HW04_Sampathirao_Anvita

Anvita Sampathirao

6/4/2019

R Markdown

#1a.

$$z = \frac{x - \mu}{\sigma}$$

```
z <- (45-70)/10
z
```

```
## [1] -2.5
```

#1b.

```
pnorm(45,mean = 70, sd =10, lower.tail = TRUE)
```

```
## [1] 0.006209665
```

#1c.

```
2*pnorm(45,mean = 70, sd =10, lower.tail = TRUE)
```

```
## [1] 0.01241933
```

#2a.

```
SMPL<- as.data.frame(matrix(1:100, nrow = 10000, ncol = 10))
RNDM<- function(d, x) {
    s <- c(d[sample(nrow(d), 1), sample(ncol(d), 1)])
    x <- x - 1
    for (i in c(1:x)) {
        row <- sample(nrow(d), 1)
        col <- sample(ncol(d), 1)
        s<- append(s, d[row, col], after = length(s))
        i <- i + 1
    }
    return(s)
}
s1 <- RNDM(SMPL, 10)
s1
```

```
##  [1]  81  83  36  95 100  15  25  24  58  85
```

#2b.

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

```
mean1 <- sum(s1)/length(s1)
mean1
```

```
## [1] 60.2
```

#2c.

$$s = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}}$$

```
standarddev<- sqrt(sum((s1-mean1)^2/(length(s1)-1)))
standarddev
```

```
## [1] 32.5672
```

#2d.

$$se = \frac{s}{\sqrt{n}}$$

```
standarderr <- standarddev/sqrt(length(s1))
standarderr
```

```
## [1] 10.29865
```

#2e.

$$CI = [\bar{x} - 1.96 * se, \bar{x} + 1.96 * se]$$

```
lowerl<-mean1-1.96*standarderr
upprl<-mean1+1.96*standarderr
lowerl
```

```
## [1] 40.01464
```

```
upprl
```

```
## [1] 80.38536
```

#2f.

$$CI = [\bar{x} - 1.83 * se, \bar{x} + 1.83 * se]$$

```
t<-qt(0.975,9)
t
```

```
## [1] 2.262157
```

```
lowerl1<-mean1-t*standarderr
upprl1<-mean1+t*standarderr
lowerl1
```

```
## [1] 36.90283
```

```r
upprl1
```

```
## [1] 83.49717
```

#2g. For a larger sample set, the distribution of the sample mean will be normally distributed. As the sample set is small (i.e.n=10) here, there will be more variations within the observations and we cannot say with confidence that the data is statistically significant.

#3a.

```r
n_sample<-c(1:100)
n<-length(n_sample)
m_1<- mean(n_sample)
sd_1<-sd(n_sample)
```

$$CI = \bar{x} \pm 1.96 * \frac{sd}{\sqrt{n}}$$

```r
lowerlimit<-m_1-1.96*sd_1/sqrt(n)
lowerlimit
```

```
## [1] 44.81375
```

```r
upperlimit<-m_1+1.96*sd_1/sqrt(n)
upperlimit
```

```
## [1] 56.18625
```

$$CI' = \bar{x} \pm 1.96 * \frac{sd}{\sqrt{n'}}$$

Also,

$$1.96 * \frac{sd}{\sqrt{n'}} = 1.96 * \frac{sd}{2 * \sqrt{n}}$$

i.e.

$$2 * \sqrt{n} = \sqrt{n'}$$

$$n' = 4 * n = 400$$

```r
n1 <- 4*n
n1 #Thus, we need 300 more observations from the original n.
```

```
## [1] 400
```

```r
lowerlimit1<-m_1-1.96*sd_1/sqrt(n1)
lowerlimit1
```

```
## [1] 47.65687
```

```r
upperlimit1<-m_1+1.96*sd_1/sqrt(n1)
upperlimit1
```

```
## [1] 53.34313
```

#3b.

$$n = (\frac{1.96 * sd}{se})^2$$

```r
number1<-round((1.96*20000/1000)^2)
number1
```

```
## [1] 1537
```

```r
number2<-round((1.96*20000/100)^2)
number2
```

```
## [1] 153664
```

#4. We pick t distribution because n<=30

```r
nruns<-1000
nsamples<-20
sample_summary<-matrix(NA,nruns,3)
for(j in 1:nruns){
  sampler <- rep(NA,nsamples)
  for(i in 1:nsamples){
    if(runif(1) < 0.5){
      sampler[i] <- rnorm(1,40,15)
    }
    else{
      sampler[i] <- rnorm(1,60,27)
    }
  }
  sample_summary[j,1] <- mean(sampler)
  standard_error <- sd(sampler)/sqrt(nsamples)
  sample_summary[j,2] <- mean(sampler) - qt(0.995,19)*standard_error
  sample_summary[j,3] <- mean(sampler) + qt(0.995,19)*standard_error
}
counter = 0
for(j in 1:nruns){
  if(50 > sample_summary[j,2] && 50 < sample_summary[j,3]){
    counter <- counter + 1
  }
}
counter
```

```
## [1] 987
```

Thus we can say with approximately 99% coonfidence that the data is statistically significant.