Gawron_Homework6

Nicholas Gawron

3/3/2021

```r
knitr::opts_chunk$set(echo = TRUE)
library(tidyverse)
library(readxl)
library(tinytex)
```

Corresponding code for a problem's part will be **ABOVE** the solution.

Problems

1. A marketing research firm estimates the proportion of potential customers preferring a certain brand of lipstick by "randomly" selecting 100 women who come by their booth in a shopping mall. Of the 100 sampled, 65 women stated a preference for brand A.

a. We would estimate the true proportion of women preferting brand A by calculating the proportion of women in the sample that had prefered the brand

b.

c.

d.

2. A school desires to estimate the average score that may be obtained on a reading comprehension exam for students in the sixth grade. The school's students are grouped into three tracks, with the faster learners in track I, the slower learners in track III, and the rest in track II. The school decides to stratify on tracks because this method should reduce the variability in test scores. The sixth grade contains 55 students in track I, 80 in track II and 65 in track III. A stratified random sample of 50 students is proportionally allocated and yields simple random samples of $n_1 = 14$, $n_2 = 20$ and $n_3 = 16$ from tracks I, II and III respectively. The data is contained in the accompanying excel file.

a. We will consider the three groups

```r
ScoreData <- (readxl::read_excel('data/HW6Data.XLS'));ScoreData
```

```
## # A tibble: 20 x 3
##    `Track I` TrackII TrackIII
##        <dbl>   <dbl>    <dbl>
##  1        80      85       42
##  2        92      82       32
##  3        68      48       36
##  4        85      75       31
##  5        72      53       65
##  6        87      73       29
##  7        85      65       43
##  8        91      78       19
##  9        90      49       53
```

```
## 10         81       69        14
## 11         62       72        61
## 12         79       81        31
## 13         61       53        42
## 14         83       59        30
## 15         NA       68        39
## 16         NA       52        32
## 17         NA       71        NA
## 18         NA       61        NA
## 19         NA       59        NA
## 20         NA       42        NA
```

```
Track1 <- (na.omit(ScoreData$`Track I`)); T1L <- length(Track1)
Track2 <-  (na.omit(ScoreData$TrackII)); T2L <- length(Track2)
Track3 <-  (na.omit(ScoreData$TrackIII)); T3L <- length(Track3)
N <- T1L + T2L + T3L

Yst <-(1/(N))*((T1L*mean(Track1))+(T2L*mean(Track2))+(T3L*mean(Track3))); Yst
```

```
## [1] 60.2
```

```
#FPC For each Strata
fpc1 <- (1-(T1L/N))
fpc2 <- (1-(T2L/N))
fpc3 <- (1-(T3L/N))

#variance
VYst = (1/(N)^2)*(fpc1*sd(Track1)^2/T1L)
```

3. We will now look at cavities.

```
n <-10;N<- 400;
Cavities <- c(1 ,4 ,1 ,0 ,3 ,2, 4 ,0 ,3 ,2)
```

4. We will consider both parts

```
Abv <-  c(48 ,48.7 ,50.1 ,43.3 ,47.5 ,49.4 ,39.9 ,52 ,46.7 ,50.5,45.6 ,49.7 ,45.3 ,46.9 ,48.5)
Bel <- c(37.8, 45, 44.2, 60, 54.2, 56.4, 59.3, 44.4, 41.8, 52.9,45.7, 57, 48.1, 58.2, 42.5, 41.1)
AllTemp <- c(Abv,Bel)

#part a
meanAll <- mean(AllTemp ) # expected value of All temps
sdA<- sd(AllTemp)
BdAll<- 2*sqrt(sdA^2/length(AllTemp)); # computes the Boundary
```

a. We will ignore the FPC for the simple random sample since the total population is all backyard pools which is trivially larger than 20 times our smaple size of $n = 31$. The sample mean for an SRS is given by: 48.4096774. The standard deviation for the sample is calculated to be 5.7234229. The boundary, given by the formula: $B = 2\sqrt{\frac{s^2}{n}}$ computes to 2.0559142. We are 95% confident that the true mean backyard pool temperature is captured by the interval: **(46.3537632,50.4655916)**. Since $50 \in (46.3537632, 50.4655916)$, there is not enough evidence to suggest the average pool temperature is different from the recommended 50 degrees.

```
  Yst <- (1/length(AllTemp)) *( (length(Abv)*mean(Abv)) +  (length(Bel)*mean(Bel)));
fpcAbv <- 1 - (length(Abv) /length(AllTemp) )
fpcBel <-1 - (length(Bel) /length(AllTemp) )
```

```r
Var <- ((1/length(AllTemp))^2)*(length(Bel))^2*fpcBel*sd(Bel)^2/length(Bel)
Bd <- 2*sqrt(Var)
```

b. We will now consider a stratified sample.

c. Treating the sample as stratified does improve our