# STAT72000 – Probability and Statistics

LECTURE 10 – POINT ESTIMATION AND INTERVAL ESTIMATION

# Recap:

- Continuous Variable Distributions
- Discrete Variable Distributions
- Population and Sample

# Agenda

- Point Estimation
- The Central Limit Theorem
- Interval Estimation

# Point Estimation

# Point Estimation

- Population: A population is a complete collection of the entire set of objects under investigation. e.g. population of families, students, birds…
- Parameter: the numbers that describe characteristics of scores in the population
- Sample: A sample is usually drawn from the population for study.
- Statistics: the numbers that describe characteristics of scores in the sample

# Point Estimation

Let $X_1, X_2, \ldots, X_n$ be random samples from a population:

- The sample mean $\overline{X}$ is used to estimate the population mean $\mu$
- The sample variance $s^2$ is used to estimate the population variance $\sigma^2$
- If we know $X \sim Bin(n, p)$, the sample proportion $\hat{p} = \frac{X}{n}$ is used to estimate the unknown population proportion p

- A quantity calculated from data, is called a statistic.
- The statistic that is used to **estimate an unknown constant(parameter) of a distribution** is called a point estimator.
- **e.g.** $\hat{p} = \frac{X}{n}$ is called a *point estimator* of p, or simply called *estimator*.

# Point Estimation

If X and Y are discrete random variables:

| Measure | Parameters (Population) | Statistic (Sample) |
|---|---|---|
| Mean: E(X) | $\mu = \sum_{i=1}^{N} X_i \cdot p(X_i)$ | $\hat{\mu} = \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ |
| Variance: V(X) | $\sigma^2 = E[(X - E(X))^2]$ $=\sum_{i=1}^{N}(X_i - \mu)^2 \cdot p(X_i)$ | $\hat{\sigma}^2 = s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$ |
| Sd. Deviation: sd(X) | $\sigma = \sqrt{\sigma^2}$ | $s=\sqrt{s^2}$; $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$ |
| Covariance: Cov(X,Y) | $\sigma_{X,Y} = E[(X - E(X)(Y - E(Y)]$ $=\sum_{i=1}^{N}(X_i - \mu_X)(Y_i - \mu_Y) \cdot p(X_i, Y_i)$ | $\hat{\sigma}_{X,Y} = s_{X,Y} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{n-1}$ |
| Correlation Coefficient: Cor(X,Y) | $r_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}$ | $\hat{r}_{X,Y} = \frac{\hat{\sigma}_{X,Y}}{\hat{\sigma}_X \hat{\sigma}_Y}$ |

# Point Estimation - Example

An automobile manufacturer has developed a new type of bumper, which is supposed to absorb impacts with less damage than previous bumpers. The manufacturer has used this bumper in a sequence of **25** controlled crashes against a wall, each at 10 mph, using one of its compact car models.

- Let X the number of crashes that result in no visible damage. If **X is observed to be x=15**.
- **Estimate** *p*, the proportion of all such crashes that result in no damage
  - Estimator: $\hat{p} = \frac{X}{n} = \frac{15}{25} = 0.6$
  - In most problem, there will be more than on estimator.

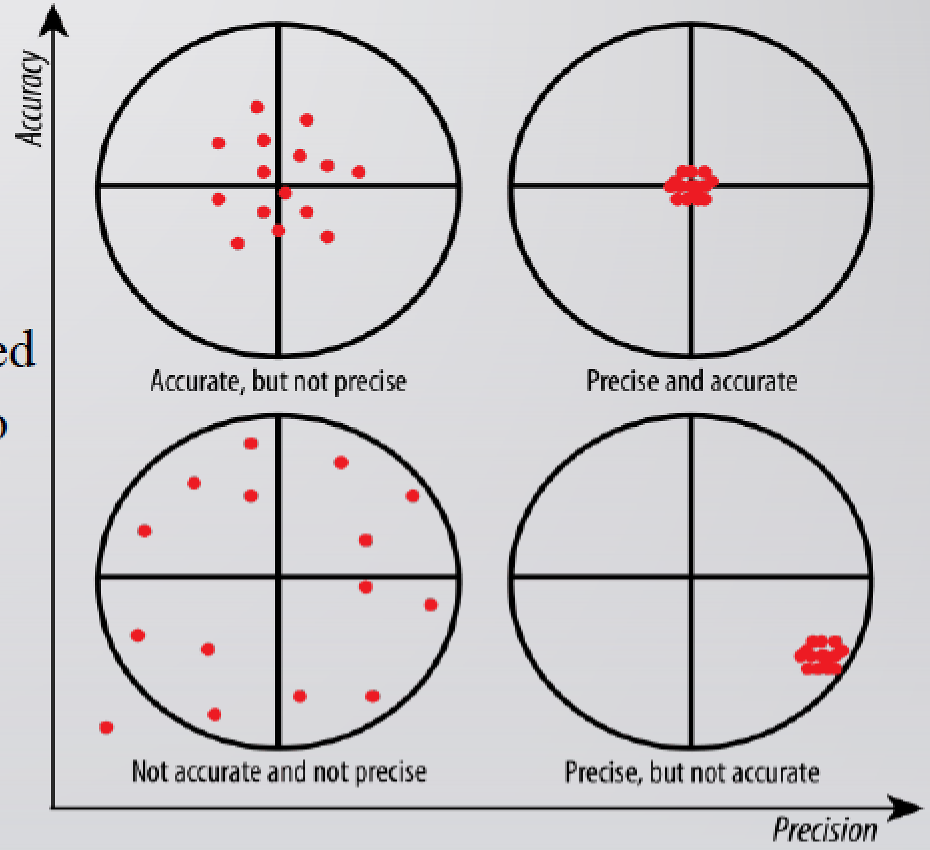
# How good is a point estimator?

- Bias of Estimator
- Variance of Estimator
- Mean Squared Error (MSE)

# Measure the Goodness of an Estimator

Ideally, an estimator should be both accurate and precise.

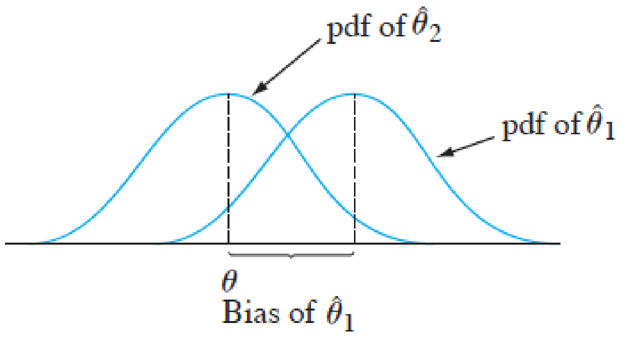
- **Accuracy**: How close an estimated value is to the true value. The accuracy of an estimator is measured by its **bias**. If $E(\hat{\theta}) = \theta$, then the estimator is said to be **unbiased** ($\theta$ is parameter, $\hat{\theta}$ is an estimator).
- **Precision**: How close estimated values are to each other; Usually refers to the stability of an estimate result. The precision is measured by its **standard deviation**, or **uncertainty**.



# Unbiased Estimator

**Definition:**

- Unbiased estimator: The point estimator $\hat{\theta}$ is an unbiased estimator for the parameter $\theta$, if $E(\hat{\theta}) = \theta$.
- Bias: if the estimator is not unbiased, then the difference $(E(\hat{\theta}) - \theta)$ is called the bias.

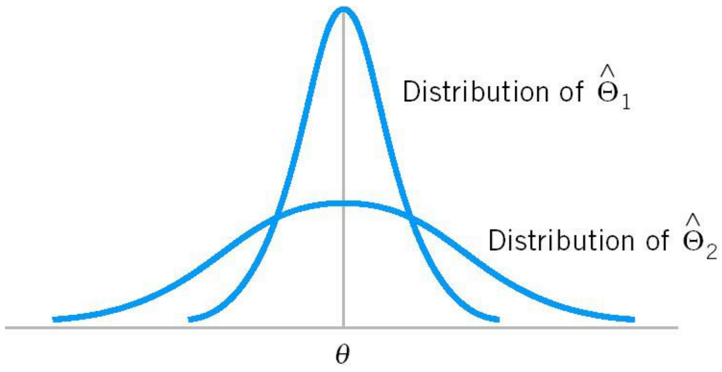



$\hat{\theta}_1$ is a biased estimator;
$\hat{\theta}_2$ is an unbiased estimator

# Estimators with Minimum Variance

The variance of an estimator is a measure of how much the estimator varies from sample to sample:

$$Var(\hat{\theta}) = \mathrm{E}(\hat{\theta}^2) - [\mathrm{E}(\hat{\theta})]^2$$




- $\hat{\theta}_1$ is a better estimator (with smaller variance than the $\hat{\theta}_2$)
- Ideally a good estimator is the **minimum variance unbiased estimator (MVUE)**

# Mean Squared Error (MSE)

The **Mean Squared Error (MSE)** is the most often used to evaluate the overall goodness of an estimator, which combines the bias and variance (uncertainty)

- Let $\theta$ be a parameter, and $\hat{\theta}$ as an estimator of $\theta$:
- $MSE = E\left[\left(\hat{\theta} - \theta\right)^2\right] = \left[E(\hat{\theta} - \theta)\right]^2 + \sigma_{\hat{\theta}}^2 = Bias^2 + Var(\hat{\theta})$

# MSE - Example

Let $X \sim Bin(n,p)$, where p is unknown.

Find the MSE of the estimator: $\hat{p} = \frac{X}{n} = \frac{\#\, of\ successes}{\#\, of\ trials}$

- Given $X \sim Bin(n,p)$, we have: $\mu_X = np, \quad \sigma_X^2 = p(1-p)$
- $MSE_{\hat{p}} = bias^2 + var(\hat{p})$
  - bias: $E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{E(X)}{n} = \frac{np}{n} = p$, **so the bias = 0**
  - variance: $var(\hat{p}) = E[\hat{p}^2] - [E(\hat{p})]^2 = E\left[\left(\frac{X}{n}\right)^2\right] - \left[E\left(\frac{X}{n}\right)\right]^2 = \frac{1}{n^2}[E(X^2) - [E(X)]^2] = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$
  - $MSE_{\hat{p}} = 0 + \frac{p(1-p)}{n} = \frac{p(1-p)}{n}$
    - When n is increased, MSE decreased.

# The Central Limit Theorem

# The Central Limit Theorem

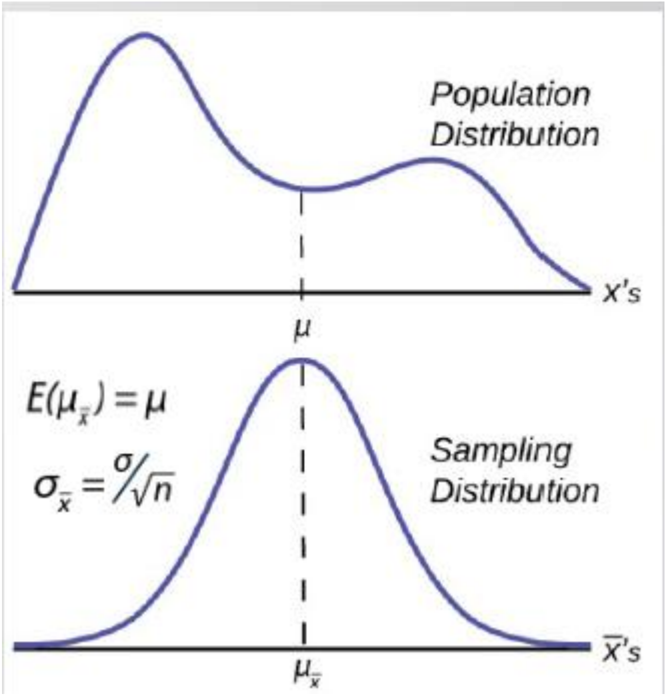
The **Central Limit Theorem** is by far the most important result in statistics!

**Main Idea:** If we draw a large enough sample from a population, then the **distribution of the sample mean is approximately normal distribution**, no matter what population the sample was drawn from.

# The Central Limit Theorem

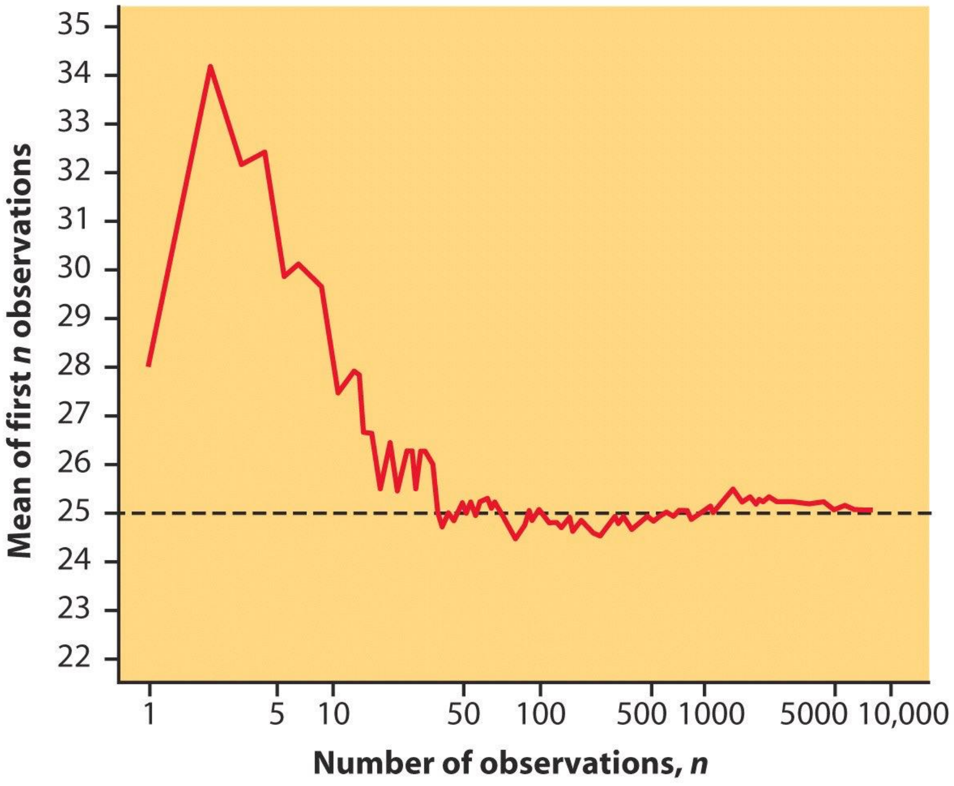
Given: Let $X_1, X_2, \ldots\ X_n$ be a simple random sample drawn from a population with mean $\mu$ and variance $\sigma^2$.

- Let the sample mean be: $\overline{X} = \frac{X_1+X_2+\cdots+X_n}{n}$
- Draw many rounds of such samples, we have $\overline{X}1, \overline{X}2, \ldots \overline{X}s$, then $\overline{X}$ approaches to a **Normal Distribution**, with mean equals to $\mu$ and variance equals $\frac{\sigma^2}{n}$
- Approximately, $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$



# Law of Large Numbers

- Draw observations at random from **any population (any types of distributions)** with finite mean $\mu$.
- If you take samples of **larger and larger size** from any population, then the **mean of the sampling** distribution, tends to get closer and closer to the true population mean, $\mu$



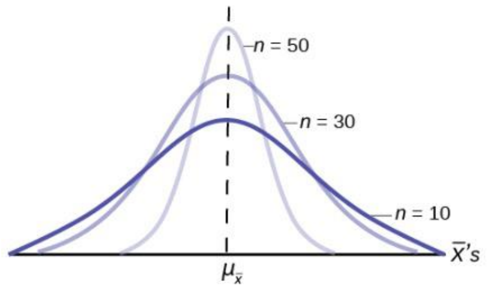
# The Central Limit Theorem

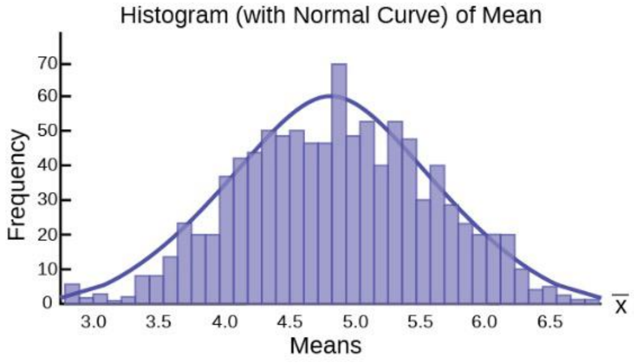
## The value of sample size n: how large is large enough?

- If the sample is drawn from a nearly **symmetric distribution**, the normal approximation can be good even for a fairly small value of n, ($n>5$).
- If the population is heavily skewed, a fairly large n may be necessary ($n>30$).

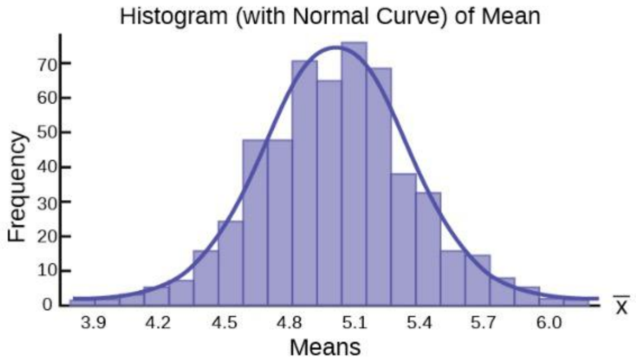
# The Central Limit Theorem

- Central Limit Theorem: As $n$ gets larger and larger, the **sample means** follow a normal distribution.
- The larger $n$ gets, the **smaller the standard deviation** of the sample means distribution gets.
- $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$




n=10




n=50

# Example

Let X denote the number of flaws in a 1-inch length of copper wire. The **probability mass function (PMF) of X** is presented in the table.

**One hundred** wires are sampled from this population. What is the probability that **the average number** of flaws per wire in this sample is less than 0.5?

| $x$ | $P(X = x)$ |
|---|---|
| 0 | 0.48 |
| 1 | 0.39 |
| 2 | 0.12 |
| 3 | 0.01 |

# Example

- $\mu = \sum xP(X = x) = 0.66$
- $\sigma^2 = \sum x^2 P(X = x) - \mu^2 = 0.5244$
- the average number of flaws of 100 wires: $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$ => $\overline{X} \sim N(0.66, \frac{0.5244}{100})$
- Find $P(\overline{X} < 0.5)$ => $Z = \frac{x-u}{\sigma} = \frac{0.5-0.66}{\sqrt{0.005244}} = -2.21$
- Therefore, $P(\overline{X} < 0.5)$ = $P(Z < -2.21) = 0.0136$ (look up the z-table)

Only very low probability (1.36%) that 100 samples' average flaws per wire is less than 0.5.

| $x$ | $P(X = x)$ |
|---|---|
| 0 | 0.48 |
| 1 | 0.39 |
| 2 | 0.12 |
| 3 | 0.01 |

# Exercise

At a college, the mean age of the students is 22.3 years, and the standard deviation is 4 years. A random sample of 64 students is drawn.

- What is the probability that the average age of these 64 students is greater than 23 years?

# Exercise

- Distribution Properties of X: $\mu = 22.3, \sigma^2 = 4^2 = 16$
- The average of 64 samples: $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$ => $\overline{X} \sim N(22.3, \ (\frac{16}{64} = 0.25))$
- Find $P(\overline{X} > 23)$
  - $Z = \frac{x-u}{\sigma} = \frac{23-22.3}{\sqrt{0.25}} = 1.4$
  - $P(\overline{X} > 23)$ => $P(Z > 1.4)$ => $P(Z < -1.4)$ => lookup the z-table = 0.0808

➢Therefore, **the probability is 0.0808** that the average age of these 64 students is greater than 23 years

# Interval Estimation

# Point Estimate v.s. Interval Estimate

- A **point estimate,** because it is a single number, by itself provides no information about the precision and reliability of estimation. – almost never exactly equal to the true values they are estimating.
- An alternative to reporting a single sensible value for the parameter being estimated, is to calculate and report an interval of plausible values - an **interval estimate** or confidence interval (CI)

# Large Sample Confidence Intervals for a Population Mean

Review: The Central Limit Theorem:

- Let $X_1, X_2, \ldots X_n$ be a random sample drawn from a population with mean $\mu$ and variance $\sigma^2$
- Then the sample mean: $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$
  - $\mu$: population mean
  - $\sigma^2$: population variance
  - $\frac{\sigma^2}{n}$ : sample-mean's variance

# Confidence Interval

**Example:** Assume that a large number of independent measurements, all using the same procedure, are made on the diameter of a mechanical component.

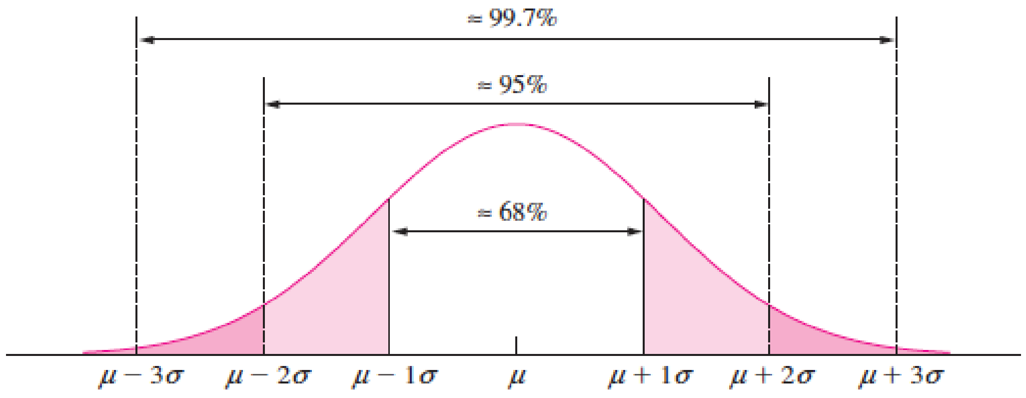
The **sample mean** of the measurements is **14.0 cm**, and the **standard deviation** of the sample mean is **0.1 cm**. Assume that the measurements are unbiased.

- The true diameter of the piston is **not** exactly equal to the sample mean of 14.0 cm
- The sample mean comes from a normal distribution, so the **standard deviation** can be used to determine how close it is likely to be to the true diameter.

# Confidence Interval

The **sample mean** of the measurements is **14.0 cm**, and the **standard deviation** of the sample mean is **0.1 cm**. Assume that the measurements are unbiased.

- The true diameter falls in the interval of **(13.7, 14.3)** with **99.7%** confidence.
- The true diameter falls in the interval of **(13.9, 14.1)** with **68%** confidence.



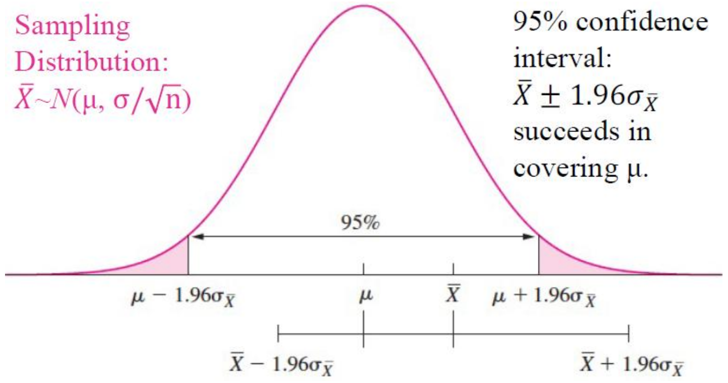
# Confidence Interval

**Example:** A quality-control engineer wants to estimate the mean of weight of boxes that have been filled with cereal by a certain machine. He draws a simple random sample of **100** boxes from the production line that have been filled by that machine.

- He computes the average weight of the 100 boxes, which is $\overline{X} = 12.05$ oz, and the sample standard deviation is $s = 0.1$ oz.
- The population mean will not be exactly equal to the sample mean 12.05oz. So it is best to conduct a confidence interval around the 12.05, that is likely to cover the population mean

# Confidence Interval

- $\mu$: population mean
- $\sigma^2$: population variance
- Given sample-mean: $\overline{X} = 12.05$ , sample variance: $s^2 = 0.1^2 = 0.01$
- According to the Central Limit Theorem: the sample mean $\overline{X}$ follows a normal distribution: $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$
- Approximate $\sigma^2$ with sample variance $s^2$
- Therefore, the sample mean $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$ => $\overline{X} \sim N(\mu, \frac{0.01}{100})$
- With confidence at 95% level, $-1.96 < \frac{\overline{X}-\mu}{\sqrt{\frac{0.01}{100}}} < 1.96$
- => The population mean 95% interval: (12.0304, 12.0696)



# Confidence Interval for **population mean**

The 95% confidence interval (CI) for **μ** can be expressed as:

$$(\overline{X} - 1.96 * \frac{\sigma}{\sqrt{n}}) < \mu < (\overline{X} + 1.96 * \frac{\sigma}{\sqrt{n}})$$

- $\left(\overline{X} - 1.96 * \frac{\sigma}{\sqrt{n}},\ \overline{X} + 1.96 * \frac{\sigma}{\sqrt{n}}\right)$ is the 95% Confidence Interval for population mean **μ**

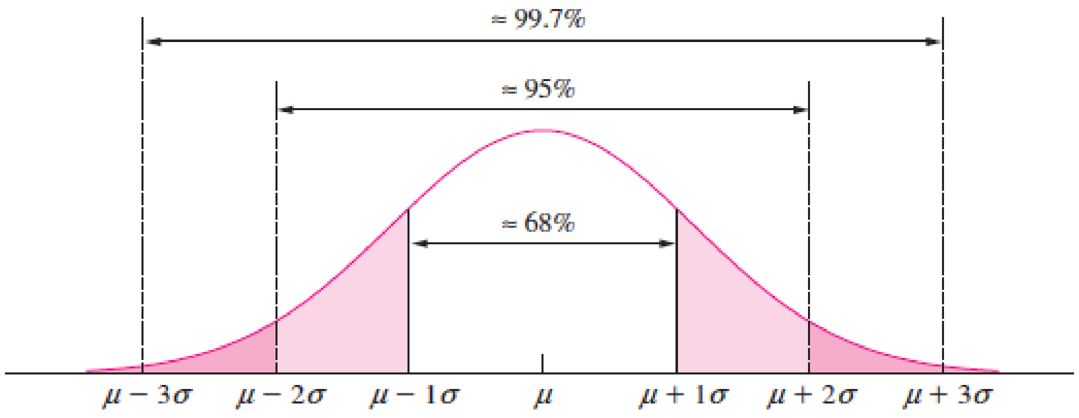
# Confidence Interval

Let $X_1, \ldots, X_n$ be a large (n>30) random sample from a population with mean $\mu$ and standard deviation $\sigma$, so that $\overline{X}$ is approximately normal.

- Then a level $100(1-\alpha)\%$ confidence interval for $\mu$ is $\overline{X} \pm Z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$
- When the value of $\sigma$ is unknown, it can be replaced with the sample standard deviation s.

Important Values of CI:

- $\overline{X} \pm 1 * \frac{\sigma}{\sqrt{n}}$ is the 68% confidence interval
- $\overline{X} \pm 1.645 * \frac{\sigma}{\sqrt{n}}$ is the 90% CI
- $\overline{X} \pm 1.96 * \frac{\sigma}{\sqrt{n}}$ is the 95% CI
- $\overline{X} \pm 2.58 * \frac{\sigma}{\sqrt{n}}$ is the 99% CI
- $\overline{X} \pm 3 * \frac{\sigma}{\sqrt{n}}$ is the 99.7% CI



# Exercise

In a sample of 50 microdrills drilling a low-carbon alloy steel, the average lifetime was 12.68 with a standard deviation of 6.83.

1) Find the 95% confidence interval for the mean lifetime
2) Find the 80% confidence interval for the mean lifetime

# Exercise

Given: n=50, $\overline{X}$=12.68, s=6.83

Question: find 95% CI and 80% CI

- Formula is $\overline{X} \pm Z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$
- 95% CI => $\alpha = 0.05$ => $z_{\alpha/2} = z_{0.025} = 1.96$
- the 95% CI is: $12.68 \pm 1.96 * \frac{6.83}{\sqrt{50}} = 12.68 \pm 1.89$
- the 95% interval is: (10.79, 14.57)

# Small-sample Confidence Interval for Population Mean

- Sample size is large (n>30), central limit theorem, population mean approximates normal distribution
- Sample size is small: sample variance may not be close to population variance, population mean may not be approximately a normal distribution.

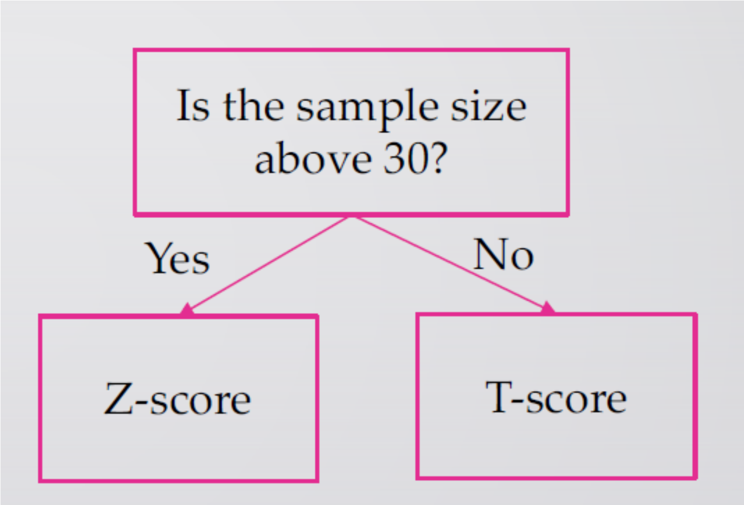
# Student's t-Distribution

- Student's t-distribution with (n-1) degrees of freedom, is denoted as $t_{n-1} = \frac{\overline{X}-\mu}{s/\sqrt{n}}$
- **t-distribution** is a type of normal distribution that is used for small sample size.

- **Definition:** Let $X_1, X_2, \dots X_n$ be a small (e.g. n<30) sample from a **normal** population with mean $\mu$. Then the quantity $\frac{\overline{X}-\mu}{s/\sqrt{n}}$ has a student's t distribution with (n-1) degrees of freedom, denoted as $t_{n-1}$

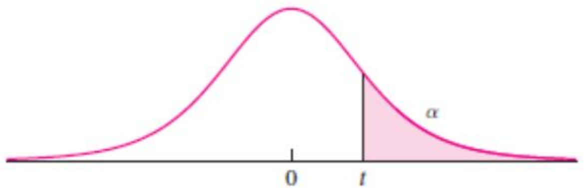
# Z-score vs. T-score

- $Z\text{-}score = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$
- $t\text{-}score = \frac{\overline{X}-\mu}{s/\sqrt{n}}$



# T-table

- v – degrees of freedom (n-1)
- $\alpha$ – the confidence level




| | α | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\nu$ | **0.40** | **0.25** | **0.10** | **0.05** | **0.025** | **0.01** | **0.005** | **0.001** | **0.0005** |
| **1** | 0.325 | 1.000 | 3.078 | 6.314 | 12.706 | 31.821 | 63.657 | 318.309 | 636.619 |
| **2** | 0.289 | 0.816 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | 22.327 | 31.599 |
| **3** | 0.277 | 0.765 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 10.215 | 12.924 |
| **4** | 0.271 | 0.741 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 7.173 | 8.610 |
| **5** | 0.267 | 0.727 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 5.893 | 6.869 |
| **6** | 0.265 | 0.718 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 | 5.208 | 5.959 |
| **7** | 0.263 | 0.711 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | 4.785 | 5.408 |
| **8** | 0.262 | 0.706 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 | 4.501 | 5.041 |
| **9** | 0.261 | 0.703 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 | 4.297 | 4.781 |
| **10** | 0.260 | 0.700 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | 4.144 | 4.587 |
| **11** | 0.260 | 0.697 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 | 4.025 | 4.437 |
| **12** | 0.259 | 0.695 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 | 3.930 | 4.318 |
| **13** | 0.259 | 0.694 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 | 3.852 | 4.221 |
| **14** | 0.258 | 0.692 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | 3.787 | 4.140 |
| **15** | 0.258 | 0.691 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 | 3.733 | 4.073 |