Exploring Assumptions 1

Code

Hide

```
library(car)
library(ggplot2)
library(pastecs)
library(psych)
```

Hide

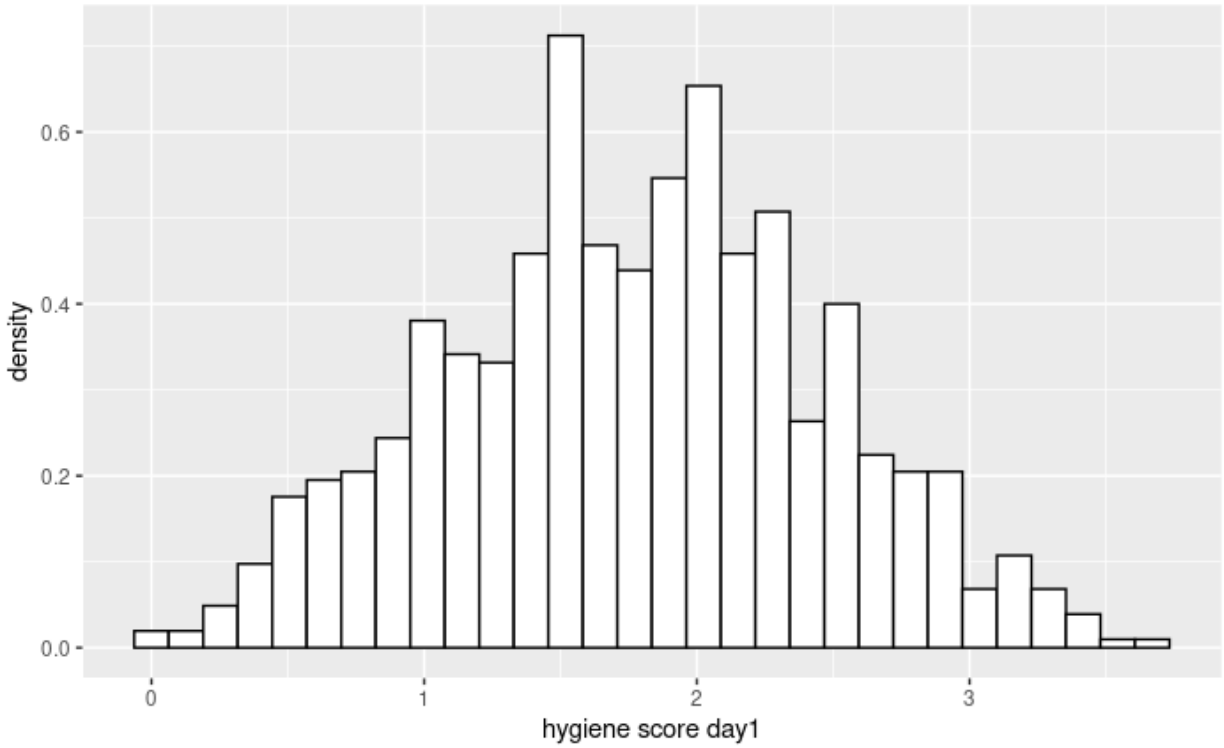
```
data<-read.delim('/home/atrides/Desktop/Applied-Statistics-with-R-master/sta
tistics_with_R/05/Data_Files/DownloadFestival1.dat',header=TRUE)
head(data, 10)
```

	ticknumb <int>	gender <chr>	day1 <dbl>	day2 <dbl>	day3 <dbl>
1	2111	Male	2.64	1.35	1.61
2	2229	Female	0.97	1.41	0.29
3	2338	Male	0.84	*NA*	*NA*
4	2384	Female	3.03	*NA*	*NA*
5	2401	Female	0.88	0.08	*NA*
6	2405	Male	0.85	*NA*	*NA*
7	2467	Female	1.56	*NA*	*NA*
8	2478	Female	3.02	*NA*	*NA*
9	2490	Male	2.29	*NA*	*NA*
10	2504	Female	1.11	0.44	0.55

1-10 of 10 rows

Hide

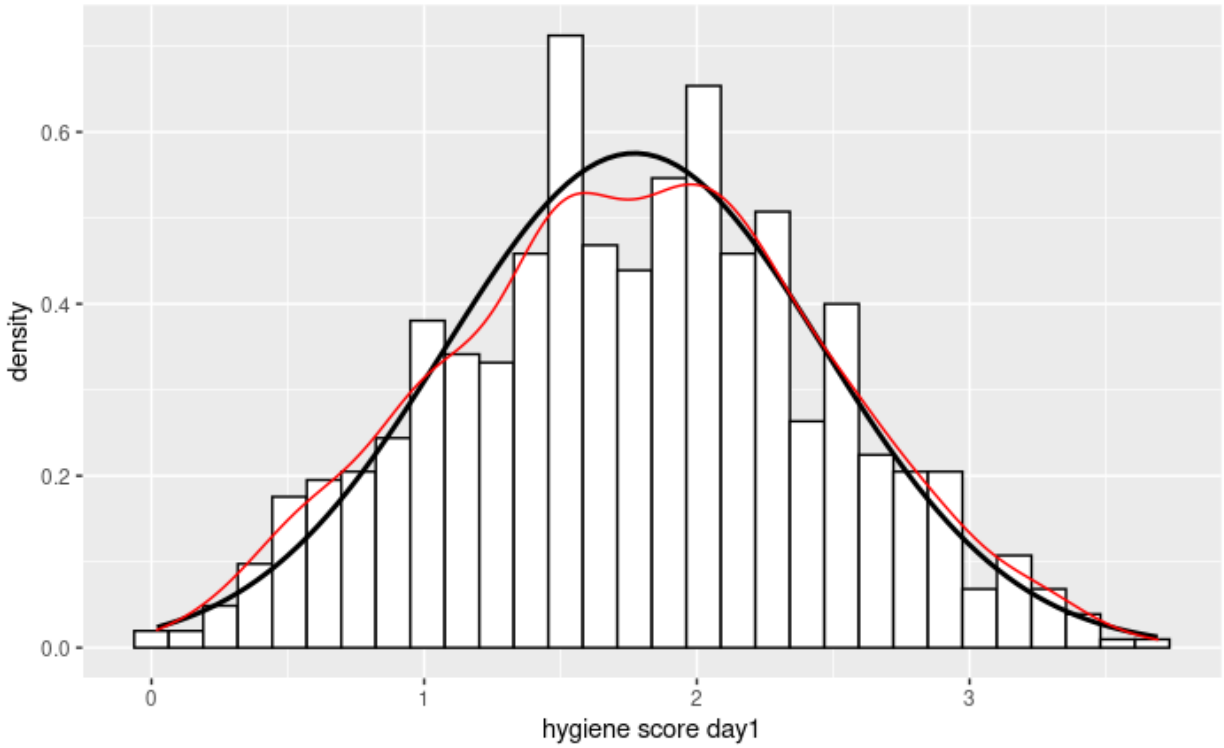
```
# Day 1
day1<-ggplot(data, aes(day1))
day1<-day1+geom_histogram(aes(y=..density..),colour='black',fill='white')+la
bs(x='hygiene score day1', y='density')
day1
```



Hide

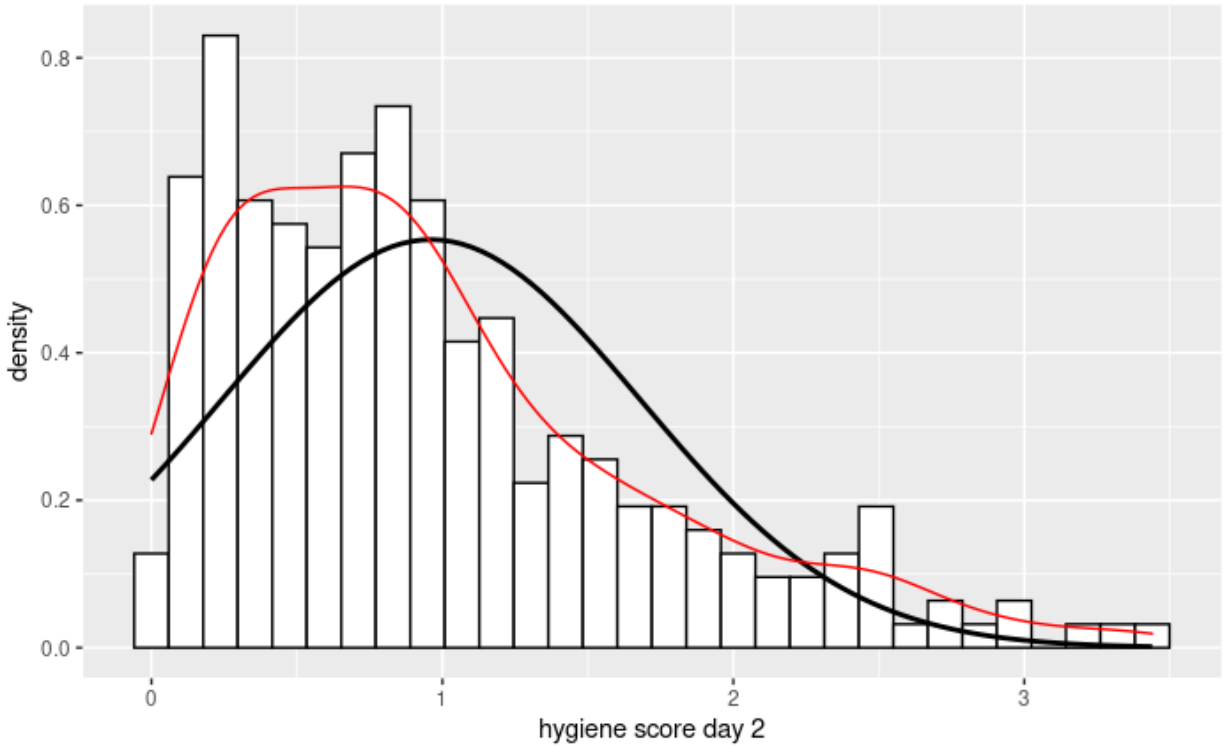
```
# adding a perfect normal distribution along with histogram + the histogram
density plot
day1<-day1+stat_function(fun=dnorm, args = list(mean=mean(data$day1,na.rm =
TRUE), sd=sd(data$day1,na.rm=TRUE)),colour='black',size=1)+
  geom_density(color='red')

day1
```



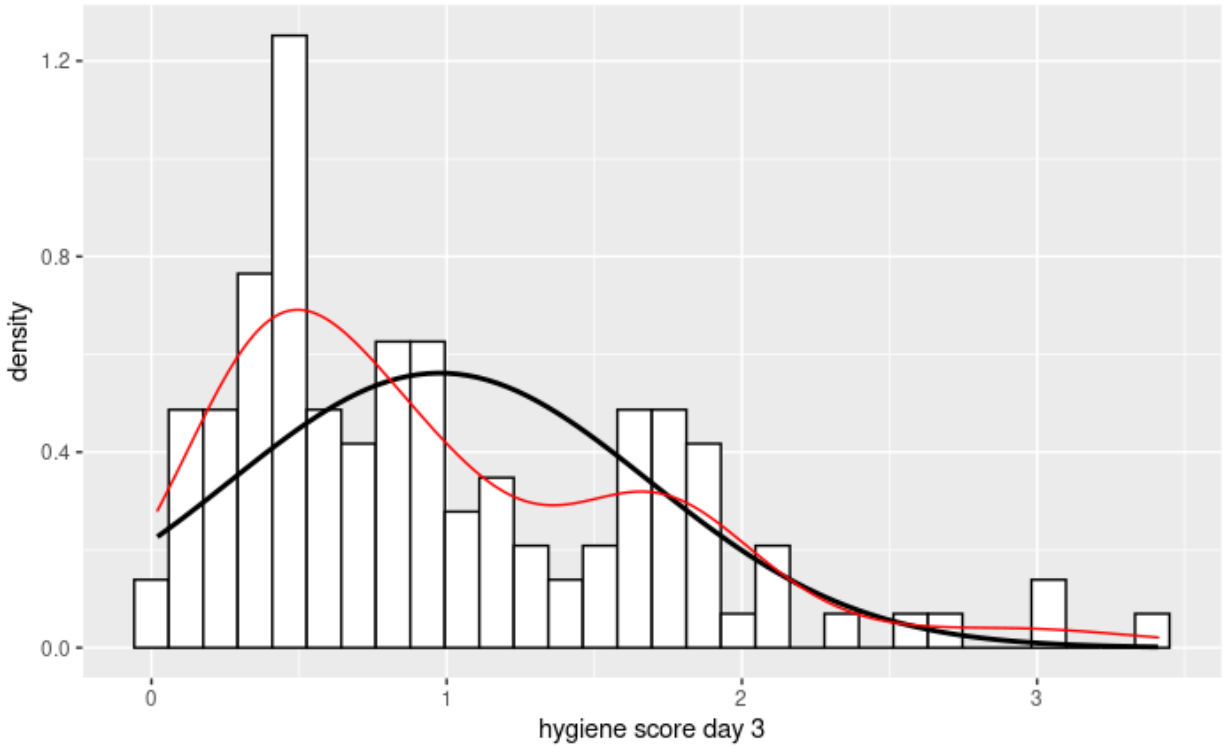
Hide

```
# Day2, normal vs real
day2<-ggplot(data, aes(day2))
day2<-day2+geom_histogram(aes(y=..density..),colour='black',fill='white')+la
bs(x='hygiene score day 2' , y='density')
day2<- day2+stat_function(fun=dnorm, args=list(mean=mean(data$day2, na.rm=TR
UE), sd=sd(data$day2, na.rm = TRUE)), colour='black', size=1)
day2<- day2+geom_density(color='red')
day2
```



Hide

```
# Day3, normal vs real
day3<-ggplot(data, aes(day3))
day3<-day3+geom_histogram(aes(y=..density..),colour='black',fill='white')+la
bs(x='hygiene score day 3' , y='density')
day3<- day3+stat_function(fun=dnorm, args=list(mean=mean(data$day3, na.rm=TR
UE), sd=sd(data$day3, na.rm = TRUE)), colour='black', size=1)
day3<- day3+geom_density(color='red')
day3
```

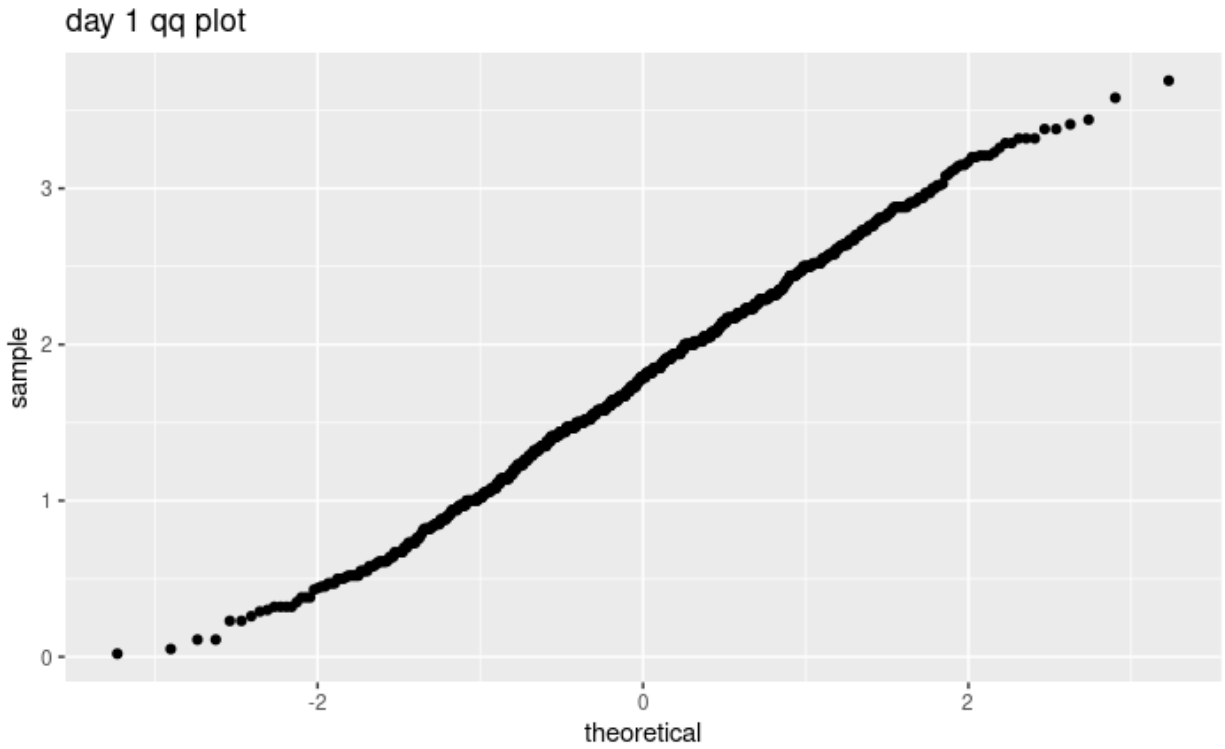


Hide

```
# Drawing Q-Plots
# https://stats.stackexchange.com/questions/348438/qq-plot-and-x-y-line
```

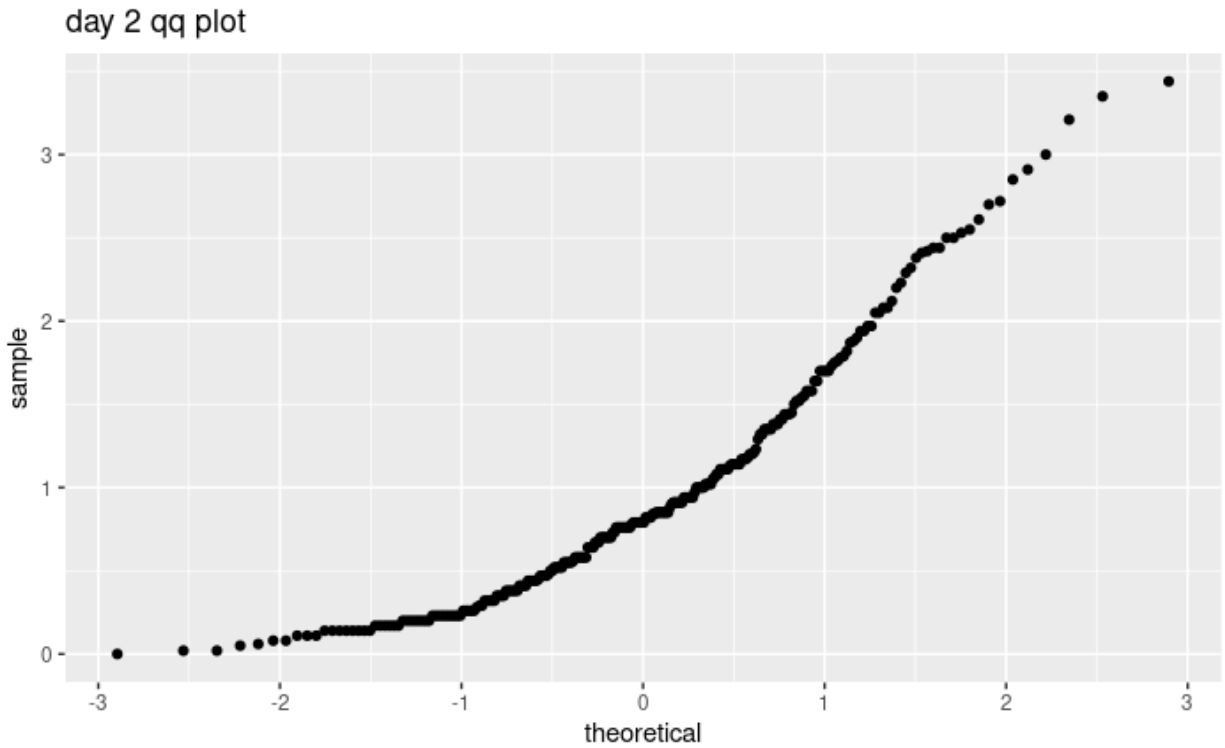
Hide

```
# Day1
qqplot.day1<-ggplot(data, aes(sample=day1))
qqplot.day1<-qqplot.day1+stat_qq()
qqplot.day1+ggtitle('day 1 qq plot')
```



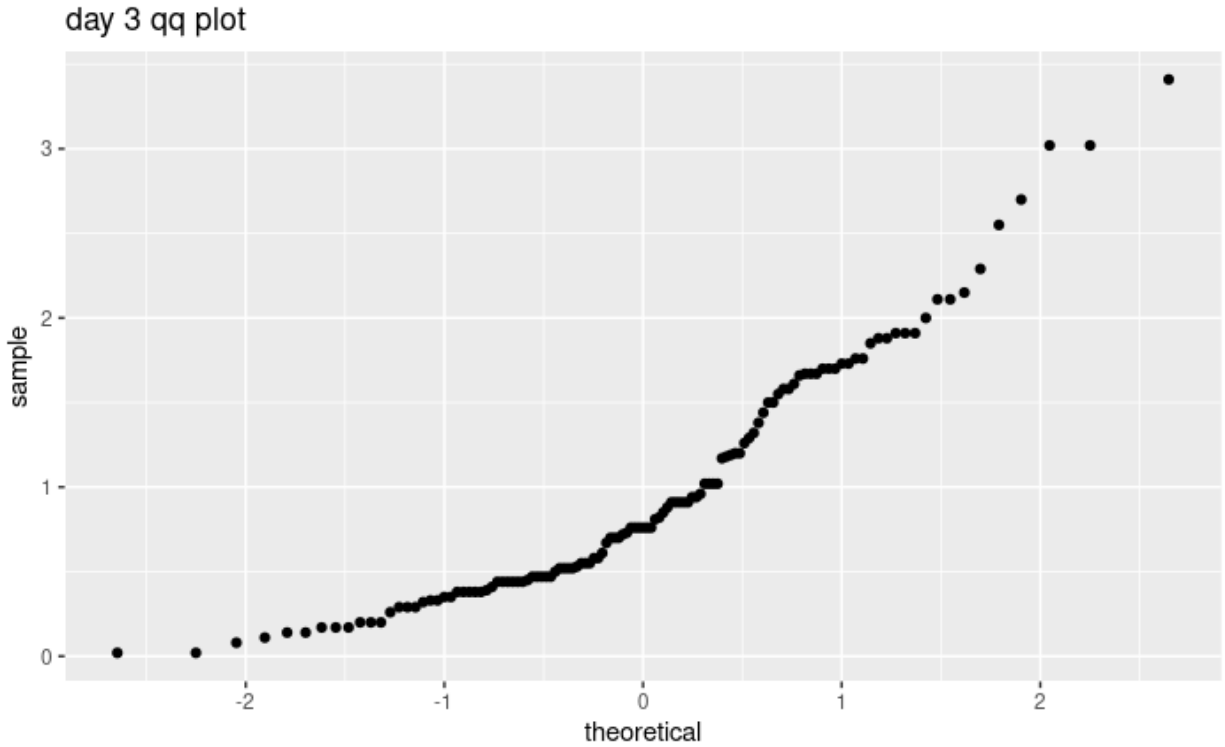
Hide

```
# Day2
qqplot.day2<-ggplot(data, aes(sample=day2))
qqplot.day2<-qqplot.day2+stat_qq()
qqplot.day2+ggtitle('day 2 qq plot')
```



Hide

```
# Day3
qqplot.day3<-ggplot(data, aes(sample=day3))
qqplot.day3<-qqplot.day3+stat_qq()
qqplot.day3+ ggtitle('day 3 qq plot')
```



Hide

```
#  Quantifying normality with number
describe(data$day1)
```

	vars	n	mean	sd	median	trimmed	mad	min	max
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
X1	1	810	1.77	0.69	1.79	1.77	0.7	0.02	3.69

1 row | 1-10 of 13 columns

Hide

```
describe(data$day2)
```

	vars	n	mean	sd	median	trimmed	mad	min	max
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
X1	1	264	0.96	0.72	0.79	0.87	0.61	0	3.44

1 row | 1-10 of 13 columns

Hide

```
describe(data$day3)
```

	vars	n	mean	sd	median	trimmed	mad	min	max
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
X1	1	123	0.98	0.71	0.76	0.9	0.61	0.02	3.41

1 row | 1-10 of 13 columns

Hide

```
# more stats at once
describe(cbind(data$day1,data$day2,data$day3))
```

	vars	n	mean	sd	median	trimmed	mad	min	max
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
X1	1	810	1.77	0.69	1.79	1.77	0.70	0.02	3.69
X2	2	264	0.96	0.72	0.79	0.87	0.61	0.00	3.44
X3	3	123	0.98	0.71	0.76	0.90	0.61	0.02	3.41

3 rows | 1-10 of 13 columns

Hide

```
# another function for describing
stat.desc(cbind(data$day1,data$day2,data$day3),norm=TRUE,basic=FALSE)
```

	V1 <dbl>	V2 <dbl>	V3 <dbl>
median	1.79000000	7.900000e-01	7.600000e-01
mean	1.77113580	9.609091e-01	9.765041e-01
SE.mean	0.02436847	4.436095e-02	6.404352e-02
CI.mean.0.95	0.04783289	8.734781e-02	1.267805e-01
var	0.48099624	5.195239e-01	5.044934e-01
std.dev	0.69353892	7.207801e-01	7.102770e-01
coef.var	0.39157862	7.501022e-01	7.273672e-01
skewness	-0.00442835	1.082811e+00	1.007813e+00
skew.2SE	-0.02577395	3.611574e+00	2.309035e+00
kurtosis	-0.42159405	7.554615e-01	5.945454e-01

1-10 of 13 rows Previous **1** 2 Next

Hide

```
# another way to do the same things
describe(data[,c('day1','day2','day3')])
```

	vars <dbl>	n <dbl>	mean <dbl>	sd <dbl>	median <dbl>	trimmed <dbl>	mad <dbl>	min <dbl>	max <dbl>
day1	1	810	1.77	0.69	1.79	1.77	0.70	0.02	3.69
day2	2	264	0.96	0.72	0.79	0.87	0.61	0.00	3.44
day3	3	123	0.98	0.71	0.76	0.90	0.61	0.02	3.41

3 rows | 1-10 of 13 columns

Hide

```
stat.desc(data[,c('day1','day2','day3')], basic = FALSE,norm=TRUE)
```

	day1 <dbl>	day2 <dbl>	day3 <dbl>
median	1.79000000	7.900000e-01	7.600000e-01
mean	1.77113580	9.609091e-01	9.765041e-01
SE.mean	0.02436847	4.436095e-02	6.404352e-02
CI.mean.0.95	0.04783289	8.734781e-02	1.267805e-01

	day1 <dbl>	day2 <dbl>	day3 <dbl>
var	0.48099624	5.195239e-01	5.044934e-01
std.dev	0.69353892	7.207801e-01	7.102770e-01
coef.var	0.39157862	7.501022e-01	7.273672e-01
skewness	-0.00442835	1.082811e+00	1.007813e+00
skew.2SE	-0.02577395	3.611574e+00	2.309035e+00
kurtosis	-0.42159405	7.554615e-01	5.945454e-01

1-10 of 13 rows Previous **1** 2 Next