WS #9 - SEs

Wednesday, October 1, 2025

Math 154 - Jo Hardin

Your Name: _____

Names of people you worked with: _____

Have your academic interests changed since you entered college? In what way?

Task:

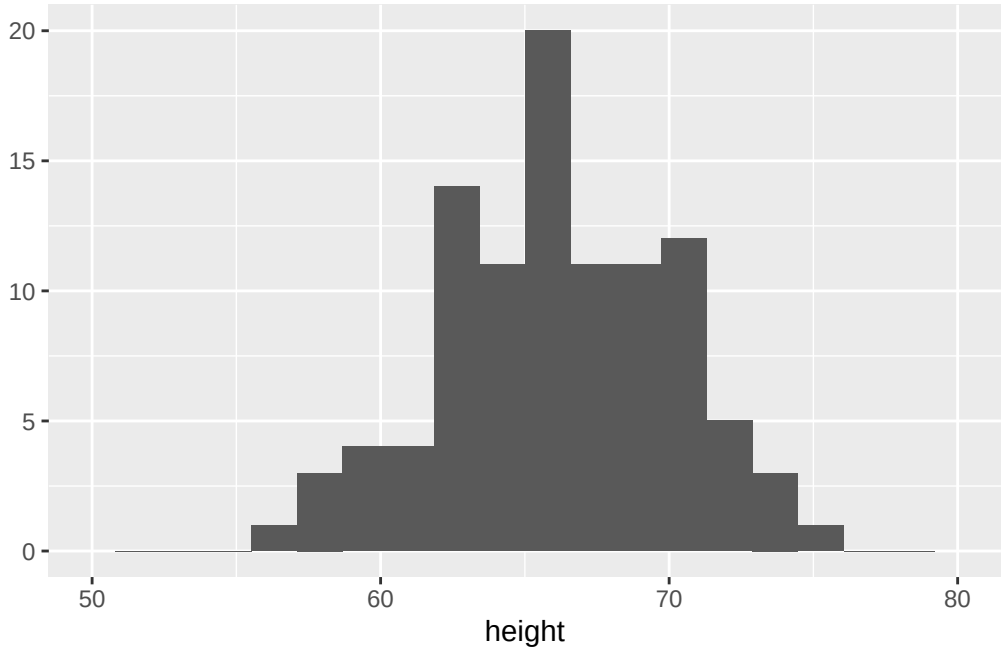
Consider a situation where you are trying to learn about the heights of Claremont Colleges students. The parameter of interest is the true median height of all undergraduates. You'll use the median as the statistic on the observed data. Sketch the following 3 plots (as densities / histograms):

1. The distribution of the heights in a sample of size 100. In inches, mark the center of the curve as well as some indication of the spread of the heights.
2. As a class project, you and your 50 best friends each collect samples of size 100. Sketch the distribution of the 50 sample medians. In inches, mark the center of the curve as well as some indication of the spread of the sample medians.
3. You all run out of time, and end up sampling only 10 undergraduates each. Sketch the distribution of the 50 sample medians. In inches, mark the center of the curve as well as some indication of the spread of the sample medians.

Solution

1. Note that the distribution is centered at 66 and goes from about 58 to 74.

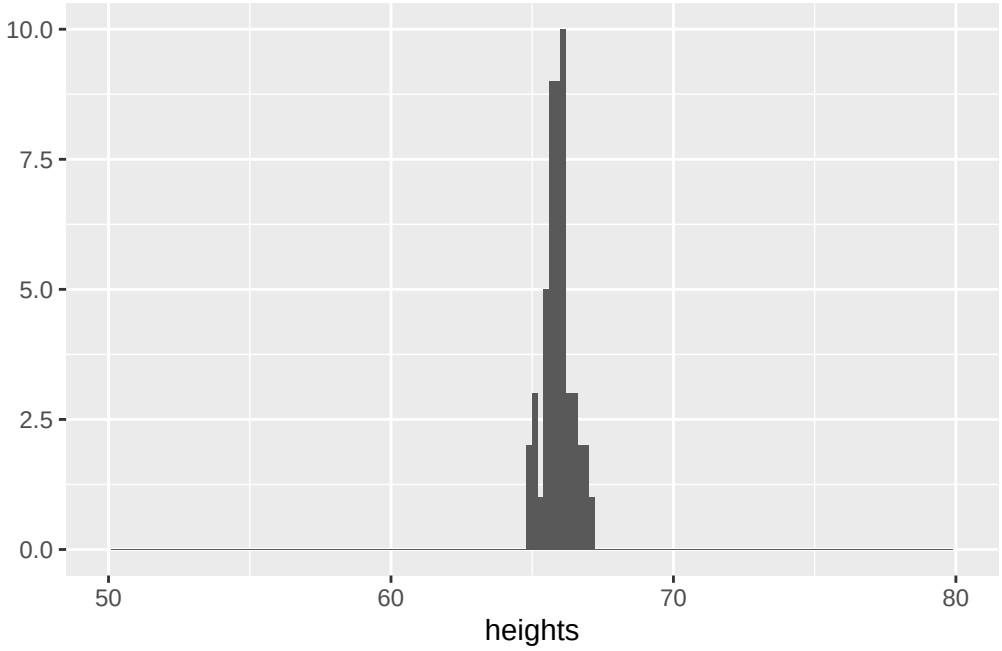
```
set.seed(47)
data.frame(height = rnorm(100, mean = 66, sd = 4)) |>
  ggplot() +
  geom_histogram(aes(x = height), bins = 20) +
  ylab("") + xlim(c(50,80))
```



2. The center of the distribution of medians is still at 66, but it is **much** narrower than the plot in #1. That's because we don't expect the medians to vary as much as we expect the individual observations to vary.

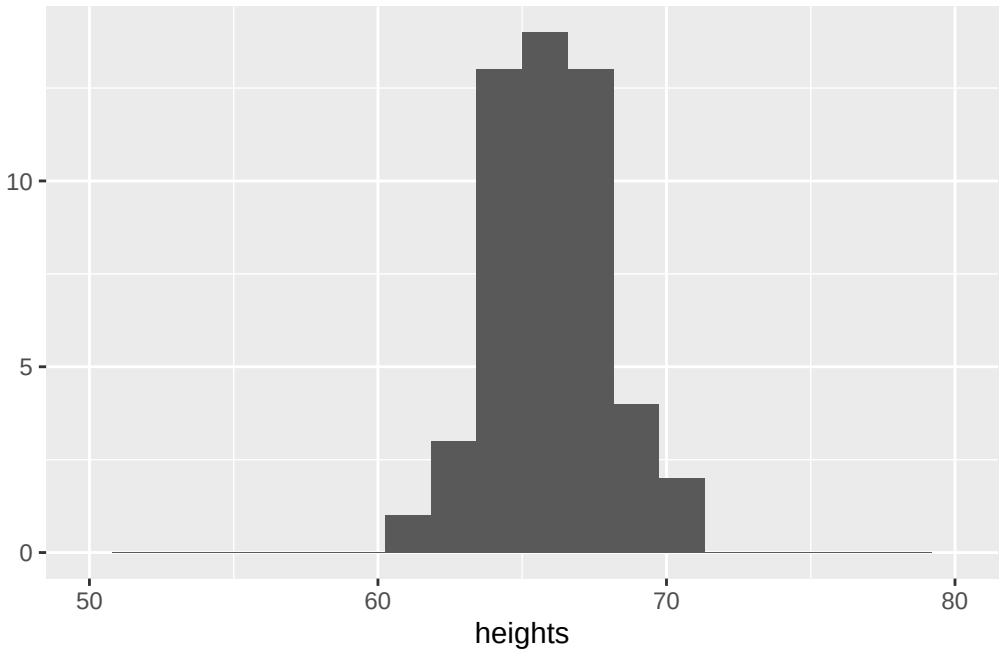
```
set.seed(47)
reps <- 50
n_people <- 100

norm_func <- function(n_obs){
  data.frame(heights = median(rnorm(n_obs, mean = 66, sd = 4)) )
  }

map(rep(n_people, reps), norm_func) |>
  list_rbind() |>
```

```
  ggplot() +
  geom_histogram(aes(x = heights), bins = 150) +
  ylab("") + xlim(c(50,80))
```



3. Again, the center is at 66. This type the spread is wider than when you sampled 100 people but narrower than when you plotted individual observations.

```
set.seed(10)
reps <- 50
n_people <- 10
map(rep(n_people, reps), norm_func) |>
  list_rbind() |>
  ggplot() +
  geom_histogram(aes(x = heights), bins = 20) +
  ylab("") + xlim(c(50,80))
```


10
5
0
50
60
70
80
heights