Lab 4: Point and interval estimation for a population mean

STAT218

library(tidyverse)  
load('data/nhanes.RData')  
load('data/temps.RData')

The goal of this lab is to learn to compute a point estimate, standard error, and interval estimate for a population mean “by hand” by performing the arithmetic directly in R. You will have an opportunity to practice interpreting these quantities as you go.

We will also use this lab for a short class activity to explore how often the interval estimate we introduced is “correct”.

### Point estimation

#### Estimate for the population mean

Since the point estimate for the population mean of a numeric variable is the sample mean, you already know how to perform the calculation in R. We’ll store this for later use:

# retrieve total cholesterol variable  
totchol <- nhanes$totchol  
  
# compute and store sample mean  
totchol.mean <- mean(totchol)  
totchol.mean

[1] 5.042938

The only novelty here is that we now interpret this as a point estimate of the population mean total cholesterol:

The mean total cholesterol of U.S. adults is estimated to be 5.043 mmol/L.

This is in contrast to the interpretation as a descriptive summary:

The average total cholesterol among the respondents in the NHANES survey was 5.043 mmol/L.

Both interpretations are valid; just different. By interpreting the sample mean as a point estimate, we are implicitly assuming that the data are a random sample from the U.S. adult population.

|  |
| --- |
| Your turn |
| Use the temps data to estimate mean body temperature.  # retrieve variable of interest  # compute and store sample mean  *Check your understanding:*   * interpret the result as a descriptive summary * interpret the result as a point estimate |

#### Standard error for the sample mean

A standard error is a measure of the sampling variability of a point estimate. Technically, it’s an estimate of the point estimate’s standard deviation across all possible random samples of a fixed size.

The standard error for the sample mean is calculated according to the formula:

Where:

* is the sample standard deviation
* is the sample size

To calculate this in R, we perform the arithmetic by hand (for now):

# store sample sd and sample size  
totchol.sd <- sd(totchol)  
totchol.n <- length(totchol)  
  
# compute standard error  
totchol.se <- totchol.sd/sqrt(totchol.n)  
totchol.se

[1] 0.01906042

This result is interpreted as follows:

The root average deviation of the sample mean from the population mean is estimated to be 0.0191 mmol/L.

|  |
| --- |
| Your turn |
| Calculate and interpret the standard error for the sample mean of the body temperature variable.  # store sample sd and sample size  # compute standard error |

### Interval estimation

#### Interval estimate for the mean

A common interval for the population mean is:

For now, we’ll calculate this by directly performing the arithmetic. Later, you’ll use commands that return interval estimates by default.

# add/subtract two standard errors from the mean  
totchol.mean + c(-2, 2)\*totchol.se

[1] 5.004817 5.081059

We’ll talk more about the exact interpretation later; for now, you should think of this as a range of plausible values for the population mean.

|  |
| --- |
| Your turn |
| Calculate an interval estimate for the mean body temperature using the body temperature data.  # interval estimate for mean body temp |

#### How often is the interval correct?

An interval “covers” the population mean if the true value is between the interval endpoints.

We can explore how often the interval covers the parameter by having everyone in the class simulate their own sample from a population with a known mean and check whether the interval they obtain from the sample covers the population mean or not.

The commands below simulate a sample and then compute an interval.

# function to simulate body temp data from a population with mean 98.6  
sample.bodytemps <- function(n){  
 rnorm(n, mean = 98.6, sd = 1)  
}  
  
# simulate a sample of body temperatures  
bodytemp <- sample.bodytemps(n = 150)  
  
# compute interval 'ingredients'  
bodytemp.mean <- mean(bodytemp)  
bodytemp.sd <- sd(bodytemp)  
bodytemp.n <- length(bodytemp)  
bodytemp.se <- bodytemp.sd/sqrt(bodytemp.n)  
  
# compute interval estimate  
bodytemp.mean + c(-2, 2)\*bodytemp.se

[1] 98.40416 98.70546

# margin of error  
2\*bodytemp.se

[1] 0.150648

|  |
| --- |
| Your turn |
| Use the example above to generate a sample of size 20 and compute an interval estimate for the mean body temperature.  Then:   1. Determine whether your interval covers the population mean. 2. Compute the margin of error used in your interval ().   Repeat with . Then fill out [this form](https://forms.office.com/r/dGtfLzavM6). |

### Practice problems

1. Vu and Harrington exercise 4.1. Additionally:
   1. Compute an interval estimate for the mean BGC of nests.
   2. Supposing a sample of 30 nests returned exactly the same summary statistics, recompute your interval in (e). Is the margin of error smaller or larger?
2. The brfss dataset contains a measurement of body weight, weight, as well as a variable, wtdesire, that is the desired weight reported by respondents.
   1. Estimate the mean difference between actual and desired weight. Report the point estimate and standard error.
   2. Does the point estimate suggest that the average U.S. adult would prefer to lose or gain weight?
   3. Compute an interval estimate for the mean difference between actual and desired weight.