

progress report

Rui Gao

11/30/2021

```
knitr::opts_chunk$set(echo = TRUE)
set.seed(2021)
yrbss_2007 <- readRDS("yrbss_2007.rds")
yrbss_2017 <- readRDS("yrbss_2017.rds")
```

```
library(tidyverse)
library(ggplot2)
```

##Task 1 a. Using repeated samples of size $n = 10, 100,$ and 1000 from the bmi variable, describe the sampling distribution of the sample mean of BMI in 2017. Include at least one plot to help describe your results. Report the means and standard deviations of the sampling distributions, and describe how they change with increasing sample size.

```
#constant
```

```
n_bmi <- 1000
pop_sd_2017 <- sd(yrbss_2017$bmi)
length_2017_bmi <- length(nchar(yrbss_2017$bmi))
```

```
df_2017_bmi <- data.frame(yrbss_2017$bmi)
df_2017_bmi$index <- 1:nrow(df_2017_bmi)
```

```
#I don't know how to select multiple values by index in column
#I tried convert it into rows,
#I also tried convert the data frame to value,
#and some other ways, but still.....
#not sure where are the mistakes
#so, I'd rather just write frames of codes here
```

```
#function of selecting numbers from the dataset
```

```
get_value_2017 <- function(n){
  with(df_2017_bmi, yrbss_2017.bmi[index ==
    c(numbers selected by# "sample.int(length_2017_bmi, n, replace = TRUE)" )])
}
```

```
#get sample mean's function
```

```
get_mean_2017 <- function(n, n_bmi, pop_sd_2017){
  replicate(n_bmi, mean(get_value_2017))
}
```

```
#generate random values for each group
```

```
ns <- c(10, 100, 1000)
values <- lapply(ns, get_value_2017, n = ns)
```

```
#plot for each group, I cannot check, so if this is wrong, I would do it separately
```

```
qplot(values)
```

```
#calculate the
means <- lapply(ns, get_mean_2017, n_bmi = n_bmi, pop_sd = pop_sd)
```

```
spread_sampdist <- sapply(means, sd)
true_se <- pop_sd_2017/ns
```

```
rbind(round(spread_sampdist, 3),
round(true_se, 3))
```

```
#conclusion
```

The means and standard deviations for each group are..., comparing them with the population's mean and standard deviation, we can notice that as the size changed,

##Task 2 1. How has the BMI of high-school students changed between 2007 and 2017? Are high-schoolers getting more overweight?

Comparing the means and medians of BMI of two different years, I will try the two sample test and Mood's test. However, since the dietary habit hasn't changed too much during the decade and the food suppliers/brands are almost the same in the USA, we cannot say the data is totally independent.

2. In 2017, are 12th graders more or less likely than 9th graders to be "physically active at least 60 minutes per day on 5 or more days"?

Finding the proportions' difference between two graders, contrasting the mean and confident interval between them. As the leaving time decrease, the 12th graders are facing the more and more enrollment pressure, does your intuition correct?

3. How much sleep do highschoolers get?

Try to calculate the mean and interval of sleeping time. The data is discrete and unaccurate, would that be a problem? The schools have basically the same schedule for students, it's hard to say the data is independent. By the way, if you need me to tell you these information, I don't even sure the answer because I don't count them and the answer would also be influenced by the near days' situations rather than the real situations.