CHAPTER 4
SAMPLING DISTRIBUTION AND ESTIMATION

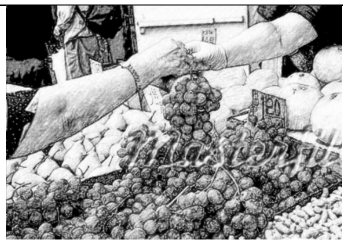
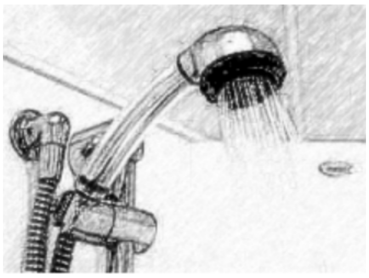
Learning Objectives:

1. *Define sampling distribution of the sample mean, $\bar{X}$.*
2. *Link the mean and standard deviation of sampling distribution of the sample mean, $\bar{X}$ to the mean and standard deviation of random variable X.*
3. *Understand the Central Limit Theorem and its role in statistical inference.*
4. *Use sampling distributions to evaluate claims ("rare events") on values of the population mean.*
5. *Distinguish between population parameter and sample statistic.*
6. *Define point estimate.*
7. *Distinguish between point and interval estimates.*
8. *Compute margin of error.*
9. *Construct confidence intervals from large samples with population standard deviation known or unknown.*
10. *Interpret confidence intervals.*
11. *Construct confidence intervals using Minitab Express by selecting Z or t-distributions.*

Content

1. Sampling Techniques

What is sampling? Observe two scenarios below:

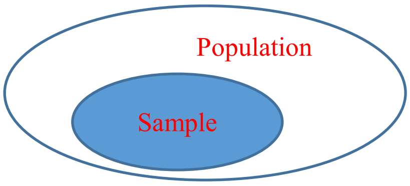
 "Hmm…I wonder if these grapes are sweet"	 "Let me try one grape… Hmm, it is sweet!"	 "Ok, I will buy this bunch of grapes then…"
 "Is the water too hot?"	 "Let me feel the water quickly… hmm, just nice…not too hot"	 "Ok, the water temperature is just nice to shower this puppy"

What do you observe in both scenarios? Hopefully we observed the same thing:

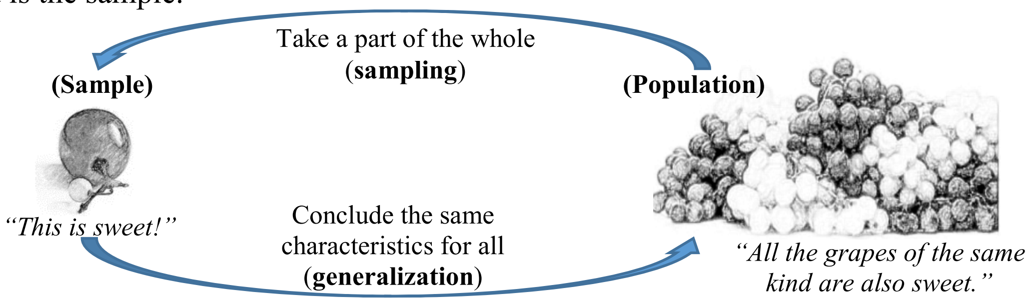
- In the first, the lady tasted one grape and concluded that all the grapes taste sweet.
- In the second, the puppy owner tested the temperature of the shower water for a brief moment and concluded that the temperature of the running water will not be too hot.

What we observe in both scenarios are examples of sampling.

Sampling is taking a part (sample) of the whole (population). So, a sample is a set of observations that is part of all the possible observations of a phenomenon.



For example, all the grapes of the same kind by the seller is the population and the one grape tasted is the sample.

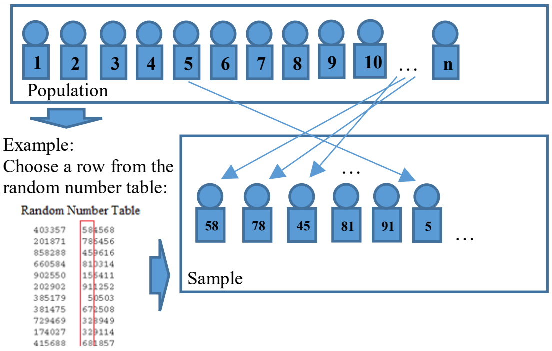
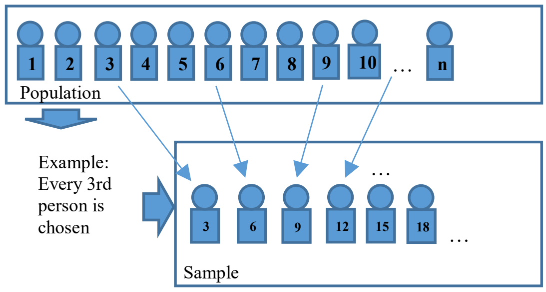
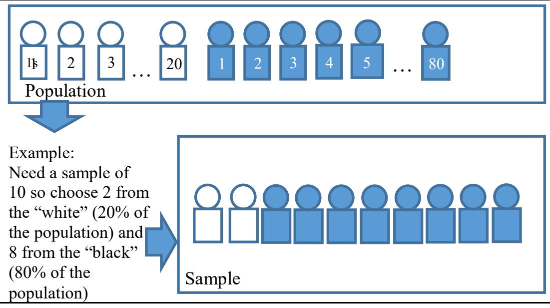
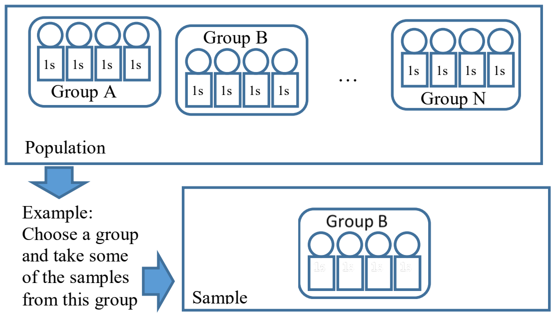


The field of statistics is ultimately concerned with generalization and prediction. In many cases, sampling is more feasible to study than the entire population.
Why? Well, the lady cannot possibly taste all the grapes before she buys nor can the dog owner let the water run forever.

Generally, here are three possible reasons why sampling is more feasible, with an example cited for each reason.

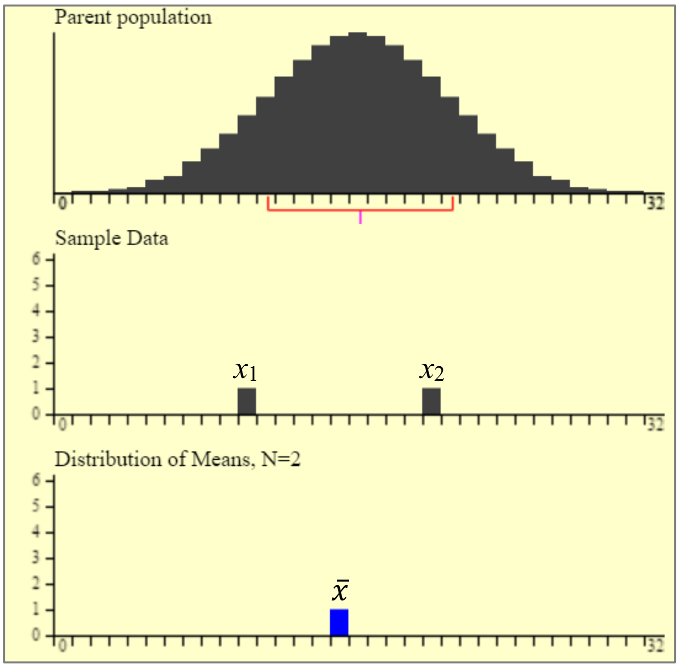
Possible Reason	Example
It is impossible to take all observations from an infinite population.	Amount of salt in the sea water in South China Sea.
Sampled objects for observation may not be returned to the population.	Impact testing on cars.
Do not have the resource (time and money) to collect all observations.	Everyone who is eligible to vote in the 2017 U.S. election and who would they vote for, candidate A or B?

There are different techniques to perform sampling; four of them are described below.

<table>
<tr><th>Sampling techniques</th><th>Illustration</th></tr>
<tr><td>Simple random sampling
A sample is selected in such a way that every item or person in the population has the same chance of being selected.</td><td>
</td></tr>
<tr><td>Systematic sampling
A random starting point is selected and then every k^{th} member of the population is selected.</td><td>
</td></tr>
<tr><td>Stratified sampling
A population is divided into subgroups and a sample is selected from all subgroups according to the proportions of each subgroup (proportion of the population).</td><td>
</td></tr>
<tr><td>Cluster sampling
A population is divided into clusters or groups. Then clusters are randomly selected and a sample is collected by randomly selecting from each cluster.</td><td>
</td></tr>
</table>

2. Distribution of Sample Means

We shall use an *applet** to demonstrate what the probability distribution of samples means look like if we are to repeatedly draw samples from the population.
[*link to applet: http://onlinestatbook.com/stat_sim/sampling_dist/index.html]

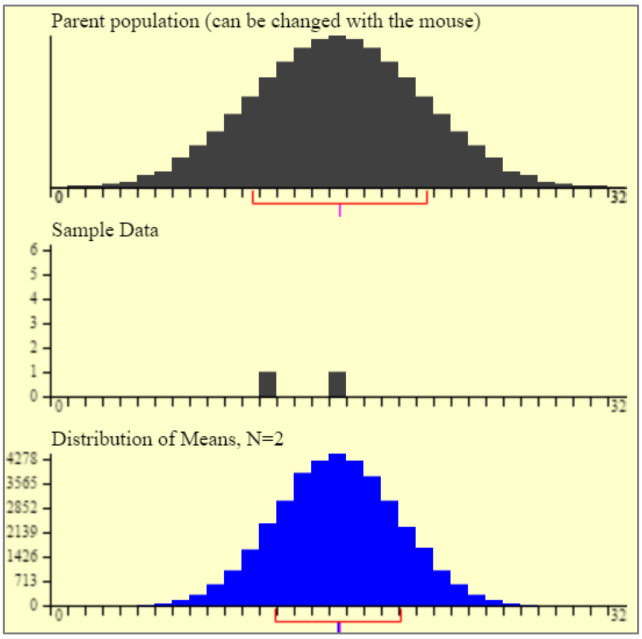


⓿ Population data X is normally distributed. i.e. $X \sim N(\mu, \sigma^2)$.

❶ Two data x_1 and x_2 is randomly drawn from the population. i.e. sample size, $n = 2$

❷ The mean of the two data x_1 and x_2 is computed, that is the sample mean $\bar{x}$.

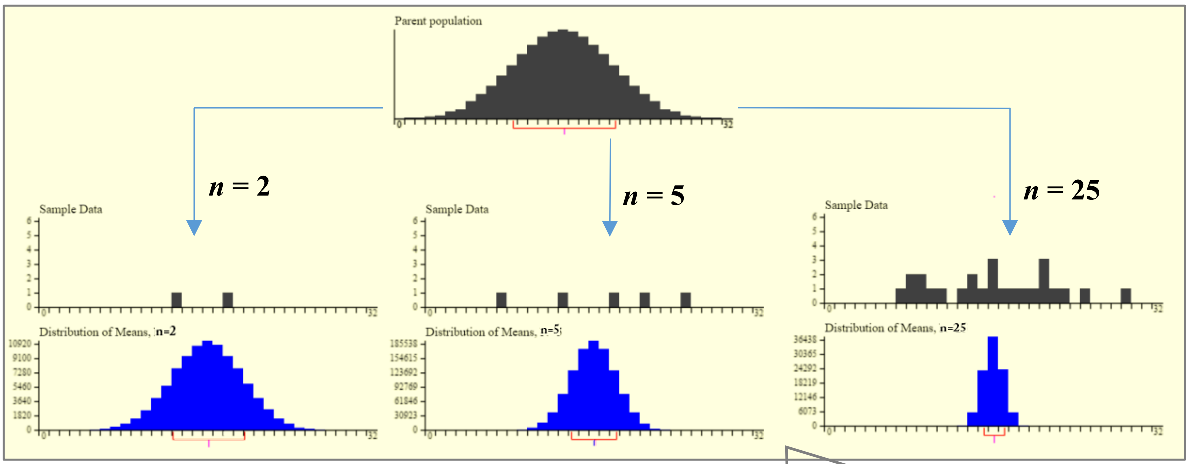
What if we repeatedly do steps ❶ and ❷ above? That is, what if we repeatedly draw 2 sample data and compute the sample mean? Surely, we will have many sample means. If we plot all these sample means on a graph, how will the distribution look like?



So, if we repeatedly draw samples of size 2 ($n = 2$) from a normal population, compute the sample means $\bar{x}$ and plot them on a graph, we will observe that the distribution of the sample means is bell-shaped!

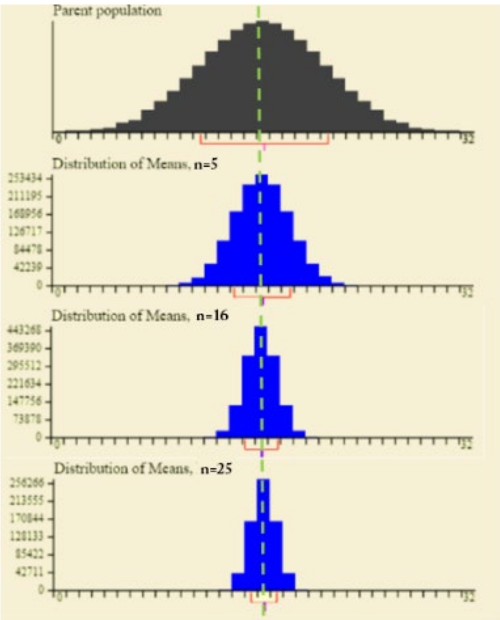
This distribution of sample means, or simply, distribution of $\bar{X}$, is also known as the **sampling distribution of the sample mean**.

In fact, the <u>sampling distribution of the sample mean is always normally distributed for any sample size n if the population is normally distributed</u>, as illustrated below:



Think-out-loud…
What do you notice about the spread of the sampling distributions as n increases?

The distribution of the population, together with the sampling distributions of the sample mean for n = 5, n = 16 and n = 25 are shown here for comparisons:



Population X is normal.
Centre: mean is μ
Spread: SD is σ

Distribution of $\overline{X}$ of sample size $n = 5$ is normal.
Centre: mean is $\mu_{\overline{X}}$
Spread: SD is $\sigma_{\overline{X}}$

Distribution of $\overline{X}$ of sample size $n = 16$ is normal.
Centre: mean is $\mu_{\overline{X}}$
Spread: SD is $\sigma_{\overline{X}}$

Distribution of $\overline{X}$ of sample size $n = 25$ is normal.
Centre: mean is $\mu_{\overline{X}}$
Spread: SD is $\sigma_{\overline{X}}$

Observe that:

#1. The centre of the distribution of $\overline{X}$ is the same as the centre of the population.
That is, $\mu_{\overline{X}} = \mu$.

#2. The spread of the distribution of $\overline{X}$ is smaller than the spread of the population.
That is, $\sigma_{\overline{X}} < \sigma$, for $n \geq 2$.

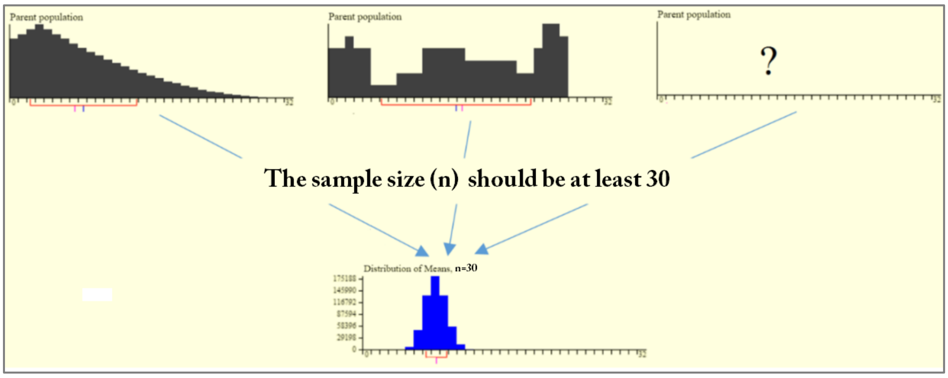
#3. Moreover, the spread of the distribution of $\overline{X}$ decreases as the sample size increases.

In fact, $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$.

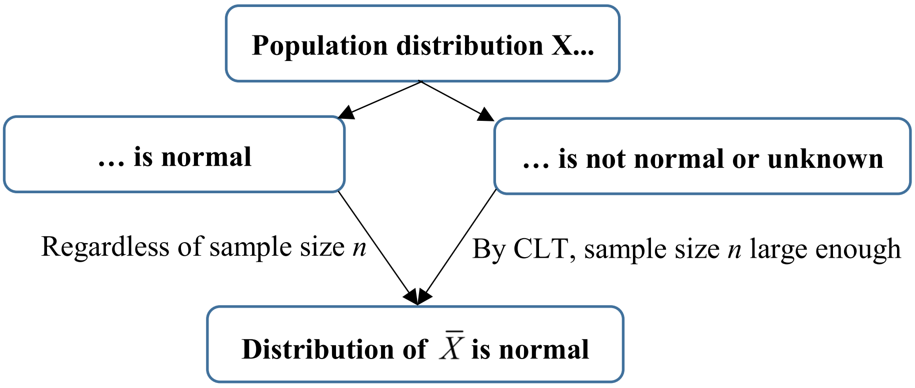
What if the population X is skewed or not normally distributed or its distribution is unknown?

The **Central Limit Theorem** (CLT) roughly states that as the sample size n increases, the sampling distribution of the sample mean will approach the normal distribution.

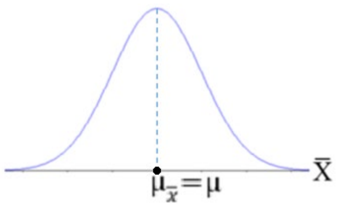
But how large should n be? The rule of thumb is that sample size of at least 30 is sufficient to assume that the sampling distribution of the sample mean is approximately normal.



Summary



Since the distribution of $\bar{X}$ is normal, we can write as: $\boxed{\bar{X} \sim \mathrm{N}\left(\mu_{\bar{X}}, \sigma_{\bar{X}}^2\right)}$



Mean of sample means, $\boxed{\mu_{\bar{X}} = \mu}$

Standard deviation of sample means, $\boxed{\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}}$

$\sigma_{\bar{X}}$ is better known as **standard error** (**SE**) of sample mean.

Then, to convert $\bar{x}$-value to z-score, the formula is: $\boxed{z = \dfrac{\bar{x} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \dfrac{\bar{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}}}$

If population SD, σ, is unknown, we can use sample SD, s, as an estimate for σ.

But, then $\dfrac{\bar{x} - \mu}{\dfrac{s}{\sqrt{n}}}$ follows a distribution known as **Student's *t*-distribution**.

Example 1: A lightbulb manufacturer claims that the lifespan of his lightbulbs follows a normal distribution with mean 750 hours and standard deviation 30 hours. A random sample of 20 lightbulbs is to be selected for testing of lifespan.

(a) Describe the sampling distribution of the sample mean lifespan of 20 lightbulbs.

(b) What is the probability that 20 lightbulbs will have a mean lifespan of 725 hours or less?

(c) Hence, is it rare to get a sample of 20 lightbulbs with a mean lifespan of 725 hours or less?

Example 2: Chocolate Delight produces chocolate bars for baking. The brand claims that their chocolate bars contain an average of 250g of cocoa content, with standard deviation of 20g. Amy, a baker, took a sample of 30 such chocolate bars and measured the cocoa content.

(a) Is the sampling distribution of the sample mean cocoa content of 30 chocolate bars normal? Why?

(b) What is the probability that the sample mean cocoa content differs from the claimed mean by at least 11g?

(c) Explain the meaning of this probability.

3. Estimating the Population Mean

In statistical inference, we use **sample statistic** (e.g. sample mean) to estimate **population parameter** (e.g. population mean or true mean) which is usually unknown.

For instance:

Population parameter	Sample statistic
Patient's overall mean blood pressure (μ)	Nurse measures patient's blood pressure several times and take average ($\bar{x}$)
Average lifetime of a Duracell battery (μ)	Average lifetime of random Duracell batteries that are tested on the rack ($\bar{x}$)

The sample mean blood pressure ($\bar{x}$) is used to estimate the true value of the patient's mean blood pressure (μ).
Likewise, the sample mean lifetime of the tested Duracell batteries ($\bar{x}$) is used to estimate the true value of the lifetime of a Duracell battery on average (μ).

But how well does the sample mean really match the true value of the population mean?
So, instead of using a single number (i.e. sample mean), we can compute a range of values along with a confidence level for that range. This range of values is called a **confidence interval**.

To estimate the unknown population mean, we shall first satisfy the following **assumptions**:

#1. Observations are independent

#2. Data are from a normal distribution, or sample size is sufficiently large ($n \geq 30$)

#3. Population standard deviation σ is known

The last assumption is not reasonable in the real world. For now we will assume that from past data, we know the population SD.

Point estimate is a single number sample statistic that can be used to estimate a single number for the population parameter. We use $\bar{x}$ as the point estimate for μ.

Sample mean $\bar{x}$ is the point estimate for population mean μ.

(When population SD σ is unknown, we can use sample SD, s, as the point estimate for σ.)

One problem with using the sample mean $\bar{x}$ to infer the population mean μ is that $\bar{x}$ can vary depending on the sample we take. So, a single number like $\bar{x}$ is not a very helpful estimate of μ without some indication of how accurate it is. Therefore, let's include a **margin of error**:

point estimate $\pm$ margin of error

Due to our assumptions, the sampling distribution of the sample mean is normal, thus:

$$\mu_{\bar{X}} = \mu \ , \ \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

If we apply the Empirical rule, then about 95% of sample means in the sampling distribution will lie within two standard deviations $\left(\sigma_{\bar{X}}\right)$ away from the mean $\left(\mu_{\bar{X}} = \mu\right)$.

So to estimate the unknown μ from $\bar{x}$ which is known of the sample we took, we say that we are approximately 95% **confident** that μ is within $\bar{x} \pm 2\sigma_{\bar{X}}$.

That is how likely it is that the interval we come up with actually contains the unknown population mean μ. In other words, if we continue to take samples, we will catch the true value of μ about 95% of the time, over many samples.

4. Constructing Confidence Intervals

4.1 Known Population SD σ

In constructing confidence intervals, we can select a level of confidence that gives the probability that the estimation method will give an interval that catches the unknown population mean (μ).

Confidence levels of 90%, 95% and 99% are usually chosen, with 95% being the most common.

Generally, confidence interval has the structure: **point estimate ± margin of error**

Specifically, for confidence interval of μ, the formula is:

$$\boxed{\bar{x} \pm z^{*}\left(\frac{\sigma}{\sqrt{n}}\right)}$$

where **critical value z^*** takes the following values, obtained from z-table, and depending on selected confidence level:

Confidence level	90%	95%	99%
z^* value			

Case Study 1: Duracell Batteries

Reference: Confidence Intervals: Against All Odds—Inside Statistics. (2013). Films Media Group.
Available at: http://fod.infobase.com/PortalPlaylists.aspx?wID=151497&xtid=111543

Battery companies, like Duracell, have always trumpeted their product's long lives in their commercials. Because the companies promise specific improvements in battery lifetimes, they need proof before the ads are aired.

At Kodack's Ultra Technologies, technicians use rigorous testing to back up the marketers' claims. Random samples of batteries are pulled from the warehouse and tested on the rack, which mimics the load of real products in a controlled environment.

Formulating questions	**What is the average lifetime of Duracell's batteries?**
Collecting data	________ random Duracell batteries are tested on the rack. Let random variable X = ________________________________ Sample mean lifetime, $\overline{x}$ = ________ mins Based on Kodack's past testing experiences, σ = ________ mins
Analysing data	μ = mean lifetime of Duracell's batteries Sampling distribution of $\overline{x}$ is normal because 95% confidence interval for μ is:
Interpreting results	

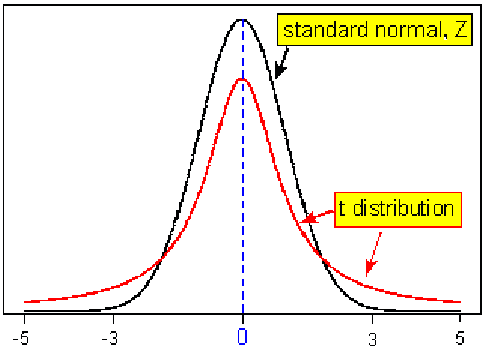
<u>4.2 Unknown Population SD</u>

Usually, when the population mean μ is unknown, the population standard deviation σ is also unknown. We can use sample standard deviation s as an estimate for the population standard deviation σ. Consequently, the underlying distribution will not be the Z-distribution but the **Student's *t*-distribution**. So, the confidence interval of μ becomes:

$$\bar{x} \pm t^*\left(\frac{s}{\sqrt{n}}\right)$$

Key features of the *t*-distribution:

- Centered at and symmetrical about 0
- More spread out than Z-curve
- Thicker tails than Z-curve
- Its shape depends on degree of freedom (d.f. = $n-1$)
- As n increases, *t*-distribution approaches Z-distribution.



Thus, if the <u>sample size n is large enough</u>, the confidence interval of μ is approximately:

$$\boxed{\bar{x} \pm z^*\left(\frac{s}{\sqrt{n}}\right)}$$

The decision to choose Z- or *t*-distribution for critical values is summarized in the table as follows:

Scenario	Critical value
If population X is normal and <u>σ known</u>. OR If population X is not normal, n is large and <u>σ known</u>.	Z*
If population X is normal and <u>σ unknown</u>. OR If population X is not normal, n is large and <u>σ unknown</u>.	t^* If n is large, then $t^* \approx Z^*$

(The use of *t*-distribution to compute confidence interval will be covered in lab session.)

Case Study 2: Subway

Retrieved from: ***https://news.yahoo.com/subway-crisis-footlong-sub-really-11-inches-174939561.html***

Subway 'crisis': Is foot-long sub really 11 inches?

NEW YORK (AP) — What's in an inch? Apparently, enough missing meat, cheese and tomatoes to cause an uproar.

Subway, the world's largest fast food chain with 38,000 locations, is facing widespread criticism after a man who appears to be from Australia posted a photo on the company's Facebook page of one of its foot-long sandwiches next to a tape measure that shows the sub is just 11 inches.

More than 100,000 people have "liked" or commented on the photo, which had the caption "Subway pls respond." Lookalike pictures popped up elsewhere on Facebook. And The New York Post conducted its own investigation that found that four out of seven foot-long sandwiches that it measured were shy of the 12 inches that makes a foot.

The original photo was no longer visible by Thursday afternoon on Subway's Facebook page, which has 19.8 million fans. A spokesman for Subway, which is based in Milford, Conn., said Subway did not remove it.

Subway also said that the length of its sandwiches may vary slightly when its bread, which is baked at each Subway location, is not made to the chain's exact specifications.
"We are reinforcing our policies and procedures in an effort to ensure our offerings are always consistent no matter which Subway restaurant you visit," read an emailed statement.

The Subway photo — and the backlash — illustrates a challenge that companies face with the growth of social media sites like Facebook, YouTube and Twitter. Before, someone in a far flung local in Australia would not be able to cause such a stir. But the power of social media means that negative posts about a company can spread from around the world in seconds.

"People look for the gap between what companies say and what they give, and when they find the gap — be it a mile or an inch — they can now raise a flag and say, 'Hey look at this,' I caught you," said Allen Adamson, managing director of branding firm Landor Associates in New York.

Subway has always offered foot-long sandwiches since it opened in 1965. A customer can order any sandwich as a foot-long. The chain introduced a $5 foot-long promotion in 2008 as the U.S. fell into the recession, and has continued offering the popular option throughout the recovery.

An attempt to contact someone with the same name and country as the person who posted the photo of the foot-long sandwich on Subway's Facebook page was not returned on Thursday.

But comments by other Facebook users about the photo ran the gamut from outrage to indifference to amusement. One commenter urged people to "chill out." Another one said she was switching to Quiznos. And one man posted a photo of his foot in a sock next to a Subway sandwich to show it was shorter than a "foot."

"I've never seen so many people in an uproar over an inch. Wow," read one Facebook post. "Let's all head to McDonald's and weigh a Quarter Pounder," suggested another poster.

Upon reading the news article, Thomas took a sample by buying a foot-long Subway sandwich per day for 30 days. He measured each sandwich before consuming, and recorded the readings as follows:

12.0	11.5	11.7	12.1	12.0	11.8	12.2	12.3	12.0	11.9
12.6	12.0	11.6	12.3	12.1	12.0	11.4	11.8	11.7	12.4
11.5	11.2	11.7	11.4	11.6	11.9	11.1	12.1	11.8	11.2

Thomas then used Minitab Express to calculate the summary statistics, given as follows:

Descriptive Statistics: Length

Statistics

Variable	N	Mean	SE Mean	StDev	Minimum	Median	Maximum
Length	30	11.8300	0.06732	0.36874	11.1000	11.8500	12.6000

Assume that the length of Subway foot-long sandwiches is normally distributed.

Let X =

μ =

(a) State the size, mean and standard deviation of Thomas's sample.

(b) If Subway foot-long sandwiches are truly 12 inches on average, that is, $\mu = 12$, what is the probability that Thomas gets $\bar{x}$ that differed from μ by at least 0.17 inches?

(c) Based on the probability calculated in part (b), does Thomas have reason to doubt that Subway foot-long sandwiches are truly 12 inches on average? Explain.

(d) Construct a 95% confidence interval for mean length of Subway foot-long sandwiches. Hence, interpret the interval. Does Thomas has reason to doubt that $\mu = 12$?

TUTORIAL 4

1. Given that $\mu = 52$, $\sigma = 35$ and $n = 35$, what is $P(X > 73)$ and $P(\overline{X} > 73)$?

2. The paint drying time varies depending on the type of paint. The label on a can of a latex-based paint claims that the drying time is on average 200 minutes with a standard deviation of 50 minutes at room temperature. A random sample of 49 latex-based paint is examined for their drying times.
   (a) Describe the sampling distribution of the sample mean drying time of 49 latex-based paint.
   (b) Find the probability that the sample mean drying time is more than 210 minutes
   (c) Find the probability that the sample mean drying time is at most 160 minutes.
   (d) Find the probability that the sample mean drying time is between 180 and 220 minutes.

3. The following is taken from the website of DigitalTrend.com:

> According to an updated study released earlier today by Nielson, Facebook is still eats up the most amount of time for a typical Web user. Clocking in at 7 hours and 45 minutes, the average person spends that much time each month wandering though endless status updates, leaving comments on new media from friends and playing social games like Zynga's CityVille. However, Facebook is still number two when it comes to the total Internet audience. Google takes the top spot again and the average user spends about 1 hour and 45 minutes on Google products each month. AOL and its variety of Web properties take second place in regards to the amount of time spent on an AOL site, nearly three hours per user per month.
>
> *(http://tinyurl.com/digitaltrendsfb, retrieved on 2 January 2015)*

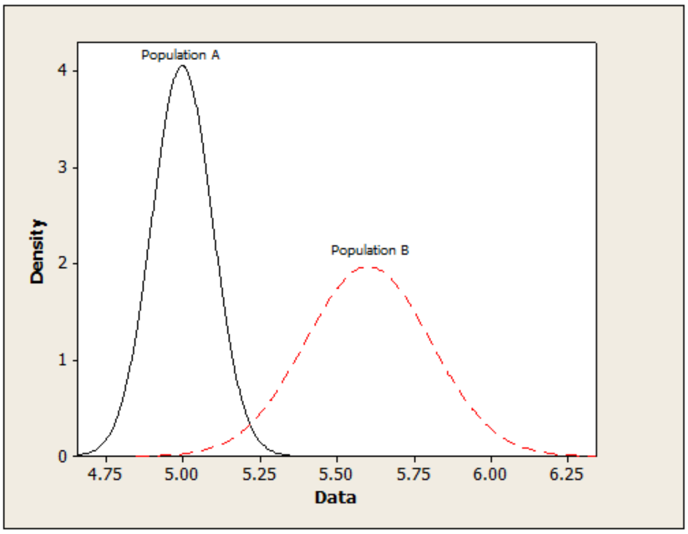
   (a) What is the average time spent (in minutes) by a Facebook user per month?
   (b) What is the probability that a Facebook user spends less than 300 minutes? Assume that time spent is normally distributed with $\sigma = 65$ minutes.
   (c) What is the probability that 40 students selected at random will spend an average of more than 500 minutes on Facebook? Assume $\sigma = 65$ minutes.

4. A machine is regulated so that it produces a mechanical part with an average diameter of 240 mm and a standard deviation of 15 mm. This machine is routinely examined to ensure that it is working at the expected level. Periodically, a sample of 35 mechanical parts are checked and the mean diameter is computed. If the sample mean is within the interval of $\mu_{\overline{X}} \pm 2\,\sigma_{\overline{X}}$, then the machine is thought to be operating properly. Otherwise, adjustments will have to be made to the machine parts.
   (a) What is the range of values for $\mu_{\overline{X}} \pm 2\,\sigma_{\overline{X}}$?
   (b) In one of the regular checks, a quality control officer found that the mean of a sample of 35 mechanical parts to be 234 mm, and concluded that the machine needs adjustment. Using the answer from part (a), is his conclusion reasonable? Explain.

5. The weight of male students attending a large polytechnic is approximately normally distributed with $\mu = 68$ kg and $\sigma = 5$ kg. If twenty male students are crowded into the lift, what is the probability that the lift's maximum capacity of 1400 kg would be exceeded? Explain the meaning of the probability computed.

6. In a chemical process the amount of a certain type of impurity in the output is difficult to control. It is claimed that the population mean amount of the impurity is 0.2 g per gram of output. It is known that the standard deviation is 0.05 g per gram of output.
   An experiment is conducted to gain more insight regarding the claim that $\mu = 0.2$ g/g. The process was run on a lab scale 50 times and the sample average $\bar{x}$ turned out to be 0.23 g per gram of output.
   (a) If $\mu = 0.2$ g/g, what is the probability that the process yields $\bar{x}$ that differed from μ by at least 0.03 g/g?
   (b) Hence, is there reason to doubt the claim that the mean amount of impurity is 0.2 g per gram of output?

7. You are given the following graphs:



$\mu = 4.998$

$\mu = 5.602$

$\sigma = 0.202$

$\sigma = 0.098$

   (a) Two values of mean and two values of standard deviation are given but not labelled. Decide which values of mean and standard deviation best describe the distributions of Population A and Population B.
   (b) A sample of 100 were taken from one of the populations. The sample mean is calculated to be 5.01. Determine whether this sample is more likely to be obtained from Population A or Population B. Justify your answer.

8. The American Management Association (AMA) wishes to have information on the mean income of middle managers in the retail industry. A random sample of 256 managers revealed a sample mean of $45,420. The standard deviation of this population is $2,050.
   (a) Define the population mean and sample mean in this context.
   (b) What is the point estimate for the population mean?
   (c) Construct a 95% confidence interval for the population mean.

9. Measurements of the diameters of a random sample of 200 ball bearings made by a certain machine showed a mean of 0.824 cm. Past data put population standard deviation of diameters at 0.042 cm.
   (a) Find the 90% confidence interval for the mean diameter of ball bearings. Interpret this interval.
   (b) If we construct confidence intervals by the same method 20 times, how many of these intervals will we expect to capture the true value of mean diameter?

10. Severe Acute Respiratory Syndrome (SARS) is a viral respiratory illness. It has the distinction of being the first new communicable disease of the 21st century. Researchers wanted to estimate the incubation period of patients with SARS. Based on interviews with 81 SARS patients, they found that the mean incubation period was 4.6 days with a standard deviation of 15.9 days. Using this information, construct a 95% confidence interval for the mean incubation period of the SARS virus.

11. The following summary gives the background characteristics of 50 participants in a research study evaluating HIV medications:

Patient characteristics	Mean	SD
Age (in years)	37.6	6.8
% male	75.1%	-
Education (in years)	13.6	2.4
CD4 cell count (in cells/ μl)	376	94

    Construct a 99% confidence interval for the mean CD4 cell count.

12. To encourage more shoppers in Orchard, the Urban Planning Authority (UPA) built a new multi-storey carpark that charges cheaply. UPA plans to pay for the structure through collected parking fees. During a two-month period (60 days), daily fees collected averaged $12,900, with a standard deviation of $1,650.
    (a) Construct a 95% confidence interval for the mean daily income this carpark will generate. Interpret this interval.
    (b) The consultant who advised UPA on this project predicted that parking revenues would average $13,500 per day. Based on your answer in part (a), do you have reason to doubt the consultant? Explain.

13. In a factory, a sample of 50 resistors are randomly selected to estimate the true mean resistance of resistors produced by the factory. The 99% confidence interval for true mean resistance is computed to be between 98.6 Ω and 101.3 Ω.
    (a) What is the mean resistance of the sample?
    (b) What is the margin of error?
    (c) What is the standard error of the sample mean resistance?

ANSWERS

1. $P(X > 73)$ cannot be determined as the distribution of X is unknown.
   $P(\overline{X} > 73) = 0.0002$

2. (a) $\bar{X} \sim N\left(200, \left(\frac{50}{7}\right)^2\right)$ (b) 0.0808 (c) ≈ 0 (d) 0.9948

3. (a) 465 mins (b) 0.0055 (c) 0.0003

4. (a) 234.92 mm to 245.07 mm
   (b) Yes, because the sample mean of 234 mm is below the minimum amount of the acceptable range.

5. 0.0367. Since the probability is close to zero (<0.05), it is rare that we could get a sample mean of 20 male students from a population of weights with $\mu = 68$ kg and $\sigma = 5$ kg that exceeds the average maximum lift capacity (in this case, it is 70kg).

6. (a) $P(\overline{X} > 0.23) \approx 0$. Since the probability is close to zero (<0.05), it is rare that the process yields $\bar{x}$ that differed from μ by at least 0.03 g/g if $\mu = 0.2$, yet the process did yield such a sample average.
   (b) Hence, it is reasonable to doubt the claim that population mean amount of the impurity is 0.2 g per gram of output.

7. (a) Population A: $\mu = 4.998$ (because the centre of distribution A is to the left of centre of distribution B) and $\sigma = 0.098$ (because the spread of distribution is smaller compared to that of distribution B).
   (b) It is more likely to be from Population A because $P(Z > \frac{5.01-4.998}{0.098/\sqrt{100}}) = 0.1112$ is higher compared to $P(Z < \frac{5.01-5.602}{0.202/\sqrt{100}}) \approx 0$.

8. (a) Population mean: mean income of middle managers in the retail industry
   Sample mean: mean income of sample of 256 middle managers
   (b) $45,420 (c) $45,169 to $45,671

9. (a) We are 90% confident that the true value of the mean diameter of ball bearing is captured between 0.8191 cm and 0.8289 cm. (b) 18 times

10. 1.14 to 8.06 days

11. 340.4 to 411.6 (using t^*) or 341.8 to 410.2 (using z^*)

12. (a) UPA is 95% confident that the mean daily income generated by the carpark is captured between $12,482.49 and $13,317.51.
    (b) Since $13,500 is not contained in the interval that UPA constructed, I have reason to doubt the consultant.

13. (a) 99.95 Ω (b) 1.35 Ω (c) 0.524 Ω

LAB 4 : Constructing Confidence Intervals

Learning Objectives:

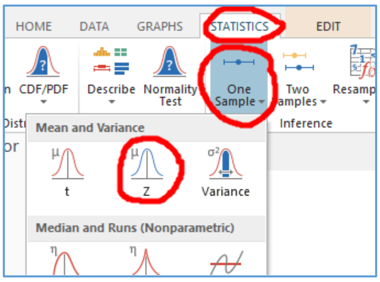
1. *Find confidence intervals using Minitab Express with raw data.*
2. *Find confidence intervals using Minitab Express with summarized statistics.*
3. *Select Z or t-distribution appropriately.*

Task 1

The boiling temperature (in °C) of a certain liquid is being studied. From past experiences, boiling temperature is known to be normally distributed with σ = 1.2 °C. A student measuring the boiling temperature on 6 different samples of the liquid observes the readings (in °C) to be 102.5, 101.7, 103.1, 100.9, 100.5 and 102.2. Construct a 95% confidence interval for the mean boiling temperature.

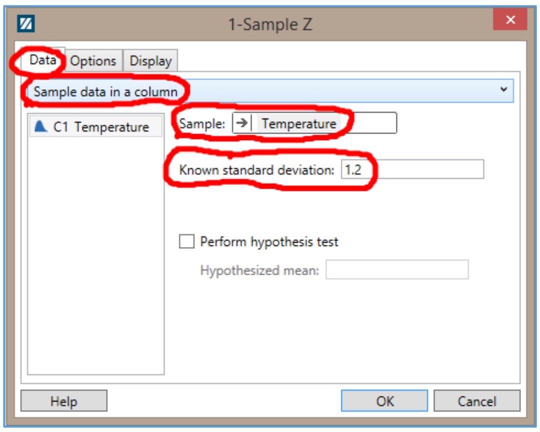
Step 1: Input the data into a column. Label the heading as *Temperature*.

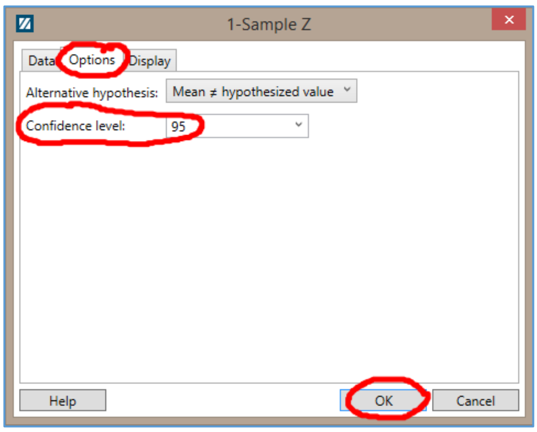
Step 2: In Minitab Express, select **STATISTICS > One Sample > Mean and Variance Z.**



Step 3a: In **Data** tab, select *Sample data in a column*.
For **Sample**, select *Temperature*.
For **Known standard deviation**, key in *1.2*.

Step 3b: In **Options** tab, for **Confidence level**, select desired level.
Click **OK**.





Step 4: The results will be displayed in the output window. Interpret the results:

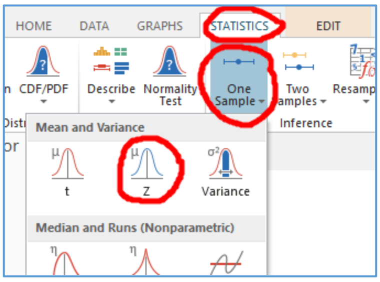
Task 2

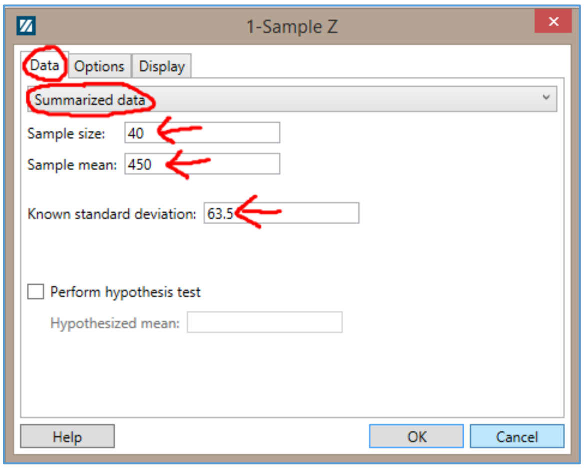
Recall the following information from Case Study 1:

40 random Duracell batteries are tested on the rack.
Let random variable X = lifetime of a Duracell battery in minutes.
Sample mean lifetime, $\bar{x}$ = 450 mins
Based on Kodack's past testing experiences, σ = 63.5 mins.

Construct a 99% confidence interval for mean lifetime of Duracell batteries.

Step 1: In Minitab Express, select
STATISTICS > One Sample > Mean and Variance Z.





Step 2a: In **Data** tab, select *Summarized data*.
For **Sample size**, key in *40*.
For **Sample mean**, key in *450*.
For **Known standard deviation**, key in *63.5*.

Step 2b: In **Options** tab, for **Confidence level**, select desired level.
Click **OK**.

Step 3: The results will be displayed in the output window. Interpret the results:

Task 3

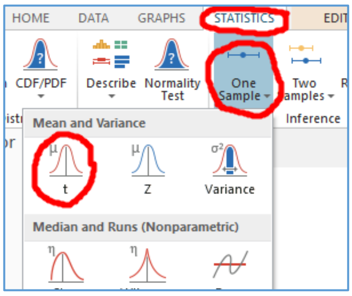
A machine is producing metal pieces that are cylindrical in shape. A sample of pieces is taken and the diameters (in cm) are:

1.01, 0.97, 1.03, 1.04, 0.99, 0.98, 0.99, 1.01, 1.03

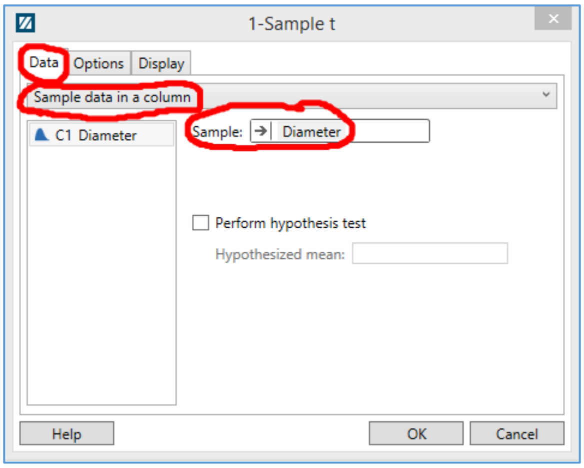
Construct a 90% confidence interval of the mean diameter of pieces from this machine, assuming an approximate normal distribution.

Step 1: Input the data into a column. Label the heading as *Diameter*.

Step 2: In Minitab Express, select
STATISTICS > **One Sample** > **Mean and Variance t**.



Step 3a: In **Data** tab, select *Sample data in a column*.
For **Sample**, select *Diameter*.



Step 3b: In **Options** tab, for **Confidence level**, select desired level.
Click **OK**.

Step 4: The results will be displayed in the output window. Interpret the results: