

Psychology 405: Psychometric Theory

Homework Problem Set #2

Department of Psychology
Northwestern University
Evanston, Illinois USA

April, 2021

Outline

The problem, part 1)

The Problem, Part 2)

The problem (Part 1)

1. A data set of 1,000 simulated cases may be found using the following commands in R

R code

2.

```
library(psych) #need to this once per session
datafilename<-
  "http://personality-project.org/r/datasets/psychometrics.prob2.txt"
dataset <- read.file(datafilename,header=TRUE) #read the data file
```

3. Or, we can read it from the clipboard

R code

```
library(psych) #need to this once per session
dataset <- read.clipboard()
```

4. Get the data set and find the basic descriptive statistics. Then plot GREV versus GREQ against each other.
5. Find the correlation matrix of all of the variables in the data set.
6. Show the scatter plot of all the variables
7. Find the multiple correlation of Verbal and Quant with MA

Read the data and describe it

R code

```
datafilename<-
  "http://personality-project.org/r/datasets/psychometrics.prob2.txt"
dataset <- read.file(datafilename) #read the data file
describe(dataset)
```

	var	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
ID	1	1000	500.50	288.82	500.50	500.50	370.65	1.0	1000.00	999.00	0.00	-1.20	9.13
GREV	2	1000	499.77	106.11	497.50	498.75	106.01	138.0	873.00	735.00	0.09	-0.07	3.36
GREQ	3	1000	500.53	103.85	498.00	498.51	105.26	191.0	914.00	723.00	0.22	0.08	3.28
GREA	4	1000	498.13	100.45	495.00	498.67	99.33	207.0	848.00	641.00	-0.02	-0.06	3.18
Ach	5	1000	49.93	9.84	50.00	49.88	10.38	16.0	79.00	63.00	0.00	0.02	0.31
Anx	6	1000	50.32	9.91	50.00	50.43	10.38	14.0	78.00	64.00	-0.14	0.14	0.31
Prelim	7	1000	10.03	1.06	10.00	10.02	1.48	7.0	13.00	6.00	-0.02	-0.01	0.03
GPA	8	1000	4.00	0.50	4.02	4.01	0.53	2.5	5.38	2.88	-0.07	-0.29	0.02
MA	9	1000	3.00	0.49	3.00	3.00	0.44	1.4	4.50	3.10	-0.07	-0.09	0.02

Find the correlations

Many ways of doing this.

R code

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

```
> cor(dataset)
```

	ID	GREV	GREQ	GREA	Ach	Anx	Prelim
ID	1.000000000	-0.007259876	0.002827727	-0.005342069	-0.004044547	-0.009117778	0.01764427
GREV	-0.007259876	1.000000000	0.728847119	0.641151694	0.005612846	0.010193145	0.42824146
GREQ	0.002827727	0.728847119	1.000000000	0.596345026	0.006846819	0.005469727	0.38308810
GREA	-0.005342069	0.641151694	0.596345026	1.000000000	0.453464501	-0.389629680	0.57243190
Ach	-0.004044547	0.005612846	0.006846819	0.453464501	1.000000000	-0.556186667	0.30334338
Anx	-0.009117778	0.010193145	0.005469727	-0.389629680	-0.556186667	1.000000000	-0.22788772
Prelim	0.017644273	0.428241464	0.383088099	0.572431898	0.303343378	-0.227887724	1.00000000
GPA	-0.004907548	0.419468609	0.366971989	0.516206843	0.276380979	-0.222404655	0.42229426
MA	-0.009435498	0.322297502	0.287387909	0.454549472	0.263466528	-0.219220334	0.35832775

	GPA	MA
ID	-0.004907548	-0.009435498
GREV	0.419468609	0.322297502
GREQ	0.366971989	0.287387909
GREA	0.516206843	0.454549472
Ach	0.276380979	0.263466528
Anx	-0.222404655	-0.219220334
Prelim	0.422294264	0.358327754
GPA	1.000000000	0.306467093
MA	0.306467093	1.000000000

Round the answer to 2 decimals

```
> round(cor(dataset),2)
```

	ID	GREV	GREQ	GREA	Ach	Anx	Prelim	GPA	MA
ID	1.00	-0.01	0.00	-0.01	0.00	-0.01	0.02	0.00	-0.01
GREV	-0.01	1.00	0.73	0.64	0.01	0.01	0.43	0.42	0.32
GREQ	0.00	0.73	1.00	0.60	0.01	0.01	0.38	0.37	0.29
GREA	-0.01	0.64	0.60	1.00	0.45	-0.39	0.57	0.52	0.45
Ach	0.00	0.01	0.01	0.45	1.00	-0.56	0.30	0.28	0.26
Anx	-0.01	0.01	0.01	-0.39	-0.56	1.00	-0.23	-0.22	-0.22
Prelim	0.02	0.43	0.38	0.57	0.30	-0.23	1.00	0.42	0.36
GPA	0.00	0.42	0.37	0.52	0.28	-0.22	0.42	1.00	0.31
MA	-0.01	0.32	0.29	0.45	0.26	-0.22	0.36	0.31	1.00

Use pairwise correlations (to handle missing data), round the output

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

	ID	GREV	GREQ	GREA	Ach	Anx	Prelim	GPA	MA
ID	1.00	-0.01	0.00	-0.01	0.00	-0.01	0.02	0.00	-0.01
GREV	-0.01	1.00	0.73	0.64	0.01	0.01	0.43	0.42	0.32
GREQ	0.00	0.73	1.00	0.60	0.01	0.01	0.38	0.37	0.29
GREA	-0.01	0.64	0.60	1.00	0.45	-0.39	0.57	0.52	0.45
Ach	0.00	0.01	0.01	0.45	1.00	-0.56	0.30	0.28	0.26
Anx	-0.01	0.01	0.01	-0.39	-0.56	1.00	-0.23	-0.22	-0.22
Prelim	0.02	0.43	0.38	0.57	0.30	-0.23	1.00	0.42	0.36
GPA	0.00	0.42	0.37	0.52	0.28	-0.22	0.42	1.00	0.31
MA	-0.01	0.32	0.29	0.45	0.26	-0.22	0.36	0.31	1.00

Using the lowerCor function, combine these previous operations

R code

```
> lowerCor(dataset)
```

	ID	GREV	GREQ	GREA	Ach	Anx	Prelm	GPA	MA
ID	1.00								
GREV	-0.01	1.00							
GREQ	0.00	0.73	1.00						
GREA	-0.01	0.64	0.60	1.00					
Ach	0.00	0.01	0.01	0.45	1.00				
Anx	-0.01	0.01	0.01	-0.39	-0.56	1.00			
Prelim	0.02	0.43	0.38	0.57	0.30	-0.23	1.00		
GPA	0.00	0.42	0.37	0.52	0.28	-0.22	0.42	1.00	
MA	-0.01	0.32	0.29	0.45	0.26	-0.22	0.36	0.31	1.00

Do the significance tests

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

```
> corr.test(dataset)
```

```
Call:corr.test(x = dataset)
```

```
Correlation matrix
```

	ID	GREV	GREQ	GREA	Ach	Anx	Prelim	GPA	MA
ID	1.00	-0.01	0.00	-0.01	0.00	-0.01	0.02	0.00	-0.01
GREV	-0.01	1.00	0.73	0.64	0.01	0.01	0.43	0.42	0.32
...									
MA	-0.01	0.32	0.29	0.45	0.26	-0.22	0.36	0.31	1.00

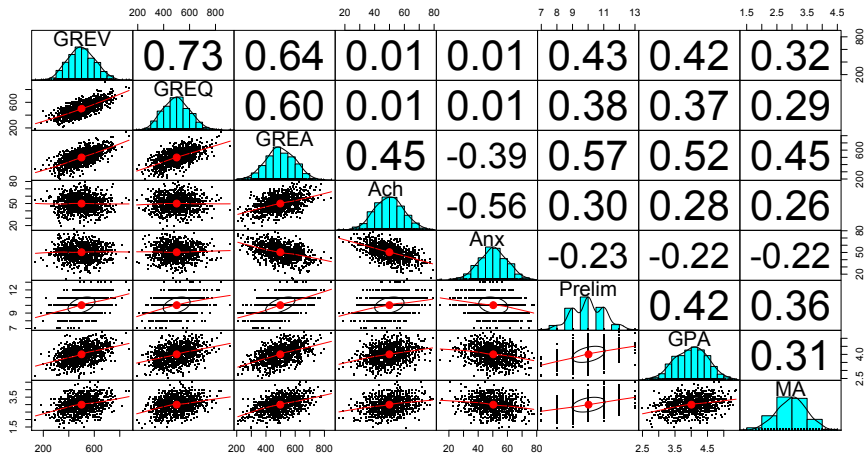
```
Sample Size
```

	ID	GREV	GREQ	GREA	Ach	Anx	Prelim	GPA	MA
ID	1000	1000	1000	1000	1000	1000	1000	1000	1000
GREV	1000	1000	1000	1000	1000	1000	1000	1000	1000
...									
MA	1000	1000	1000	1000	1000	1000	1000	1000	1000

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

	ID	GREV	GREQ	GREA	Ach	Anx	Prelim	GPA	MA
ID	0.00	1.00	1.00	1	1	1	1	1	1
GREV	0.82	0.00	0.00	0	1	1	0	0	0
GREQ	0.93	0.00	0.00	0	1	1	0	0	0
GREA	0.87	0.00	0.00	0	0	0	0	0	0
Ach	0.90	0.86	0.83	0	0	0	0	0	0
Anx	0.77	0.75	0.68	0	0	0	0	0	0
Prelim	0.77	0.75	0.68	0	0	0	0	0	0
GPA	0.77	0.75	0.68	0	0	0	0	0	0
MA	0.77	0.75	0.68	0	0	0	0	0	0

`pairs.panels(dataset[-1],pch=".",gap=0)`



Multiple regression from data

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

```
> 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

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

Call:

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

Residuals:

	Min	1Q	Median	3Q	Max
	-1.37779	-0.29802	0.01818	0.30276	1.38537

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

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	997	217.25				
2	996	193.22	1	24.035	123.89	< 2.2e-16 ***

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

Use SetCor to do the analyses and see the plots

R code

```
Model1 <- setCor(MA ~ GREV + GREQ, data= dataset)
Model2 <- setCor(MA ~ GREV + GREQ + GREA,data=dataset)
```

```
Model1
Call: setCor(y = MA ~ GREV + GREQ, data = dataset)
```

Multiple Regression from raw data

```
DV = MA
intercept = -173.34
      slope  se    t      p lower.ci upper.ci VIF
GREV  0.24 0.04 5.52 4.4e-08    0.16    0.33 2.13
GREQ  0.11 0.04 2.57 1.0e-02    0.03    0.20 2.13
```

```
Multiple Regression
      R   R2  Ruw R2uw Shrunken R2 SE of R2 overall F df1 df2      p SE residual
MA 0.33 0.11 0.33 0.11      0.11    0.02   61.46   2 997 6.78e-26      0.94
> Model2
Call: setCor(y = MA ~ GREV + GREQ + GREA, data = dataset)
```

Multiple Regression from raw data

```
DV = MA
intercept = -232.48
      slope  se    t      p lower.ci upper.ci VIF
GREV  0.05 0.04  1.23 2.2e-01   -0.03    0.14 2.48
GREQ  0.00 0.04 -0.10 9.2e-01   -0.09    0.08 2.27
GREA  0.42 0.04 11.13 3.3e-27    0.35    0.50 1.81
```

```
Multiple Regression
      R   R2  Ruw R2uw Shrunken R2 SE of R2 overall F df1 df2      p SE residual
MA 0.46 0.21 0.4 0.16      0.21    0.02   87.32   3 996 3.65e-50      0.89
```

Multiple DVs simultaneously

R code

```
> R <- lowerCor(dataset[-1])  
> set.cor(y=6:8,x=1:5,data=R)
```

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

Multiple Regression from matrix input

Beta weights

	Prelim	GPA	MA
GREV	0.14	0.20	0.10
GREQ	0.04	0.05	0.03
GREA	0.40	0.29	0.31
Ach	0.11	0.12	0.10
Anx	-0.01	-0.05	-0.05

Multiple R

Prelim	GPA	MA
0.59	0.54	0.47

Multiple R2

Prelim	GPA	MA
0.34	0.29	0.22

Various estimates of between set correlations

Squared Canonical Correlations

[1] 0.4943 0.0036 0.0017

Chisq of canonical correlations