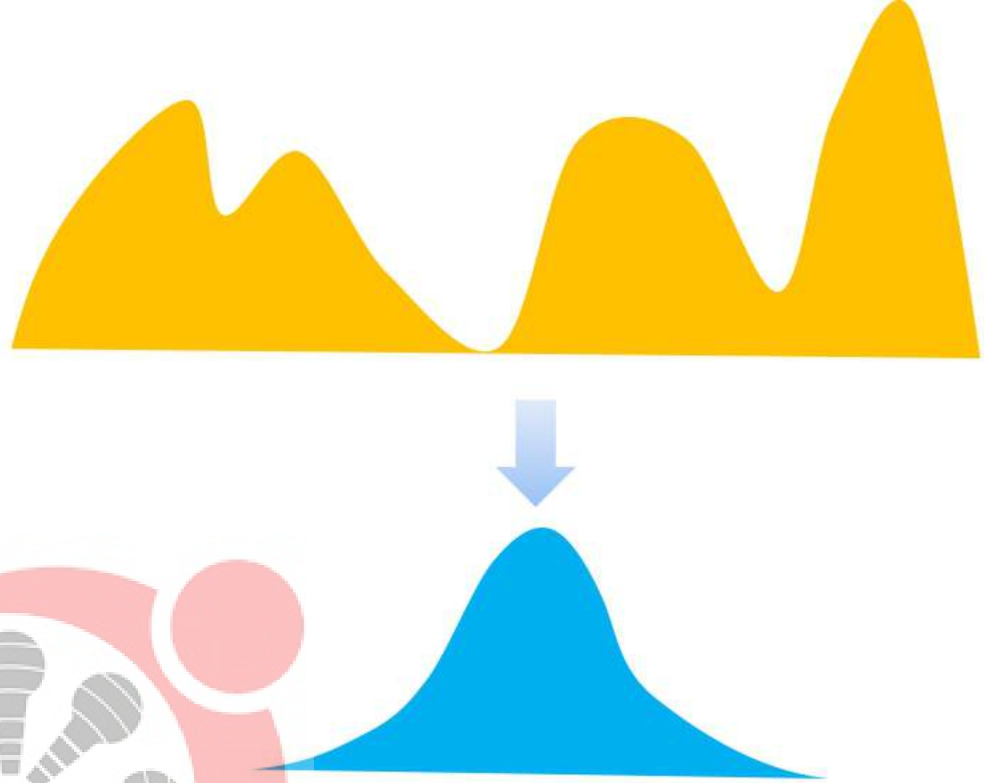


Sampling Distribution of Means

Sampling Distribution

- The core goal of inferential statistics is to be able to make intelligent conclusions about the population parameters by looking at sample statistics.

Eg: Estimate the mean height of the students in a class, from a small sample.

Sampling Distribution of means

- The sampling distribution of means is what you get if you consider all possible samples of size *n taken from the same* population and form a distribution of their means.

- Each randomly selected sample is an independent observation.

Central Limit Theorem

- http://onlinestatbook.com/stat_sim/sampling_dist/

- As sample size goes large and number of buckets are high, the means will follow a normal distribution with same mean (μ) and $\frac{1}{n}$ of variance (σ^2).

Using the Central Limit Theorem

Let us say the mean number of Gems per packet is 10, and the variance is 1. If you take a sample of 30 packets, what is the probability that the sample mean is 9.5 Gems per packet or fewer?

Using the Central Limit Theorem

We known that $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \mu = 10, \sigma^2 = 1$ and n = 30. We

need the value of P($\bar{X}$ < 9.5) when $\bar{X}$ ~ N(10,0.0333).

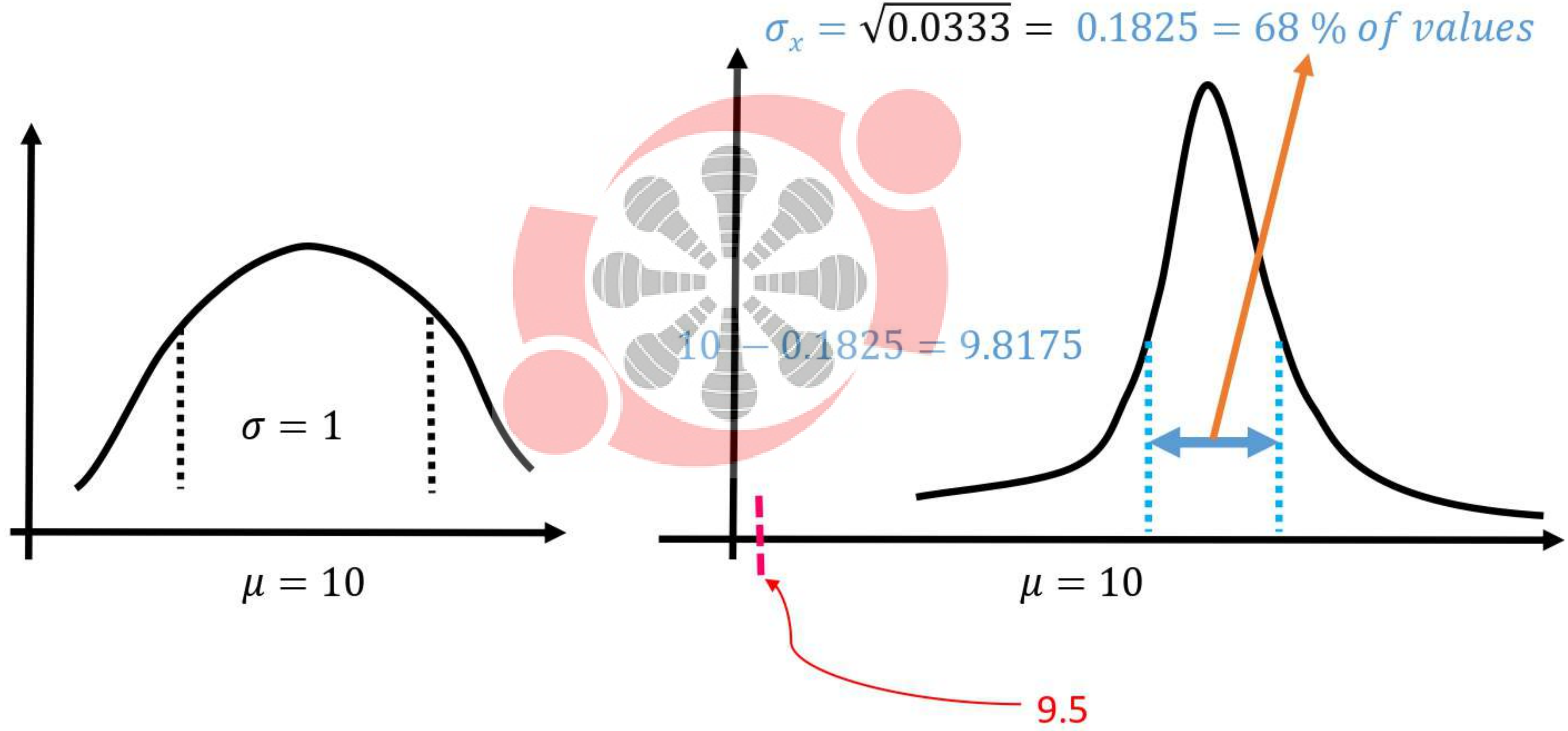
$$Z = \frac{9.5 - 10}{\sqrt{0.0333}} = -2.7472$$

P(Z < z) = P(Z < -2.7472)

This doesn't exit in probability tables. What does it mean ?

Using the Central Limit Theorem

How do we visualize it ?

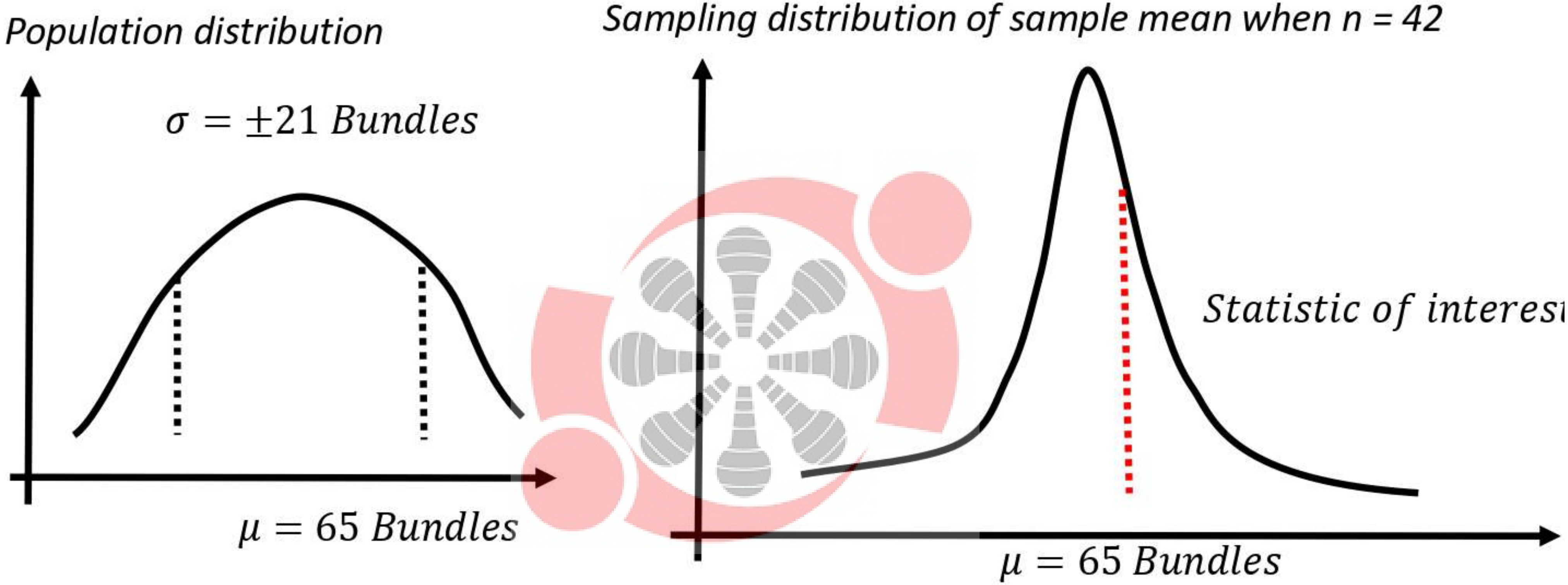


Using the Central Limit Theorem

Accountant of a news paper company states that the average employee uses 65 bundles of paper in a year.

A random sample of 42 employees is monitored for one year to check the paper usage. If the population standard deviation of annual usage is 21 Bundles, what is the probability that the sample mean will be > 70 bundles?

Sampling Distribution

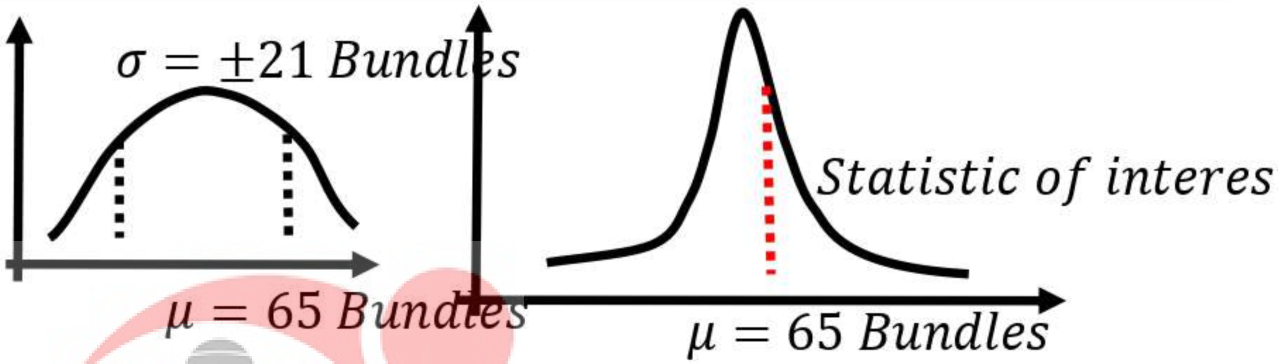


Step 1 : List all known parameters and values

Step2 : Calculate others, or estimate if cannot be calculated

Step3: Find probabilities using tables, Excel or R

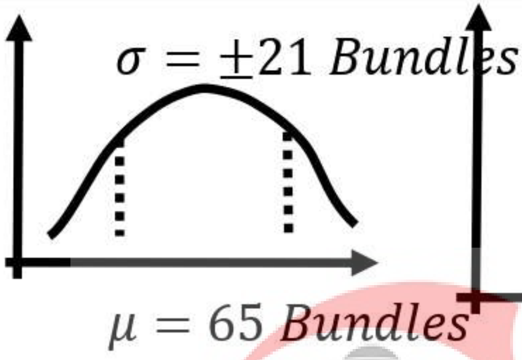
Sampling Distribution

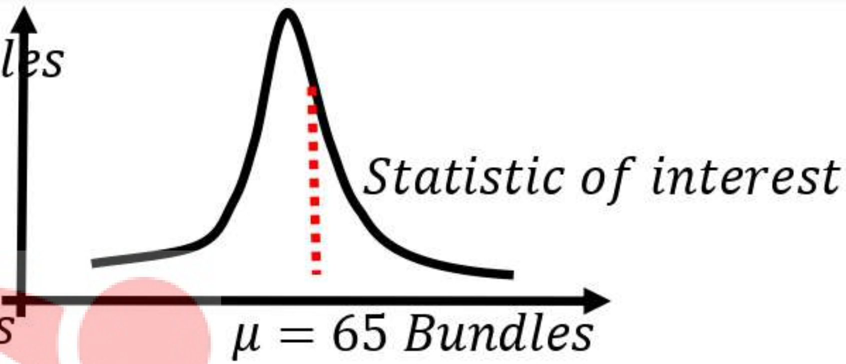


Step-1 : List all known parameters and values

- Population mean, μ= 65 Bundles
- Population standard deviation, σ = 21 Bundles
- Sample size, n = 42 Employees
- Sample mean, $\bar{x}$= > 70 Bundles
- Mean of sample means, $\mu_{\bar{x}}$= μ = 65 Bundles

Sampling Distribution





Step-2: Calculate others or estimate, if cannot be calculated.

- Standard deviation of sample means, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{21}{\sqrt{42}} = 3.24$
- $\therefore Z = \frac{70 - 65}{3.24} = 1.54$

Step-3:

→ Please calculate these for

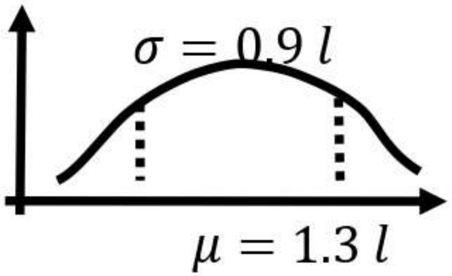
- > 85 Bundles
- > 63 Bundles < 64 Bundles
- < 45 Bundles

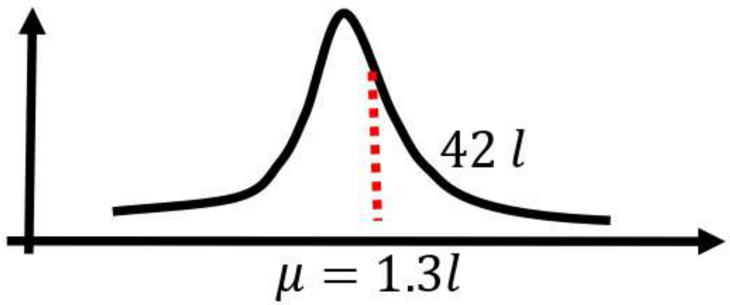
Sampling Distribution

According to the World.by map the daily consumption of petrol in India for an average person is 1.3liters with a standard deviation of 0.9liters. You are planning a racing competition for 25 people and bring 42liters of petrol. What is the probability that you run out of petrol?

μ = 1.3, σ = 0.4

$P(run\ out) \Rightarrow P(use$ > 42$l) \Rightarrow P(average$ petrol $use\ per$ person > 1.68l)





Sampling Distribution

$$\mu_{\overline{x}} = \mu = 1.3l, \sigma_{\overline{x}}^2 = \frac{\sigma^2}{n} = \frac{0.81}{25}$$

$$\Rightarrow \sigma_{\overline{x}}^2 = 0.0324$$

$$\Rightarrow z = \frac{1.68 - 1.3}{0.18} = 2.111$$

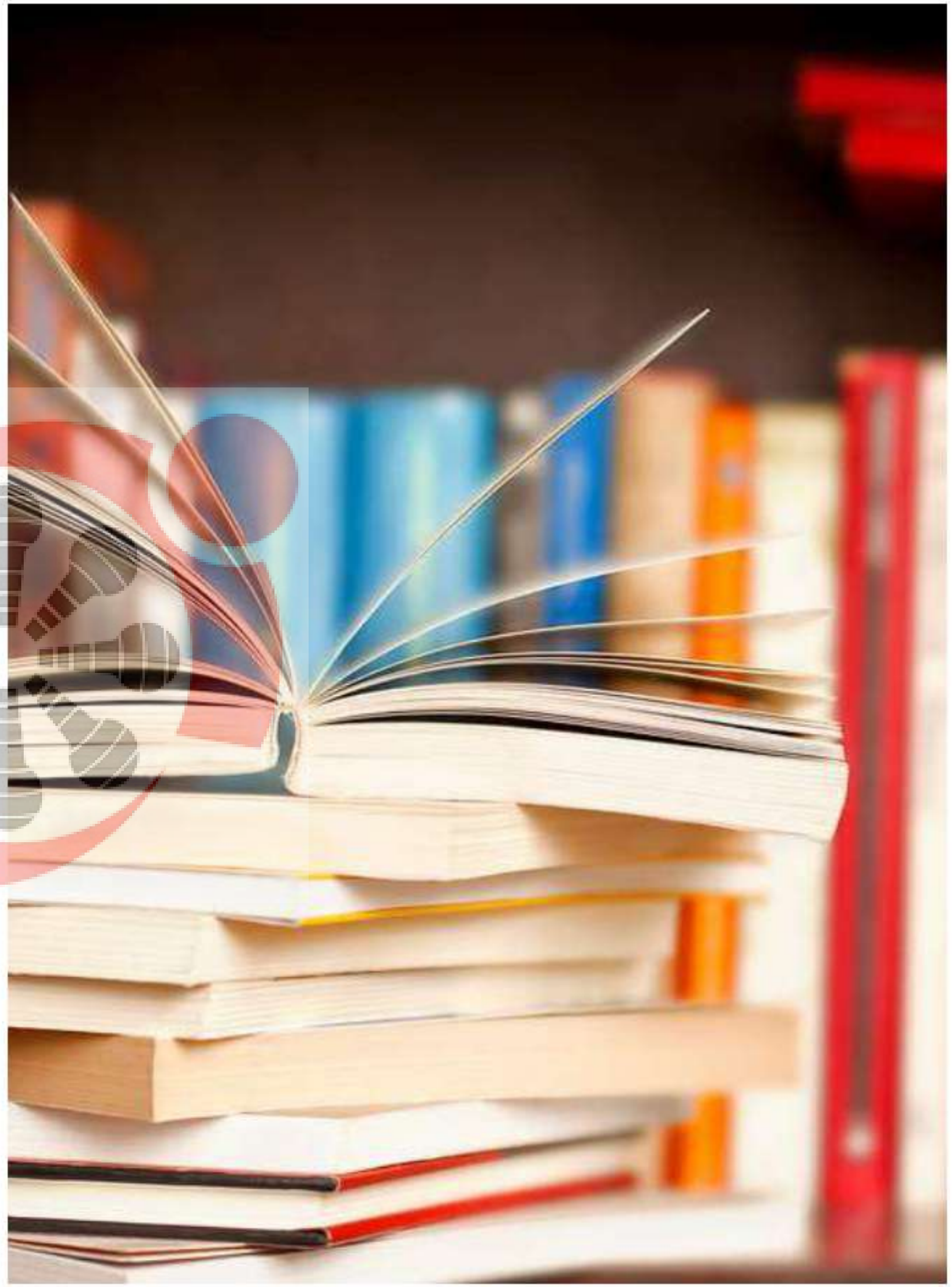
$$\Rightarrow P(\bar{X} < 2.111) = 0.9826$$

$\Rightarrow$The probability of running out of water is

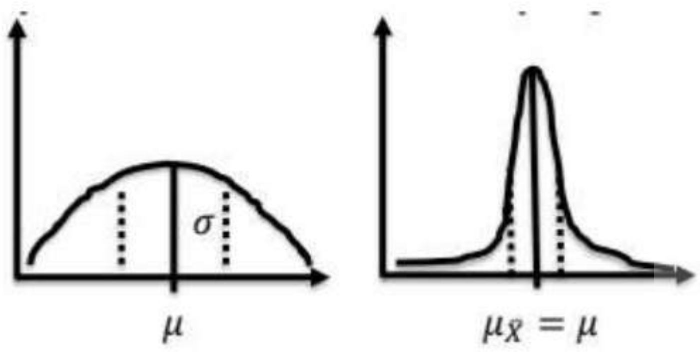
1 – 0.9826 = 0.0174 or 1.74 %

Using the Central Limit Theorem

You sample 64 books from your note book manufacturers of 400,000 books. The mean weight of the sample is 1.12kg with a 0.4kg sample standard deviation. What is the probability that the mean weight of all 400,000 apples is between 1 and 1.3kg?



Population distribution



Sampling distribution of sample mean when n = 64

$\sigma_{\bar{x}}$ = σ / sqrt(n)

=> $\sigma_{\bar{x}}$ = σ / 8

What are we trying to find out?

We need to know if population mean, μ, is within ±0.12kg of the sample mean, $\bar{X}$.

This is the same as saying that we need to know if sample mean, $\bar{X}$, is within ±0.12kg of the population mean, μ. Since μ = $\mu_{\bar{x}}$, we can now use the sampling distribution of the means.