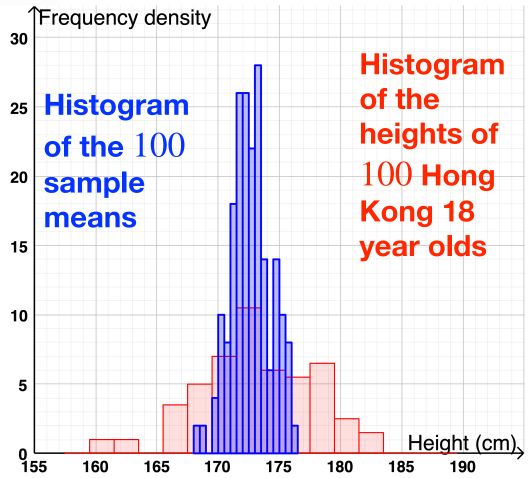
Distribution of the sample mean

Starter

1. $10\,\%$ of the chocolates produced in a factory are mis-shapes. A sample of 1000 chocolates is taken.
   (a) Find the probability that the number of mis-shapes is between 90 and 115 inclusive:
      (i) using a binomial distribution
      (ii) using a Normal approximation.
   (b) Calculate the percentage error of the Normal approximation.

Notes

In a previous lesson, we drew a histogram based on the heights of 1000 Hong Kong 18 year olds and then a normal curve was superimposed over the histogram.



From the original data, a small sample of, say, 10 values could be taken and the mean calculated for the sample. This process could be repeated, say 100 times, with each sample having a mean. A ***histogram of these sample means*** could be produced.

From the histogram it is possible to see that the distribution of the mean from the samples:

- is normally distributed about the mean of the original data
- Has a much lower standard deviation

In fact, the standard deviation of the sample means depends on the size of the sample.

Distribution of the sample mean

If a number of random samples of size n are taken from a normal distribution with mean μ and variance σ^2 such that $X \sim \mathrm{N}(\mu, \sigma^2)$, then the distribution of the sample means of the samples will be normally distributed such that $\bar{X}_n \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$.

The standard deviation of this distribution, $\frac{\sigma}{\sqrt{n}}$, is called the ***standard error of the mean***.

N.B. Calculations for $\bar{X}_n \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$ use the same methods as $X \sim \mathrm{N}(\mu, \sigma^2)$ but use $\frac{\sigma}{\sqrt{n}}$ for the standard deviation value.

E.g. 1 (a) If $X \sim \mathrm{N}(75, 8)$, find $P(\bar{X}_9 < 74)$ (b) If $X \sim \mathrm{N}(12, 3^2)$, find $P(\bar{X}_7 > 12.5)$.

Working: (a) $X \sim \mathrm{N}(75, 8) \;\Rightarrow\; \bar{X}_9 \sim \mathrm{N}\left(75, \frac{8}{9}\right)$

$$\text{Standard error} = \frac{\sigma}{\sqrt{n}} = \frac{2\sqrt{2}}{3}$$

$$P(\bar{X}_9 < 74) = 0.144 \text{ (3 s.f.)}$$

E.g. 2 A random sample of size 15 is taken from a normal distribution with mean 60 and standard deviation 4. Find the probability that the mean of the sample is less than 58.

E.g. 3 The heights of a particular species of plant follow a normal distribution with mean 21 and standard deviation $\sqrt{90}$. A random sample of 10 plants is taken and the mean height calculated. Find the probability that this mean lies between 18 cm and 27 cm.

When a calculation requires the use of the standard normal distribution, the formula is $z = \dfrac{x-\mu}{\frac{\sigma}{\sqrt{n}}}$.

E.g. 4 A large number of random samples of size n are taken from the distribution of X where $X \sim \text{N}(74, 36)$ and the sample means are calculated. If $P(\bar{X} > 72) = 0.854$, estimate the value of n.

Working: $X \sim \text{N}(74, 36) \quad \Rightarrow \quad \bar{X}_n \sim \text{N}\left(74, \dfrac{36}{n}\right)$

Standard error $= \dfrac{\sigma}{\sqrt{n}} = \dfrac{6}{\sqrt{n}}$

$P(\bar{X} > 72) = 0.854 \quad \Rightarrow \quad P(\bar{X} < 72) = 1 - 0.854 = 0.146$

Let z be the $z-$value corresponding to 72.

$P(Z < z) = 0.146 \quad \Rightarrow \quad z = -1.0537$

Substitute into $z = \dfrac{x-\mu}{\frac{\sigma}{\sqrt{n}}}$: $\quad -1.0537 = \dfrac{72-74}{\frac{6}{\sqrt{n}}}$

$$-1.0537 \times 6 = -2 \times \sqrt{n}$$
$$n = 9.99 \approx 10$$

The value of n is 10.

E.g. 5 A normal distribution has a mean of 30 and a variance of 5. Find n such that the probability that the average of n observations exceeds 30.5 is less than $1\,\%$.

Video: Distribution of sample mean A
Video: Distribution of sample mean B

Solutions to Starter and E.g.s

Exercise
p400 18A Qu 1i, 2i, 3-8,(9-10 red)

Summary
Distribution of the sample mean:

$$X \sim \text{N}(\mu, \sigma^2) \quad \Rightarrow \quad \bar{X}_n \sim \text{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

The standard deviation of $\bar{X}_n$ is $\dfrac{\sigma}{\sqrt{n}}$ and is called the ***standard error of the mean***.

For calculations requiring the standard normal distribution, use the formula $z = \dfrac{x-\mu}{\frac{\sigma}{\sqrt{n}}}$.