I05a - Sampling distribution

STAT 5870 (Engineering)
Iowa State University

August 28, 2024

Sampling distribution

The sampling distribution of a statistic is the distribution of the statistic *over different realizations of the data*.

Find the following sampling distributions:

- If $Y_i \stackrel{ind}{\sim} N(\mu, \sigma^2)$,

$$\overline{Y} \qquad \text{and} \qquad \frac{\overline{Y} - \mu}{S/\sqrt{n}}.$$

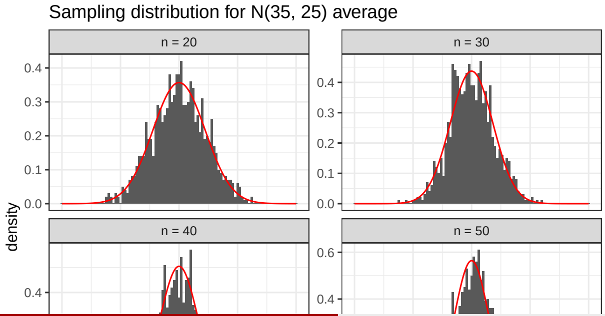
- If $Y \sim Bin(n, p)$,

$$\frac{Y}{n}.$$

Normal model

Let $Y_i \stackrel{ind}{\sim} N(\mu, \sigma^2)$, then $\overline{Y} \sim N(\mu, \sigma^2/n)$.

```
Warning:  The dot-dot notation (`..density..`) was deprecated in ggplot2 3.4.0.
i Please use `after_stat(density)` instead.
This warning is displayed once every 8 hours.
Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
generated.
```

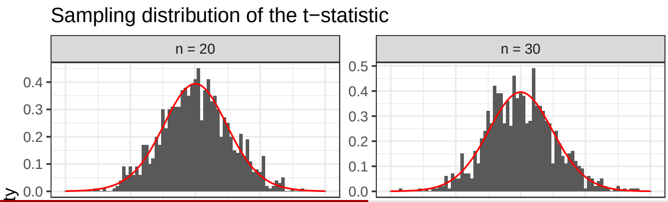


Normal model

Let $Y_i \stackrel{ind}{\sim} N(\mu, \sigma^2)$, then the t-statistic

$$T = \frac{\overline{Y} - \mu}{S/\sqrt{n}} \sim t_{n-1}.$$

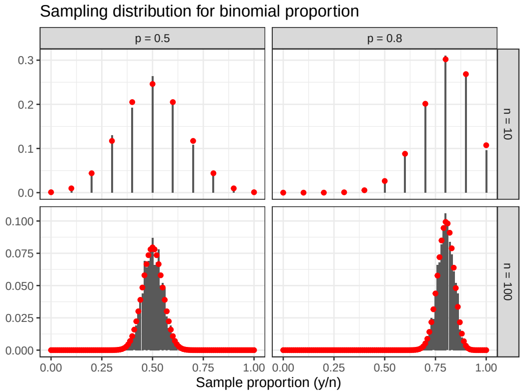
```
Warning:  Returning more (or less) than 1 row per `summarise()` group was
deprecated in dplyr 1.1.0.
i Please use `reframe()` instead.
i When switching from `summarise()` to `reframe()`, remember that `reframe()`
always returns an ungrouped
 data frame and adjust accordingly.
Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
generated.
```



Binomial model

Let $Y \sim Bin(n, p)$, then

$$P\left(\frac{Y}{n} = p\right) = P(Y = np), \qquad p = 0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1.$$



Approximate sampling distributions

Recall that from the Central Limit Theorem (CLT):

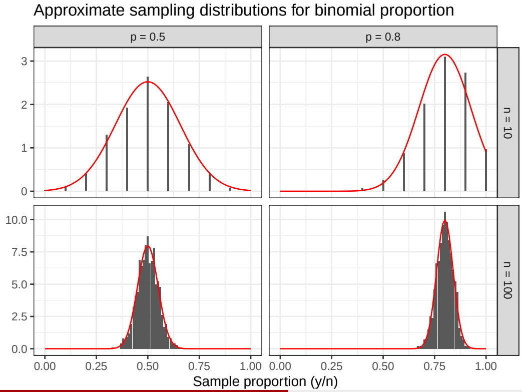
$$S = \sum_{i=1}^{n} X_i \dot{\sim} N(n\mu, n\sigma^2) \qquad \text{and} \qquad \overline{X} = S/n \dot{\sim} N(\mu, \sigma^2/n)$$

for independent X_i with $E[X_i] = \mu$ and $Var[X_i] = \sigma^2$.

Approximate sampling distribution for binomial proportion

If $Y = \sum_{i=1}^{n} X_i$ with $X_i \stackrel{ind}{\sim} Ber(p)$, then

$$\frac{Y}{n} \dot{\sim} N\left(p, \frac{p[1-p]}{n}\right).$$



Summary

Sampling distributions:

- If $Y_i \stackrel{ind}{\sim} N(\mu, \sigma^2)$,
  - $\overline{Y} \sim N(\mu, \sigma^2/n)$ and
  - $\frac{\overline{Y}-\mu}{S/\sqrt{n}} \sim t_{n-1}$.
- If $Y \sim Bin(n, p)$,
  - $P\left(\frac{Y}{n} = p\right) = P(Y = np)$ and
  - $\frac{Y}{n} \dot\sim N\left(p, \frac{p[1-p]}{n}\right)$.
- If X_i independent with $E[X_i] = \mu$ and $Var[X_i] = \sigma^2$, then

$$S = \sum_{i=1}^{n} X_i \dot\sim N(n\mu, n\sigma^2)$$

and

$$\overline{X} = S/n \dot\sim N(\mu, \sigma^2/n)$$