Lecture 15 Sampling Distribution And The Central Limit Theorem, part I

BIO210 Biostatistics

Xi Chen

Spring, 2022

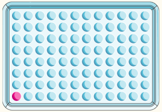
School of Life Sciences
Southern University of Science and Technology

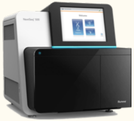
南方科技大学生命科学学院
SUSTech · SCHOOL OF LIFE SCIENCES

Use Sample Statistics To Estimate Population Parameters

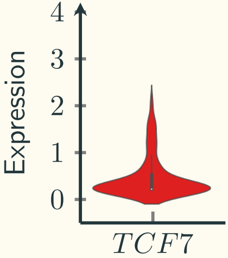
$$\text{Sample(s)} \xrightarrow[\text{Sampling distribution}]{?} \text{Population}$$

Intuition of Sampling Distribution

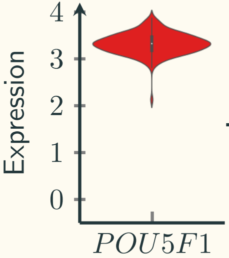




Experiment_name
- plate_1
  - cell_1
  - cell_2
  - cell_3
  - ⋮
- plate_2
  - cell_1
  - cell_2
  - cell_3
  - ⋮



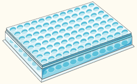
→ Not T cells



→ Probably ES cells

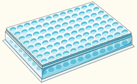
Intuition of Sampling Distribution

100 plates (samples)



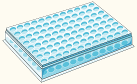
$n = 96$

$\bar{x} = 3.07$



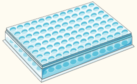
$n = 96$

$\bar{x} = 2.97$



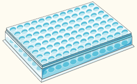
$n = 96$

$\bar{x} = 2.80$



$n = 96$

$\bar{x} = 15$

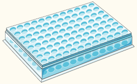


$n = 96$

$\bar{x} = 3.50$

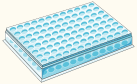
$\vdots$

$\vdots$



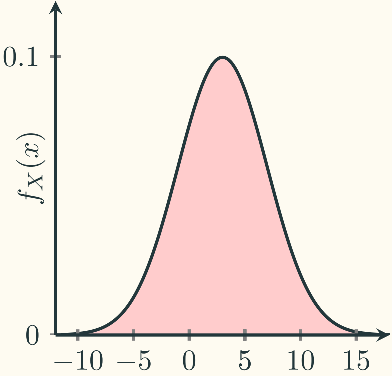
$n = 96$

$\bar{x} = -10$



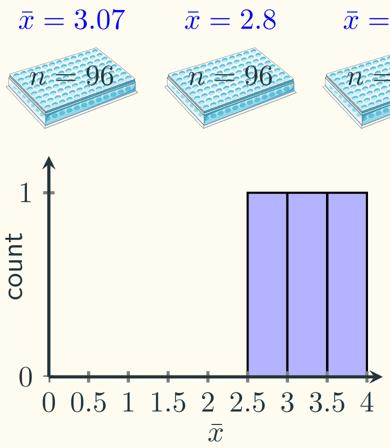
$n = 96$

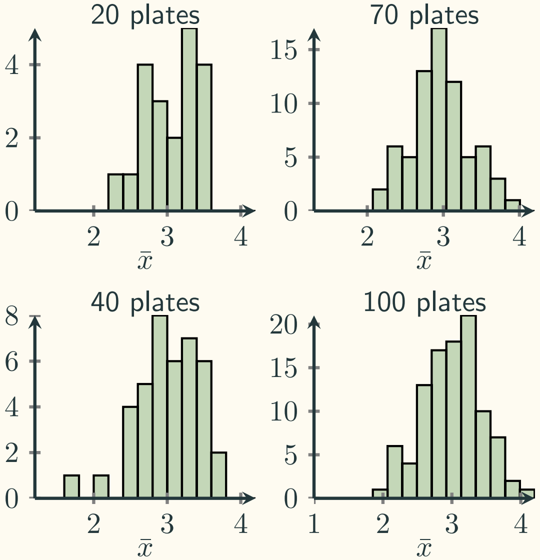
$\bar{x} = 3.27$



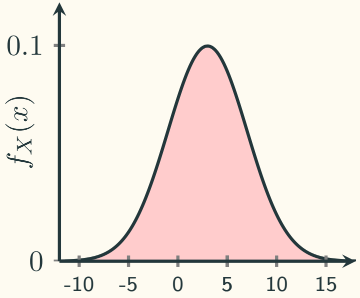
$POU5F1 \sim \mathcal{N}(\mu = 3, \sigma^2 = 4^2)$

Intuition of Sampling Distribution



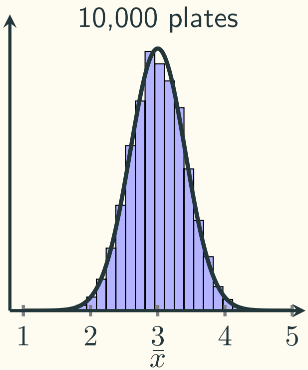


Sampling Distribution of The Sample Mean



$POU5F1 \sim \mathcal{N}(\mu = 3, \sigma^2 = 4^2)$

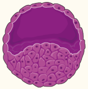
Sample	**Sample mean**
$n = 96$	$\longmapsto \bar{x} = 3.05$
$n = 96$	$\longmapsto \bar{x} = 3.13$
$n = 96$	$\longmapsto \bar{x} = 3.15$
$n = 96$	$\longmapsto \bar{x} = 2.95$
$n = 96$	$\longmapsto \bar{x} = 2.87$
$n = 96$	$\longmapsto \bar{x} = 3.20$
$\vdots$	$\vdots \quad \vdots$
$n = 96$	$\longmapsto \bar{x} = 2.50$
$\vdots$	$\vdots \quad \vdots$



$\bar{X} \sim$

Sampling distribution of the sample mean

i.i.d. Random Variables



$expression \sim D(\mu, \sigma^2)$

X_1	X_2	X_3	$\cdots$	$\cdots$	X_{n-1}	X_n	$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$	
x_1	x_2	x_3			x_{n-1}	x_n	$\bar{x}$	
2.3	2.5	3.3	$\cdots$	$\cdots$	4.2	2.9	3.0	sample 1
4.3	1.5	3.8	$\cdots$	$\cdots$	3.0	2.5	3.2	sample 2
2.3	1.8	5.8	$\cdots$	$\cdots$	2.0	4.5	2.8	sample 3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	

$X_1, X_2, ..., X_n$ are independent and identically distributed (**i.i.d.**) random variables.

$$X_1 \sim D(\mu, \sigma^2)$$
$$X_2 \sim D(\mu, \sigma^2)$$
$$X_3 \sim D(\mu, \sigma^2)$$
$$\vdots$$
$$X_{n-1} \sim D(\mu, \sigma^2)$$
$$X_n \sim D(\mu, \sigma^2)$$

$$\bar{X} \sim ?(?, ?)$$

The Central Limit Theorem

By Pierre Simon de Laplace in 1810.

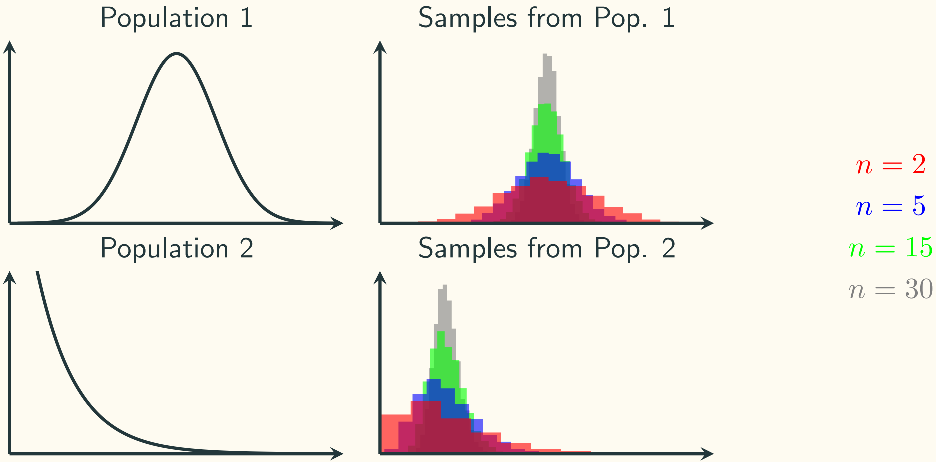
Theorem

The sampling distribution of the sample mean of n independent and identically distributed (i.i.d.) random variables is approximately normal, even if original variables themselves are not normally distributed, provided that n is large enough.

$$\bar{X} \dot{\sim} \mathcal{N}(\mu_{\bar{X}}, \sigma^2_{\bar{X}}),\ where\ \mu_{\bar{X}} = \mu, \sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$$

$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$: **standard error**.

The Central Limit Theorem



Exercise

Trip planning: In general, a person drinks 2 L of water when active outdoors with a standard deviation of 0.7 L. You are planning a full day nature trip for 50 people and will bring 110 L of water. What is the probability that you will run out?

Event of interest: $\{$ run out of water $\}$

Equivalent event #1: $\{$ 50 people drink more than 100 L $\}$

Equivalent event #2: $\{$ average water use per person $>$ 2.2 L $\}$