**4.30**

“One way of seeing that this is a biased estimator of the standard deviation of the population is to start from the result that s2 is an unbiased estimator for the variance σ2 of the underlying population if that variance exists and the sample values are drawn independently with replacement. The square root is a nonlinear function, and only linear functions commute with taking the expectation. Since the square root is a strictly concave function, it follows from Jensen's inequality that the square root of the sample variance is an underestimate.”[[1]](#footnote-1)

The function below can show that sample variance is biased towards underestimating the standard deviation. It takes 10,000 random samples of a population and compares the square root of the variance to the known standard deviation:

|  |
| --- |
| > f <- function(population) {  + greater <- 0  + less <- 0  +  + for(i in 1:10000) {  + n <- sample(1:length(population), 1)  +  + if(sd(population) > sqrt(var(sample(population, n))))  + less <- less + 1  + else  + greater <- greater + 1  + }  +  + print(less)  + print(greater)  + } |
| |  | | --- | |  | |

**4.31**

The function below calculates the known variance of the population and calculates the sample variance using a known mean:

> g <- function(population) {

+

+ m <- mean(population)

+ greater <- 0

+ less <- 0

+

+ for(i in 1:10000) {

+ sampVar <- 0

+

+ n <- sample(1:length(population), 1)

+ for(i in (sample(population, n)))

+ sampVar <- sampVar + ((i - m)\*(i - m))/n

+

+ if(var(population) > sampVar)

+ less <- less + 1

+ else

+ greater <- greater + 1

+ }

+

+ print(less)

+ print(greater)

+ }

Showing S2 is not independent from Xbar in an exponential population:

> samples <- replicate(1000, rexp(50))

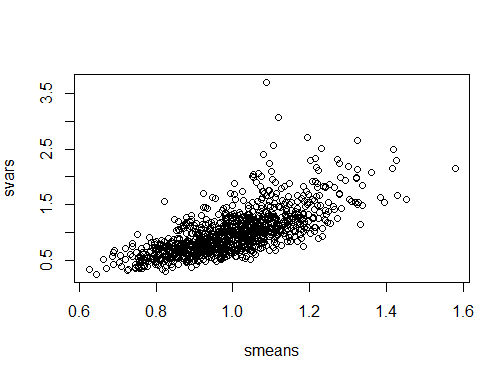
> smeans <- apply(samples,2,mean)

> svars <- apply(samples,2,var)

> cor(smeans,svars)

[1] 0.7036323

> plot(smeans,svars)



1. https://en.wikipedia.org/wiki/Unbiased\_estimation\_of\_standard\_deviation [↑](#footnote-ref-1)