STAT 501 R HW1

McBeth Ahortor

9/4/2021

Problems Using R (Q.1.)

Generating a random sample of size 24 from a normal distribution with mean = 21 and sd = 3 and then adding a 25th observation of 52.2

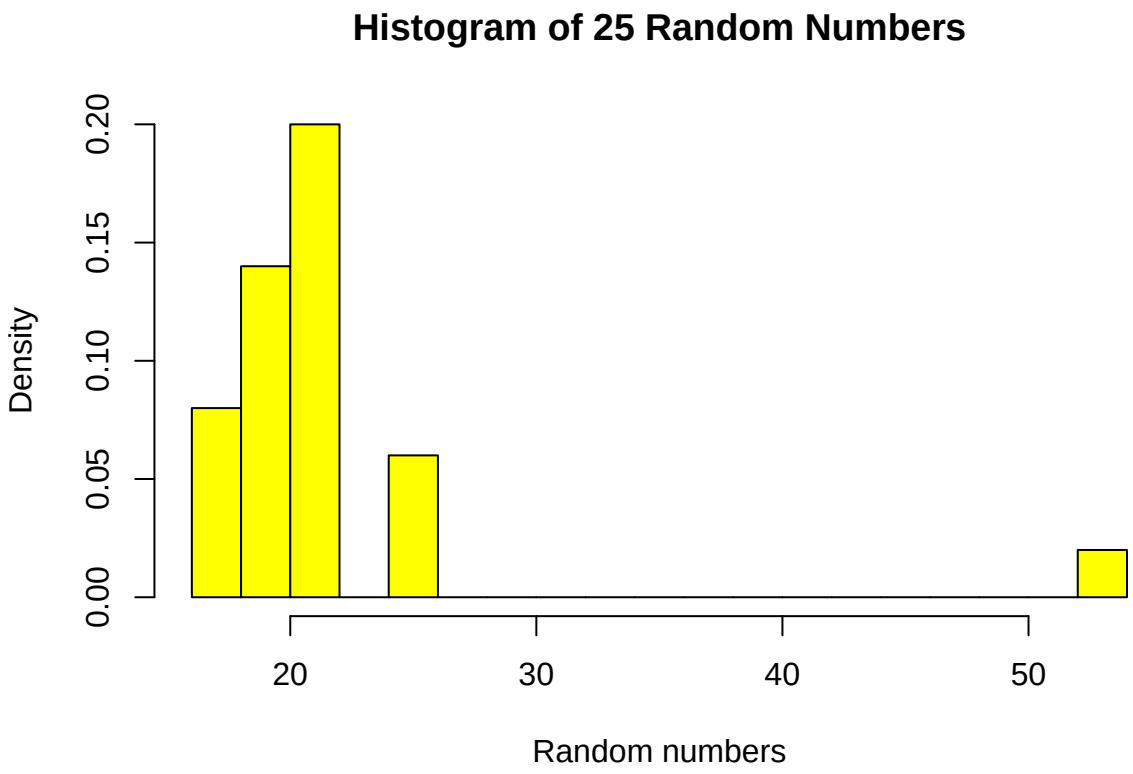
```
set.seed(3)
x <- rnorm(24, mean = 21, sd = 3)
x
```

```
##  [1] 18.11420 20.12242 21.77636 17.54360 21.58735 21.09037 21.25625 24.34983
##  [9] 17.34343 24.80211 18.76566 17.60634 18.85092 21.75796 21.45614 20.07703
## [17] 18.14095 19.05527 24.67294 21.59943 19.26455 18.17310 20.38882 16.00058
```

```
y <- c(x,52.2)
y
```

```
##  [1] 18.11420 20.12242 21.77636 17.54360 21.58735 21.09037 21.25625 24.34983
##  [9] 17.34343 24.80211 18.76566 17.60634 18.85092 21.75796 21.45614 20.07703
## [17] 18.14095 19.05527 24.67294 21.59943 19.26455 18.17310 20.38882 16.00058
## [25] 52.20000
```

```
# Concatenating 52.2 to the 24 random sample
hist(y, freq = FALSE, nclass = 20, col = "yellow", xlab = "Random numbers", main = "Histogram of 25 Ran
```



a. Mean and standard deviation of the 25 random numbers

```
u <- mean(y)
u
```

```
## [1] 21.43982
```

```
#mean
s <- sd(y)
#Standard deviation
s
```

```
## [1] 6.814987
```

b. Intervals

For k=1,

```
k1 <- c(u-1*s,u+1*s)

k1
```

```
## [1] 14.62484 28.25481
```

For k=2,

```
k2 <- c(u-2*s,u+2*s)
k2
```

```
## [1]  7.809851 35.069798
```

For k=3,

```
k3 <- c(u-3*s,u+3*s)
k3
```

```
## [1]  0.9948644 41.8847845
```

C. Counting the number of measurements

```
t1 <-print( sum(u-1*s< y&y < u+1*s))
```

```
## [1] 24
```

```
t2 <-print( sum(u-2*s< y&y < u+2*s))
```

```
## [1] 24
```

```
t3 <-print( sum(u-3*s< y&y < u+3*s))
```

```
## [1] 24
```

```
print("24 measurements fell in the intervals k1, k2 and k3")
```

```
## [1] "24 measurements fell in the intervals k1, k2 and k3"
```

For the empirical rule,

```
25*(0.68) # We expect 68% of the measurements to fall within the interval
```

```
## [1] 17
```

```
25*(0.95) # We expect 95% of the measurements to fall within the interval
```

```
## [1] 23.75
```

```
25*(1) # We expect almost all of the measurements to fall within the interval
```

```
## [1] 25
```

d. Range

Finding the range for the 25 random sample,

```
r <- range(y)
R <-diff(r)
R
```

```
## [1] 36.19942
```

```
print('The range = 36.19942')
```

```
## [1] "The range = 36.19942"
```

Per empirical rule, the standard deviation can be approximated as range/4

```
sbar = R/4
sbar
```

```
## [1] 9.049856
```

```
print('The approximated standard error = 9.049856')
```

```
## [1] "The approximated standard error = 9.049856"
```

e. Removing the outlier

```
Y <- y[-25]
Y
```

```
##  [1] 18.11420 20.12242 21.77636 17.54360 21.58735 21.09037 21.25625 24.34983
##  [9] 17.34343 24.80211 18.76566 17.60634 18.85092 21.75796 21.45614 20.07703
## [17] 18.14095 19.05527 24.67294 21.59943 19.26455 18.17310 20.38882 16.00058
```

```
# Calculating the mean and standard deviation without the outlier;
ybar <-mean(Y)
S <- sd(Y)

print('The mean = 20.15815 and the standard deviation = 2.368674')
```

```
## [1] "The mean = 20.15815 and the standard deviation = 2.368674"
```

Yes, there is a wide difference in the two standard deviations and the two means. the disparity is as a result of the outlier.

‘‘‘