American University of Beirut

STAT 210: Elementary Statistics for Sciences

2022–2023 Fall

Siamak Taati

Chapter 7

Sampling distributions

Part 2

Distribution of sample statistics

Sample mean

Mean and standard deviation of sample mean

If $X_1, X_2, \ldots, X_n$ is a random sample of size n from a population with mean μ and standard deviation σ, then the sample mean

$$\overline{X} := \frac{X_1 + X_2 + \cdots + X_n}{n}$$

is a random variable with

- $\boxed{\mathbb{E}[\overline{X}] = \mu}$ (valid for both types of sample)

  $\overline{X}$ is an unbiased estimator for μ.

- $\boxed{\mathbb{SD}[\overline{X}] = \sigma/\sqrt{n}}$ (assuming the sample is with replacement)

  The larger the sample, the more accurate the estimator.

Sample mean

Example (Weight of orangutans)

Mean weight of a random sample of 50 orangutans

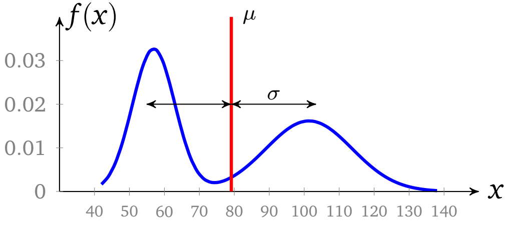
$$\overline{W} := \frac{W_1 + W_2 + W_3 + \cdots + W_{50}}{50}$$

is a RV with

$$\mathbb{E}[\overline{W}] = \mu = \boxed{79.15\,\text{Kg}}$$

$$\mathbb{SD}[\overline{W}] = \frac{\sigma}{\sqrt{50}} = \frac{24.28}{\sqrt{50}} \approx \boxed{3.43\,\text{Kg}}$$

We can illustrate these using a computer simulation.



Population distribution of weight of orangutans

Sample mean

Example (Weight of orangutans)

Drawing a random sample of 50 from the population of orangutans is a random experiment which we can simulate using a computer.

To visualize the distribution of $\overline{W}$, we can simulate this experiment many many times and calculate the mean weight $\overline{W}$ of the sampled orangutans in each simulation.

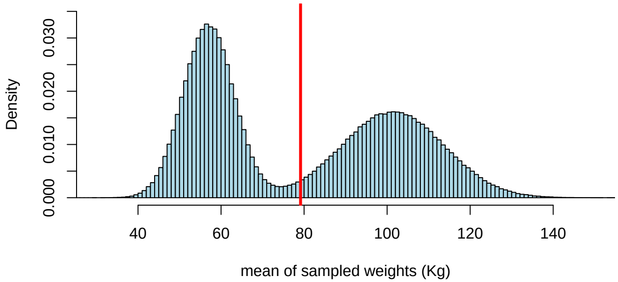
simulation	W_1	W_2	W_3	$\cdots$	W_{50}	$\overline{W}$
1	60.53	95.88	61.01	$\cdots$	90.54	74.6260
2	64.06	65.79	61.92	$\cdots$	94.01	76.7592
3	60.73	100.03	90.35	$\cdots$	53.36	80.2934
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
1,000,000	102.51	77.68	117.99	$\cdots$	60.03	80.5370

We can then plot the histogram of the values of $\overline{W}$ in all these repeated simulations. The histogram would approximate the probability density function of $\overline{W}$.

Sample mean

Example (Weight of orangutans)

The following histograms are based on 1,000,000 repeated simulations.

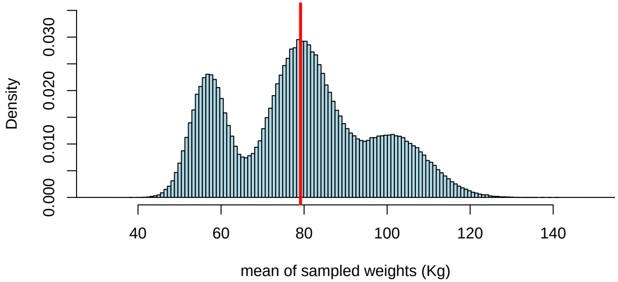


Approximate distribution of $\overline{W}$ for a sample of size $n = 1$

Sample mean

Example (Weight of orangutans)

The following histograms are based on 1,000,000 repeated simulations.

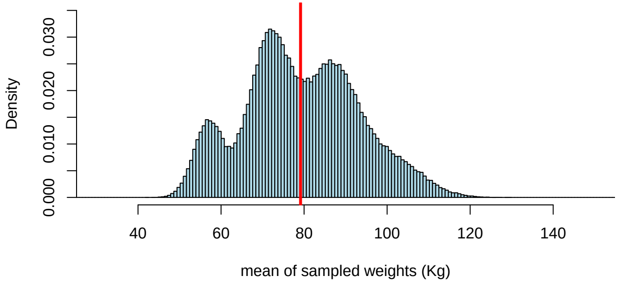


Approximate distribution of $\overline{W}$ for a sample of size $n = 2$

Sample mean

Example (Weight of orangutans)

The following histograms are based on 1,000,000 repeated simulations.

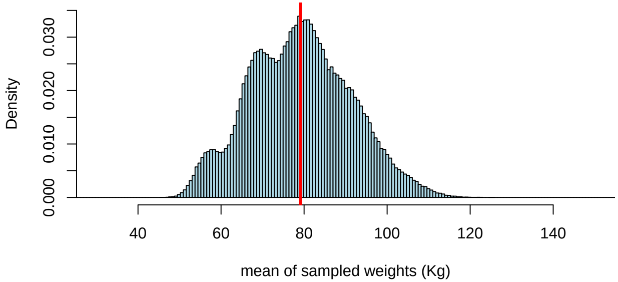


Approximate distribution of $\overline{W}$ for a sample of size $n = 3$

Sample mean

Example (Weight of orangutans)

The following histograms are based on 1,000,000 repeated simulations.

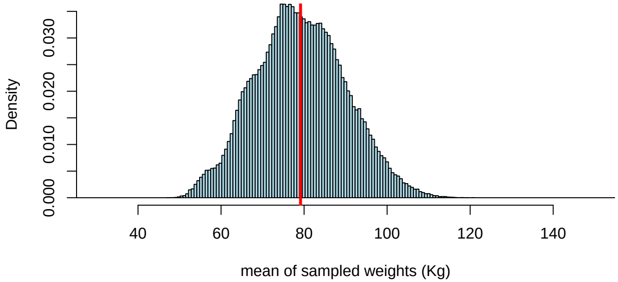


Approximate distribution of $\overline{W}$ for a sample of size $n = 4$

Sample mean

Example (Weight of orangutans)

The following histograms are based on 1,000,000 repeated simulations.

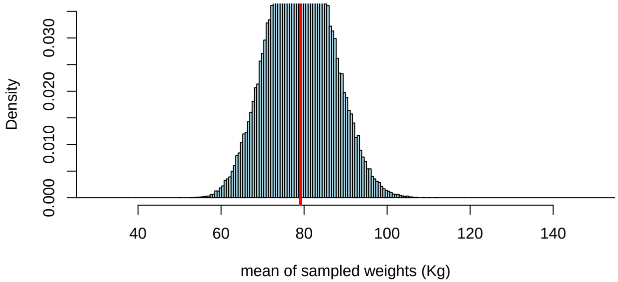


Approximate distribution of $\overline{W}$ for a sample of size $n = 5$

Sample mean

Example (Weight of orangutans)

The following histograms are based on 1,000,000 repeated simulations.

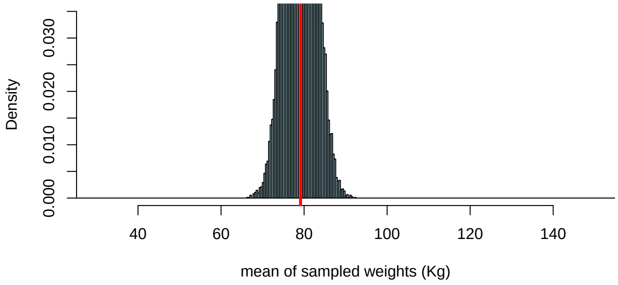


Approximate distribution of $\overline{W}$ for a sample of size $n = 10$

Sample mean

Example (Weight of orangutans)

The following histograms are based on 1,000,000 repeated simulations.

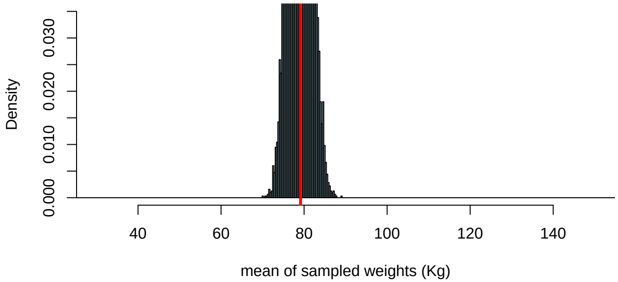


Approximate distribution of $\overline{W}$ for a sample of size $n = 50$

Sample mean

Example (Weight of orangutans)

The following histograms are based on 1,000,000 repeated simulations.

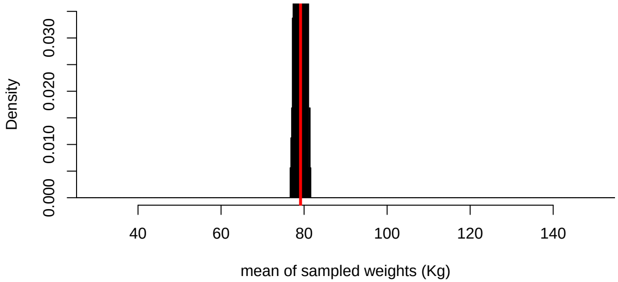


Approximate distribution of $\overline{W}$ for a sample of size $n = 100$

Sample mean

Example (Weight of orangutans)

The following histograms are based on 1,000,000 repeated simulations.



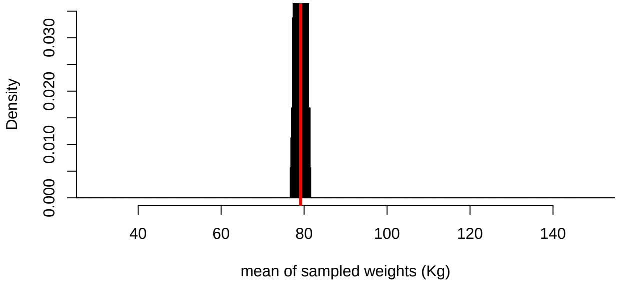
Approximate distribution of $\overline{W}$ for a sample of size $n = 1000$

Sample mean

Example (Weight of orangutans)

The following histograms are based on 1,000,000 repeated simulations.

Ⓠ What do these simulations tell us about the distribution of $\overline{W}$?



Approximate distribution of $\overline{W}$ for a sample of size $n = 1000$

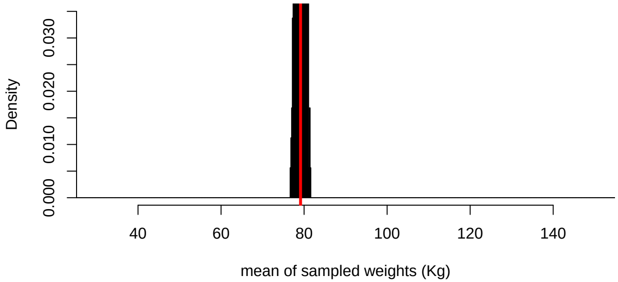
Sample mean

Example (Weight of orangutans)

The following histograms are based on 1,000,000 repeated simulations.

Ⓠ What do these simulations tell us about the distribution of $\overline{W}$?

$\boxed{\text{A}}$ ① The distribution seems to be centered at μ.



Approximate distribution of $\overline{W}$ for a sample of size $n = 1000$

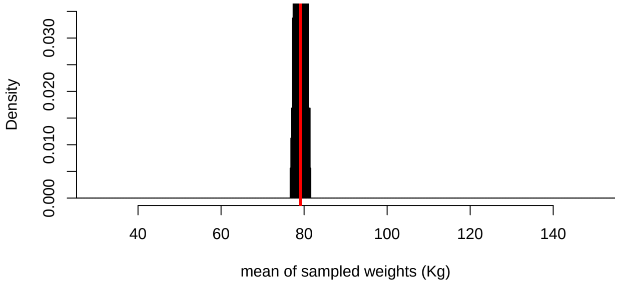
Sample mean

Example (Weight of orangutans)

The following histograms are based on 1,000,000 repeated simulations.

Q What do these simulations tell us about the distribution of $\overline{W}$?

A
1. The distribution seems to be centered at μ.
2. The larger the sample size, the more concentrated around μ.



Approximate distribution of $\overline{W}$ for a sample of size $n = 1000$

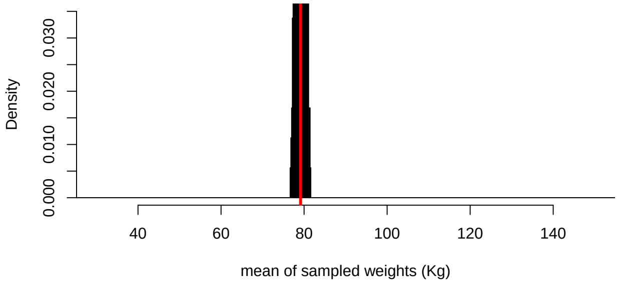
Sample mean

Example (Weight of orangutans)

The following histograms are based on 1,000,000 repeated simulations.

Q What do these simulations tell us about the distribution of $\overline{W}$?

A
1. The distribution seems to be centered at μ.
2. The larger the sample size, the more concentrated around μ.
3. What about the shape of the distribution?



Approximate distribution of $\overline{W}$ for a sample of size $n = 1000$

Sample mean

Example (Weight of orangutans)

To better see the shape of the distribution of $\overline{W}$, we can standardize it

$$Z := \frac{\overline{W} - \mu}{\sigma/\sqrt{n}}$$

so that it has mean 0 and standard deviation 1, irrespective of the sample size.

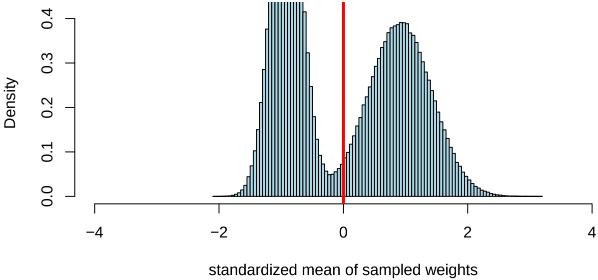
simulation	W_1	W_2	W_3	$\cdots$	W_{50}	$\overline{W}$	Z
1	60.53	95.88	61.01	$\cdots$	90.54	74.6260	-1.3194
2	64.06	65.79	61.92	$\cdots$	94.01	76.7592	-0.6973
3	60.73	100.03	90.35	$\cdots$	53.36	80.2934	0.3335
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
1,000,000	102.51	77.68	117.99	$\cdots$	60.03	80.5370	0.4045

The distribution of Z will have the same shape as the distribution of $\overline{W}$, except for its center and spread.

Sample mean

Example (Weight of orangutans)

The following histograms are based on 1,000,000 repeated simulations.

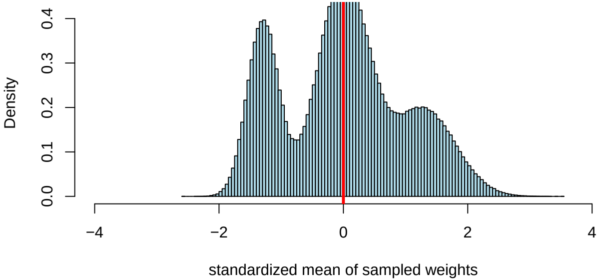


Approximate distribution of $Z = \frac{\overline{W} - \mu}{\sigma/\sqrt{n}}$ for a sample of size $n = 1$

Sample mean

Example (Weight of orangutans)

The following histograms are based on 1,000,000 repeated simulations.

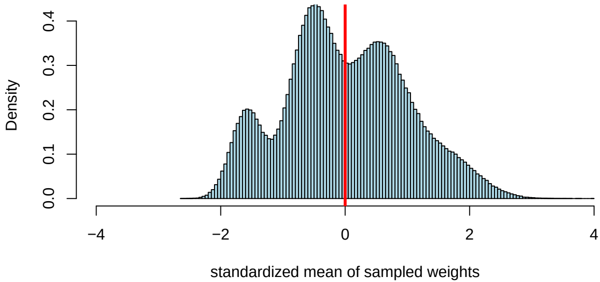


Approximate distribution of $Z = \frac{\overline{W} - \mu}{\sigma / \sqrt{n}}$ for a sample of size $n = 2$

Sample mean

Example (Weight of orangutans)

The following histograms are based on 1,000,000 repeated simulations.

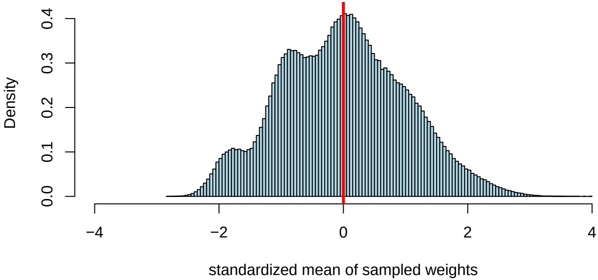


Approximate distribution of $Z = \frac{\overline{W} - \mu}{\sigma / \sqrt{n}}$ for a sample of size $n = 3$

Sample mean

Example (Weight of orangutans)

The following histograms are based on 1,000,000 repeated simulations.

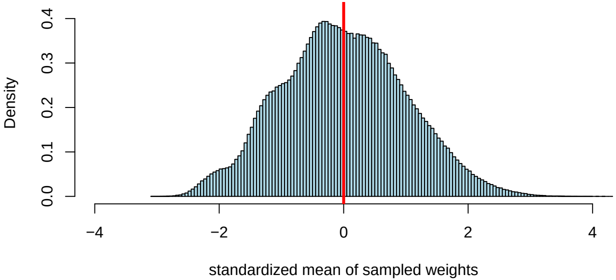


Approximate distribution of $Z = \frac{\overline{W} - \mu}{\sigma / \sqrt{n}}$ for a sample of size $n = 4$

Sample mean

Example (Weight of orangutans)

The following histograms are based on 1,000,000 repeated simulations.

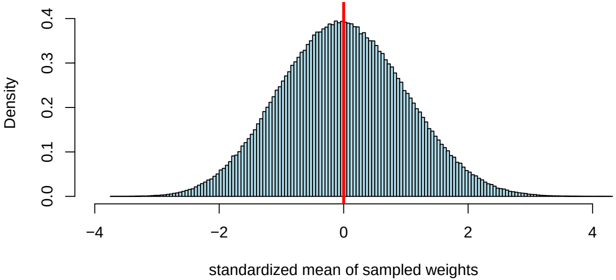


Approximate distribution of $Z = \frac{\overline{W} - \mu}{\sigma / \sqrt{n}}$ for a sample of size $n = 5$

Sample mean

Example (Weight of orangutans)

The following histograms are based on 1,000,000 repeated simulations.

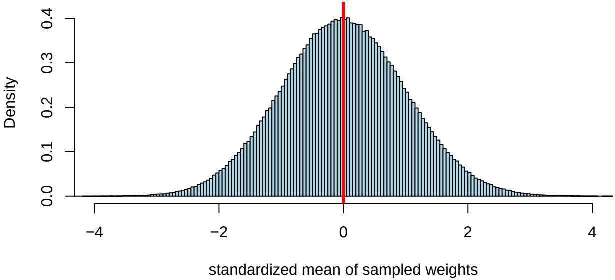


Approximate distribution of $Z = \frac{\overline{W}-\mu}{\sigma/\sqrt{n}}$ for a sample of size $n = 10$

Sample mean

Example (Weight of orangutans)

The following histograms are based on 1,000,000 repeated simulations.

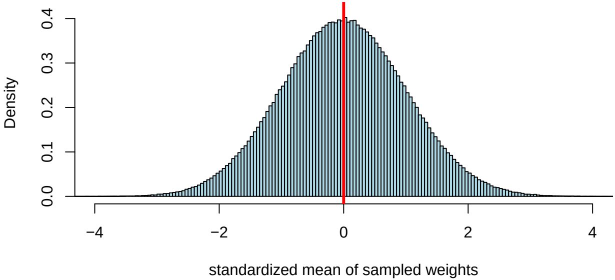


Approximate distribution of $Z = \frac{\overline{W} - \mu}{\sigma / \sqrt{n}}$ for a sample of size $n = 50$

Sample mean

Example (Weight of orangutans)

The following histograms are based on 1,000,000 repeated simulations.



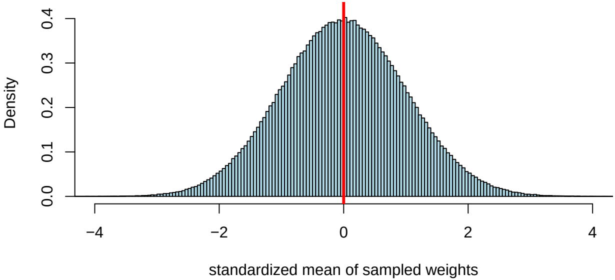
Approximate distribution of $Z = \frac{\overline{W} - \mu}{\sigma / \sqrt{n}}$ for a sample of size $n = 100$

Sample mean

Example (Weight of orangutans)

The following histograms are based on 1,000,000 repeated simulations.

Surprise!!
The larger the sample, the closer Z appears to be to standard normal!



Approximate distribution of $Z = \frac{\overline{W}-\mu}{\sigma/\sqrt{n}}$ for a sample of size $n = 100$

Sample mean

Distribution of sample mean (central limit theorem)

Let $X_1, X_2, \ldots, X_n$ be an independent random sample of size n from a population with mean μ and standard deviation σ. Let

$$\overline{X} := \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Sample mean

Distribution of sample mean (central limit theorem)

Let $X_1, X_2, \ldots, X_n$ be an independent random sample of size n from a population with mean μ and standard deviation σ. Let

$$\overline{X} := \frac{X_1 + X_2 + \cdots + X_n}{n}$$

- If the sample size is large, then $\overline{X}$ is approximately normal, irrespective of the population distribution.
  (The larger n, the closer is the distribution of $\overline{X}$ to normal.)

Sample mean

Distribution of sample mean (central limit theorem)

Let $X_1, X_2, \ldots, X_n$ be an independent random sample of size n from a population with mean μ and standard deviation σ. Let

$$\overline{X} := \frac{X_1 + X_2 + \cdots + X_n}{n}$$

- If the sample size is large, then $\overline{X}$ is approximately normal, irrespective of the population distribution.
  (The larger n, the closer is the distribution of $\overline{X}$ to normal.)
- If the population distribution is normal, then $\overline{X}$ is exactly normal, even if the sample size is not large.

Sample mean

Distribution of sample mean (central limit theorem)

Let $X_1, X_2, \ldots, X_n$ be an independent random sample of size n from a population with mean μ and standard deviation σ. Let

$$\overline{X} := \frac{X_1 + X_2 + \cdots + X_n}{n}$$

- If the sample size is large, then $\overline{X}$ is approximately normal, irrespective of the population distribution.
  (The larger n, the closer is the distribution of $\overline{X}$ to normal.)
- If the population distribution is normal, then $\overline{X}$ is exactly normal, even if the sample size is not large.

Remark 1

Q What about the parameters of this normal distribution?

Sample mean

Distribution of sample mean (central limit theorem)

Let $X_1, X_2, \ldots, X_n$ be an independent random sample of size n from a population with mean μ and standard deviation σ. Let

$$\overline{X} := \frac{X_1 + X_2 + \cdots + X_n}{n}$$

- If the sample size is large, then $\overline{X}$ is approximately normal, irrespective of the population distribution.
  (The larger n, the closer is the distribution of $\overline{X}$ to normal.)
- If the population distribution is normal, then $\overline{X}$ is exactly normal, even if the sample size is not large.

Remark 1

Q What about the parameters of this normal distribution?

A We already know $\mathbb{E}[\overline{X}] = \mu$ and $\mathbb{SD}[\overline{X}] = \sigma/\sqrt{n}$.

Sample mean

Distribution of sample mean (central limit theorem)

Let $X_1, X_2, \ldots, X_n$ be an independent random sample of size n from a population with mean μ and standard deviation σ. Let

$$\overline{X} := \frac{X_1 + X_2 + \cdots + X_n}{n}$$

- If the sample size is large, then $\overline{X}$ is approximately normal, irrespective of the population distribution.
  (The larger n, the closer is the distribution of $\overline{X}$ to normal.)
- If the population distribution is normal, then $\overline{X}$ is exactly normal, even if the sample size is not large.

Remark 2

Q How large is large enough?

Sample mean

Distribution of sample mean (central limit theorem)

Let $X_1, X_2, \ldots, X_n$ be an independent random sample of size n from a population with mean μ and standard deviation σ. Let

$$\overline{X} := \frac{X_1 + X_2 + \cdots + X_n}{n}$$

- If the sample size is large, then $\overline{X}$ is approximately normal, irrespective of the population distribution.
  (The larger n, the closer is the distribution of $\overline{X}$ to normal.)
- If the population distribution is normal, then $\overline{X}$ is exactly normal, even if the sample size is not large.

Remark 2

Q How large is large enough?

A Rule of thumb: $n \geq 30$ (But beware!!)

Sample mean

Distribution of sample mean (central limit theorem)

Let $X_1, X_2, \ldots, X_n$ be an independent random sample of size n from a population with mean μ and standard deviation σ. Let

$$\overline{X} := \frac{X_1 + X_2 + \cdots + X_n}{n}$$

- If the sample size is large, then $\overline{X}$ is approximately normal, irrespective of the population distribution.
  (The larger n, the closer is the distribution of $\overline{X}$ to normal.)
- If the population distribution is normal, then $\overline{X}$ is exactly normal, even if the sample size is not large.

Remark 3

The second statement is a consequence of the following general principle:

The sum of any two independent normal RVs is again normal.

Sample mean

Example (Weight of orangutans)

$\boxed{\mu = 79.15\,\text{Kg}}$ $\boxed{\sigma = 24.28\,\text{Kg}}$

Researchers in Borneo take a <u>random sample of 50 orangutans</u> and measure their weights. Let $\overline{W}$ denote the sample mean.

Sample mean

Example (Weight of orangutans)

$\boxed{\mu = 79.15\,\text{Kg}}$ $\boxed{\sigma = 24.28\,\text{Kg}}$

Researchers in Borneo take a random sample of 50 orangutans and measure their weights. Let $\overline{W}$ denote the sample mean.

Q What is the chance that the sample mean $\overline{W}$ is between 75 Kg and 85 Kg?

Sample mean

Example (Weight of orangutans)

$\boxed{\mu = 79.15\,\text{Kg}}$ $\boxed{\sigma = 24.28\,\text{Kg}}$

Researchers in Borneo take a <u>random sample of 50 orangutans</u> and measure their weights. Let $\overline{W}$ denote the sample mean.

Q What is the chance that the sample mean $\overline{W}$ is between 75 Kg and 85 Kg?

$\boxed{\text{A}}$ As we saw from the simulation, the sample mean is <u>approximately normal</u>.

Sample mean

Example (Weight of orangutans)

$$\boxed{\mu = 79.15\,\text{Kg}} \qquad \boxed{\sigma = 24.28\,\text{Kg}}$$

Researchers in Borneo take a random sample of 50 orangutans and measure their weights. Let $\overline{W}$ denote the sample mean.

Q What is the chance that the sample mean $\overline{W}$ is between 75 Kg and 85 Kg?

A As we saw from the simulation, the sample mean is approximately normal. We already know that

$$\mathbb{E}[\overline{W}] = \mu = \boxed{79.15\,\text{Kg}} \qquad \mathbb{SD}[\overline{W}] = \frac{\sigma}{\sqrt{50}} \approx \boxed{3.43\,\text{Kg}}$$

Sample mean

Example (Weight of orangutans)

$$\boxed{\mu = 79.15\,\text{Kg}} \qquad \boxed{\sigma = 24.28\,\text{Kg}}$$

Researchers in Borneo take a random sample of 50 orangutans and measure their weights. Let $\overline{W}$ denote the sample mean.

Q: What is the chance that the sample mean $\overline{W}$ is between 75 Kg and 85 Kg?

A: As we saw from the simulation, the sample mean is approximately normal. We already know that

$$\mathbb{E}[\overline{W}] = \mu = \boxed{79.15\,\text{Kg}} \qquad \mathbb{SD}[\overline{W}] = \frac{\sigma}{\sqrt{50}} \approx \boxed{3.43\,\text{Kg}}$$

Therefore, using the applet for normal distribution,

$$\mathbb{P}(75 < \overline{W} < 85) = \mathbb{P}(\overline{W} < 85) - \mathbb{P}(\overline{W} \leq 75)$$

$$\approx$$

Sample mean

Example (Weight of orangutans)

$$\boxed{\mu = 79.15\,\text{Kg}} \qquad \boxed{\sigma = 24.28\,\text{Kg}}$$

Researchers in Borneo take a random sample of 50 orangutans and measure their weights. Let $\overline{W}$ denote the sample mean.

Q What is the chance that the sample mean $\overline{W}$ is between 75 Kg and 85 Kg?

A As we saw from the simulation, the sample mean is approximately normal. We already know that

$$\mathbb{E}[\overline{W}] = \mu = \boxed{79.15\,\text{Kg}} \qquad \mathbb{SD}[\overline{W}] = \frac{\sigma}{\sqrt{50}} \approx \boxed{3.43\,\text{Kg}}$$

Therefore, using the applet for normal distribution,

$$\begin{aligned}\mathbb{P}(75 < \overline{W} < 85) &= \mathbb{P}(\overline{W} < 85) - \mathbb{P}(\overline{W} \leq 75) \\ &\approx 0.95595 - 0.11316 \approx \boxed{0.84279} \approx 84\%\end{aligned}$$

Sample mean

Example (Gestation period of rodents)

The gestation period of a species of rodents is normally distributed with the average of 7 weeks and standard deviation of 2.1 days.

Sample mean

Example (Gestation period of rodents)

The gestation period of a species of rodents is normally distributed with the average of 7 weeks and standard deviation of 2.1 days.

Researchers take a <u>random sample of 5 pregnant rodents</u> from this species and measure the duration of their pregnancies.

Sample mean

Example (Gestation period of rodents)

The gestation period of a species of rodents is normally distributed with the average of 7 weeks and standard deviation of 2.1 days.

Researchers take a random sample of 5 pregnant rodents from this species and measure the duration of their pregnancies.

Q What is the chance that the average gestation period among these 5 subjects is larger than 51 days?

Sample mean

Example (Gestation period of rodents)

The gestation period of a species of rodents is normally distributed with the average of 7 weeks and standard deviation of 2.1 days.

Researchers take a random sample of 5 pregnant rodents from this species and measure the duration of their pregnancies.

Q What is the chance that the average gestation period among these 5 subjects is larger than 51 days?

A Let $\overline{X}$ denote the mean gestation period of the sampled rodents.

Sample mean

Example (Gestation period of rodents)

The gestation period of a species of rodents is normally distributed with the average of 7 weeks and standard deviation of 2.1 days.

Researchers take a random sample of 5 pregnant rodents from this species and measure the duration of their pregnancies.

Q What is the chance that the average gestation period among these 5 subjects is larger than 51 days?

A Let $\overline{X}$ denote the mean gestation period of the sampled rodents.

Since the population is normally distributed, so is $\overline{X}$.

Sample mean

Example (Gestation period of rodents)

The gestation period of a species of rodents is normally distributed with the average of 7 weeks and standard deviation of 2.1 days.

Researchers take a random sample of 5 pregnant rodents from this species and measure the duration of their pregnancies.

Q What is the chance that the average gestation period among these 5 subjects is larger than 51 days?

$\boxed{\text{A}}$ Let $\overline{X}$ denote the mean gestation period of the sampled rodents.

Since the population is normally distributed, so is $\overline{X}$.

The population has mean $\mu = 49$ days and standard deviation $\sigma = 2.1$ days. Hence,

$$\mathbb{E}[\overline{X}] = \qquad\qquad \mathbb{SD}[\overline{X}] =$$

Sample mean

Example (Gestation period of rodents)

The gestation period of a species of rodents is normally distributed with the average of 7 weeks and standard deviation of 2.1 days.

Researchers take a random sample of 5 pregnant rodents from this species and measure the duration of their pregnancies.

Q What is the chance that the average gestation period among these 5 subjects is larger than 51 days?

A Let $\overline{X}$ denote the mean gestation period of the sampled rodents.

Since the population is normally distributed, so is $\overline{X}$.

The population has mean $\mu = 49$ days and standard deviation $\sigma = 2.1$ days. Hence,

$$\mathbb{E}[\overline{X}] = \mu = 49\,\text{days} \qquad \mathbb{SD}[\overline{X}] = \frac{\sigma}{\sqrt{5}} = \frac{2.1}{\sqrt{5}} \approx 0.9391$$

Sample mean

Example (Gestation period of rodents)

The gestation period of a species of rodents is normally distributed with the average of 7 weeks and standard deviation of 2.1 days.

Researchers take a random sample of 5 pregnant rodents from this species and measure the duration of their pregnancies.

Q What is the chance that the average gestation period among these 5 subjects is larger than 51 days?

$\boxed{\text{A}}$ Let $\overline{X}$ denote the mean gestation period of the sampled rodents.

Since the population is normally distributed, so is $\overline{X}$.

The population has mean $\mu = 49$ days and standard deviation $\sigma = 2.1$ days. Hence,

$$\mathbb{E}[\overline{X}] = \mu = 49 \text{ days} \qquad \mathbb{SD}[\overline{X}] = \frac{\sigma}{\sqrt{5}} = \frac{2.1}{\sqrt{5}} \approx 0.9391$$

Therefore, using the applet for normal distribution,

$$\mathbb{P}(\overline{X} > 51) \approx$$

Sample mean

Example (Gestation period of rodents)

The gestation period of a species of rodents is normally distributed with the average of 7 weeks and standard deviation of 2.1 days.

Researchers take a random sample of 5 pregnant rodents from this species and measure the duration of their pregnancies.

Q What is the chance that the average gestation period among these 5 subjects is larger than 51 days?

A Let $\overline{X}$ denote the mean gestation period of the sampled rodents.

Since the population is normally distributed, so is $\overline{X}$.

The population has mean $\mu = 49$ days and standard deviation $\sigma = 2.1$ days. Hence,

$$\mathbb{E}[\overline{X}] = \mu = 49\,\text{days} \qquad \mathbb{SD}[\overline{X}] = \frac{\sigma}{\sqrt{5}} = \frac{2.1}{\sqrt{5}} \approx 0.9391$$

Therefore, using the applet for normal distribution,

$$\mathbb{P}(\overline{X} > 51) \approx \boxed{0.0166} \approx 1.7\%$$

Sample proportion

Example (Birds with blue feather)

About 24% of the birds from a species found in Lebanon have blue feather.

Sample proportion

Example (Birds with blue feather)

About 24% of the birds from a species found in Lebanon have blue feather.

We take a random sample of 100 birds from this species.

Sample proportion

Example (Birds with blue feather)

About 24% of the birds from a species found in Lebanon have blue feather.

We take a random sample of 100 birds from this species.

Q What is the chance that at least 30% of the sampled birds have blue feather?

Sample proportion

Example (Birds with blue feather)

About 24% of the birds from a species found in Lebanon have blue feather.

We take a random sample of 100 birds from this species.

Q What is the chance that at least 30% of the sampled birds have blue feather?

A1 Let X denote the number of sampled birds with blue feather. We are looking for $\mathbb{P}(X/100 \geq 0.30) = \mathbb{P}(X \geq 30)$.

Sample proportion

Example (Birds with blue feather)

About 24% of the birds from a species found in Lebanon have blue feather.

We take a random sample of 100 birds from this species.

Q What is the chance that at least 30% of the sampled birds have blue feather?

A1 Let X denote the number of sampled birds with blue feather. We are looking for $\mathbb{P}(X/100 \geq 0.30) = \mathbb{P}(X \geq 30)$.

Q What is the distribution of X?

Sample proportion

Example (Birds with blue feather)

About 24% of the birds from a species found in Lebanon have blue feather.

We take a random sample of 100 birds from this species.

Q What is the chance that at least 30% of the sampled birds have blue feather?

A1 Let X denote the number of sampled birds with blue feather.
We are looking for $\mathbb{P}(X/100 \geq 0.30) = \mathbb{P}(X \geq 30)$.

- Q What is the distribution of X?
- A Binomial with parameters $p = 0.24$ and $n = 100$.

Sample proportion

Example (Birds with blue feather)

About 24% of the birds from a species found in Lebanon have blue feather.

We take a random sample of 100 birds from this species.

Q What is the chance that at least 30% of the sampled birds have blue feather?

A1 Let X denote the number of sampled birds with blue feather. We are looking for $\mathbb{P}(X/100 \geq 0.30) = \mathbb{P}(X \geq 30)$.

Q What is the distribution of X?

A Binomial with parameters $p = 0.24$ and $n = 100$.

Therefore, using the applet for binomial distribution,

$$\mathbb{P}(X/100 \geq 0.30) = \mathbb{P}(X \geq 30) \approx$$

Sample proportion

Example (Birds with blue feather)

About 24% of the birds from a species found in Lebanon have blue feather.

We take a random sample of 100 birds from this species.

Q What is the chance that at least 30% of the sampled birds have blue feather?

A1 Let X denote the number of sampled birds with blue feather. We are looking for $\mathbb{P}(X/100 \geq 0.30) = \mathbb{P}(X \geq 30)$.

Q What is the distribution of X?

A Binomial with parameters $p = 0.24$ and $n = 100$.

Therefore, using the applet for binomial distribution,

$$\mathbb{P}(X/100 \geq 0.30) = \mathbb{P}(X \geq 30) \approx \boxed{0.10087} \approx 10.1\%$$

Sample proportion

Example (Birds with blue feather)

About 24% of the birds from a species found in Lebanon have blue feather.

We take a random sample of 100 birds from this species.

Q What is the chance that at least 30% of the sampled birds have blue feather?

$\boxed{\text{A2}}$ Let R denote the proportion of sampled birds with blue feather. We are looking for $\mathbb{P}(R \geq 0.30)$.

Sample proportion

Example (Birds with blue feather)

About 24% of the birds from a species found in Lebanon have blue feather.

We take a random sample of 100 birds from this species.

Q What is the chance that at least 30% of the sampled birds have blue feather?

A2 Let R denote the proportion of sampled birds with blue feather. We are looking for $\mathbb{P}(R \geq 0.30)$.

Q What can we say about the distribution of R?

Sample proportion

Example (Birds with blue feather)

About 24% of the birds from a species found in Lebanon have blue feather.

We take a random sample of 100 birds from this species.

Q What is the chance that at least 30% of the sampled birds have blue feather?

A2 Let R denote the proportion of sampled birds with blue feather. We are looking for $\mathbb{P}(R \geq 0.30)$.

Q What can we say about the distribution of R?

A Approximately normal with

$$\mathbb{E}[R] = \boxed{0.24} \qquad \mathbb{SD}[R] = \sqrt{\frac{0.24 \times (1-0.24)}{100}} \approx \boxed{0.04271}$$

Sample proportion

Example (Birds with blue feather)

About 24% of the birds from a species found in Lebanon have blue feather.

We take a random sample of 100 birds from this species.

Q What is the chance that at least 30% of the sampled birds have blue feather?

A2 Let R denote the proportion of sampled birds with blue feather. We are looking for $\mathbb{P}(R \geq 0.30)$.

Q What can we say about the distribution of R?

A Approximately normal with

$$\mathbb{E}[R] = \boxed{0.24} \qquad \mathbb{SD}[R] = \sqrt{\frac{0.24 \times (1 - 0.24)}{100}} \approx \boxed{0.04271}$$

Therefore, using the applet for normal distribution,

$$\mathbb{P}(R \geq 0.30) \approx$$

Sample proportion

Example (Birds with blue feather)

About 24% of the birds from a species found in Lebanon have blue feather.

We take a random sample of 100 birds from this species.

Q What is the chance that at least 30% of the sampled birds have blue feather?

$\boxed{\text{A2}}$ Let R denote the proportion of sampled birds with blue feather. We are looking for $\mathbb{P}(R \geq 0.30)$.

Q What can we say about the distribution of R?

$\boxed{\text{A}}$ Approximately normal with

$$\mathbb{E}[R] = \boxed{0.24} \qquad \mathbb{SD}[R] = \sqrt{\frac{0.24\times(1-0.24)}{100}} \approx \boxed{0.04271}$$

Therefore, using the applet for normal distribution,

$$\mathbb{P}(R \geq 0.30) \approx \boxed{0.08004} \approx 8.0\%$$

Sample proportion

Example (Birds with blue feather)

About 24% of the birds from a species found in Lebanon have blue feather.

We take a random sample of 100 birds from this species.

Q What is the chance that at least 30% of the sampled birds have blue feather?

A2 Let R denote the proportion of sampled birds with blue feather. We are looking for $\mathbb{P}(R \geq 0.30)$.

Q What can we say about the distribution of R?

A Approximately normal with

$$\mathbb{E}[R] = \boxed{0.24} \qquad \mathbb{SD}[R] = \sqrt{\frac{0.24 \times (1 - 0.24)}{100}} \approx \boxed{0.04271}$$

Therefore, using the applet for normal distribution,

$$\mathbb{P}(R \geq 0.30) \approx \boxed{0.08004} \approx 8.0\%$$

Not a perfect approximation, but not bad either!

Sample proportion

Distribution of sample proportion I

Suppose the proportion of members in a population that have a certain characteristic $\circledast$ is p. $[0 < p < 1]$

Sample proportion

Distribution of sample proportion I

Suppose the proportion of members in a population that have a certain characteristic $\circledast$ is p. $[0 < p < 1]$

Suppose we take an random sample of size n with replacement from this population, and let

$$X := \langle \text{number of sampled members with characteristic } \circledast \rangle$$

Sample proportion

Distribution of sample proportion I

Suppose the proportion of members in a population that have a certain characteristic $\circledast$ is p. $[0 < p < 1]$

Suppose we take an random sample of size n with replacement from this population, and let

$$X := \langle \text{number of sampled members with characteristic } \circledast \rangle$$

Then, X is a binomial RV with parameters p and n.

Sample proportion

Distribution of sample proportion I

Suppose the proportion of members in a population that have a certain characteristic $\circledast$ is p. $[0 < p < 1]$

Suppose we take an random sample of size n with replacement from this population, and let

$$X := \langle \text{number of sampled members with characteristic } \circledast \rangle$$

Then, X is a binomial RV with parameters p and n.
In particular,

$$\mathbb{E}[X] = \qquad\qquad \mathbb{SD}[X] =$$

as we already know.

Sample proportion

Distribution of sample proportion I

Suppose the proportion of members in a population that have a certain characteristic $\circledast$ is p. $[0 < p < 1]$

Suppose we take an random sample of size n with replacement from this population, and let

$$X := \langle \text{number of sampled members with characteristic } \circledast \rangle$$

Then, X is a binomial RV with parameters p and n.
In particular,

$$\mathbb{E}[X] = np \qquad \mathbb{SD}[X] =$$

as we already know.

Sample proportion

Distribution of sample proportion I

Suppose the proportion of members in a population that have a certain characteristic $\circledast$ is p. $[0 < p < 1]$

Suppose we take an random sample of size n with replacement from this population, and let

$$X := \langle \text{number of sampled members with characteristic } \circledast \rangle$$

Then, X is a binomial RV with parameters p and n.
In particular,

$$\mathbb{E}[X] = np \qquad \mathbb{SD}[X] = \sqrt{np(1-p)}$$

as we already know.

Sample proportion

Distribution of sample proportion II

Suppose the proportion of members in a population that have a certain characteristic $\circledast$ is p. $[0 < p < 1]$

Suppose we take an random sample of size n with replacement from this population, and let

$$\widehat{p} := \langle \text{proportion of sampled members with characteristic } \circledast \rangle$$

Sample proportion

Distribution of sample proportion II

Suppose the proportion of members in a population that have a certain characteristic $\circledast$ is p. $[0 < p < 1]$

Suppose we take an random sample of size n with replacement from this population, and let

$$\widehat{p} := \langle \text{proportion of sampled members with characteristic } \circledast \rangle$$

If n is large, and pn and $(1-p)n$ are not too small, then $\widehat{p}$ is approximately normal with

$$\mathbb{E}[\widehat{p}] = \qquad \qquad \mathbb{SD}[\widehat{p}] =$$

Sample proportion

Distribution of sample proportion II

Suppose the proportion of members in a population that have a certain characteristic $\circledast$ is p. $[0 < p < 1]$

Suppose we take an random sample of size n with replacement from this population, and let

$$\widehat{p} := \langle \text{proportion of sampled members with characteristic } \circledast \rangle$$

If n is large, and pn and $(1-p)n$ are not too small, then $\widehat{p}$ is approximately normal with

$$\mathbb{E}[\widehat{p}] = p \qquad \mathbb{SD}[\widehat{p}] =$$

Sample proportion

Distribution of sample proportion II

Suppose the proportion of members in a population that have a certain characteristic $\circledast$ is p. $[0 < p < 1]$

Suppose we take an random sample of size n with replacement from this population, and let

$$\widehat{p} := \langle \text{proportion of sampled members with characteristic } \circledast \rangle$$

If n is large, and pn and $(1-p)n$ are not too small, then $\widehat{p}$ is approximately normal with

$$\mathbb{E}[\widehat{p}] = p \qquad \mathbb{SD}[\widehat{p}] = \sqrt{\frac{p(1-p)}{n}}$$

Sample proportion

Distribution of sample proportion II

Suppose the proportion of members in a population that have a certain characteristic $\circledast$ is p. $[0 < p < 1]$

Suppose we take an random sample of size n with replacement from this population, and let

$$\widehat{p} := \langle \text{proportion of sampled members with characteristic } \circledast \rangle$$

If n is large, and pn and $(1-p)n$ are not too small, then $\widehat{p}$ is approximately normal with

$$\mathbb{E}[\widehat{p}] = p \qquad \mathbb{SD}[\widehat{p}] = \sqrt{\frac{p(1-p)}{n}}$$

Remark 1

Q How large should pn and $(1-p)n$ be?

Sample proportion

Distribution of sample proportion II

Suppose the proportion of members in a population that have a certain characteristic $\circledast$ is p. $[0 < p < 1]$

Suppose we take an random sample of size n with replacement from this population, and let

$$\widehat{p} := \langle \text{proportion of sampled members with characteristic } \circledast \rangle$$

If n is large, and pn and $(1-p)n$ are not too small, then $\widehat{p}$ is approximately normal with

$$\mathbb{E}[\widehat{p}] = p \qquad \mathbb{SD}[\widehat{p}] = \sqrt{\frac{p(1-p)}{n}}$$

Remark 1

Q How large should pn and $(1-p)n$ be?

A Rule of thumb: $pn > 5$ and $(1-p)n > 5$ (But beware!!)

Sample proportion

Distribution of sample proportion II

Suppose the proportion of members in a population that have a certain characteristic $\circledast$ is p. $[0 < p < 1]$

Suppose we take an random sample of size n with replacement from this population, and let

$$\widehat{p} := \langle \text{proportion of sampled members with characteristic } \circledast \rangle$$

If n is large, and pn and $(1-p)n$ are not too small, then $\widehat{p}$ is approximately normal with

$$\mathbb{E}[\widehat{p}] = p \qquad \mathbb{SD}[\widehat{p}] = \sqrt{\frac{p(1-p)}{n}}$$

Remark 2

Q What is the connection with sample mean?

Sample proportion

Distribution of sample proportion II

Suppose the proportion of members in a population that have a certain characteristic $\circledast$ is p. $[0 < p < 1]$

Suppose we take an random sample of size n with replacement from this population, and let

$$\widehat{p} := \langle \text{proportion of sampled members with characteristic } \circledast \rangle$$

If n is large, and pn and $(1-p)n$ are not too small, then $\widehat{p}$ is approximately normal with

$$\mathbb{E}[\widehat{p}] = p \qquad \mathbb{SD}[\widehat{p}] = \sqrt{\frac{p(1-p)}{n}}$$

Remark 2

Q What is the connection with sample mean?

$\boxed{\text{A}}$ ... (can you guess?)