Final Project

Nora Quick

2021-11-30

Task 1: Simulation Study

```
yrbss_2007 <- readRDS("yrbss_2007.rds")
yrbss_2017 <- readRDS("yrbss_2017.rds")
```

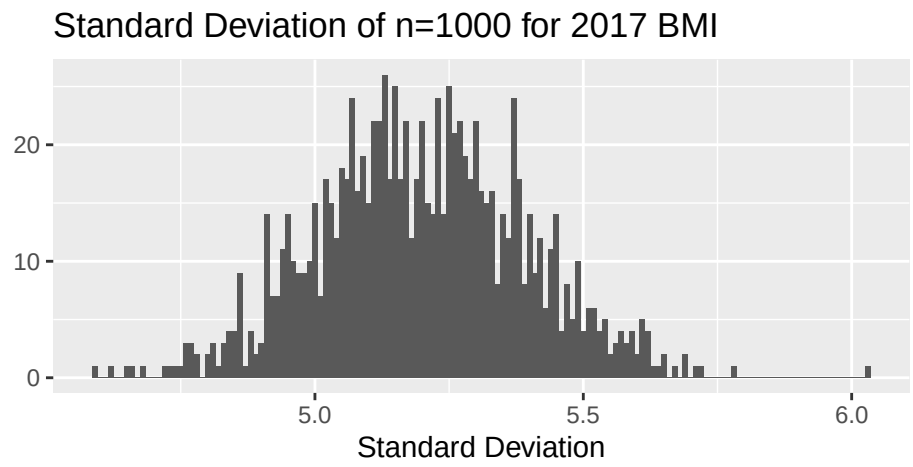
a) Using repeated samples of size n = 10, 100, and 1000 from the bmi variable, describe the sampling distribution of the *sample mean* of BMI in 2017. Include at least one plot to help describe your results. Report the means and standard deviations of the sampling distributions, and describe how they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

```
bmi_2017 <- yrbss_2017$bmi
```

```
mean_10 <- replicate(1000, mean(sample(x = bmi_2017, size = 10)))
mean_100 <- replicate(1000, mean(sample(x = bmi_2017, size = 100)))
mean_1000 <- replicate(1000, mean(sample(x = bmi_2017, size = 1000)))
```

```
sd_10 <- replicate(1000, sd(sample(x = bmi_2017, size = 10)))
sd_100 <- replicate(1000, sd(sample(x = bmi_2017, size = 100)))
sd_1000 <- replicate(1000, sd(sample(x = bmi_2017, size = 1000)))
```

```
qplot(sd_1000, binwidth = 0.01) +
  labs(
    title = "Standard Deviation of n=1000 for 2017 BMI",
    x = "Standard Deviation"
  )
```



- As the sample size increases the distribution becomes more normal around the mean value / average standard deviation.

b) Repeat the simulation in part (a), but this time use the *25th percentile* as the sample statistic. In R, `quantile(x, prob = 0.25)` will give you the 25th percentile of the values in `x`.

c) Repeat the simulation in part (a), but this time use the *sample minimum* as the sample statistic.

d) Describe the sampling distribution of the *difference in the sample median BMI between 2017 and 2007*, by using repeated samples of size n_1 = 5, n_2 = 5, n_1 = 10, n_2 = 10 and n_1 = 100, n_2 = 100. Report the means and standard deviations of the sampling distributions, and describe how they change with the different sample sizes.

e) *Summarize your results.*

Task 2: Data Analysis

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

- What is the sample mean difference between the BMI of high-school students between 2007 and 2017? Is the mean BMI for high-school students in 2017 skewed to the right/higher?

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"?

- Is there a larger count of TRUEs to qn79 for 12th graders than there are 9th graders in the 2017 survey?

3) How much sleep do highschoolers get?

- What is the mean of q88 for all high-schoolers? (Add 2007 mean with 2017 mean. Divide by 2)?