Confidence Interval, Sigma Unknown

Confidence Interval, Sigma Unknown

```
# This is a small demonstration of getting a
#   confidence interval for the mean
#   of a population with unknown standard deviation.
#
#  First, we will get a population
#  In this case we will get a large population
source("gnrnd5.R")
gnrnd5(182651734104,285002867)
```

```
## style= 4   size= 7342    seed= 82651   num digits= 1   alt_sign= 1
##  mean= 286.7   st dev= 28.5

## [1] "DONE "
```

```
#let us look at the head and tail values
head(L1)
```

```
## [1] 314.3 350.3 317.4 278.4 282.0 295.4
```

```
tail(L1)
```

```
## [1] 216.1 295.0 242.1 288.1 311.1 299.1
```

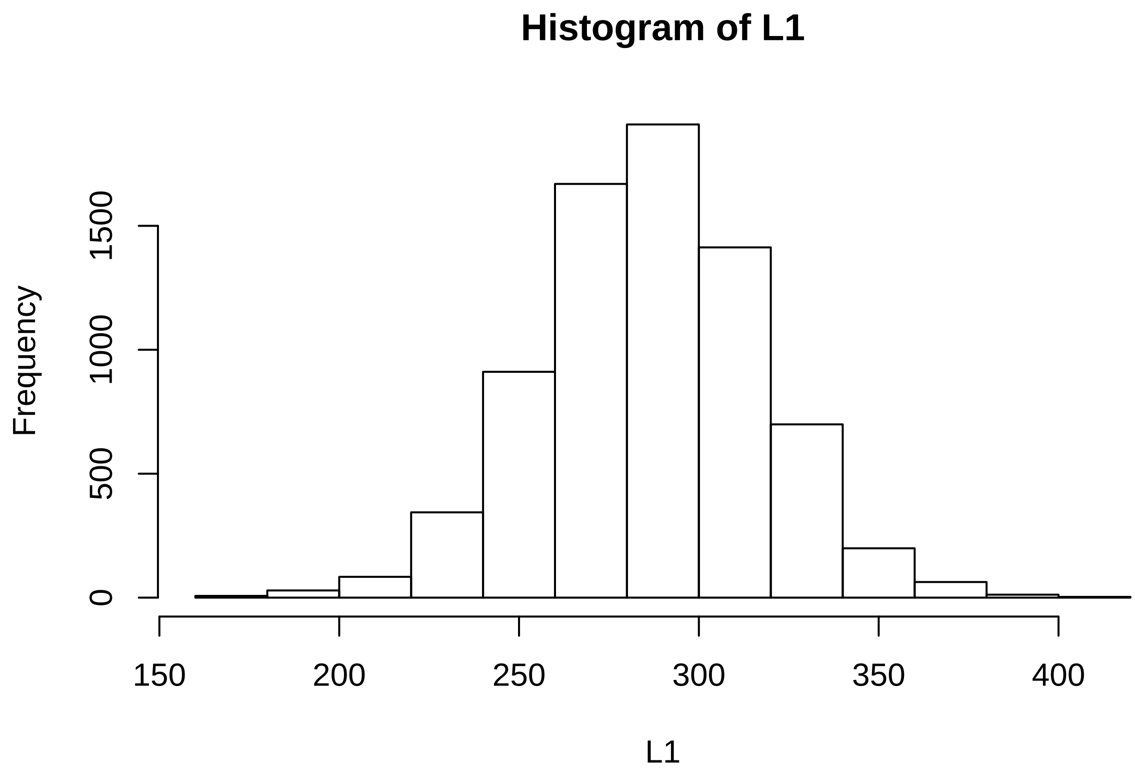
```
min(L1)
```

```
## [1] 163.9
```

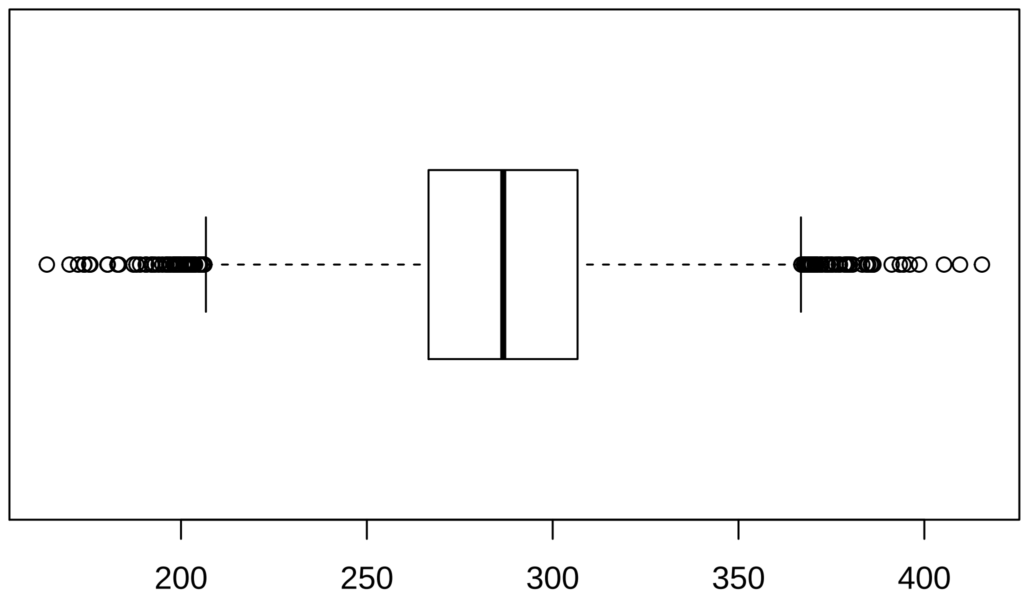
```
max(L1)
```

```
## [1] 415.5
```

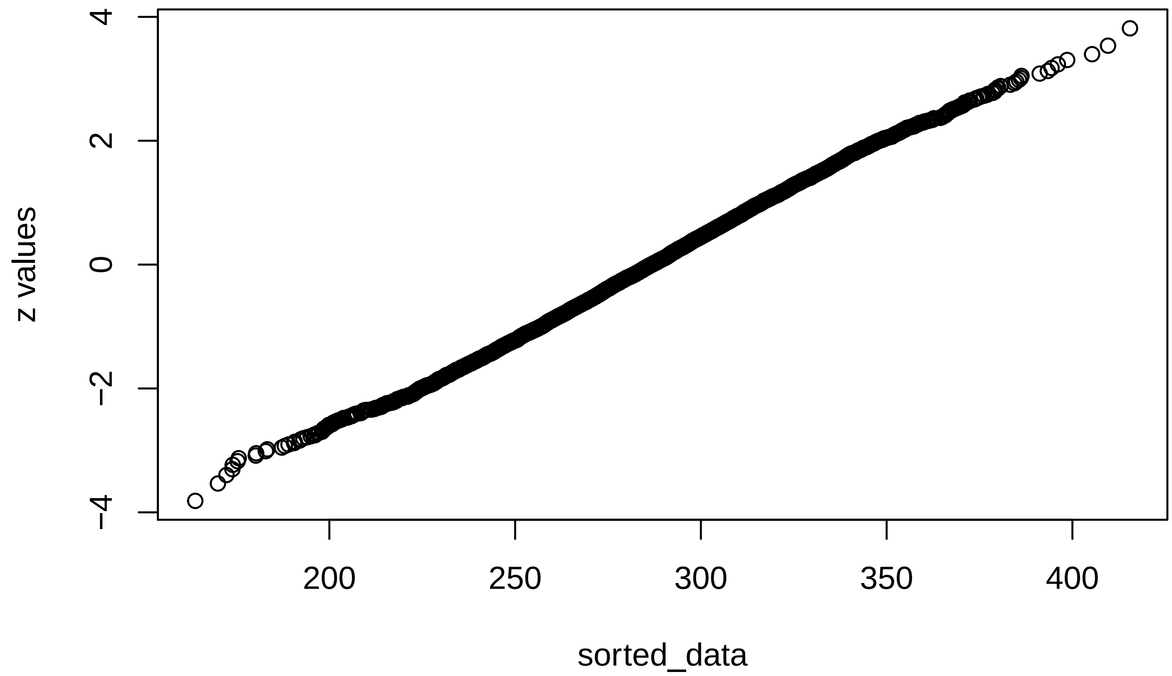
```
# now, we could find the standard deviation of the
# population, but this is supposed to be an example
# of finding confidence intervals for the mean when
# we do not know the population standard deviation.
#
# just a quick look at L1
hist(L1)
```




```
boxplot(L1, horizontal=TRUE)
```




```r
source("assess_normality.R")
assess_normality( L1 )
```




```r
#
#  L1 sure looks like a Normal distribution.
#
```

```r
# ##########################
# ##  Problem: find the 95% confidence interval
# ##      for the mean of the population when we
# ##      do not know the population standard deviation.
# ##########################

# take a simple random sample of size 23
#
#  Be careful:  Every time we do this we get
#               a different random sample
#
L2 <- as.integer( runif(23, 1, 7343) )
# L2  holds the index values of our simple random sample
L2
```

```
##  [1] 1625  933 6739 5564 3692  677 3304 2257 5387 1276 5311 1147 4287 5106
## [15] 2637  272 1835 7024 1855 4929 5690 2125 5775
```

```r
L3 <- L1[ L2 ]   # L3 holds the simple random sample
L3
```

```
##  [1] 245.3 250.7 271.9 237.5 274.6 292.7 338.5 242.6 274.9 258.7 235.1
## [12] 289.0 341.7 293.2 265.8 291.6 306.6 262.0 311.6 346.4 241.9 313.7
## [23] 367.6
```

```r
# we will get the mean of L3
xbar <- mean(L3)
xbar
```

```
## [1] 284.9391
```

```r
#
# and we will get the standard deviation of the sample
sx <- sd( L3 )
sx
```

```
## [1] 38.12598
```

```r
#  Remember that the distribution of sample means
#  will be the Student's-t distribution with n-1
#  degrees of freedom, in this case 22 degrees of
#  freedom.  And the distribution of the sample means
#  will have the same mean as the population
#  and standard deviation equal to the population
#  standard deviation divided by the square root of
#  the sample size. However, for the Student's-t we
#  the standard deviation of the sample divided by
#  the square root of the size of the sample.
#
#  The long way to generate the confidence interval
#  is to find the t-value in a Student's-t
#  distribution with 22 degrees of freedom
#  such that there is 95% of the
#  area between -t and t.  That means that 2.5% is
#  less than -t and 2.5% is greater that t.
#  We can find that z value via qt.
#
t <- qt(0.025, 22, lower.tail=FALSE)
t
```

```
## [1] 2.073873
```

```r
#
#  Then our margin of error is z*sigma/sqrt(23)
#
moe <- t*sx/sqrt(23)
moe
```

```
## [1] 16.48691
```

```r
#
#  and our confidence interval is between
#  xbar-moe and xbar+moe
#
xbar - moe
```

```
## [1] 268.4522
```

```
xbar + moe
```

```
## [1] 301.426
```

```
#   Of course, we could use the function
#   ci_unknown() to do this in
#   one easy step.
#
source("../ci_unknown.R")
ci_unknown( sx, 23, xbar, 0.95 )
```

```
##     CI Low    CI High        MOE  Std Error
## 268.452219 301.426042  16.486912   7.949817
```

go back and execute lines 34-90 many more times. Each time you get a different random sample. Therefore, each time you get a different confidence interval. Note that the MOE changes each time because the standard deviation of the sample changes for each sample. By the way, the true mean of the population is about 286.62002.