

13.1: Sampling Error (Section @ref{samplingerror})

Here we will repeatedly sample from the NHANES Height variable in order to obtain the sampling distribution of the mean.

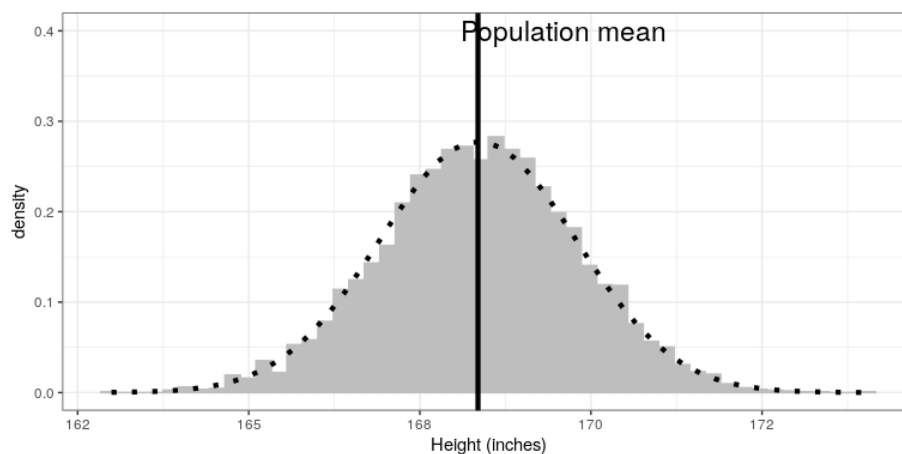
```
sampSize <- 50 # size of sample
nsamps <- 5000 # number of samples we will take

# set up variable to store all of the results
sampMeans <- tibble(meanHeight=rep(NA,nsamps))

# Loop through and repeatedly sample and compute the mean
for (i in 1:nsamps) {
  sampMeans$meanHeight[i] <- NHANES_adult %>%
    sample_n(sampSize) %>%
    summarize(meanHeight=mean(Height)) %>%
    pull(meanHeight)
}
```

Now let's plot the sampling distribution. We will also overlay the sampling distribution of the mean predicted on the basis of the population mean and standard deviation, to show that it properly describes the actual sampling distribution.

```
# pipe the sampMeans data frame into ggplot
sampMeans %>%
  ggplot(aes(meanHeight)) +
  # create histogram using density rather than count
  geom_histogram(
    aes(y = ..density..),
    bins = 50,
    col = "gray",
    fill = "gray"
  ) +
  # add a vertical line for the population mean
  geom_vline(xintercept = mean(NHANES_adult$Height),
1.5      size=) +
  # add a label for the line
  annotate(
    "text",
169.6    x = ,
.4      y = ,
    label = "Population mean",
    size=6
  ) +
  # label the x axis
  labs(x = "Height (inches)") +
  # add normal based on population mean/sd
  stat_function(
    fun = dnorm, n = sampSize,
    args = list(
      mean = mean(NHANES_adult$Height),
      sd = sd(NHANES_adult$Height)/sqrt(sampSize)
    ),
1.5    size = ,
    color = "black",
    linetype='dotted'
  )
```



This page titled [13.1: Sampling Error \(Section @ref\(samplingerror\)\)](#) is shared under a [CC BY-NC 4.0](#) license and was authored, remixed, and/or curated by [Russell A. Poldrack](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.