HYD298 - Homework #1

Andrew Brown

January 11, 2016

7.12 Derive Kirby's bound, Equation 7.45, on the estimate of the coefficient of skewness by computing the sample estimates of the skewness of the most skewed sample it would be possible to observe. Derive also the upper bound $(n-1)^{1/2}$ for the estimate of the population coefficient of variation when all the observations must be nonnegative.

A) Coefficient of Skewness

```
n=100000 #large n (needed for part B, mostly)
foo=c(1,rep(0,n-1)) #most skewed sample
```

Maximized sample estimate of coefficient of skewness:

```
coef_skew=function(xx) {
  #Calculates sample estimate of skewness coef
  nn=length(xx)
  (nn*sum((xx-mean(xx))^3)/((nn-1)*(nn-2)))/(sd(xx)^3)
}
coef_skew(foo)
```

```
## [1] 316.2278
```

Theoretical upper-bound (Kirby 1974; Bobee and Robitaille 1975)

```
## [1] 316.2278
```

B) Population Coefficient of Variation

```
mu=mean(foo) #population mean
sigma=sd(foo)*(n-1)/n #population sd
```

Maximized population coefficient of variation:

```
sigma/mu
```

```
## [1] 316.2246
```

Theoretical upper bound

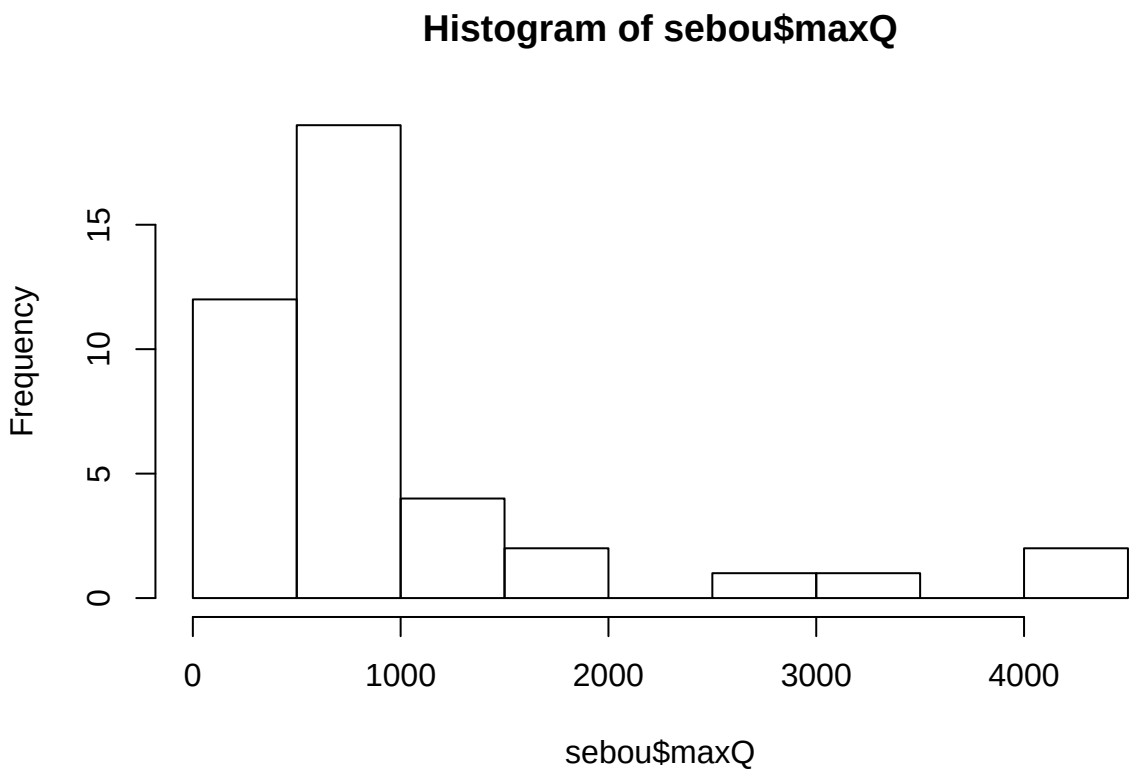
```
sqrt(length(foo)-1)
```

```
## [1] 316.2262
```

7.17 The following data represent a sequence of annual flood flows, the maximum flow rate observed each year, for the Sebou River at the Azib Soltane gaging station in Morocco.

A) Construct a histogram of the Sebou flood flow data to see what the flow distribution looks like.

```
sebou=read.csv("question2_sebou.csv")
hist(sebou$maxQ)
```



B) Calculate the mean, variance, and sample skew. Based on Table 7.3, does the sample skew appear to be significantly different from zero?

Mean fo Sebou River dataset:

```
mean(sebou$maxQ)
```

```
## [1] 1004.488
```

Variance for Sebou River dataset:

```
var(sebou$maxQ)
```

```
## [1] 866469.3
```

Sample skew coefficient for Sebou River dataset:

```r
library(moments)
```

```
## Warning: package 'moments' was built under R version 3.2.3
```

```r
coef_skew(sebou$maxQ)#*(sd(sebou$maxQ)^3)
```

```
## [1] 2.369083
```

```r
agostino.test(sebou$maxQ)
```

```
##
##  D'Agostino skewness test
##
## data:  sebou$maxQ
## skew = 2.2815, z = 4.7029, p-value = 2.565e-06
## alternative hypothesis: data have a skewness
```

The Sebou River dataset has 41 observations. The calculated coefficient of skewness for this sample was 2.36, which is greater than the expected value (2.32, with std. dev of 0.77) for a sample of n=50 from a population with skewness coefficient of 3.00 (the largest value for skewness available in Table 7.3).

Therefore, the skewness is likely to be significantly greater than zero. This can be confirmed by a visual check of the data in the histogram from part A which is clearly right-skewed. Further, the D'Agostino's K^2 test suggests that the null hypothesis of normal distribution (skewness coefficient = 0) should be rejected.