SPSS and R Examples of Descriptive Statistics of Psychological Measurement Evaluations

These are the course evaluations I received from my Psychological Measurement class. Below is a list of the questions. Responses were made on a 6-point scale, from 1=strongly disagree to 6=strongly agree. For the analyses shown below, I computed an average of the 11 questions (EVALAVE).

1. The instructor uses class time effectively.
2. The instructor answered questions in a clear and understandable way.
3. The instructor was prepared for class.
4. The instructor used examples effectively in explaining difficult concepts.
5. The instructor raised challenging questions for discussion.
6. The instructor welcomed students questions or comments.
7. The instructor seemed to enjoy teaching.
8. The grading policies were clearly stated.
9. The course material was presented clearly.
10. Relevant information in addition to the textbook was presented.
11. I learned a lot in this course.

SPSS

```
output close *.
*The above command clears the output file (notice asterisk above refers to current output file, which diff
ers from its use in the present line which indicates that everything until the period is ignored by SPSS).

get file='c:\jason\spsswin\uvclass\pmclass evals.sav'.

compute evalave=mean(q1 to q11).

frequencies vars=evalave
   /histogram=normal.

descriptives vars=evalave
  /statistics=all.
```

Frequencies

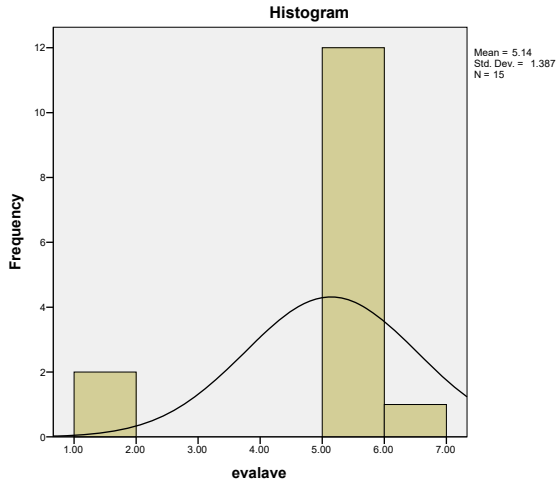
Statistics

evalave

N	Valid	15
	Missing	0

evalave

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.73	1	6.7	6.7	6.7
	1.82	1	6.7	6.7	13.3
	5.00	1	6.7	6.7	20.0
	5.40	1	6.7	6.7	26.7
	5.55	1	6.7	6.7	33.3
	5.64	3	20.0	20.0	53.3
	5.70	1	6.7	6.7	60.0
	5.73	2	13.3	13.3	73.3
	5.82	2	13.3	13.3	86.7
	5.91	1	6.7	6.7	93.3
	6.00	1	6.7	6.7	100.0
	Total	15	100.0	100.0	



Descriptives

Descriptive Statistics

	N	Range	Minimum	Maximum	Sum	Mean		Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
evalave	15	4.27	1.73	6.00	77.10	5.1400	.35808	1.38683	1.923	-2.276	.580	3.933	1.121
Valid N (listwise)	15												

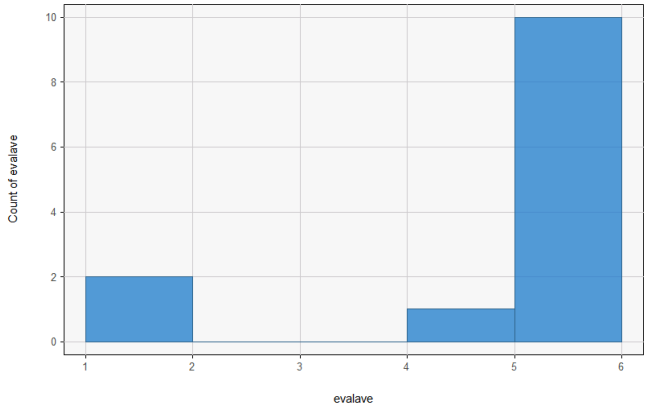
I encourage you to begin using syntax as much as possible, but the above output can also be obtained through the windows: **Analyze** –> **Descriptive Statistics** -> **Frequencies**, then move over the desired variable to the Variable(s) box and click "ok."

R

```
> #clear active frame from previous analyses
> rm(d)
>
> #note: all lines starting with hashtag/lb sign are ignored by R
>
> #I read an SPSS data file (.sav) in with the haven package, which must be installed
> #on the first use on a computer.  It must be loaded for use after that using the library() function
> #install.packages("haven")
> library(haven)
> d = read_sav("c:/jason/spsswin/uvclass/pmclass evals.sav")
>
> #the lessR package could also be used
> #the lessR package allows you to perform some R commands simply
> #library(lessR)
> #d = Read("c:/jason/spsswin/uvclass/pmclass evals.sav",quiet=TRUE)
> #two ways to look at the data set
> #d # typing the name of the data frame lists all values in the console
> #View(d)  #data view-like appearance (lessR and many other packages)
>
> #base R method of computing scale mean (requires all items present—no missing data)
> #d$evalave = (d$q1+d$q2+d$q3+d$q4+d$q5+d$q6+d$q7+d$q8+d$q9+d$q10+d$q11)/11
>
> #somewhat simpler and more flexible methods
> #list-wise deletion approach (matches the base R approach above)
> #d$evalave<-rowMeans(d[, c("q1", "q2", "q3","q4","q5","q6","q7","q8","q9","q10","q11")], na.rm=F)
>
> #available information approach (matches SPSS method) by adding na.rm=T
> d$evalave<-rowMeans(d[, c("q1", "q2", "q3","q4","q5","q6","q7","q8","q9","q10","q11")], na.rm=T)
>
>
> #the freq function in the summarytools package has a nice frequency table
> #I ask for a frequency table only for the evalave variable here,
> #but using freq(d) gives tables for all of the variables
> library(summarytools)
> freq(d$evalave)
```

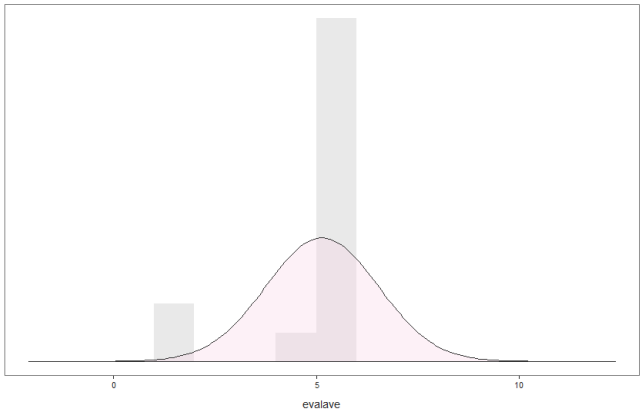
```
                      Freq   % Valid   % Valid Cum.   % Total   % Total Cum.
--------------------- ------ --------- -------------- --------- --------------
      1.72727272727273      1      6.67           6.67      6.67           6.67
      1.81818181818182      1      6.67          13.33      6.67          13.33
                     5      1      6.67          20.00      6.67          20.00
                   5.4      1      6.67          26.67      6.67          26.67
      5.54545454545455      1      6.67          33.33      6.67          33.33
      5.63636363636364      3     20.00          53.33     20.00          53.33
                   5.7      1      6.67          60.00      6.67          60.00
      5.72727272727273      2     13.33          73.33     13.33          73.33
      5.81818181818182      2     13.33          86.67     13.33          86.67
      5.90909090909091      1      6.67          93.33      6.67          93.33
                     6      1      6.67         100.00      6.67         100.00
                 <NA>      0                              0.00         100.00
                 Total    15    100.00         100.00    100.00         100.00
> #an alternative, which you can also use is a lessR function pivot() - output not shown
> # I prefer the summarytools format but it does not make much difference
> #library(lessR)
> #pivot(data=d,table,variable=evalave)

> #obtain frequency histogram from lessR package of just evalave variable
> # lessR is used for these plots, you can use variable names without data frame name
> library(lessR)
> Histogram(evalave)
```

Histogram



Density plot

```
> Histogram(evalave,density=TRUE,type=c("normal"))
```



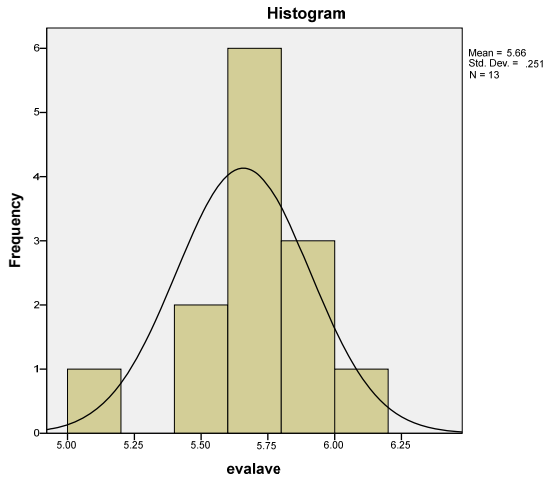
Let's see what the distribution looks like after eliminating the two outliers. Ordinarily, you would not want to remove outliers from a data set unless you have a good reason to do so (e.g., there is an error, the participant did not belong in the study to begin with—did not meet study inclusion criteria). This time, we will add a boxplot and stem-and-leaf plot.

SPSS

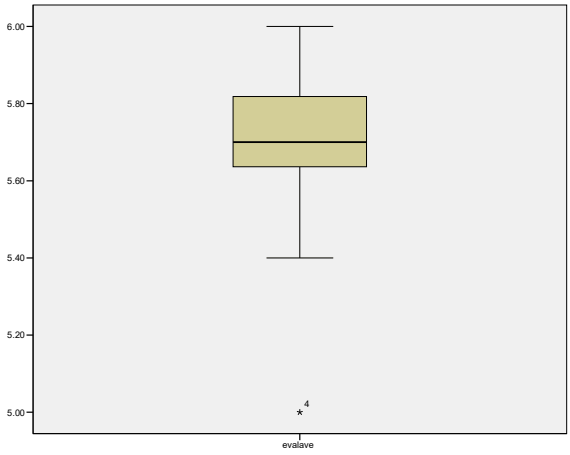
```
*the following statement selects out the two outliers.
select if evalave gt 4.

frequencies vars=evalave
   /histogram=normal.
```

evalave

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	5.00	1	7.7	7.7	7.7
	5.40	1	7.7	7.7	15.4
	5.55	1	7.7	7.7	23.1
	5.64	3	23.1	23.1	46.2
	5.70	1	7.7	7.7	53.8
	5.73	2	15.4	15.4	69.2
	5.82	2	15.4	15.4	84.6
	5.91	1	7.7	7.7	92.3
	6.00	1	7.7	7.7	100.0
	Total	13	100.0	100.0	




```
examine variables=evalave
  /plot=boxplot stemleaf
 /statistics=none.
```




```
evalave Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     1.00 Extremes    (=<5.0)
     1.00        5 .  4
    10.00        5 .  5666777889
     1.00        6 .  0

 Stem width:      1.00
 Each leaf:        1 case(s)

descriptives vars=evalave
  /statistics=all.
```

Descriptive Statistics

	N	Range	Minimum	Maximum	Sum	Mean		Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
evalave	13	1.00	5.00	6.00	73.55	5.6580	.06963	.25105	.063	-1.487	.616	3.401	1.191
Valid N (listwise)	13												

Use the Frequencies command to get the median (50th percentile) and the 25th and 75th percentiles.

```
frequencies vars=evalave
   /statistics=skewness kurtosis
   /ntiles=4.
```

Statistics

evalave

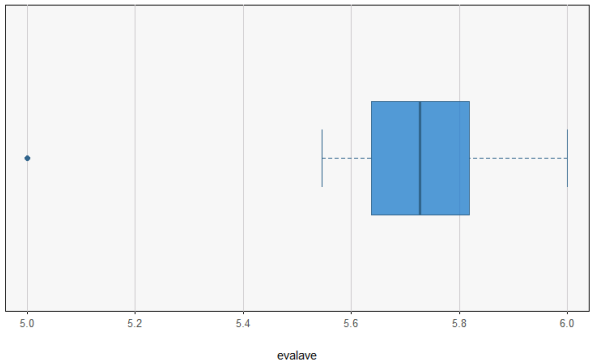
N	Valid	13
	Missing	0
Skewness		-1.487
Kurtosis		3.401
Percentiles	25	5.5909
	50	5.7000
	75	5.8182

R

The R commands are the same as I illustrated above. To save space I just show how to remove cases, obtain the descriptive statistics, and generate a boxplot and stem and leaf plot.

```
> #select non-extreme cases for boxplot and new density plot (base R function)
> d = subset(d,evalave > 4))

#BoxPlot is a lessR function
> BoxPlot(evalave)
```

Boxplot




```
> #the psych package's describe provides nice descriptives output (I add 25th and 75th percentiles)
> library(psych)
> describe(d$evalave,quant=c(.25,.75))
  vars  n mean   sd median trimmed  mad min max range  skew kurtosis   se Q0.25 Q0.75
1    1 13 5.66 0.25    5.7    5.69 0.18   5   6     1 -1.16     1.09 0.07  5.64  5.82


> #the stem-and-leaf command is base R, so need to specify data set name
> #by prefixing the variable name with the data set plus $
> stem(d$evalave)


--- evalave ---

  5 | 0
  5 | 566677889
  6 | 0
```

Sample Write-up *[based on second analysis]*
Descriptive statistics were obtained for the course evaluation measure. Results indicated high ratings of the instructor overall, *M* = 5.66 and *SD* =.25. The range was equal to 1.00, with minimum score of 5.00 and a maximum score of 6.00. The distribution was skewed to the left and leptokurtic. The skewness value was equal to -1.49 and the kurtosis value was equal to 3.40. A frequency histogram, using a normal curve overlay and a boxplot, were obtained. Both plots suggested an outlier with a value of approximately 5.00. The boxplot depicted that the median was approximately 5.7, the range was between approximately 5.50 and 6.0 (without the outlier), and 50% of the scores occurred between approximately 5.65 and 5.85.[1]

[Note that descriptive statistics are often relegated to a table and/or are very briefly given in the description of the sample, where the descriptive statistics are given just for sociodemographic variables. Percents are usually used for categorical variables (e.g., 53% 11th graders, 47% 12th graders) and mean and standard deviations for continuous variables (e.g, instructor ratings as shown above) and there is little or no discussion of the shape of the distribution].

[1] Note: the SPSS FREQUENCIES command and the R `describe` function provide the exact values for the median and the 25 and 75 percentiles, which could be reported instead. Remember that for the boxplot SPSS uses hinges and rounded median, and R may compute these more exactly for the box plot.