Psychology 405: Psychometric Theory

Homework Problem Set #2

Department of Psychology
Northwestern University
Evanston, Illinois USA

April, 2013
Outline

1. The problem, part 1)
2. The Problem, Part 2)
A data set of 1,000 simulated cases may be found using the following commands in R

```r
datafilename <- "http://personality-project.org/R/datasets/psychometrics.prob2.txt"
dataset <- read.table(datafilename, header=TRUE) #read the data file
```

Get the data set and find the basic descriptive statistics. Then plot GREV versus GREQ against each other.

Find the correlation matrix of all of the variables in the data set.

Show the scatter plot of all the variables

Find the multiple correlation of Verbal and Quant with MA
### Read the data and describe it

```r
# read the data file
dataset <- read.table(datafilename, header=TRUE)
describe(dataset)
```

<table>
<thead>
<tr>
<th>var</th>
<th>n</th>
<th>mean</th>
<th>sd</th>
<th>median</th>
<th>trimmed</th>
<th>mad</th>
<th>min</th>
<th>max</th>
<th>range</th>
<th>skew</th>
<th>kurtosis</th>
<th>se</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>1</td>
<td>1000</td>
<td>500.50</td>
<td>288.82</td>
<td>500.50</td>
<td>370.65</td>
<td>1.0</td>
<td>1000.00</td>
<td>999.00</td>
<td>0.00</td>
<td>-1.20</td>
<td>9.13</td>
</tr>
<tr>
<td>GREV</td>
<td>2</td>
<td>1000</td>
<td>499.77</td>
<td>106.11</td>
<td>497.50</td>
<td>106.01</td>
<td>138.0</td>
<td>873.00</td>
<td>735.00</td>
<td>0.09</td>
<td>-0.07</td>
<td>3.36</td>
</tr>
<tr>
<td>GREQ</td>
<td>3</td>
<td>1000</td>
<td>500.53</td>
<td>103.85</td>
<td>498.00</td>
<td>105.26</td>
<td>191.0</td>
<td>914.00</td>
<td>723.00</td>
<td>0.22</td>
<td>0.08</td>
<td>3.28</td>
</tr>
<tr>
<td>GREA</td>
<td>4</td>
<td>1000</td>
<td>498.13</td>
<td>100.45</td>
<td>495.00</td>
<td>99.33</td>
<td>207.0</td>
<td>848.00</td>
<td>641.00</td>
<td>-0.02</td>
<td>0.02</td>
<td>3.18</td>
</tr>
<tr>
<td>Ach</td>
<td>5</td>
<td>1000</td>
<td>49.93</td>
<td>9.84</td>
<td>50.00</td>
<td>49.88</td>
<td>10.38</td>
<td>79.00</td>
<td>63.00</td>
<td>0.00</td>
<td>0.02</td>
<td>0.31</td>
</tr>
<tr>
<td>Anx</td>
<td>6</td>
<td>1000</td>
<td>50.32</td>
<td>9.91</td>
<td>50.00</td>
<td>50.43</td>
<td>10.38</td>
<td>78.00</td>
<td>64.00</td>
<td>-0.14</td>
<td>0.14</td>
<td>0.31</td>
</tr>
<tr>
<td>Prelim</td>
<td>7</td>
<td>1000</td>
<td>10.03</td>
<td>1.06</td>
<td>10.00</td>
<td>10.02</td>
<td>1.48</td>
<td>7.0</td>
<td>13.00</td>
<td>6.00</td>
<td>-0.02</td>
<td>-0.01</td>
</tr>
<tr>
<td>GPA</td>
<td>8</td>
<td>1000</td>
<td>4.00</td>
<td>0.50</td>
<td>4.02</td>
<td>4.01</td>
<td>0.53</td>
<td>2.5</td>
<td>5.38</td>
<td>2.88</td>
<td>-0.07</td>
<td>-0.29</td>
</tr>
<tr>
<td>MA</td>
<td>9</td>
<td>1000</td>
<td>3.00</td>
<td>0.49</td>
<td>3.00</td>
<td>3.00</td>
<td>0.44</td>
<td>1.4</td>
<td>4.50</td>
<td>3.10</td>
<td>-0.07</td>
<td>-0.02</td>
</tr>
</tbody>
</table>
```
Find the correlations

Many ways of doing this.

```r
cor(dataset)
round(cor(dataset),2)
round(cor(dataset,use="pairwise"),2)
lowerCor(dataset)

> cor(dataset)
```

```
       ID   GREV   GREQ   GREA      Ach  Anx Prelim
ID 1.0000000 -0.0072  0.003  -0.0053 -0.0040 0.018  0.0176
GREV 0.00725  1.0000  0.7288  0.6411  0.0056 0.0102  0.4282
GREQ 0.00282  0.7288  1.0000  0.5963  0.0068 0.0055  0.3831
GREA -0.00534 0.6411  0.5963  1.0000 -0.3896 -0.5562  0.5724
Ach -0.00405 0.0056  0.0068  0.4534  1.0000 -0.5562  0.3034
Anx -0.00912 0.0054  0.0055 -0.3896 -0.5562 1.0000  0.2279
Prelim 0.0176  0.4282  0.3831  0.5724  0.3034 -0.2279  1.0000
GPA -0.0049  0.4195  0.3669  0.5162  0.2763  0.2224  0.3065
MA -0.0094  0.3223  0.2874  0.4545  0.2635 -0.2192  0.3065
       GPA    MA
ID -0.0049 0.0094
GREV 0.4195 0.3223
GREQ 0.3669 0.2874
GREA 0.5162 0.4545
Ach  0.2763 0.2635
Anx -0.2224 -0.2192
GPA  0.3065 0.3065
MA  0.3065 0.3065`
```
Round the answer to 2 decimals

\[
\text{> \, \text{\texttt{round(cor(dataset),2)}}}
\]

<table>
<thead>
<tr>
<th></th>
<th>ID</th>
<th>GREV</th>
<th>GREQ</th>
<th>GREA</th>
<th>Ach</th>
<th>Anx</th>
<th>Prelim</th>
<th>GPA</th>
<th>MA</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>1.00</td>
<td>-0.01</td>
<td>0.00</td>
<td>-0.01</td>
<td>0.00</td>
<td>-0.01</td>
<td>0.02</td>
<td>0.00</td>
<td>-0.01</td>
</tr>
<tr>
<td>GREV</td>
<td>-0.01</td>
<td>1.00</td>
<td>0.73</td>
<td>0.64</td>
<td>0.01</td>
<td>0.01</td>
<td>0.43</td>
<td>0.42</td>
<td>0.32</td>
</tr>
<tr>
<td>GREQ</td>
<td>0.00</td>
<td>0.73</td>
<td>1.00</td>
<td>0.60</td>
<td>0.01</td>
<td>0.01</td>
<td>0.38</td>
<td>0.37</td>
<td>0.29</td>
</tr>
<tr>
<td>GREA</td>
<td>-0.01</td>
<td>0.64</td>
<td>0.60</td>
<td>1.00</td>
<td>0.45</td>
<td>-0.39</td>
<td>0.57</td>
<td>0.52</td>
<td>0.45</td>
</tr>
<tr>
<td>Ach</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.45</td>
<td>1.00</td>
<td>-0.56</td>
<td>0.30</td>
<td>0.28</td>
<td>0.26</td>
</tr>
<tr>
<td>Anx</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>-0.39</td>
<td>-0.56</td>
<td>1.00</td>
<td>-0.23</td>
<td>-0.22</td>
<td>-0.22</td>
</tr>
<tr>
<td>Prelim</td>
<td>0.02</td>
<td>0.43</td>
<td>0.38</td>
<td>0.57</td>
<td>0.30</td>
<td>-0.23</td>
<td>1.00</td>
<td>0.42</td>
<td>0.36</td>
</tr>
<tr>
<td>GPA</td>
<td>0.00</td>
<td>0.42</td>
<td>0.37</td>
<td>0.52</td>
<td>0.28</td>
<td>-0.22</td>
<td>0.42</td>
<td>1.00</td>
<td>0.31</td>
</tr>
<tr>
<td>MA</td>
<td>-0.01</td>
<td>0.32</td>
<td>0.29</td>
<td>0.45</td>
<td>0.26</td>
<td>-0.22</td>
<td>0.36</td>
<td>0.31</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Use pairwise correlations (to handle missing data), round the output

```r
> round(cor(dataset,use="pairwise"),2)

<table>
<thead>
<tr>
<th></th>
<th>ID</th>
<th>GREV</th>
<th>GREQ</th>
<th>GREA</th>
<th>Ach</th>
<th>Anx</th>
<th>Prelim</th>
<th>GPA</th>
<th>MA</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>1.00</td>
<td>-0.01</td>
<td>0.00</td>
<td>-0.01</td>
<td>0.00</td>
<td>-0.01</td>
<td>0.02</td>
<td>0.00</td>
<td>-0.01</td>
</tr>
<tr>
<td>GREV</td>
<td>-0.01</td>
<td>1.00</td>
<td>0.73</td>
<td>0.64</td>
<td>0.01</td>
<td>0.01</td>
<td>0.43</td>
<td>0.42</td>
<td>0.32</td>
</tr>
<tr>
<td>GREQ</td>
<td>0.00</td>
<td>0.73</td>
<td>1.00</td>
<td>0.60</td>
<td>0.01</td>
<td>0.01</td>
<td>0.38</td>
<td>0.37</td>
<td>0.29</td>
</tr>
<tr>
<td>GREA</td>
<td>-0.01</td>
<td>0.64</td>
<td>0.60</td>
<td>1.00</td>
<td>0.45</td>
<td>-0.39</td>
<td>0.57</td>
<td>0.52</td>
<td>0.45</td>
</tr>
<tr>
<td>Ach</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.45</td>
<td>1.00</td>
<td>-0.56</td>
<td>0.30</td>
<td>0.28</td>
<td>0.26</td>
</tr>
<tr>
<td>Anx</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>-0.39</td>
<td>-0.56</td>
<td>1.00</td>
<td>-0.23</td>
<td>-0.22</td>
<td>-0.22</td>
</tr>
<tr>
<td>Prelim</td>
<td>0.02</td>
<td>0.43</td>
<td>0.38</td>
<td>0.57</td>
<td>0.30</td>
<td>-0.23</td>
<td>1.00</td>
<td>0.42</td>
<td>0.36</td>
</tr>
<tr>
<td>GPA</td>
<td>0.00</td>
<td>0.42</td>
<td>0.37</td>
<td>0.52</td>
<td>0.28</td>
<td>-0.22</td>
<td>0.42</td>
<td>1.00</td>
<td>0.31</td>
</tr>
<tr>
<td>MA</td>
<td>-0.01</td>
<td>0.32</td>
<td>0.29</td>
<td>0.45</td>
<td>0.26</td>
<td>-0.22</td>
<td>0.36</td>
<td>0.31</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Using the `lowerCor` function, combine these previous operations

```r
> lowerCor(dataset)
```

<table>
<thead>
<tr>
<th></th>
<th>ID</th>
<th>GREV</th>
<th>GREQ</th>
<th>GREA</th>
<th>Ach</th>
<th>Anx</th>
<th>Prelm</th>
<th>GPA</th>
<th>MA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ID</strong></td>
<td><strong>1.00</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GREV</td>
<td>-0.01</td>
<td><strong>1.00</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GREQ</td>
<td>0.00</td>
<td>0.73</td>
<td><strong>1.00</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GREA</td>
<td>-0.01</td>
<td>0.64</td>
<td>0.60</td>
<td><strong>1.00</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ach</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.45</td>
<td><strong>1.00</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anx</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>-0.39</td>
<td>-0.56</td>
<td><strong>1.00</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prelm</td>
<td>0.02</td>
<td>0.43</td>
<td>0.38</td>
<td>0.57</td>
<td>0.30</td>
<td>-0.23</td>
<td><strong>1.00</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPA</td>
<td>0.00</td>
<td>0.42</td>
<td>0.37</td>
<td>0.52</td>
<td>0.28</td>
<td>-0.22</td>
<td>0.42</td>
<td><strong>1.00</strong></td>
<td></td>
</tr>
<tr>
<td>MA</td>
<td>-0.01</td>
<td>0.32</td>
<td>0.29</td>
<td>0.45</td>
<td>0.26</td>
<td>-0.22</td>
<td>0.36</td>
<td>0.31</td>
<td><strong>1.00</strong></td>
</tr>
</tbody>
</table>
Do the significance tests

corr.test does one pair at a time, corr.test does it for all variables

> corr.test(dataset)

Call: corr.test(x = dataset)

Correlation matrix

<table>
<thead>
<tr>
<th></th>
<th>ID</th>
<th>GREV</th>
<th>GREQ</th>
<th>GREA</th>
<th>Ach</th>
<th>Anx</th>
<th>Prelim</th>
<th>GPA</th>
<th>MA</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>1.00</td>
<td>-0.01</td>
<td>0.00</td>
<td>-0.01</td>
<td>0.00</td>
<td>-0.01</td>
<td>0.02</td>
<td>0.00</td>
<td>-0.01</td>
</tr>
<tr>
<td>GREV</td>
<td>-0.01</td>
<td>1.00</td>
<td>0.73</td>
<td>0.64</td>
<td>0.01</td>
<td>0.01</td>
<td>0.43</td>
<td>0.42</td>
<td>0.32</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>MA</td>
<td>-0.01</td>
<td>0.32</td>
<td>0.29</td>
<td>0.45</td>
<td>0.26</td>
<td>-0.22</td>
<td>0.36</td>
<td>0.31</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Sample Size

<table>
<thead>
<tr>
<th></th>
<th>ID</th>
<th>GREV</th>
<th>GREQ</th>
<th>GREA</th>
<th>Ach</th>
<th>Anx</th>
<th>Prelim</th>
<th>GPA</th>
<th>MA</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
</tr>
<tr>
<td>GREV</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>MA</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
</tr>
</tbody>
</table>

Probability values (Entries above the diagonal are adjusted for multiple tests.)

<table>
<thead>
<tr>
<th></th>
<th>ID</th>
<th>GREV</th>
<th>GREQ</th>
<th>GREA</th>
<th>Ach</th>
<th>Anx</th>
<th>Prelim</th>
<th>GPA</th>
<th>MA</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GREV</td>
<td>0.82</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GREQ</td>
<td>0.93</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GREA</td>
<td>0.87</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
pairs.panels(dataset[-1], pch=".", gap=0)
1. Find the multiple R of GREV and GREQ predicting MA score
2. Find the multiple R of GREV, GREQ, GREA prediction prelims
Multiple R of GREV and GREQ predicting MA score

```r
> mod1 <- lm(MA ~ GREV + GREQ, data = dataset)
> summary(mod1)

Call:
  lm(formula = MA ~ GREV + GREQ, data = dataset)

Residuals:
               Min        1Q      Median        3Q        Max
-1.47912 -0.31272   0.01348   0.31216   1.41784

Coefficients:
                               Estimate Std. Error t value Pr(>|t|)
(Intercept)                     2.1690912  0.0771484  28.116  < 2e-16 ***
GREV                             0.0011211  0.0002033   5.515  4.44e-08 ***
GREQ                             0.0005328  0.0002077   2.565   0.0105 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1  1

Residual standard error: 0.4668 on 997 degrees of freedom
Multiple R-squared:  0.1098,  Adjusted R-squared:  0.108
F-statistic: 61.46 on 2 and 997 DF,  p-value: < 2.2e-16
```
Now, add GREA to the model

```r
> mod2 <- lm(MA~GREV+ GREQ+ GREA,data=dataset)
> summary(mod2)
```

Call:
`lm(formula = MA ~ GREV + GREQ + GREA, data = dataset)`

Residuals:

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
<td>-1.37779</td>
<td>-0.29802</td>
<td>0.01818</td>
<td>0.30276</td>
<td>1.38537</td>
</tr>
</tbody>
</table>

Coefficients:

|                  | Estimate | Std. Error | t value | Pr(>|t|) |
|------------------|----------|------------|---------|----------|
| (Intercept)      | 1.844e+00 | 7.842e-02  | 23.518  | <2e-16 *** |
| GREV             | 2.555e-04 | 2.070e-04  | 1.235   | 0.217    |
| GREQ             | -1.986e-05| 2.022e-04  | -0.098  | 0.922    |
| GREA             | 2.076e-03 | 1.865e-04  | 11.131  | <2e-16 *** |

---

Signif. codes:  0 *** 0.001 ** 0.01 * 0.05  0.1  1

Residual standard error: 0.4404 on 996 degrees of freedom
Multiple R-squared: 0.2082, Adjusted R-squared: 0.2059
F-statistic: 87.32 on 3 and 996 DF, p-value: < 2.2e-16
Compare these two models

> anova(mod1, mod2)

Analysis of Variance Table

Model 1: MA ~ GREV + GREQ
Model 2: MA ~ GREV + GREQ + GREA

<table>
<thead>
<tr>
<th>Res.Df</th>
<th>RSS</th>
<th>Df</th>
<th>Sum of Sq</th>
<th>F</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>217.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>193.22</td>
<td>1</td>
<td>24.035</td>
<td>123.89</td>
<td>&lt; 2.2e-16 ***</td>
</tr>
</tbody>
</table>

Signif. codes:  0 ***  0.001 **  0.01 *  0.05 .  0.1  

---
Multiple DVs simultaneously

\[
\begin{align*}
R & \leftarrow \text{lowerCor(dataset[[-1])} \\
\text{set.cor}(y=6:8,x=1:5, data=R)
\end{align*}
\]

Call: set.cor(y = 6:8, x = 1:5, data = R)

Multiple Regression from matrix input

Beta weights

<table>
<thead>
<tr>
<th></th>
<th>Prelim</th>
<th>GPA</th>
<th>MA</th>
</tr>
</thead>
<tbody>
<tr>
<td>GREV</td>
<td>0.14</td>
<td>0.20</td>
<td>0.10</td>
</tr>
<tr>
<td>GREQ</td>
<td>0.04</td>
<td>0.05</td>
<td>0.03</td>
</tr>
<tr>
<td>GREA</td>
<td>0.40</td>
<td>0.29</td>
<td>0.31</td>
</tr>
<tr>
<td>Ach</td>
<td>0.11</td>
<td>0.12</td>
<td>0.10</td>
</tr>
<tr>
<td>Anx</td>
<td>-0.01</td>
<td>-0.05</td>
<td>-0.05</td>
</tr>
</tbody>
</table>

Multiple R

<table>
<thead>
<tr>
<th></th>
<th>Prelim</th>
<th>GPA</th>
<th>MA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.59</td>
<td>0.54</td>
<td>0.47</td>
</tr>
</tbody>
</table>

Multiple R2

<table>
<thead>
<tr>
<th></th>
<th>Prelim</th>
<th>GPA</th>
<th>MA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.34</td>
<td>0.29</td>
<td>0.22</td>
</tr>
</tbody>
</table>

Various estimates of between set correlations

Squared Canonical Correlations

\[
[1] 0.4943 0.0036 0.0017
\]

Chisq of canonical correlations