Lecture 15 Sampling Distribution And The Central Limit Theorem

BIO210 Biostatistics

Xi Chen

Spring, 2023

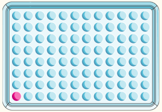
School of Life Sciences
Southern University of Science and Technology

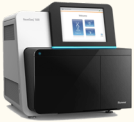
南方科技大学生命科学学院
SUSTech · SCHOOL OF
LIFE SCIENCES

Use Sample Statistics To Estimate Population Parameters

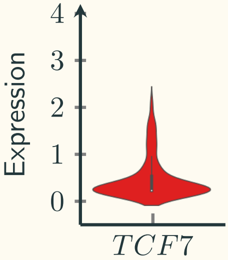
$$\text{Sample(s)} \xrightarrow[\text{Sampling distribution}]{?} \text{Population}$$

Intuition of Sampling Distribution

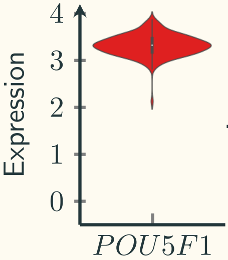




Experiment_name
- plate_1
  - cell_1
  - cell_2
  - cell_3
  - ⋮
- plate_2
  - cell_1
  - cell_2
  - cell_3
  - ⋮

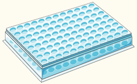


→ Not T cells



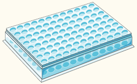
→ Probably ES cells

100 plates (samples)



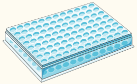
$n = 96$

$\bar{x} = 3.07$



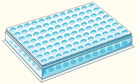
$n = 96$

$\bar{x} = 2.97$



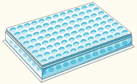
$n = 96$

$\bar{x} = 2.80$



$n = 96$

$\bar{x} = 4.95$

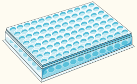


$n = 96$

$\bar{x} = 3.60$

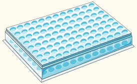
$\vdots$

$\vdots$



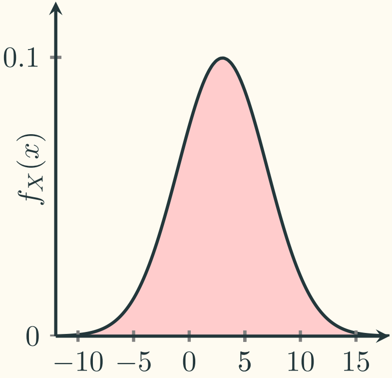
$n = 96$

$\bar{x} = 1.05$



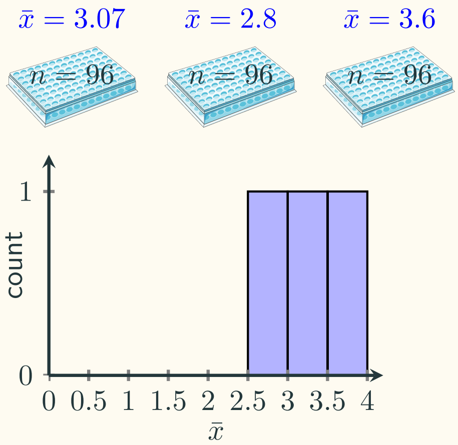
$n = 96$

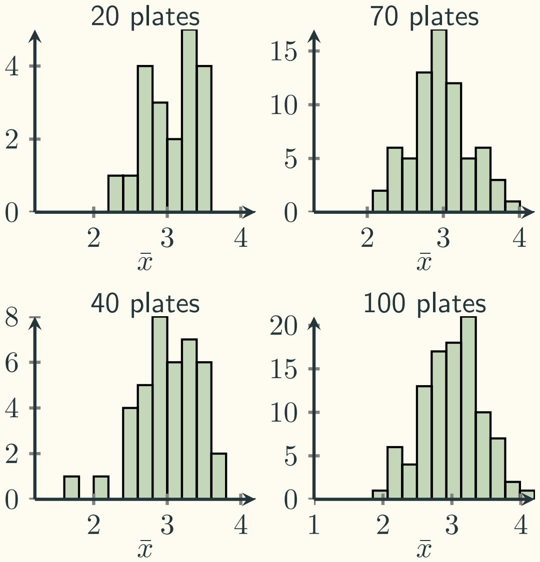
$\bar{x} = 3.27$



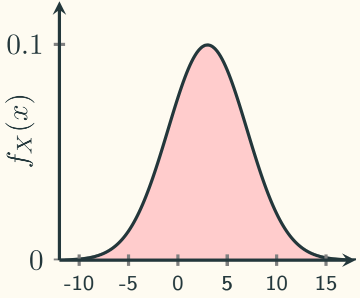
$POU5F1 \sim \mathcal{N}(\mu = 3, \sigma^2 = 4^2)$

Intuition of Sampling Distribution



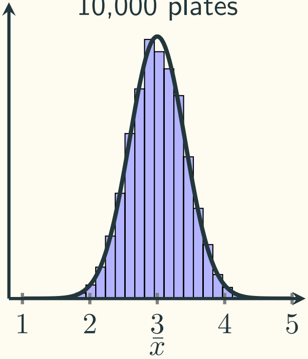


Sampling Distribution of The Sample Mean



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

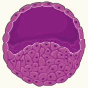
Sample		**Sample mean**
$n = 96$	$\longmapsto$	$\bar{x} = 3.05$
$n = 96$	$\longmapsto$	$\bar{x} = 3.13$
$n = 96$	$\longmapsto$	$\bar{x} = 3.15$
$n = 96$	$\longmapsto$	$\bar{x} = 2.95$
$n = 96$	$\longmapsto$	$\bar{x} = 2.87$
$n = 96$	$\longmapsto$	$\bar{x} = 3.20$
$\vdots$	$\vdots$	$\vdots$
$n = 96$	$\longmapsto$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



$\bar{X} \sim$

Sampling distribution of the sample mean

i.i.d. Random Variables



Gene A expression $\sim \mathcal{D}(\mu, \sigma^2)$

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.

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

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

The Central Limit Theorem

By Pierre Simon de Laplace in 1810.

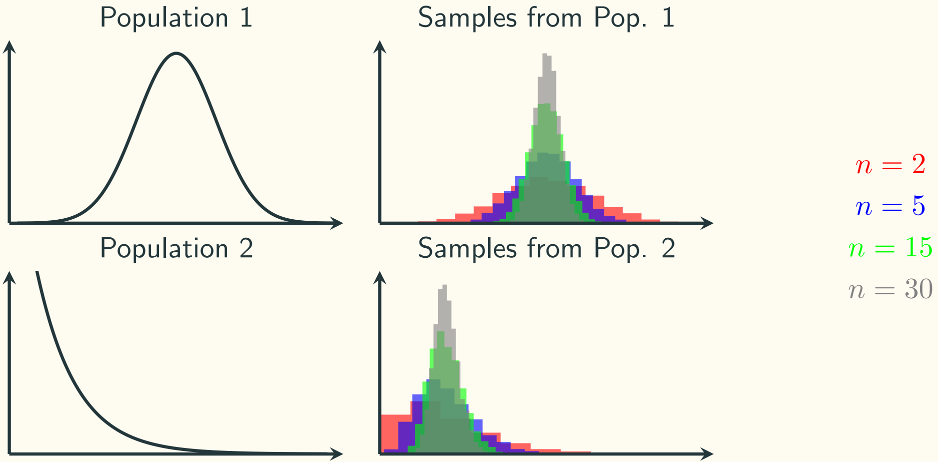
Theorem

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**.

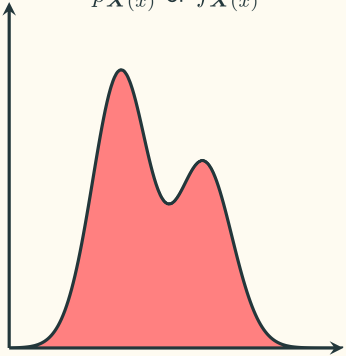
The Central Limit Theorem



Three Distributions

Population Distribution

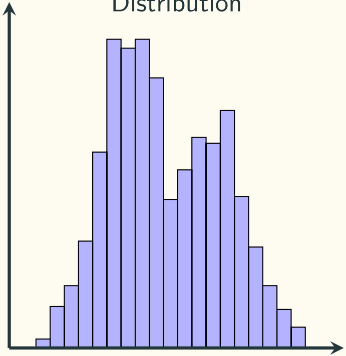
$p_{\boldsymbol{X}(x)}$ or $f_{\boldsymbol{X}(x)}$



Population parameters

$\mathbb{E}\left[X\right] = \mu,\ \mathbb{V}\mathrm{ar}\left(X\right) = \sigma^2$

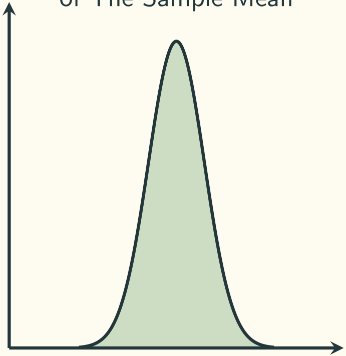
Sample Distribution



Sample statistics

$\bar{x},\ s^2$

Sampling Distribution of The Sample Mean

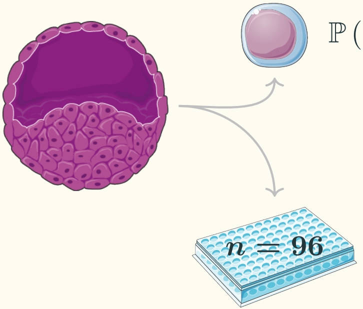


Parameters

$\mathbb{E}\left[\bar{X}\right] = \mu_{\bar{X}},\ \mathbb{V}\mathrm{ar}\left(\bar{X}\right) = \sigma^2_{\bar{X}}$

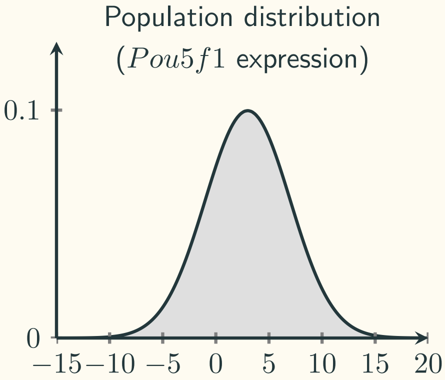
Practice: $Pou5f1$ Expression

Based on the previous research, the expression of $Pou5f1$ in all ES cells follow a normal distribution with $\mu = 3$ and $\sigma^2 = 4^2$.

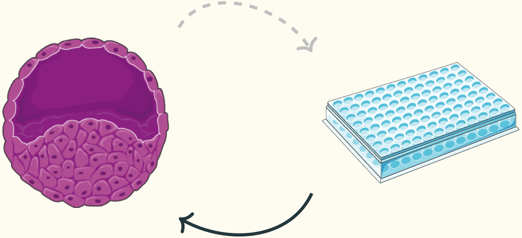


$\mathbb{P}\left(x \geqslant 8\right)$?

$\mathbb{P}\left(\bar{x} \geqslant 3.8\right)$?



Estimation



Use info. from the sample
to do a point estimation

Population parameter

μ, σ^2

Sample statistics

$\bar{x}, s^2$

- **Estimator**

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

- **Estimate**

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Unbiased Estimator

We say the following estimators are unbiased estimators:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Because:

$$\mathbb{E}\left[\bar{X}\right] = \mu$$

$$\mathbb{E}\left[S^2\right] = \sigma^2$$