

OBJECTIVES

- Insert a population data and summarize its statistics.
- Withdraw samples from the population.
- Investigate the sampling distribution of the sample mean for different sample sizes.
- Investigate the relation between the sample size and the standard error (sampling variability of the mean).

STEP 1: Revision

```
install.packages("dplyr", dependencies = TRUE)
install.packages("ggplot2", dependencies = TRUE)
install.packages("statsr", dependencies = TRUE)

library(statsr)
library(dplyr)
library(ggplot2)

data(ames)

ames %>% summarise(mu = mean(area), pop_med = median(area), sigma = sd(area), pop_iqr =
IQR(area), pop_min = min(area), pop_max = max(area), pop_q1 = quantile(area, 0.25), pop_q3
= quantile(area, 0.75))

# A tibble: 1 x 8
      mu pop_med sigma pop_iqr pop_min pop_max pop_q1 pop_q3
  <dbl> <dbl> <dbl> <dbl> <int> <int> <dbl> <dbl>
1 1500.  1442  506.   617.   334   5642  1126  1743.
```

STEP 2: the step's objective is to take a random sample from the population

```
sampl <- ames %>% sample_n(size = 50)

sampl %>% summarise(mu_sample = mean(area), sample_med = median(area), sigma_sample =
sd(area), sample_iqr = IQR(area), sample_min = min(area), sample_max = max(area),
sample_q1 = quantile(area, 0.25), sample_q3 = quantile(area, 0.75))

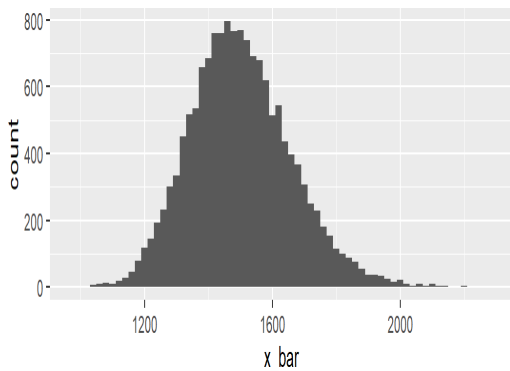
# A tibble: 1 x 8
  mu_sample sample_med sigma_sample sample_iqr sample_min sample_max sample_q1 sample_q3
  <dbl> <dbl> <dbl> <dbl> <int> <int> <dbl> <dbl>
1 1459.  1444.  477.   614.   848   3112  1093.  1707.
```

STEP 3: the step's objective is to draw many samples

```
sample_means10 <- ames %>%
rep_sample_n(size = 10, reps =
15000, replace = TRUE) %>%

summarise(x_bar = mean(area))

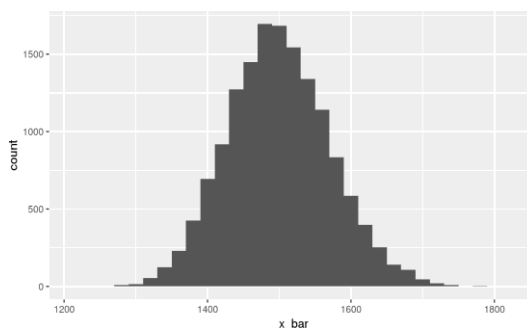
ggplot(data = sample_means10,
aes(x = x_bar)) +
geom_histogram(binwidth = 20)
```



```
sample_means100 <- ames %>%
rep_sample_n(size = 100, reps =
15000, replace = TRUE) %>%

summarise(x_bar = mean(area))

ggplot(data = sample_means100,
aes(x = x_bar)) +
geom_histogram(binwidth = 20)
```



```
sample_means1000 <- ames %>%
rep_sample_n(size = 1000, reps =
15000, replace = TRUE) %>%

summarise(x_bar = mean(area))

ggplot(data = sample_means1000,
aes(x = x_bar)) +
geom_histogram(binwidth = 20)
```

