3701 HW2

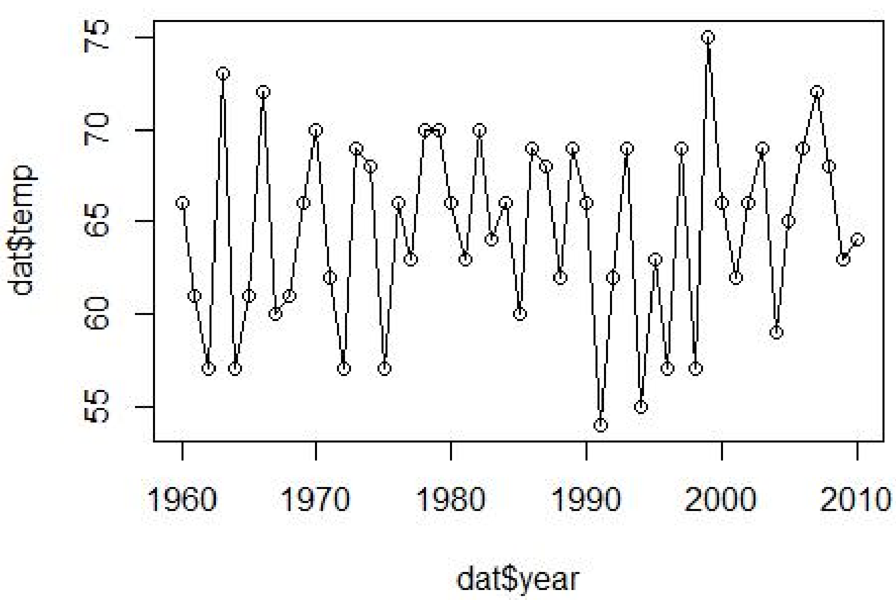
Dongnan Liu

2024-02-17

```
#1(b)
dat=read.table("C:/Users/DELL/Desktop/msptemp.txt")
dat$temp
```

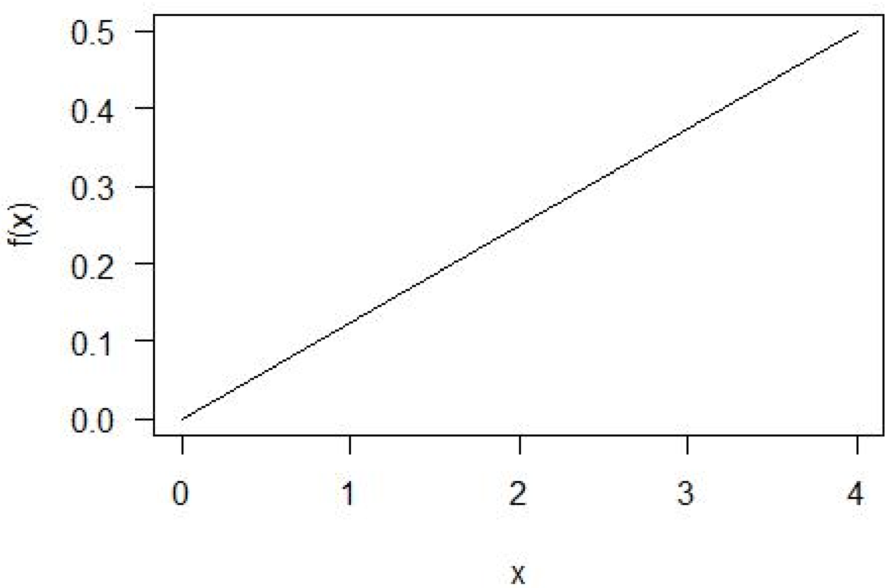
```
##  [1] 66 61 57 73 57 61 72 60 61 66 70 62 57 69 68 57 66 63 70 70 66 63
 70 64 66
## [26] 60 69 68 62 69 66 54 62 69 55 63 57 69 57 75 66 62 66 69 59 65 69
 72 68 63
## [51] 64
```

```
plot(dat$year,dat$temp,type="o")
```

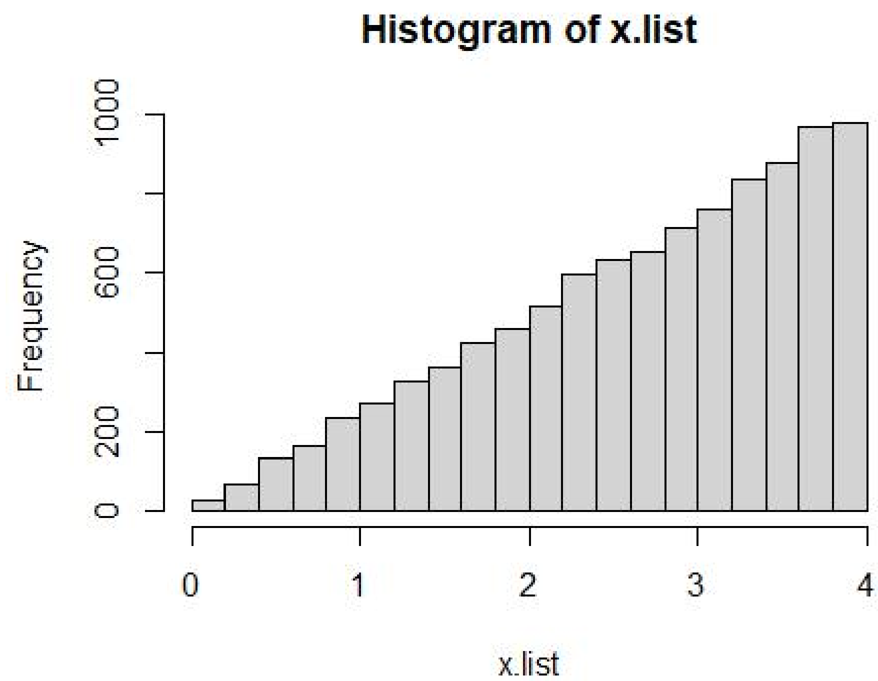



```
#I didn't find visual evidence that the distribution of the response (mi
nimum
# temperature in the Twin Cities on July 25) is changing with time, beca
use I didn't
# see that the temperatures increase or decrease when the time went by.
#1(c)
set.seed(3701)
```

```r
mu=mean(dat$temp)
sigma=sd(dat$temp)
n=51
alpha=0.05
#get moe
moe = (sigma/sqrt(n))*qt(1-alpha/2, n-1)
CI=mu+c(-1,1)*moe
CI

## [1] 63.13864 65.99861

#1(d)
#The interpretation is not correct since Confidence intervals do not pro
vide
#probabilities for the true value of the parameter being estimated (in t
his
#case, the population mean μ), but rather they provide a range of plausi
ble
#values for the parameter.It's correct to say that we are 95% confident
#that the true population mean falls within the interval.

#1(e)
set.seed(3701)
n=51
mu=64
sigma=5
alpha=0.05
tperc = qt(1-alpha/2, n-1)
x.list=round(rnorm(n=n,mean=64,sd=5))
xbar=mean(x.list)
s=sd(x.list)
moe=tperc*s/sqrt(n)
CI=xbar+c(-1,1)*moe
CI

## [1] 62.80758 65.50615

#yet-to-be observed (random) 95%:Before generating the data and computin
g the
#confidence interval, the probability that a randomly chosen 95% confide
nce
#interval will capture the actual mean value of 64 is indeed 0.95.

#2(a)
theta=4
x.vals=c(0,theta)
y.vals=2*theta^(-2)*x.vals
plot(x.vals, y.vals, t="l", xlab="x", ylab="f(x)", las=1)
```




```
#2(c)
#create function include n and theta
rtri2=function(n,theta)
  {
  u.list=runif(n)
  x.list=theta*sqrt(u.list)
  return(x.list)
}
set.seed(3701)
x.list=rtri2(n=1e4, theta=4)
hist(x.list)
```




```
# We can see that the histogram of these realizations is very similar to

#its shape to the graph of f (·; 4).

#2(d)
set.seed(3701)
reps=1e4
theta=4
x.vals=theta*sqrt(runif(reps))
mean(x.vals)

## [1] 2.676486

#the theor_mean is calculated by myself in another page
theor_mean=2/3*theta
c(mean(x.vals),theor_mean)

## [1] 2.676486 2.666667

# We can see that the simulation-based estimate close to formula.

#2(e)
set.seed(3701)
n.list=c(5, 10, 15, 100)
reps=1e4
alpha=0.05
theta=4
```

```
EV=2/3*theta
est.coverage.prob=numeric(length(n.list))
for(k in 1:length(n.list))
{ n=n.list[k]
captured=numeric(reps)
for(r in 1:reps)
{
  x.list=rtri2(n=n, theta=theta)
  xbar=mean(x.list)
  sigma=sd(x.list)
  moe = (sigma/sqrt(n))*qt(1-alpha/2, n-1)
  captured[r]=((xbar-moe) <= EV) & (EV <= (xbar+moe))
}
est.coverage.prob[k]=mean(captured)
}
#combine the n.list and est.coverage.prob
cbind(n.list, est.coverage.prob)

##      n.list est.coverage.prob
## [1,]      5            0.9302
## [2,]     10            0.9364
## [3,]     15            0.9435
## [4,]    100            0.9514

#We can see from the results that the simulated estimate of the coverage

#probability are approximatly 95%(although some of them smaller than 0.9
5)

#3(a)
set.seed(3701)
mu=68
reps=1e4
sigma=3
alpha=0.05
n=10
Lperc=qchisq(1-alpha/2,n-1)
Rperc=qchisq(alpha/2,n-1)
capture.list=numeric(reps)
for(r in 1:reps){
  x.list=rnorm(n,mu,sigma)
  chi.var=var(x.list)
  #get left quantile
  left=(n-1)*chi.var/Lperc
  #get right quantile
  right=(n-1)*chi.var/Rperc
  capture.list[r]=1*((left <= (sigma^2)) & ( (sigma^2) <= right))
}
mean(capture.list)
```

```
## [1] 0.9531

mean(capture.list)+c(-1,1)*(1/sqrt(reps))

## [1] 0.9431 0.9631

# I agree that the 1 − α interval is in the conservative approximate 95%

#confidence interval for the coverage probability.

#3(b)
set.seed(3701)
mu <- 1
reps <- 1e4
sigma <- 1
alpha.list <- c(0.01, 0.05)
n.list <- c(10, 50, 500)
#use for loop to iterate the alpha.list
for (alpha in alpha.list) {
  #use for loop to iterate the n.list
  for (n in n.list) {
    capture.list <- numeric(reps)
    for (r in 1:reps) {
      x.list <- -mu * log(1 - runif(n))
      var <- var(x.list)
      #get left quantile
      Lperc <- qchisq(1 - alpha / 2, n - 1)
      #get right quantile
      Rperc <- qchisq(alpha / 2, n - 1)
      left <- (n - 1) * var / Lperc
      right <- (n - 1) * var / Rperc
      capture.list[r] <- 1 * ((left <= sigma^2) & (sigma^2 <= right))
    }
    CI <- mean(capture.list) + c(-1, 1) * (1 / sqrt(reps))
    cat("n=", n, "alpha=", alpha, "CI is", CI, "\n")
  }
}

## n= 10 alpha= 0.01 CI is 0.8801 0.9001
## n= 50 alpha= 0.01 CI is 0.8209 0.8409
## n= 500 alpha= 0.01 CI is 0.7958 0.8158
## n= 10 alpha= 0.05 CI is 0.7589 0.7789
## n= 50 alpha= 0.05 CI is 0.7059 0.7259
## n= 500 alpha= 0.05 CI is 0.661 0.681

#We can see none of the 1 − α intervals are in the
#corresponding conservative approximate 95% confidence intervals for
#the coverage probabilities.
```

Stat3701 hw2:

1.(a) We assume that our n measurements of a response $x_1, \dots, x_{51}$ are a realization of the sequence of random variables $X_1, \dots, X_{51}$ that are iid with distribution F with unknown mean and standard deviation. 66 degrees is a realization of X_1 with distribution F with unknown mean and standard deviation. (From textbook).

2. (b)

$$F(t) = \int_0^t \frac{2}{\theta^2} x \, dt = \frac{2}{\theta^2} \cdot \frac{t^2}{2} = \frac{t^2}{\theta^2} \quad (0 \leqslant t \leqslant \theta)$$

$$u = F(F^{-1}(u)) \Rightarrow u = \frac{(F^{-1}(u))^2}{\theta^2} \Rightarrow F^{-1}(u) = \theta\sqrt{u}$$

$$F(t) = \begin{cases} 0, & t < 0 \\ \frac{t^2}{\theta^2}, & 0 \leqslant t \leqslant \theta \\ \theta\sqrt{u}, & t > \theta \end{cases} \qquad X = F^{-1}(u) = \theta\sqrt{u}$$

$\sim \text{unif}(0,1)$

2. (d)

$$E(x) = \int x f(x) \, dx = \int_0^\theta x \cdot f(x;\theta) \, dx = \int_0^\theta \frac{2}{\theta^2} \cdot x^2 \, dx$$

$$= \frac{2}{\theta^2} \cdot \frac{1}{3} x^3 \Big]_0^\theta = \frac{2}{\theta^2} \cdot \frac{1}{3}\theta^3 = \frac{2}{3}\theta$$

$$\theta = 4 \quad \therefore \quad E(x) = \frac{2}{3} \times 4 = \frac{8}{3} \approx 2.67$$