R: Working with distributions to get confidence intervals

BMR 617

February 10\textsuperscript{th} 2020
Working with qnorm

• In the course work, we set a confidence level of \( C \leq 0.99 \), and then computed the critical \( z \)-value as \( z \leq qnorm((1+C)/2) \)

• Let’s see why this was

• We want to choose \( z \) so that 99\% (or in general, a proportion \( C \)) of all values lie between \(-z\) and \( z\).

• As always, it’s easier to think about this with a picture...
Area = C
Area = \frac{1-C}{2}

Area = C
Area = $\frac{1-C}{2}$

Area = $\frac{1-C}{2}$

Area = C

-z

2
The \texttt{qnorm}(p) function gives the z-value for which $P(X \leq z) = p$.

The probability we need here is $rac{1 - C}{2} + C = \frac{1 - C + 2C}{2} = \frac{1 + C}{2}$.

So we can set

$z \leftarrow \texttt{qnorm}((1+C)/2)$
Creating functions

• To calculate the confidence interval when we know the population standard deviation, we used a series of commands:

```r
C <- 0.99 # confidence
z <- qnorm((1+C)/2)
sigma <- 500 # "known"
n <- length(x)
xbar <- mean(x)
lower <- xbar - z * (sigma / sqrt(n))
upper <- xbar + z * (sigma / sqrt(n))
```

• To calculate a confidence interval for different data and/or confidence levels, we need to run all these commands again.

• A more usable way to do this in R is to create a function
Our confidence interval function

• To compute the confidence interval, we needed three inputs:
  • The data, \( x \)
  • The confidence level, \( C \)
  • The population standard deviation, \( \sigma \)

• We can create a function which takes these inputs and \textit{returns} the confidence interval
Confidence interval function

```r
confInt <- function(x, sigma, C=0.95) {
  n <- length(x)
  z <- qnorm((1+C)/2)
  xbar <- mean(x)
  m <- z * (sigma / sqrt(n))
  lower <- xbar - m
  upper <- xbar + m
  return(c(lower, upper))
}
```
Using the confidence interval function

```r
load('birthweight.rdata')
confInt(birthweight$birthweight, 500, 0.99)
[1] 2996.165 3226.555

confInt(birthweight$birthweight, 500, 0.90)
[1] 3037.80 3184.92

confInt(birthweight$birthweight, 500)
[1] 3023.708 3199.012
```
Confidence interval when population standard deviation is unknown

• In the course, it was stated that the margin of error when the population standard deviation is unknown is

\[
m = \frac{t^* s}{\sqrt{n}}
\]

• Here \( t^* \) is the critical value of the t-distribution with the specified confidence level

• The t-distribution depends on a number of degrees of freedom, which is one less than the sample size

• We can create a function to compute this in R, instead of relying on part of the T-test

• The \( qt \) function does the same as the \( qnorm \) function, but for the t-distribution
Creating the new function

• `confIntT <- function(x, conf=0.95) {
    n <- length(x)
    t <- qt((1+conf)/2, df= n - 1)
    xbar <- mean(x)
    s <- sd(x)
    m <- t * s / sqrt(n)
    lower <- xbar - m
    upper <- xbar + m
    return(c(lower, upper))
}

• Hint: start with `confIntT <- edit(confInt)` to edit the previous function!