Skewness and Kurtosis

Ungrouped data

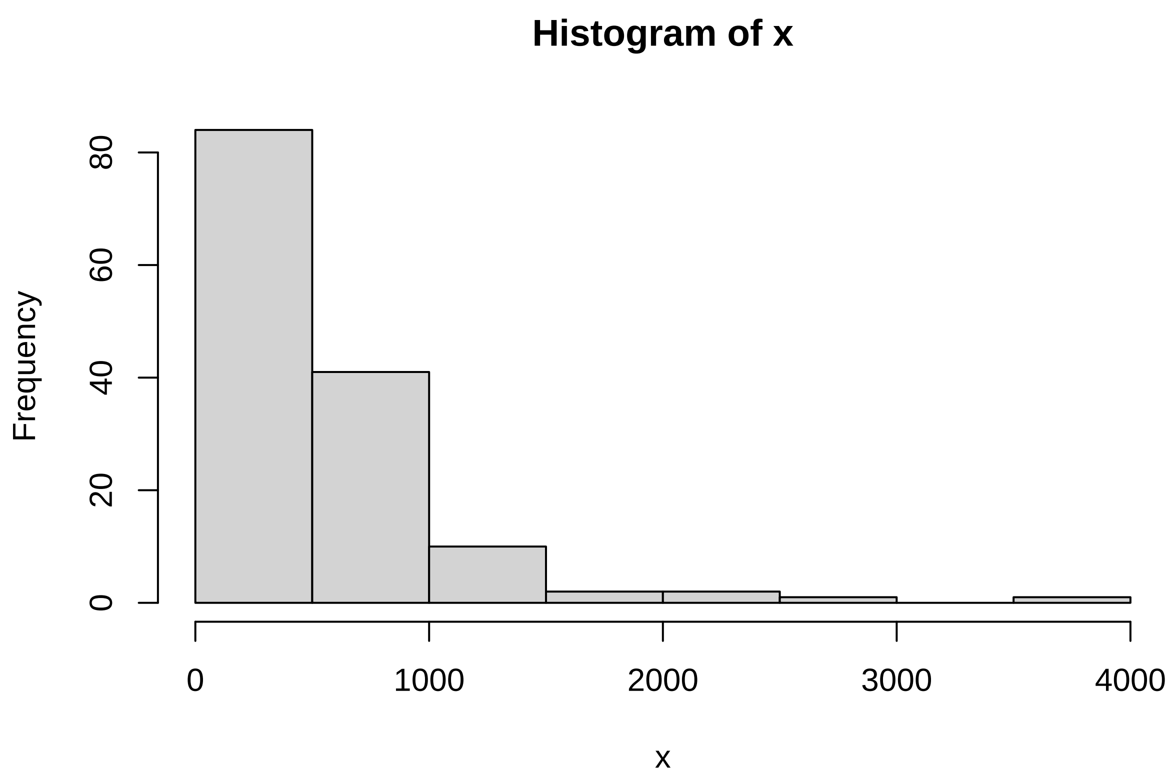
```r
data("rivers")
View(rivers)
x=rivers
xt=table(x)
xt
```

```
x
 135  202  210  215  217  230  233  237  246  250  255  259  260  265  268  270
   1    1    2    1    1    2    1    1    1    3    1    1    2    1    1    1
 276  280  281  286  290  291  300  301  306  310  314  315  320  325  327  329
   1    3    1    1    1    1    3    1    1    2    1    1    1    1    1    1
 330  332  336  338  340  350  352  360  375  377  380  383  390  392  407  410
   1    1    1    1    1    4    1    4    1    1    2    1    2    1    1    1
 411  420  424  425  430  431  435  444  445  450  460  465  470  490  500  505
   1    2    1    1    1    1    1    1    1    1    2    1    1    1    2    1
 524  525  529  538  540  545  560  570  600  605  610  618  620  625  630  652
   1    2    1    1    1    1    1    1    3    1    1    1    1    1    1    1
 671  680  696  710  720  730  735  760  780  800  840  850  870  890  900  906
   1    1    1    1    2    1    2    1    1    1    1    1    1    1    2    1
 981 1000 1038 1054 1100 1171 1205 1243 1270 1306 1450 1459 1770 1885 2315 2348
   1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1
2533 3710
   1    1
```

```r
mode=which(xt==max(xt))
mode #mode is not unique
```

```
350 360
 38  40
```

```r
hist(x)
```




```r
#Here, distribution is Positive skew so we can not use mode = 3*median - 2* mean
#we have to use Bowley's measure of skewness
summary(x)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  135.0   310.0   425.0   591.2   680.0  3710.0 
```

```r
d1=quantile(x,0.75)-quantile(x,0.50)
d1
```

```
75% 
255 
```

```r
d2=quantile(x,0.50)-quantile(x,0.25)
d2
```

```
50% 
115 
```

```r
skb=(d1-d2)/(d1+d2)
skb
```

```
      75%
0.3783784
```

```r
n1=mean(x)
n2=sum(x^2)/length(x)
n2
```

```
[1] 591677.6
```

```r
m2=n2-n1^2
m2
```

```
[1] 242178.6
```

```r
n3=sum(x^3)/length(x)
n3
```

```
[1] 1015589802
```

```r
m3=n3-3*n2*n1+2*n1^3
m3
```

```
[1] 379454890
```

```r
gama1=m3/m2^1.5
gama1 #The distribution is highly positively skew.
```

```
[1] 3.183879
```

Grouped data (Discrete data)

```r
y=c(1,2,3,4,5,7,8,10,11,12,13,17)
f=c(4,2,7,5,3,8,6,3,2,2,2,1)
fr.dist=data.frame(y,f)
fr.dist
```

```
    y f
1   1 4
2   2 2
3   3 7
4   4 5
5   5 3
6   7 8
7   8 6
8  10 3
9  11 2
10 12 2
11 13 2
12 17 1
```

```r
x=rep(y,f)
summary(x)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  1.000   3.000   7.000   6.378   8.000  17.000 
```

```r
#Karl Pearson's Coefficient of skeness
md=which(f==max(f))
md
```

```
[1] 6
```

```r
mode=y[md]
mode
```

```
[1] 7
```

```r
sk1=(mean(x)-mode)/sd(x)
sk1
```

```
[1] -0.1641024
```

```r
# Bowly's coefficient of skewness
d1=quantile(x,0.75)-quantile(x,0.50)
d1
```

```
75% 
  1 
```

```r
d2=quantile(x,0.50)-quantile(x,0.25)
d2
```

```
50% 
  4 
```

```r
sk2=(d1-d2)/(d1+d2)
sk2
```

```
 75% 
-0.6 
```

Grouped data (continuous data)

```r
midx=seq(2.1,3.9,0.3)
midx
```

```
[1] 2.1 2.4 2.7 3.0 3.3 3.6 3.9
```

```r
f=c(1,5,11,14,16,2,1)
fr.dist=data.frame(midx,f)
fr.dist
```

```
  midx  f
1  2.1  1
2  2.4  5
3  2.7 11
4  3.0 14
5  3.3 16
6  3.6  2
7  3.9  1
```

```r
n=sum(f)
n
```

```
[1] 50
```

```r
w=0.3
#Karl Pearson's coefficient of skewness
#Mean
mn=sum(midx*f)/n
mn
```

```
[1] 2.994
```

```r
#Mode
m1=which(f==max(f))
m1
```

```
[1] 5
```

```r
l=midx[m1]-w/2
l
```

```
[1] 3.15
```

```r
fm=f[m1]
fm
```

```
[1] 16
```

```r
f1=f[m1-1]
f1
```

```
[1] 14
```

```r
f2=f[m1+1]
f2
```

```
[1] 2
```

```r
mode=l+((fm-f1)/(2*fm-f1-f2)*w)
mode
```

```
[1] 3.1875
```

```r
#Variance
var=sum((f*(midx-mn)^2))/n
var
```

```
[1] 0.131364
```

```r
sd=var^0.5
sd
```

```
[1] 0.3624417
```

```r
sk1=(mn-mode)/sd
sk1
```

```
[1] -0.5338789
```

```r
#Bowley's coefficient of skewness
cl=cumsum(f)
cl
```

```
[1]  1  6 17 31 47 49 50
```

```r
#Lower quartile
m1=min(which(cl>=n/4))
m1
```

```
[1] 3
```

```r
l=midx[m1]-w/2
l
```

```
[1] 2.55
```

```r
cf=cl[m1-1]
cf
```

```
[1] 6
```

```r
fr=f[m1]
fr
```

```
[1] 11
```

```r
q1=l+((n/4-cf)/fr)*w
q1
```

```
[1] 2.727273
```

```r
#Median
m1=min(which(cl>=n/2))
m1
```

```
[1] 4
```

```r
l=midx[m1]-w/2
l
```

```
[1] 2.85
```

```r
cf=cl[m1-1]
cf
```

```
[1] 17
```

```r
fr=f[m1]
fr
```

```
[1] 14
```

```r
q2=l+((n/2-cf)/fr)*w
q2
```

```
[1] 3.021429
```

```r
#Upper quartile
m1=min(which(cl>=3*n/4))
m1
```

```
[1] 5
```

```r
l=midx[m1]-w/2
l
```

```
[1] 3.15
```

```r
cf=cl[m1-1]
cf
```

```
[1] 31
```

```r
fr=f[m1]
fr
```

```
[1] 16
```

```r
q3=l+((3*n/4-cf)/fr)*w
q3
```

```
[1] 3.271875
```

```r
#Bowly's coefficient of skewness
d1=q3-q2
d2=q2-q1
sk2=(d1-d2)/(d1+d2)
sk2
```

```
[1] -0.08025933
```

```r
#Third central moments
m3=sum(f*(midx-mn)^3)/n
m3
```

```
[1] -0.007895232
```

```r
gama1=m3/var^1.5
gama1
```

```
[1] -0.1658251
```

```r
#Distribution is negatively skew
```

Measures of kurtosis

Ungrouped data

```r
#A quality control engineer in determining whether a machine is properly adjusted to dispense 16 ounces
x=c(15.9, 16.2, 16.0, 15.6, 16.2, 15.9, 16.0, 15.6, 15.6, 16.0, 16.2, 15.6, 15.9, 16.2, 15.6, 16.2, 15.8
x
```

```
 [1] 15.9 16.2 16.0 15.6 16.2 15.9 16.0 15.6 15.6 16.0 16.2 15.6 15.9 16.2 15.6
[16] 16.2 15.8 16.0 15.8 15.9 16.2 15.8 15.8 16.2 16.0 15.9 16.2 16.2 16.0 15.6
```

```
n=length(x)
n
```

```
[1] 30
```

```
mn=mean(x)
mn
```

```
[1] 15.93667
```

```
#Fourth central moments
m4=sum((x-mn)^4)/n
m4
```

```
[1] 0.004062022
```

```
#Second central moments
m2=sum((x-mn)^2)/n
m2
```

```
[1] 0.04698889
```

```
beta2=m4/(m2^2)
beta2
```

```
[1] 1.839721
```

```
gama2=beta2-3
gama2
```

```
[1] -1.160279
```

```
#The distribution is platykurtic
```

Measures of kurtosis

Grouped data (Continuous)

```
midx=seq(2.1,3.9,0.3)
midx
```

```
[1] 2.1 2.4 2.7 3.0 3.3 3.6 3.9
```

```
f=c(1,5,11,14,16,2,1)
f
```

```
[1]  1  5 11 14 16  2  1
```

```r
fr.dist=data.frame(midx,f)
fr.dist
```

```
  midx  f
1  2.1  1
2  2.4  5
3  2.7 11
4  3.0 14
5  3.3 16
6  3.6  2
7  3.9  1
```

```r
w=0.3
n=sum(f)
mn=sum(midx*f)/n
mn
```

```
[1] 2.994
```

```r
var=sum((f*(midx-mn)^2))/n
var
```

```
[1] 0.131364
```

```r
#Fourth central moments
m4=sum(f*(midx-mn)^4)/n
m4
```

```
[1] 0.04854414
```

```r
beta2=m4/var^2
beta2
```

```
[1] 2.813093
```

```r
gama2=beta2-3
gama2
```

```
[1] -0.1869071
```

```r
#The distribution is platykurtic
```