Exercise set 7

Bjørn Christian Weinbach

22nd October, 2020

Clear R environment

```r
rm(list = ls())
```

Problem 0

$N = 10$ measurments of the measurements of the hardness of a new material (alloy) gave the following data:

168	185	164	182	169
181	172	185	172	180

A typical approach is to assume that the hardness measurements follow a $N(\mu;\sigma^2)$ distribution.

Calculate the empirical mean $\hat{\mu}$ and standard deviation

$\hat{\sigma}$.

```r
data <- c(168, 185, 164, 182, 169, 181, 172, 186, 172, 180)

print(paste("Empirical mean: ", mean(data)))
```

```
## [1] "Empirical mean:  175.9"
```

```r
print(paste("Empirical sd: ", sd(data)))
```

```
## [1] "Empirical sd:  7.79529772790409"
```

Calculate the standard deviation of $\hat{\mu}$

According to the wikipedia article on standard error accessed 22nd October, 2020 (Wikipedia contributors 2020a) For each random variable, the sample mean is a good estimator of the population mean, where a "good" estimator is defined as being efficient and unbiased. Of course the estimator will likely not be the true value of the population mean since different samples drawn from the same distribution will give different sample means and hence different estimates of the true mean. Thus the sample mean is a random variable, not a constant, and consequently has its own distribution.

We can therefore calculate the standard deviation of $\hat{\mu}$, also called the standard error of the mean as follows:

$$\sigma_{\bar{X}} = \frac{\sigma}{N} \tag{1}$$

(Wikipedia contributors 2020b)

Since the population mean is seldom known. We can estimate this like so:

$$\sigma_{\bar{X}} \approx \frac{s}{N} \tag{2}$$

Where s is the sample standard deviation.

In R:

```
print(paste("Standard error: ", sd(data)/sqrt(length(data))))
```

```
## [1] "Standard error:  2.46508958593124"
```

Calculate 95% confidence intervals for μ and σ

Since we have a unknown μ and σ for the population and the sample size is small, we will use a T values for our confidence interval for the population mean.

The confidence interval is

$$[\bar{x} + t_{n-1,\alpha/2} \cdot \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1,1-\alpha/2} \cdot \frac{s}{\sqrt{n}}] \tag{3}$$

in R:

```
# Calculate sample mean, sd and no of observations
xbar <- mean(data)
s <- sd(data)
n <- length(data)

# Alpha
a <- 0.05

# Calculate confidence interval. qt is T-distribution
left <- xbar + qt(a/2, n-1) * s/sqrt(n)
right <- xbar + qt(1 - (a/2), n-1) * s/sqrt(n)
ci <- c(left, right)

# Display confidence interval
ci
```

```
## [1] 170.3236 181.4764
```

To calculate the confidence interval for σ we use the formula in the text of the exercise set.

```
left <- sqrt((n-1)*s / qchisq(1 - a/2, df=n-1))
right <- sqrt((n-1)*s / qchisq(a/2, df=n-1))
ci <- c(left, right)
ci
```

```
## [1] 1.920440 5.097115
```

Bootstrapping $\hat{\mu}$

Now we resort to bootstrapping to estimate the calculations we have above from our data. We first do this for the statistic $\hat{\mu}$

```
library(boot)

# Function for statistic - sample mean
meanestfunc <- function(data,i)
  mean(data[i])

boot.obj <- boot(data=data, statistic = meanestfunc, R=5000)
boot.obj
```

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = data, statistic = meanestfunc, R = 5000)
##
##
## Bootstrap Statistics :
##     original   bias    std. error
## t1*    175.9 -0.03672    2.380375
```

```r
boot.ci(boot.obj,type=c("norm","basic","perc","bca"))
```

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 5000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = boot.obj, type = c("norm", "basic", "perc",
##     "bca"))
##
## Intervals :
## Level      Normal              Basic
## 95%   (171.3, 180.6 )   (171.3, 180.6 )
##
## Level     Percentile            BCa
## 95%   (171.2, 180.5 )   (171.1, 180.4 )
## Calculations and Intervals on Original Scale
```

We see that our calculations for the standard error is close to the estimated standard error by bootstrapping. The confidence interval is also close to our interval. Though the calculated one used t-values and is therefore slightly wider.

Bootstrapping $\hat{\sigma}$

Bibliography

Wikipedia contributors. 2020a. "Sample Mean and Covariance — Wikipedia, the Free Encyclopedia." https://en.wikipedia.org/w/index.php?title=Sample_mean_and_covariance&oldid=938430490#Variance_of_the_sampling_distribution_of_the_sample_mean.

———. 2020b. "Standard Error — Wikipedia, the Free Encyclopedia." https://en.wikipedia.org/w/index.php?title=Standard_error&oldid=978217434.