

Psychology 405: Psychometric Theory

Homework 1: answers

William Revelle

Department of Psychology
Northwestern University
Evanston, Illinois USA



April, 2017

Outline

Preliminaries

Assignment

Analysis

Some descriptive graphics

Preliminaries for the first run

1. Install R from CRAN
 - Current release is now 3.3.3 but 3.3.2 is fine
2. Make sure that the *psych* is installed
 - Just need to do this once for each computer
3. Could install other packages as desired (see previous slides)
4. Each time you start R you need to make the *psych* active using the `library(psych)` command

Assignment 1

1. Read data from a file or website
 - An example data set is at <https://personality-project.org/revelle/syllabi/405/405.example1.data.txt>
2. Find basic descriptive statistics
 - Measures of central tendency
 - Measures of dispersion
3. Show basic descriptive graphics
 - Measures of central tendency
 - Measures of dispersion
 - Measures of relationship

Get the data

```
> library(psych) #only need to do once per run
#copy the data into the clipboard from the web server by using your browser
#https://personality-project.org/revelle/syllabi/405/405.example1.data.txt
> prob1 <- read.clipboard() #read from the clipboard
># or
> fn <- "https://personality-project.org/revelle/syllabi/405/405.example1.data.txt"
> prob1 <- read.file(fn)
>prob1 #show the data
```

	A	B	C	D	E	F
1	1	0	16	7	1	1
2	2	7	17	62	12	2
3	3	0	9	0	5	4
4	4	7	18	35	18	8
5	5	7	13	5	28	16
6	6	8	11	10	78	32
7	7	9	13	14	0	64
8	8	2	10	48	46	128
9	9	7	16	0	23	256
10	10	3	10	13	23	512
11	11	4	14	8	11	1024
12	12	4	12	9	34	2048
13	13	3	22	5	10	4096
14	14	0	10	59	5	8192
15	15	5	13	96	24	16384
16	16	7	22	97	43	32768

Alternative ways to read the data

1. If the data are in a text file, `.txt`, `.csv`, or `.sav`
 - find the file using `read.file()` which calls `file.choose` and then reads as appropriate
 - `read.table` from the file name
 - This file can be on a remote server, in which case the file name specifies the URL.
2. If the data are in a SPSS file
 - find the file using `file.choose`
 - install the *foreign* package
 - the data using the `read.spss` command (setting `data.frame = TRUE`)
3. `file.read` will convert SPSS or `.csv` files

Summary statistics using base R

Not as user friendly as one would want

```
> summary(prob1)
```

A	B	C	D	E	F
Min. : 1.00	Min. :0.000	Min. : 9.00	Min. : 0.00	Min. : 0.00	Min. : 1
1st Qu.: 4.75	1st Qu.:2.750	1st Qu.:10.75	1st Qu.: 6.50	1st Qu.: 8.75	1st Qu.: 14
Median : 8.50	Median :4.500	Median :13.00	Median :11.50	Median :20.50	Median : 192
Mean : 8.50	Mean :4.562	Mean :14.12	Mean :29.25	Mean :22.56	Mean : 4096
3rd Qu.:12.25	3rd Qu.:7.000	3rd Qu.:16.25	3rd Qu.:50.75	3rd Qu.:29.50	3rd Qu.: 2560
Max. :16.00	Max. :9.000	Max. :22.00	Max. :97.00	Max. :78.00	Max. :32768

Summary descriptives using the describe function

More to the taste of most psychologists.
 Can save the output as an object to be processed by some other function.

```
> describe(prob1)
```

	var	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
A	1	16	8.50	4.76	8.5	8.50	5.93	1	16	15	0.00	-1.20	1.19
B	2	16	4.56	3.01	4.5	4.57	3.71	0	9	9	-0.25	-1.21	0.75
C	3	16	14.12	4.08	13.0	13.93	4.45	9	22	13	0.63	-0.23	1.02
D	4	16	29.25	33.17	11.5	26.50	13.34	0	97	97	0.95	0.08	8.29
E	5	16	22.56	20.36	20.5	20.21	17.79	0	78	78	1.16	2.47	5.09
F	6	16	4095.94	8806.17	192.0	2340.43	282.44	1	32768	32767	2.27	7.91	2201.54

summary and describe of log values

```
> summary(log(prob1))
> describe((log(prob1)))
```

A	B	C	D	E	M
Min. :0.000	Min. : -Inf	Min. :2.197	Min. : -Inf	Min. : -Inf	M
1st Qu.:1.554	1st Qu.: 1	1st Qu.:2.374	1st Qu.: 2	1st Qu.: 2	1
Median :2.138	Median : 1	Median :2.565	Median : 2	Median : 3	M
Mean :1.917	Mean : -Inf	Mean :2.611	Mean : -Inf	Mean : -Inf	M
3rd Qu.:2.505	3rd Qu.: 2	3rd Qu.:2.788	3rd Qu.: 4	3rd Qu.: 3	3
Max. :2.773	Max. : 2	Max. :3.091	Max. : 5	Max. : 4	M

```
> describe(log(prob1))
```

	var	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
A	1	16	1.92	0.79	2.14	1.99	0.69	0.0	2.77	2.77	-0.95	0.88	0.20
B	2	16	-Inf	NaN	1.50	-Inf	0.66	-Inf	2.20	Inf	NaN	NaN	NaN
C	3	16	2.61	0.28	2.56	2.61	0.35	2.2	3.09	0.89	0.27	-0.82	0.07
D	4	16	-Inf	NaN	2.43	-Inf	1.44	-Inf	4.57	Inf	NaN	NaN	NaN
E	5	16	-Inf	NaN	3.01	2.66	0.98	-Inf	4.36	Inf	NaN	NaN	NaN
F	6	16	5.20	3.30	5.20	5.20	4.11	0.0	10.40	10.40	0.00	-1.20	0.83

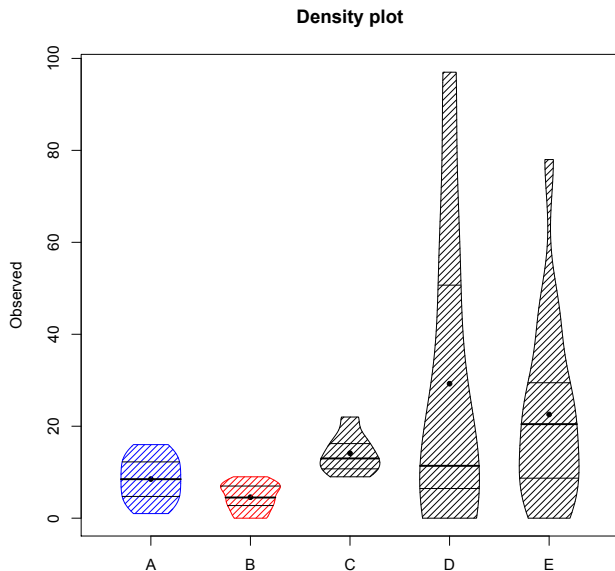
summary and describe of log values of prob +1

```
> log1prob <- log(prob1+1)
> summary(log1prob)
> boxplot(log1prob, main="boxplots of log transformed data")
> describe(log1prob)
```

A	B	C	D	E
Min. :0.6931	Min. :0.000	Min. :2.303	Min. :0.000	Min. :0.000
1st Qu.:1.7462	1st Qu.:1.314	1st Qu.:2.463	1st Qu.:2.008	1st Qu.:2.246
Median :2.2499	Median :1.701	Median :2.639	Median :2.518	Median :3.061
Mean :2.0941	Mean :1.486	Mean :2.684	Mean :2.674	Mean :2.698
3rd Qu.:2.5835	3rd Qu.:2.079	3rd Qu.:2.848	3rd Qu.:3.942	3rd Qu.:3.414
Max. :2.8332	Max. :2.303	Max. :3.135	Max. :4.585	Max. :4.369

	var	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
A	1	16	2.09	0.63	2.25	2.14	0.63	0.69	2.83	2.14	-0.72	0.00	0.16
B	2	16	1.49	0.81	1.70	1.53	0.56	0.00	2.30	2.30	-0.93	-0.02	0.20
C	3	16	2.68	0.26	2.64	2.68	0.32	2.30	3.14	0.83	0.29	-0.80	0.07
D	4	16	2.67	1.42	2.52	2.73	1.33	0.00	4.58	4.58	-0.38	-0.27	0.36
E	5	16	2.70	1.17	3.06	2.77	0.92	0.00	4.37	4.37	-0.81	0.74	0.29
F	6	16	5.30	3.16	5.20	5.26	4.02	0.69	10.40	9.70	0.08	-1.27	0.79

A violin plot of the first 5 variables `violinBy(my.data[1:5])`



A SPLOM (scatter plot matrix) pairs.panels(my.data)

