Homework 03 - STAT440

Joseph Sepich (jps6444)

09/11/2020

```r
set.seed(42)
```

Problem 1

In this problem we will use simulation to explore the accuracy of confidence intervals. Consider a sample of n i.i.d. normal random variables with mean μ and standard deviation σ. Assume μ is unknown, but σ is known.

Part a

Generate such a sample $\mu = 1$ and $\sigma = 0.5$ for the cases $n = \{10, 100, 1000\}$. In reality this step would be done by nature, but here we can simulate it ourselves.

```r
n <- c(10, 100, 1000)
samples <- vector(mode = "list", length = length(n))

mu <- 1
sigma <- 0.5

for (i in 1:length(n)) {
  samples[[i]] <- rnorm(n[i], mean = mu, sd = sigma)
}
```

Part b

Build a 95% confidence interval for each of these samples. Remember σ is assumed to be known but not μ.

Recall that when σ is known we can create our confidence interval for μ by using a transformation to the standard normal variable resulting in:

$$[\overline{X} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}]$$

Here we use $\alpha = 0.05$, so $\frac{\alpha}{2} = 0.025$. $z_{0.025} = 1.96$.

```r
z_multi <- 1.96
intervals <- vector("list", length(n))
```

```r
for (i in 1:length(n)) {
  x_bar <- mean(samples[[i]])
  diff <- z_multi * sigma / sqrt(n[i])
  intervals[[i]] <- c(x_bar - diff, x_bar + diff)
}
```

Part c

```r
num <- 1000
n <- 100
diff <- z_multi * sigma / sqrt(n)
count_in <- 0

for (i in 1:num) {
  sample <- rnorm(n, mean = mu, sd = sigma)
  x_bar <- mean(sample)
  if(x_bar - diff < mu & x_bar + diff > mu) {
    count_in <- count_in + 1
  }
}
print(count_in)
```

```
## [1] 952
```

Out of 1000 simulations of sampling 100 values from the normal distribution with $\mu = 1$ and $\sigma = 0.5$, we get μ in our 95% confidence interval 952 times, or 95.2% of the time. This makes perfect sense from the definition of a confidence interval, which states that for a given confidence level, the level represents the frequency with which a confidence interval at that level should contain the given parameter. In plain English this means that for a confidence interval at x% confidence, we should expect x% of the confidence intervals constructed at that level to contain the parameter.

Part d

If the parameter σ was not known, then we would have to use the student's t distribution to calculate the confidence interval.

Problem 2