Expectation of sampling distribution of sample mean

$$
\begin{aligned}
E(\underbrace{\tfrac{1}{n}(Y_1 + Y_2 + \ldots + Y_n)}_{\text{sample mean}}) &= \frac{1}{n} E(Y_1 + Y_2 + \ldots + Y_n) \\
&= \frac{1}{n}\left(E(Y_1) + E(Y_2) + \ldots + E(Y_n)\right) \\
&= \frac{1}{n}(\mu + \mu + \ldots + \mu) \\
&= \frac{1}{n} n\mu \\
&= \mu
\end{aligned}
$$

The sampling distribution of the sample mean is centered around the population mean.

↑ regardless of population shape

Variance of sampling distribution of sample mean

$$
\begin{aligned}
Var(\tfrac{1}{n}(Y_1 + Y_2 + \ldots + Y_n)) &= \frac{1}{n^2}\left(Var\left(Y_1 + \ldots + Y_n\right)\right) \\
&= \frac{1}{n^2}\left(Var(Y_1) + \ldots + Var(Y_n)\right) \quad \text{because } Y_i \text{ iid } \equiv \text{independent} \\
&= \frac{1}{n^2}\left(\sigma^2 + \ldots + \sigma^2\right) \\
&= \frac{1}{n^2} n\sigma^2 \\
&= \frac{1}{n}\sigma^2 = \frac{\sigma^2}{n}
\end{aligned}
$$

The variance of the sampling distribution of the sample mean is smaller than the population variance ($n > 1$), and decrease with increasing n.

Your Turn

Let X be a random variable with an unknown distribution.

I obtain $X_1, \ldots, X_{10}$ i.i.d samples from the distribution. I get:

$5, 3, 7, 4, 4, 3, 7, 2, 7, 3$

How would you estimate $P(X \leq 5)$?

$$\frac{\#\text{ obs. values} \leq 5}{10} = \frac{7}{10} = 0.7$$

estimate higher moments $\longrightarrow$ $\hat{P}(X \leq 5)$

"fitting distribution"

$$P(X \leq 3) = \frac{4}{10} = 0.4$$

Example: Commute times

```
library(tidyverse)
n <- 5
n_sim <- 1000

# Generate many samples
samples <- rerun(.n = n_sim,
  sample(class_data$commute_time, size = n))

# Do something to each sample
sample_means <- map_dbl(samples, ~ mean(.x))
```

how many times

always a list

what?

.x placeholder for an element of .x

map(.x,

for each element in .x, do something

rerun & map-dbl — functions from purrr package