
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.03506    0.05807  -0.604    0.548
## low          0.15224    0.16425   0.927    0.357
## lowmiddle    0.06384    0.08754   0.729    0.468
## uppermiddle  0.07613    0.08870   0.858    0.394
##
## Residual standard error: 0.3073 on 71 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.01904,    Adjusted R-squared:  -0.02241
## F-statistic: 0.4593 on 3 and 71 DF,  p-value: 0.7116
```

Altruism level doesn't change monotonically with respect to income level. Note coefficients on low and upper middle income category is insignificant.

```
reg3 <- lm(altruism ~ low + lowmiddle + uppermiddle, data = country)
summary(reg3)
```

```
##
## Call:
## lm(formula = altruism ~ low + lowmiddle + uppermiddle, data = country)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.81529 -0.20748 -0.06104  0.18358  0.85061
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.10159    0.06214  -1.635   0.1065
## low          0.01511    0.17577   0.086   0.9318
## lowmiddle    0.15744    0.09369   1.681   0.0973 .
## uppermiddle  0.10380    0.09493   1.093   0.2779
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3288 on 71 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.04225,    Adjusted R-squared:  0.001784
## F-statistic: 1.044 on 3 and 71 DF,  p-value: 0.3785
```

Let's try to find the relationship between altruism level and GDP per capita. It turns out that the two are uncorrelated.

```
reg4 <- lm(altruism ~ gdppc, data = country)
summary(reg4)
```

```
##
## Call:
## lm(formula = altruism ~ gdppc, data = country)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9011 -0.2028 -0.0663  0.1894  0.9494
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.326e-02  5.248e-02  -0.824    0.412
## gdppc        2.875e-07  2.359e-06   0.122    0.903
##
## Residual standard error: 0.3477 on 73 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.0002035,  Adjusted R-squared:  -0.01349
## F-statistic: 0.01486 on 1 and 73 DF,  p-value: 0.9033
```

Trust level increases in income

```
reg5 <- lm(trust ~ low + lowmiddle + uppermiddle, data = country)
summary(reg5)
```

```
##
## Call:
## lm(formula = trust ~ low + lowmiddle + uppermiddle, data = country)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.58442 -0.16939 -0.01402  0.17059  0.68532
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.06569    0.05075   1.294   0.1997
## low         -0.31304    0.14355  -2.181   0.0325 *
## lowmiddle   -0.14199    0.07651  -1.856   0.0676 .
## uppermiddle -0.08493    0.07752  -1.096   0.2770
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2686 on 71 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.08589,    Adjusted R-squared:  0.04727
## F-statistic: 2.224 on 3 and 71 DF,  p-value: 0.09284
```

Positive reciprocity is uncorrelated with income

```
reg6 <- lm(posrecip ~ low + lowmiddle + uppermiddle, data = country)
summary(reg6)
```

```
##
## Call:
## lm(formula = posrecip ~ low + lowmiddle + uppermiddle, data = country)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.06336 -0.14999  0.07109  0.21881  0.63204
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.03130    0.06561  -0.477    0.635
## low         -0.10463    0.18558  -0.564    0.575
## lowmiddle   -0.03076    0.09891  -0.311    0.757
## uppermiddle  0.05668    0.10022   0.566    0.573
##
## Residual standard error: 0.3472 on 71 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.01518,    Adjusted R-squared:  -0.02643
## F-statistic: 0.3649 on 3 and 71 DF,  p-value: 0.7786
```

Negative reciprocity increases in income

```
reg7 <- lm(negrecip ~ low + lowmiddle + uppermiddle, data = country)
summary(reg7)
```

```
##
## Call:
## lm(formula = negrecip ~ low + lowmiddle + uppermiddle, data = country)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.48443 -0.15680 -0.04821  0.15278  0.63649
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.10242    0.05121   2.000   0.0493 *
## low         -0.19827    0.14483  -1.369   0.1753
## lowmiddle   -0.13930    0.07720  -1.805   0.0754 .
## uppermiddle -0.14828    0.07822  -1.896   0.0621 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.271 on 71 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.07091,    Adjusted R-squared:  0.03165
## F-statistic: 1.806 on 3 and 71 DF,  p-value: 0.1538
```

We now investigate the relationship between Hofstede cultural dimensions and preferences. The definitions of Hofstede cultural dimensions are given below:

PDI (power distance index) measures the degree to which the less powerful members of a society accept and expect that power is distributed unequally

IDV (individualism vs collectivism)

MAS (masculinity vs femininity)

UAI (uncertainty avoidance index) measures the degree to which the members of a society feel uncomfortable with uncertainty and ambiguity.

LTOWVS (long term orientation vs short term normative orientation: societies who score low on this dimension, for example, prefer to maintain time-honoured traditions and norms while viewing societal change with suspicion.

Surprisingly, country-level patience is uncorrelated with long term vs short term orientation and uncertainty avoidance index (which is also a measure for risk preference and risk preference is correlated with time preference as shown in the correlation matrix). Masculine societies are more impatient, in line with the gender differences in patience at individual level.

```
reg8 <- lm(patience ~ gdppc + uai + ltowvs + mas, data = country)
summary(reg8)
```

```
##
## Call:
## lm(formula = patience ~ gdppc + uai + ltowvs + mas, data = country)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.21930 -0.12010  0.00614  0.09756  0.33961
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.016e-01  1.702e-01   0.597   0.5568
## gdppc        1.933e-05  2.154e-06   8.976 1.24e-08 ***
## uai         -3.302e-03  1.962e-03  -1.683   0.1072
## ltowvs       1.159e-03  1.424e-03   0.814   0.4248
## mas         -3.766e-03  1.588e-03  -2.371   0.0274 *
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1531 on 21 degrees of freedom
##   (50 observations deleted due to missingness)
## Multiple R-squared:  0.8742, Adjusted R-squared:  0.8502
## F-statistic: 36.47 on 4 and 21 DF,  p-value: 3.596e-09
```

Risk taking is uncorrelated with UAI. Note that 50 observations are dropped due to missing values. That's probably why we didn't find proper correlations.

```
reg9 <- lm(risktaking ~ gdppc + uai + idv + ltowvs + mas, data = country)
summary(reg9)
```

```
##
## Call:
## lm(formula = risktaking ~ gdppc + uai + idv + ltowvs + mas, data = country)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.40549 -0.10184  0.02477  0.10998  0.20336
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.432e-01  1.919e-01    0.746   0.4641
## gdppc        6.690e-06  2.735e-06    2.446   0.0238 *
## uai          7.868e-04  2.158e-03    0.365   0.7193
## idv         -2.440e-03  2.062e-03   -1.183   0.2506
## ltowvs      -4.119e-03  1.454e-03   -2.832   0.0103 *
## mas         -1.869e-03  1.755e-03   -1.065   0.2996
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1557 on 20 degrees of freedom
##   (50 observations deleted due to missingness)
## Multiple R-squared:  0.4352, Adjusted R-squared:  0.294
## F-statistic: 3.082 on 5 and 20 DF,  p-value: 0.0319
```

```
reg10 <- lm(posrecip ~ gdppc + pdi + idv + ltowvs + mas, data = country)
summary(reg10)
```

```
##
## Call:
## lm(formula = posrecip ~ gdppc + pdi + idv + ltowvs + mas, data = country)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.91028 -0.09274  0.02078  0.13253  0.49880
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.197e-01  3.810e-01    0.839    0.411
## gdppc        4.129e-06  6.457e-06    0.639    0.530
## pdi         -3.775e-03  4.391e-03   -0.860    0.400
## idv         -1.552e-03  3.684e-03   -0.421    0.678
## ltowvs      -3.357e-03  2.780e-03   -1.207    0.241
```

```
## mas          -7.660e-04  3.137e-03  -0.244    0.810
##
## Residual standard error: 0.3008 on 20 degrees of freedom
##   (50 observations deleted due to missingness)
## Multiple R-squared:  0.1962, Adjusted R-squared:  -0.004768
## F-statistic: 0.9763 on 5 and 20 DF,  p-value: 0.4561
```

```r
reg11 <- lm(negrecip ~ gdppc + pdi + idv + ltowvs + mas, data = country)
summary(reg11)
```

```
##
## Call:
## lm(formula = negrecip ~ gdppc + pdi + idv + ltowvs + mas, data = country)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.42797 -0.15902 -0.01917  0.18966  0.46405
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.549e-01  3.210e-01  -1.106    0.282
## gdppc        5.402e-06  5.441e-06   0.993    0.333
## pdi          3.179e-03  3.700e-03   0.859    0.400
## idv         -2.270e-03  3.104e-03  -0.731    0.473
## ltowvs       4.188e-03  2.343e-03   1.787    0.089 .
## mas          8.305e-04  2.643e-03   0.314    0.757
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2535 on 20 degrees of freedom
##   (50 observations deleted due to missingness)
## Multiple R-squared:  0.2058, Adjusted R-squared:  0.007221
## F-statistic: 1.036 on 5 and 20 DF,  p-value: 0.4236
```

```r
reg12 <- lm(trust ~ gdppc + pdi + idv + ltowvs + mas, data = country)
summary(reg12)
```

```
##
## Call:
## lm(formula = trust ~ gdppc + pdi + idv + ltowvs + mas, data = country)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.42309 -0.11588 -0.03899  0.13022  0.45753
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.910e-01  2.773e-01   1.771   0.0919 .
## gdppc       -5.492e-06  4.700e-06  -1.168   0.2564
## pdi         -6.036e-03  3.196e-03  -1.889   0.0735 .
## idv          4.490e-03  2.682e-03   1.675   0.1096
## ltowvs      -1.913e-03  2.024e-03  -0.945   0.3559
## mas         -3.681e-03  2.283e-03  -1.612   0.1226
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## 
## Residual standard error: 0.219 on 20 degrees of freedom
##   (50 observations deleted due to missingness)
## Multiple R-squared:  0.3559, Adjusted R-squared:  0.1949
## F-statistic:  2.21 on 5 and 20 DF,  p-value: 0.09359
```

Need to find a more complete dataset for cultural variables.