02-24-00201 Probability and Statistics II

DR. AHMED TAYEL

Department of Engineering Mathematics and Physics, Faculty of Engineering, Alexandria University

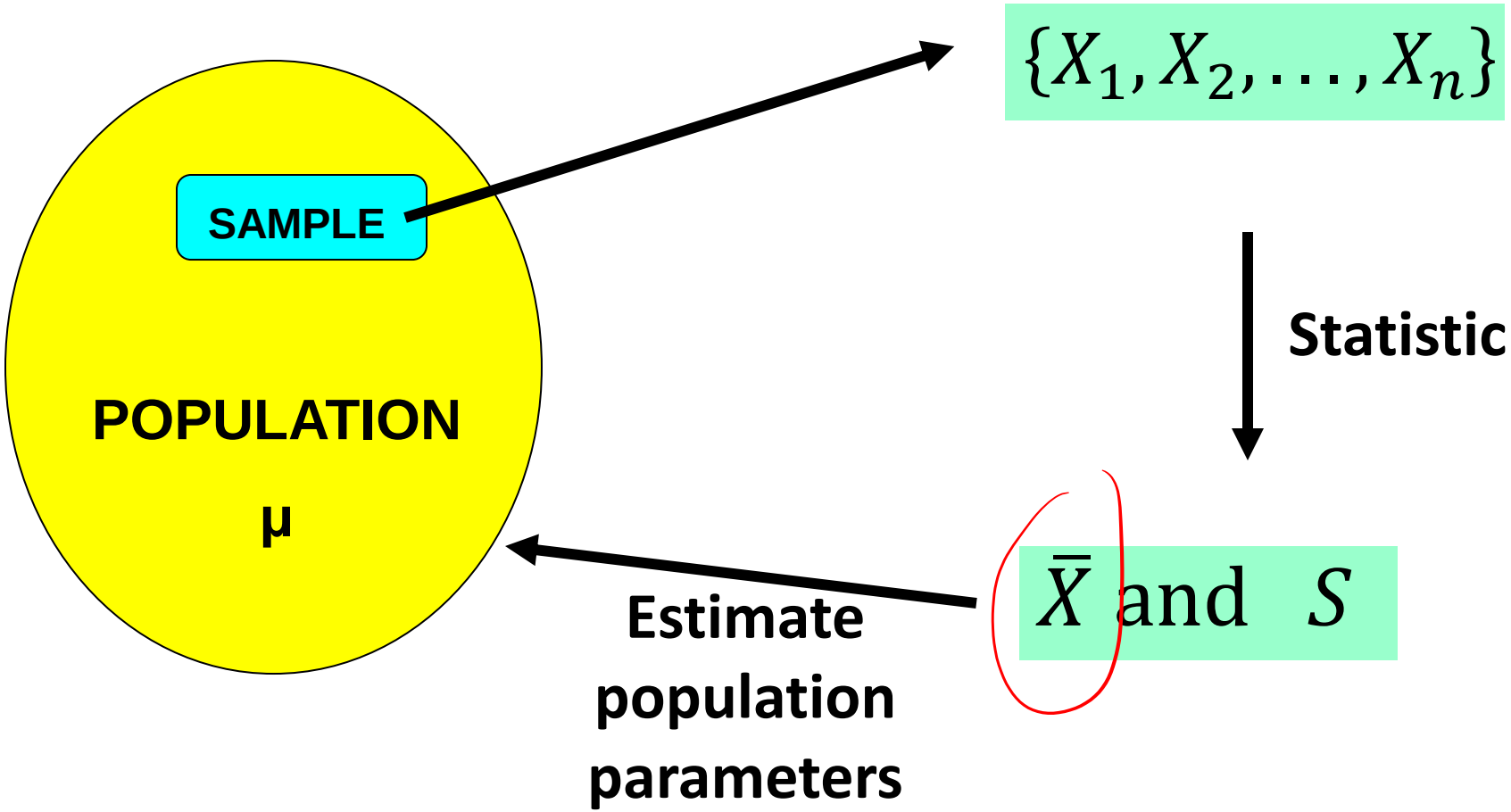
ahmed.tayel@alexu.edu.eg

Outline

1. Review.
2. Distributions derived from the Normal distribution.
    1. The Chi-squared distribution **[Previous lecture]**.
        1. Distribution of $\boldsymbol{s^2}$.
    2. The t-distribution **[This lecture]**.
    3. The F-distribution **[This lecture]**.

1. Review

The Sampling Process



Central Tendency in the Sample

Definition:

If $X_1, X_2, \ldots, X_n$ represents a random sample of size n, then the sample mean is defined to be the statistic:

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

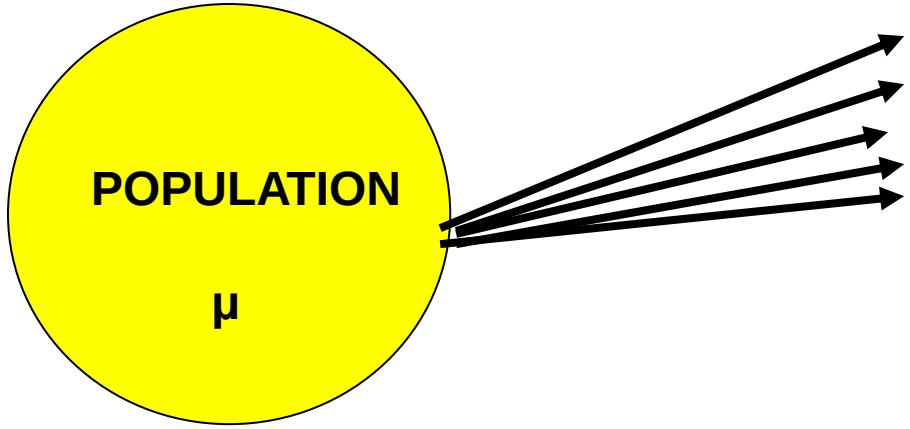
Variability in the Sample

Definition:

If $X_1, X_2, \ldots, X_n$ represents a random sample of size n, then the sample variance is defined to be the statistic:

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1} = \frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2}{n-1} \ (\text{unit})^2$$

The Sampling Distribution



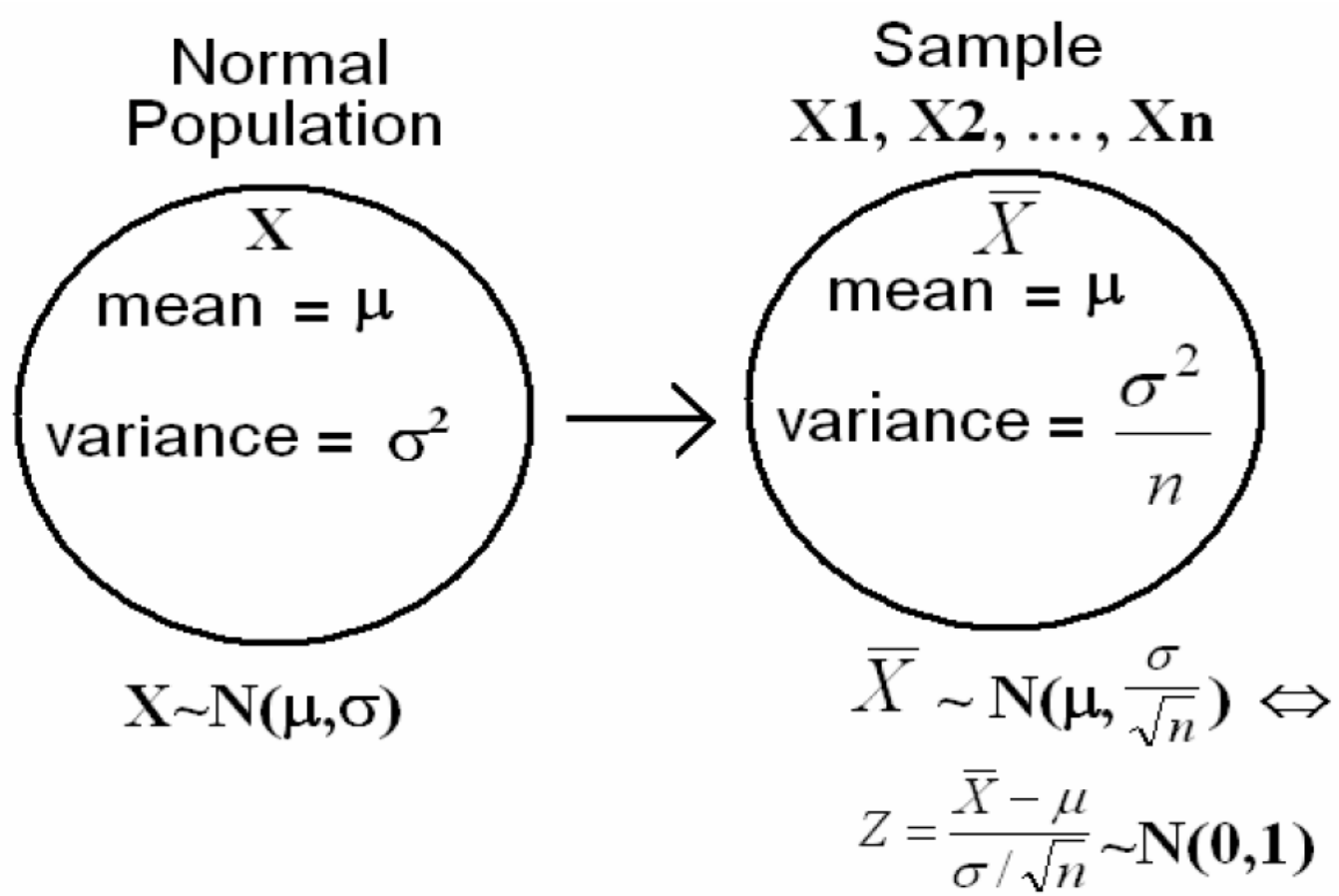
Repeated Sampling

$\bar{X}$ is also a random variable

$\{\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots, \bar{x}_N\}$

Sampling Distribution

- If $X_1, X_2, \ldots, X_n$ is a random sample of size n from $N(\mu,\sigma)$, then $\bar{X} \sim N(\mu_{\bar{X}}, \sigma_{\bar{X}})$ or $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$.
- $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}}) \Leftrightarrow Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$



What if the population is not normally distributed?

Theorem: (Central Limit Theorem)

If $X_1, X_2, \ldots, X_n$ is a random sample of size n from any distribution (population) with mean μ and finite variance σ^2, then, if the sample size n is large, the random variable

$$n \geq 30$$

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

is approximately standard normal random variable, i.e.,

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1) \text{ approximately.}$$

Case of σ^2 is unknown

If $n \geq 30$, the central limit theorem (CLT) is still valid.

If we replace σ^2 by s^2

$$Z = \frac{\bar{X} - \mu}{s / \sqrt{n}} \sim N(0,1)$$

Summary

Given the random sample $X_1, X_2, \ldots, X_n$.
We have **three cases** for the distribution of $\bar{X}$

[1]

- Sample taken from a normal population
- σ^2 is known

$$\bar{X} \sim Norm\left(\mu, \frac{\sigma^2}{n}\right)$$

[2]

- $n \geq 30$ and sample taken from any distribution

$$\bar{X} \sim Norm\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\bar{X} \sim Norm\left(\mu, \frac{s^2}{n}\right)$$

[3]

- Sample taken from a normal population
- $n < 30$
- σ^2 is unknown
- This lecture
- t-distribution

2. The t-distribution

σ unknown

Normal population

- Large sample
- s is a good estimate to σ
- does not vary much from sample to sample

- Small sample
- s^2 fluctuates considerably from sample to sample

$n \geq 30$

By CLT

$$Z \approx \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

$n < 30$

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim ??$$

Student's ***t*** Distribution

Student's t Distribution

- Let $Z \sim N(0,1)$ $\quad V \sim \chi^2_k$ $\quad Z, V$ independent
- Define: $T = \dfrac{Z}{\sqrt{V/k}}$ → $\boldsymbol{t}$ distribution with $\boldsymbol{k}$ degrees of freedom
- Application: $X_1, \ldots, X_n \sim iid\ N(\mu, \sigma^2)$
- $Z = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \sqrt{n}\left(\dfrac{\bar{X}-\mu}{\sigma}\right) \sim N(0,1)$
- $V = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ $\quad Z, V$ Independent
- $T = \dfrac{\sqrt{n}\left(\frac{\bar{X}-\mu}{\sigma}\right)}{\sqrt{\frac{(n-1)S^2}{\sigma^2}\Big/(n-1)}} = \sqrt{n}\left(\dfrac{\bar{X}-\mu}{S}\right) = \left(\dfrac{\bar{X}-\mu}{S/\sqrt{n}}\right) \sim t_{(n-1)}$

  $\boldsymbol{t}$ distribution with $\boldsymbol{n-1}$ degrees of freedom

Normal distn

σ unknown → Normal Pop.

$n \geq 30$: t-distn ≈ approx. Z-dist. (CLT)

$n < 30$: t-distn

Student's *t* Distribution

If the population standard deviation, σ, is ***unknown***, replace σ with the sample standard deviation, *s*. If the population is normal, the resulting statistic:

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

has a ***t distribution* with *(n - 1) degrees of freedom***.

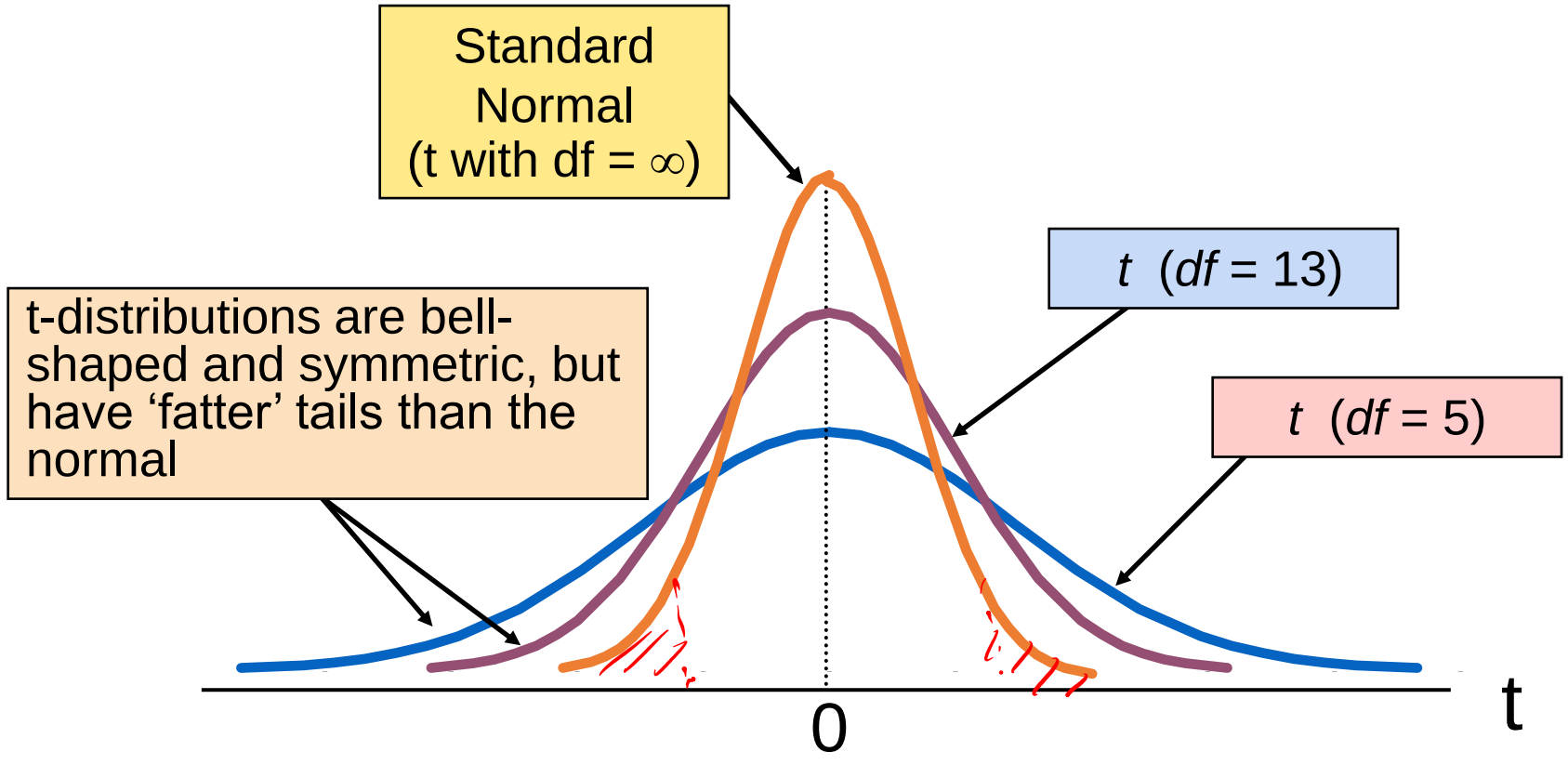
Population must be Normal

St. normal

$$T = \frac{Z}{\sqrt{\chi^2_k / k}}$$

Student's *t* Distribution

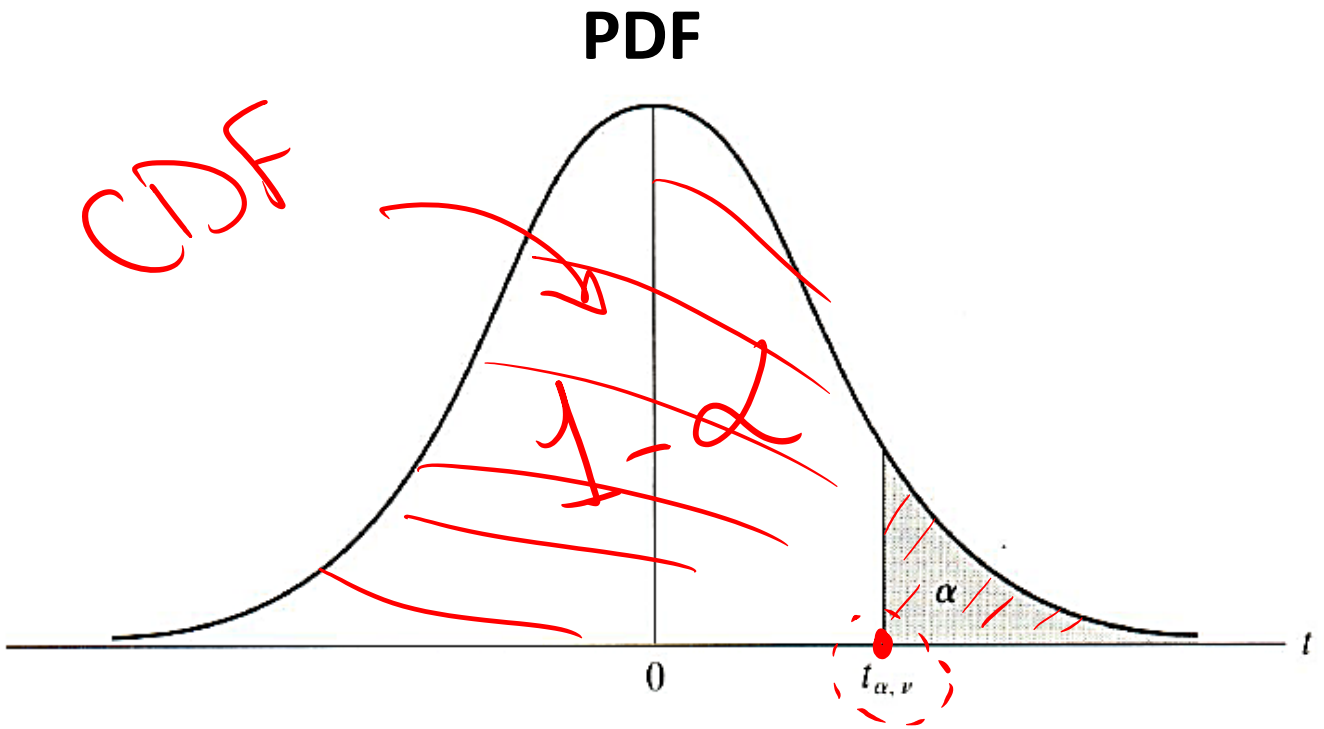
- The expected value of *t* is 0.
- The variance of *t* is greater than 1, but approaches 1 as the number of degrees of freedom increases.
- When the sample size is small (<30) we use t distribution.



Student's t Distribution

area under PDF after the point $t_{\alpha,k}$

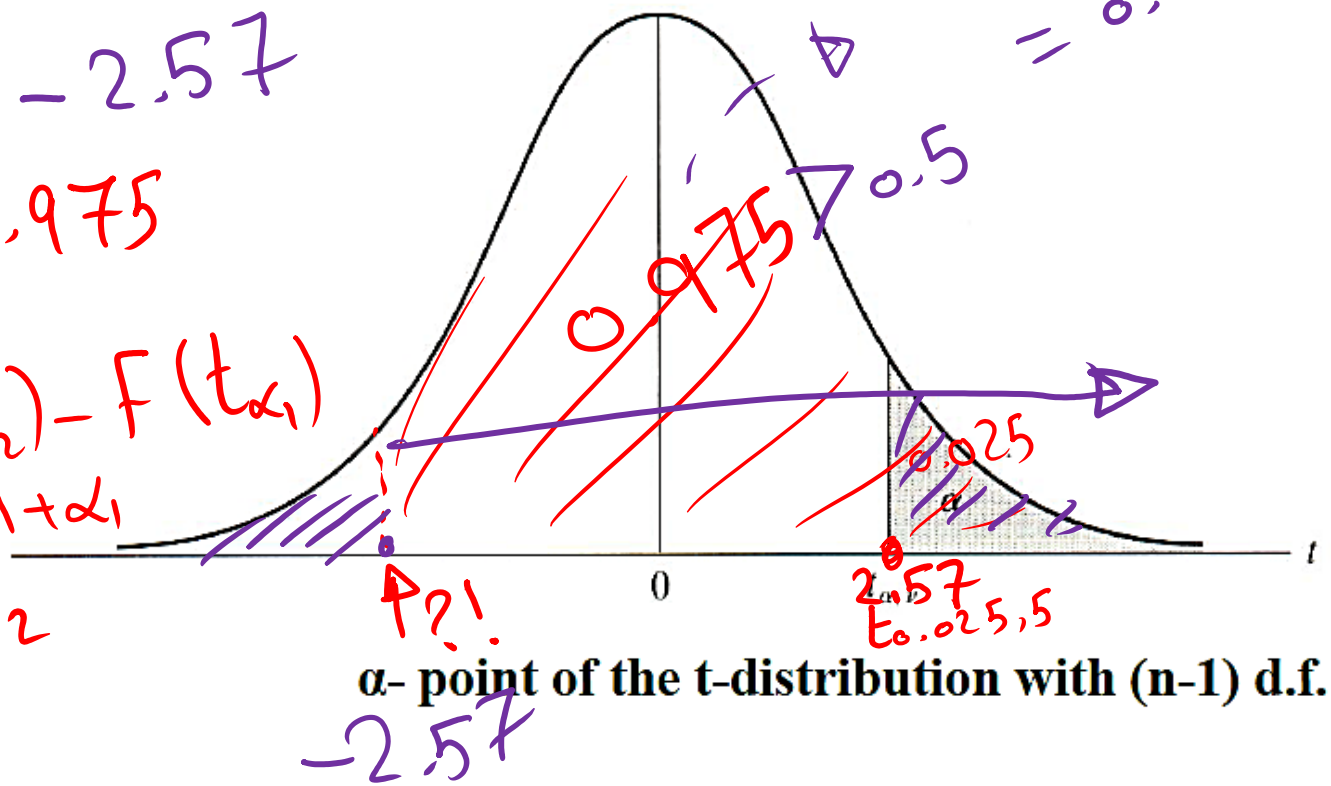
- $t_{\alpha,k}$
- $P\left(U > t_{\alpha,k}\right) = \alpha$
- $F_{CDF}\left(t_{\alpha,k}\right) = 1 - \alpha$
- $t_{\alpha,k} = F_{CDF}^{-1}(1-\alpha)$

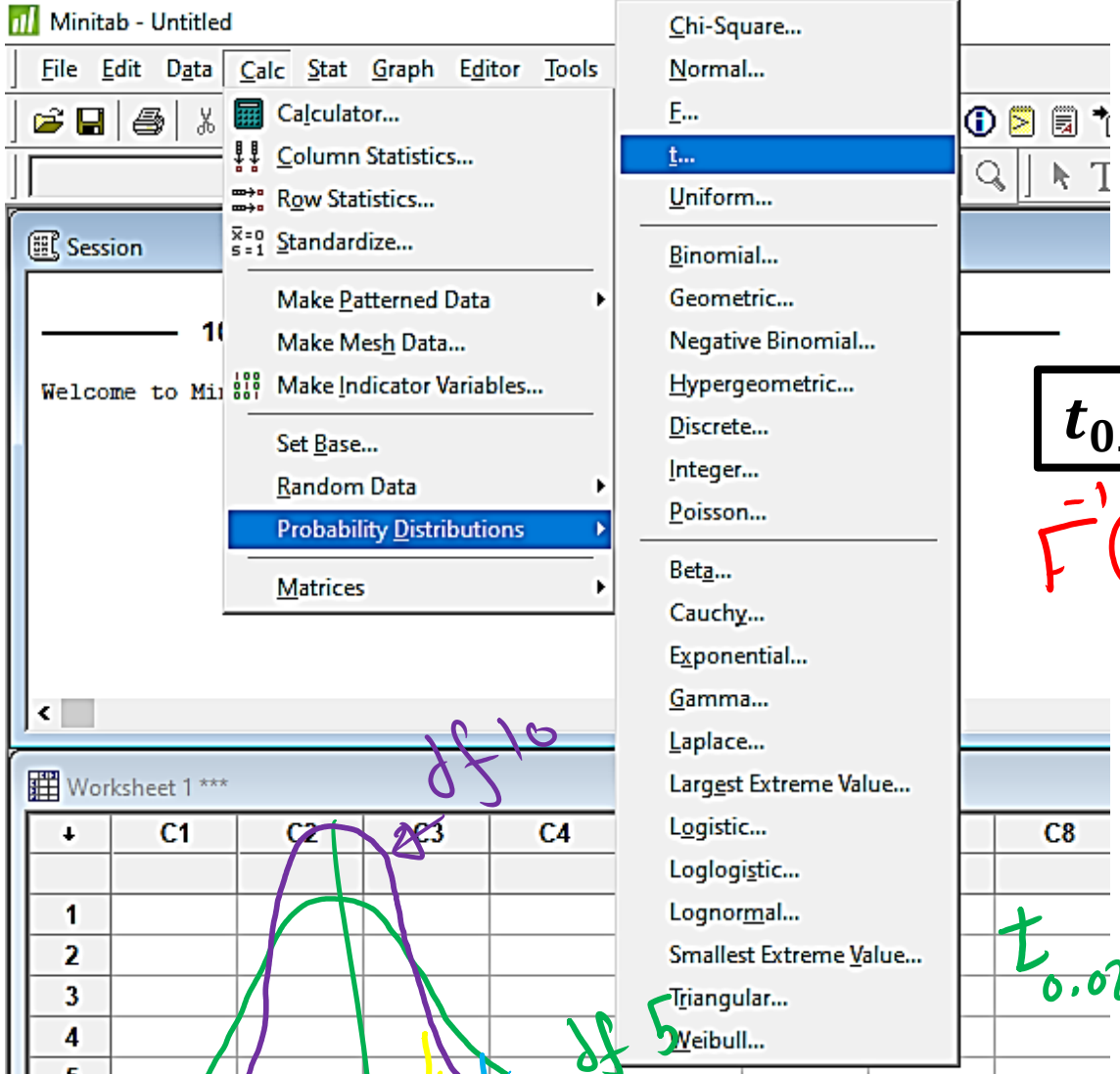


α- point of the t-distribution with (n-1) d.f.

Student's t Distribution

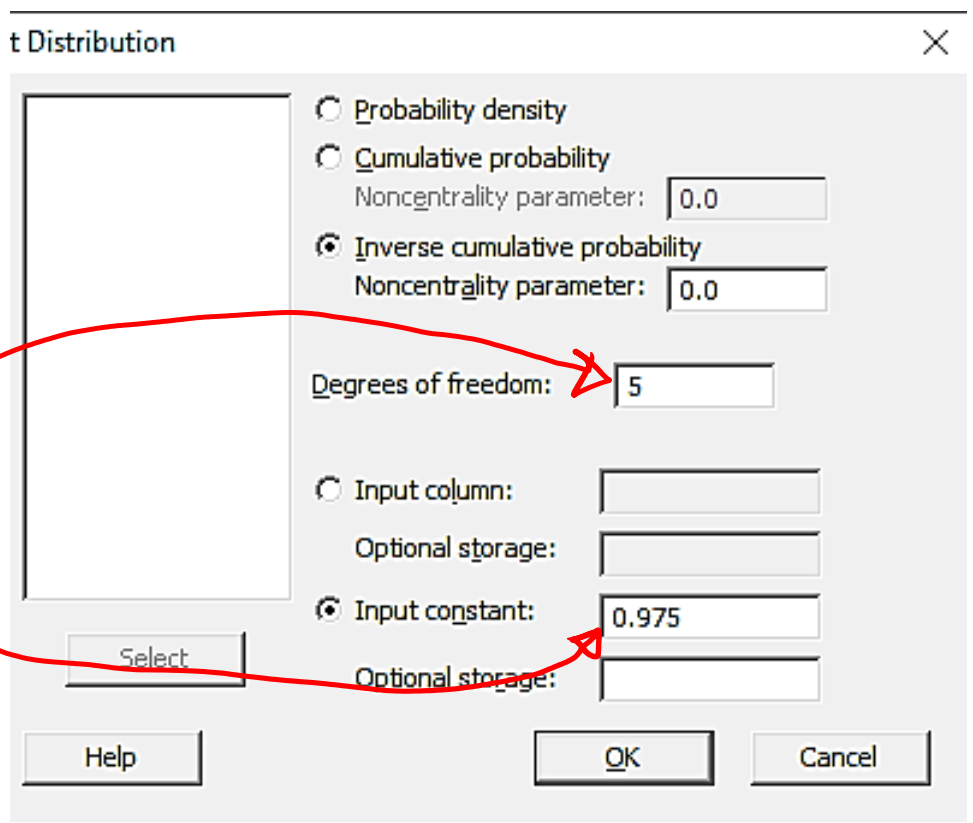
- $t_{0.5\,,\,5} = 0$
- $t_{0.025\,,\,5} = 2.57 \rightarrow t_{0.975\,,\,5} =$
- $F\left(t_{0.025\,,\,5}\right) =$
- $\Pr\left(t_{\alpha_1} < T < t_{\alpha_2}\right) =$
- $t_{0.025\,,\,5} = ?$
- Using MINITAB or R





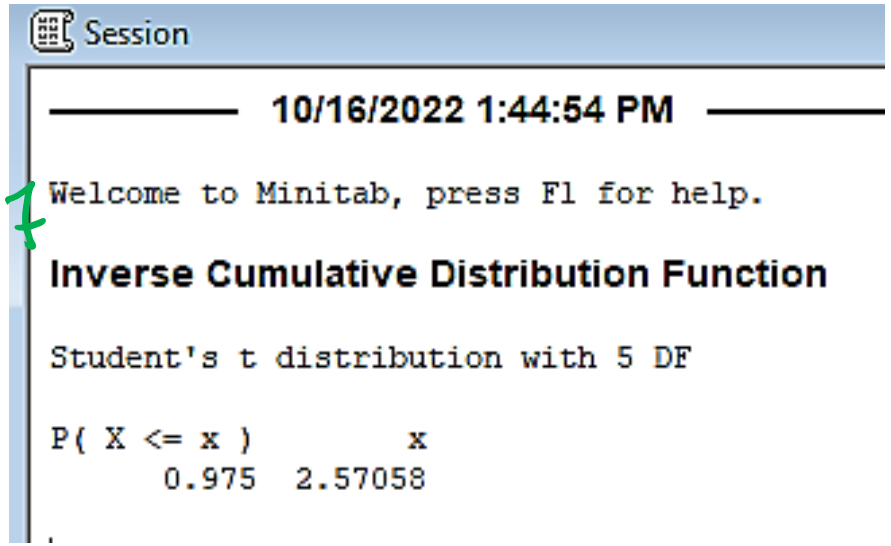
$t_{0.025,5}$

$F^{-1}(1-0.025)(0.975)$

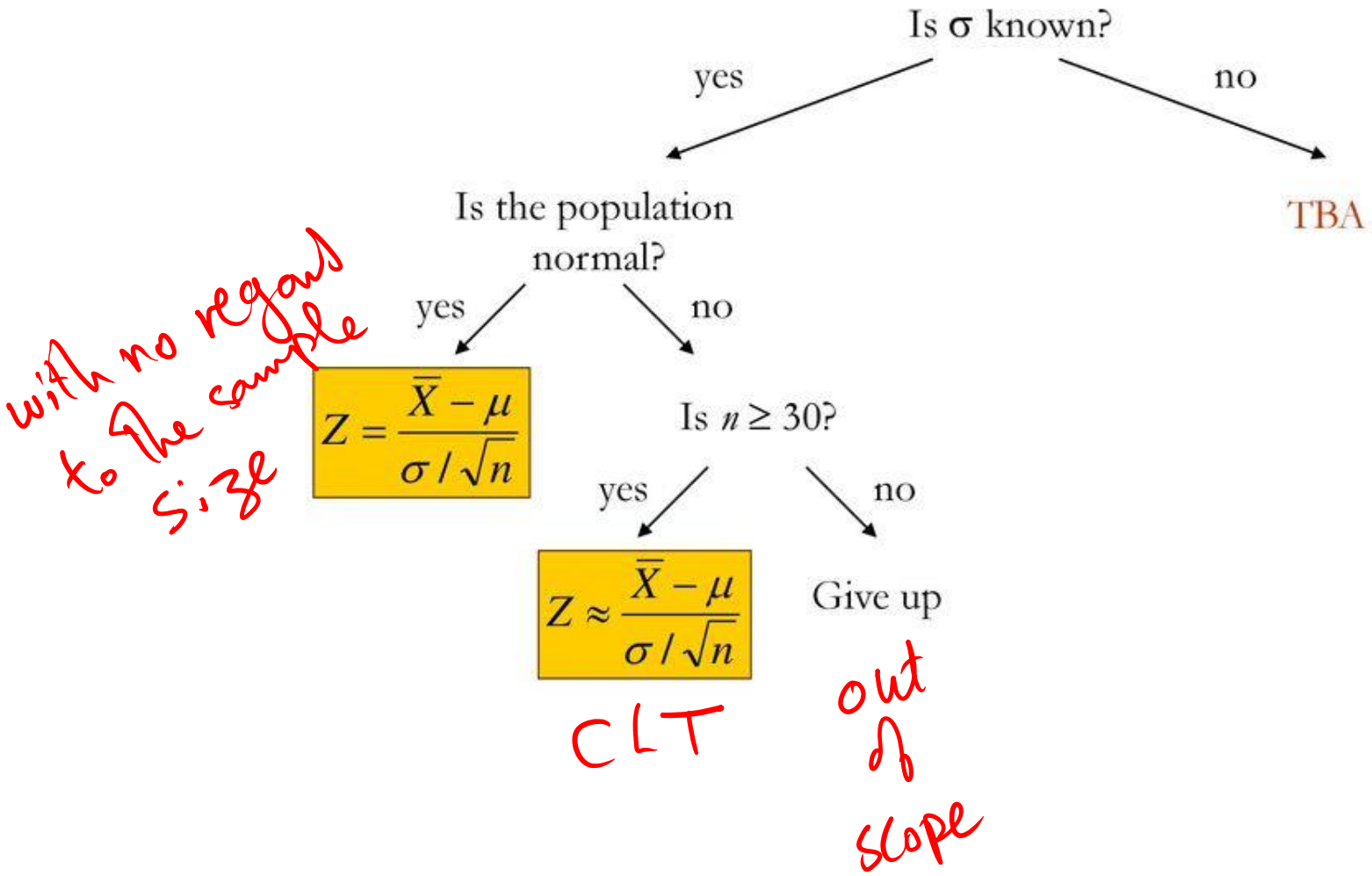


$t_{0.025,5} = 2.57$

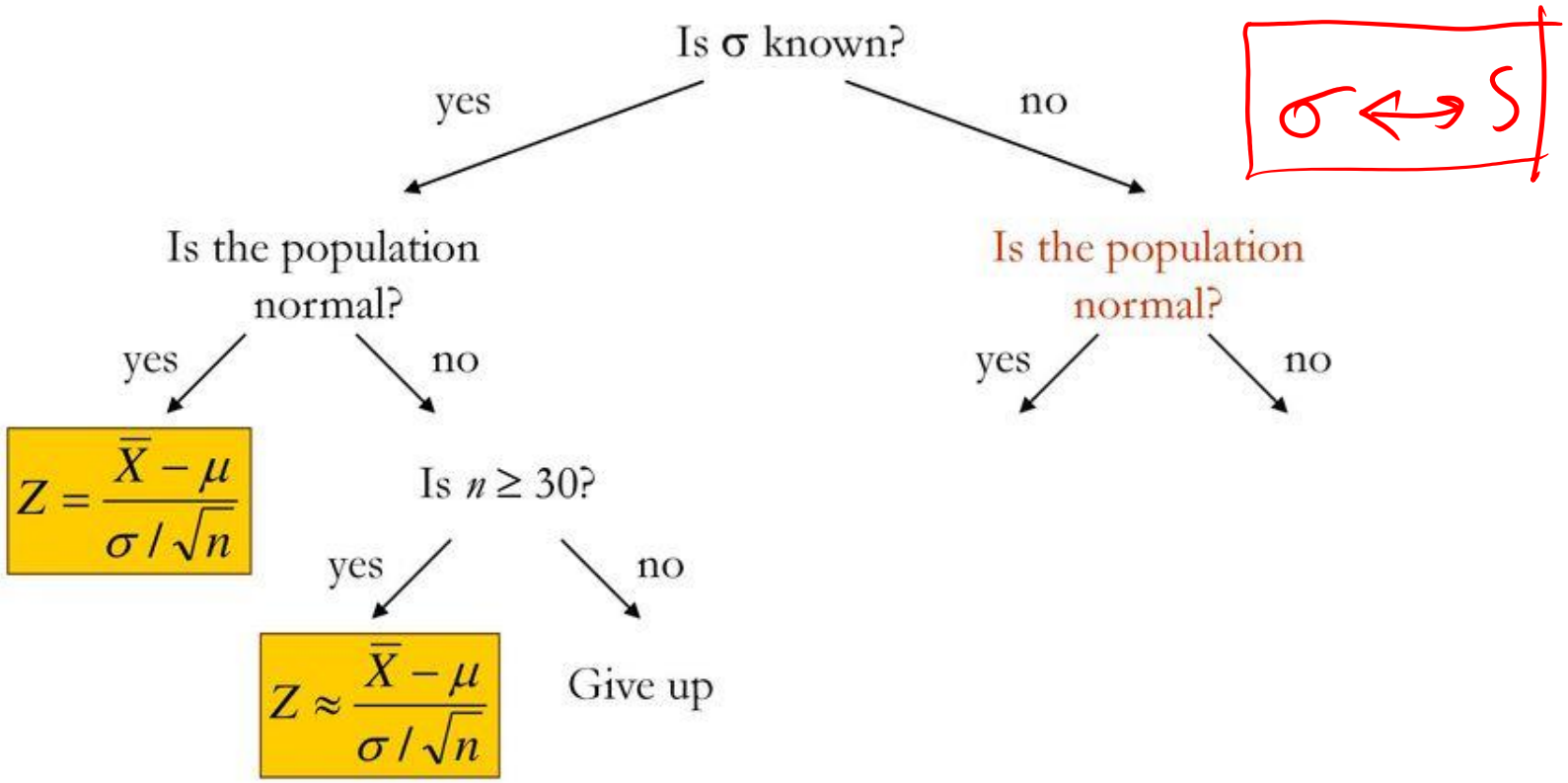
$t_{0.025,10} < 2.57$



The Decision Tree

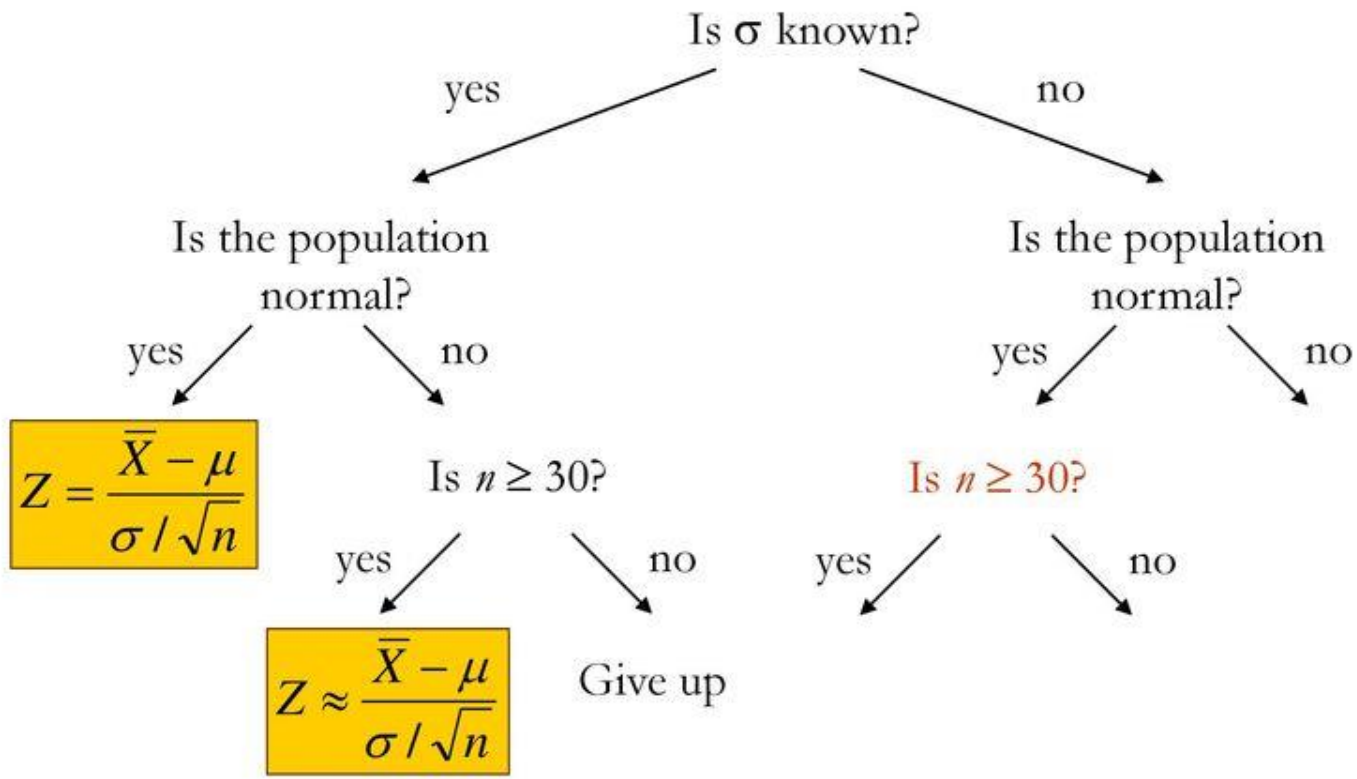


The Decision Tree

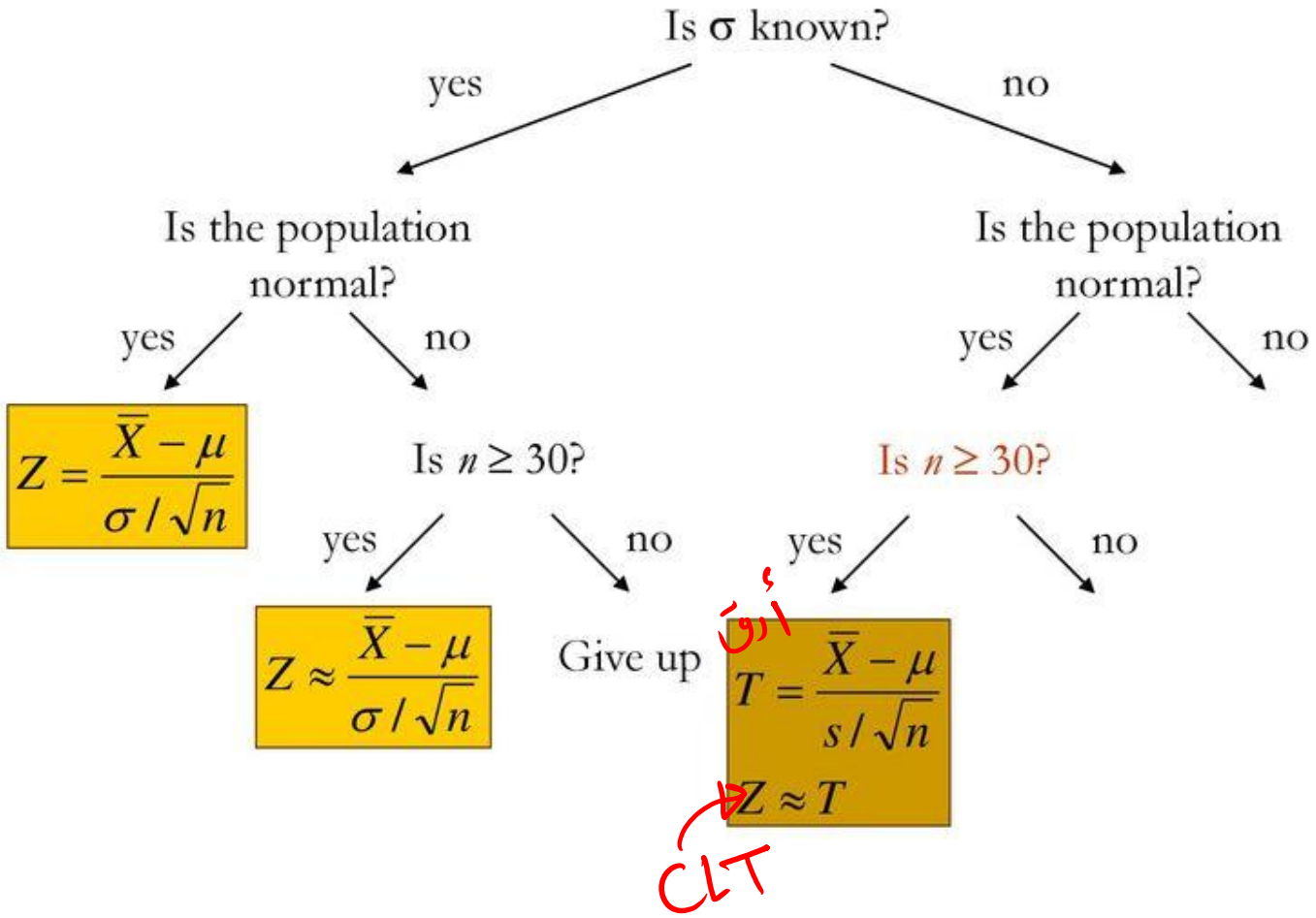


$\sigma \leftrightarrow s$

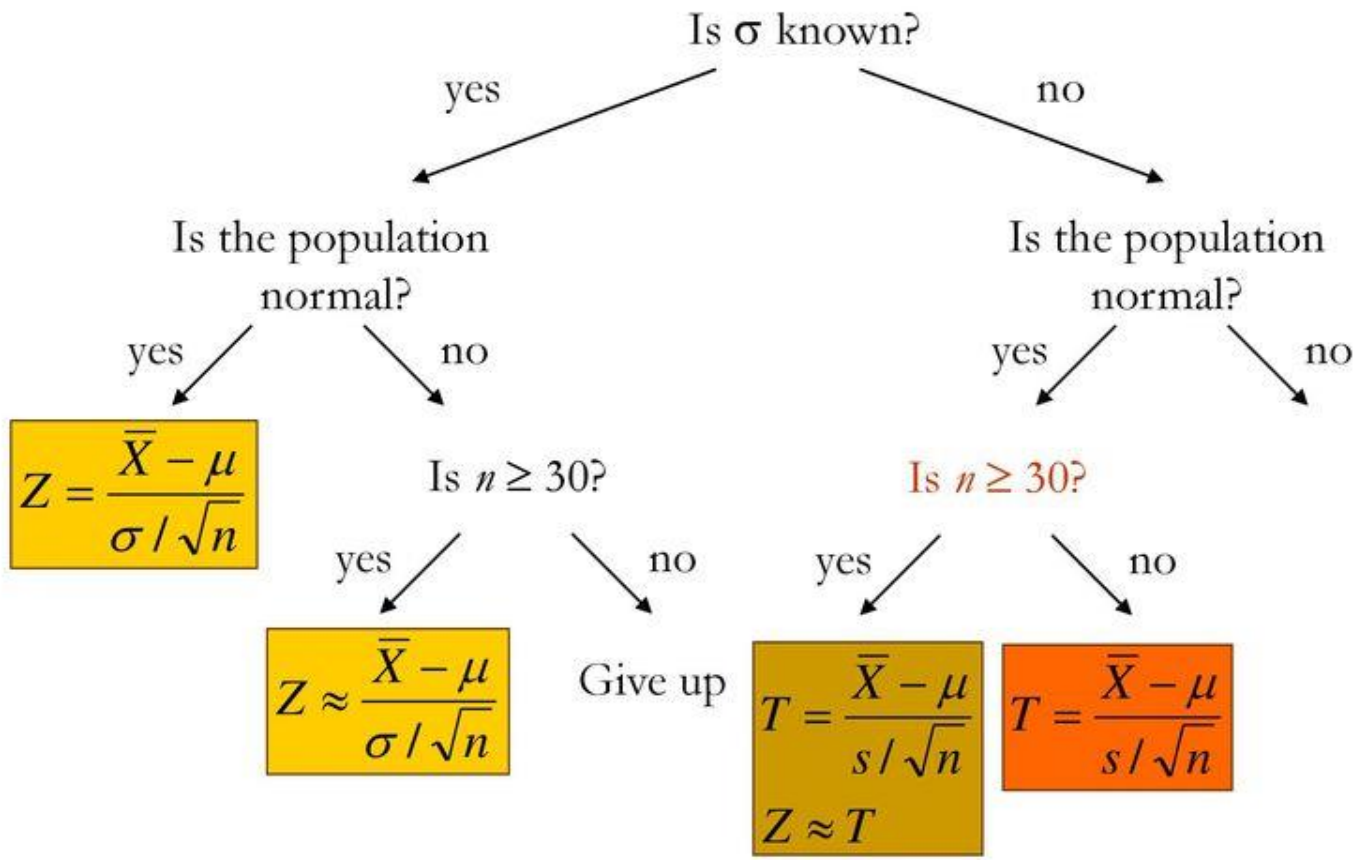
The Decision Tree



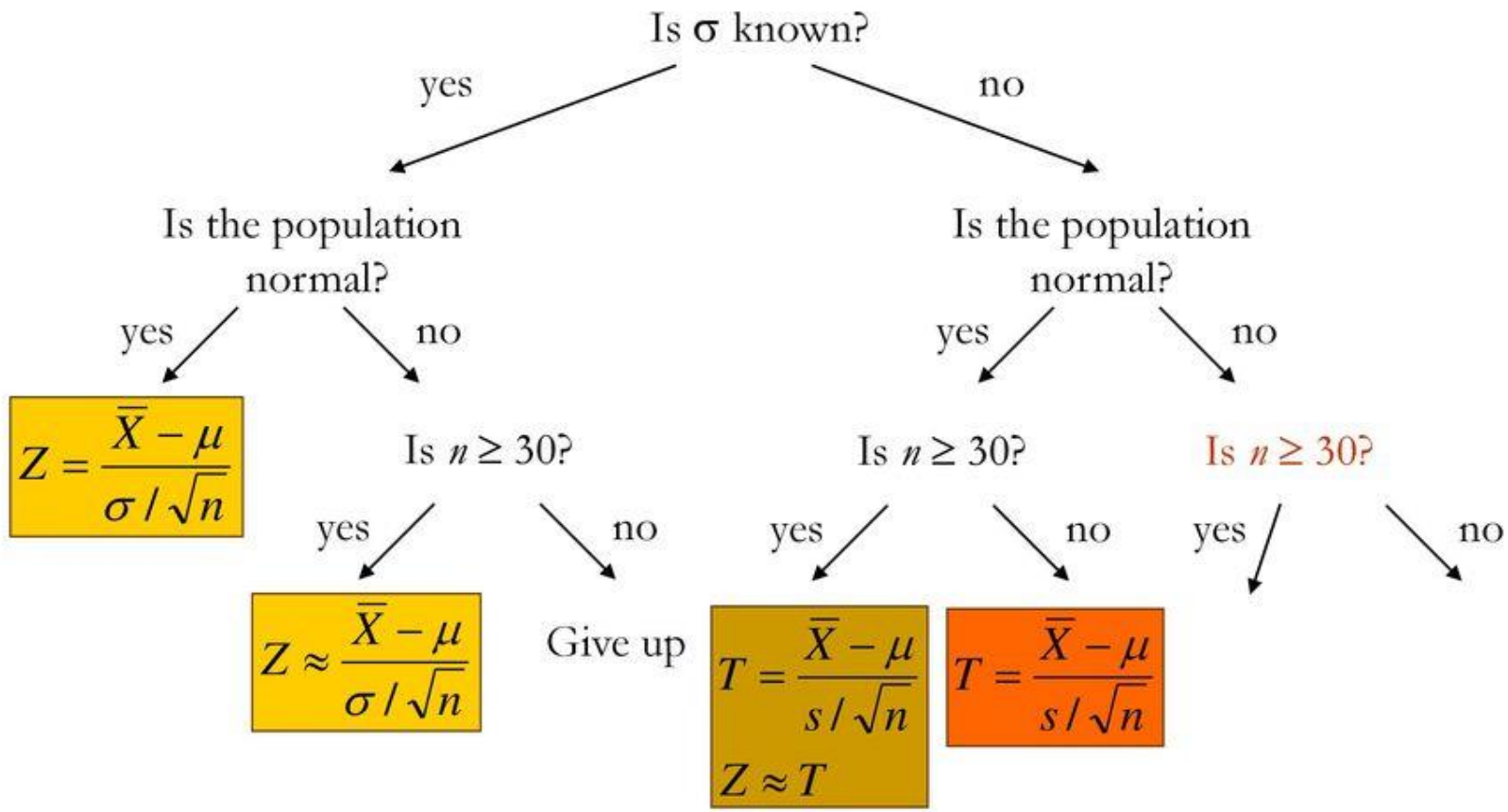
The Decision Tree



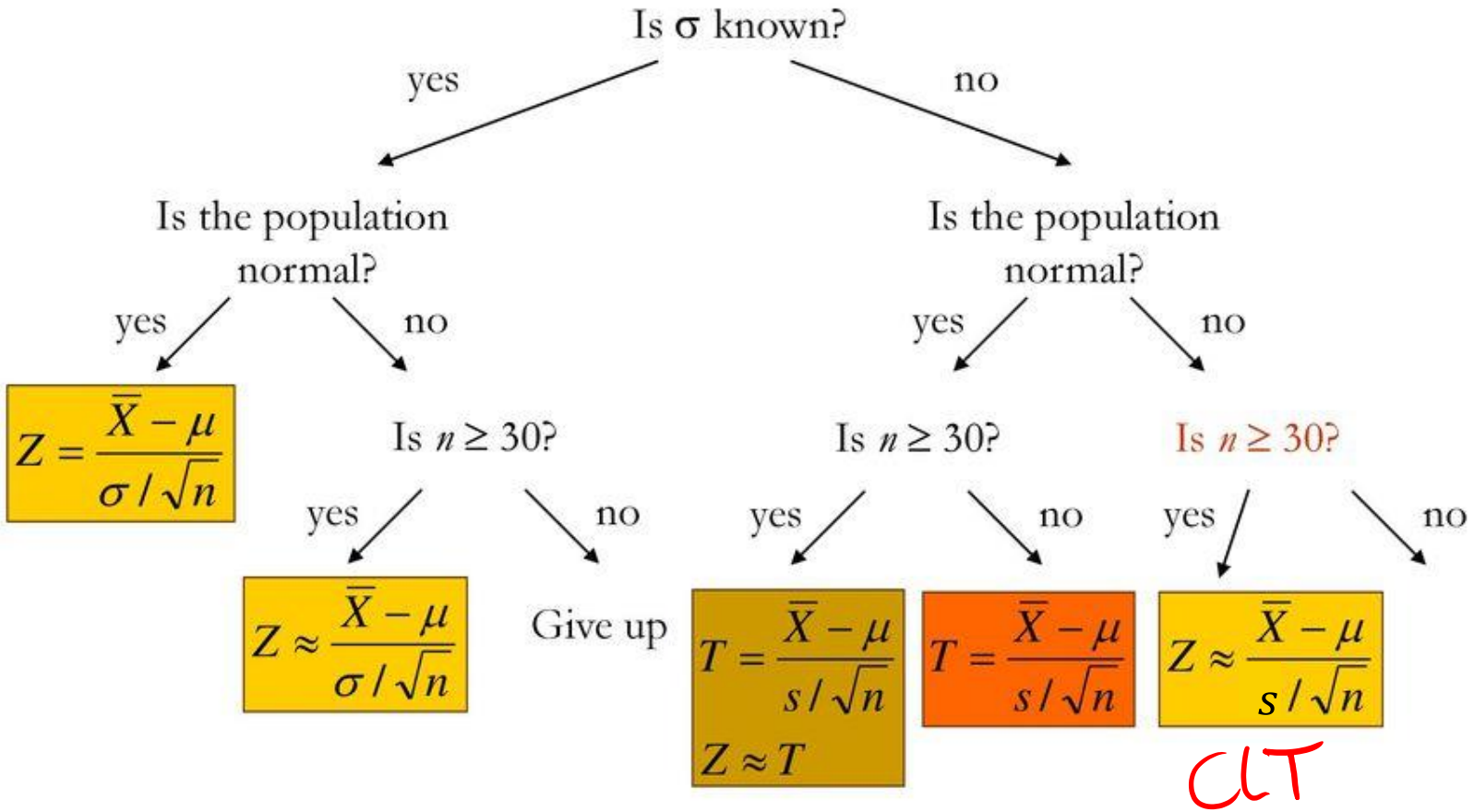
The Decision Tree



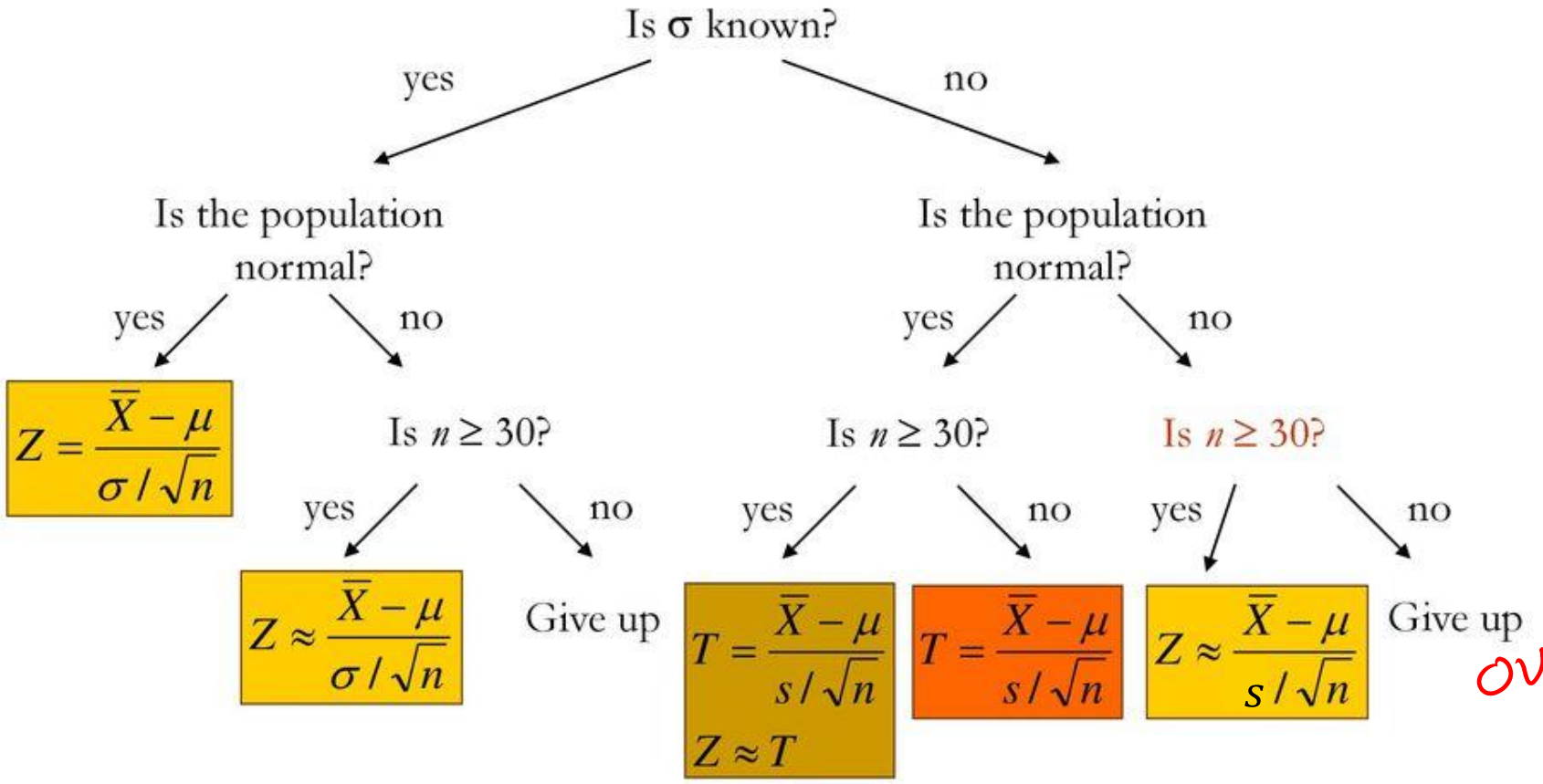
The Decision Tree



The Decision Tree

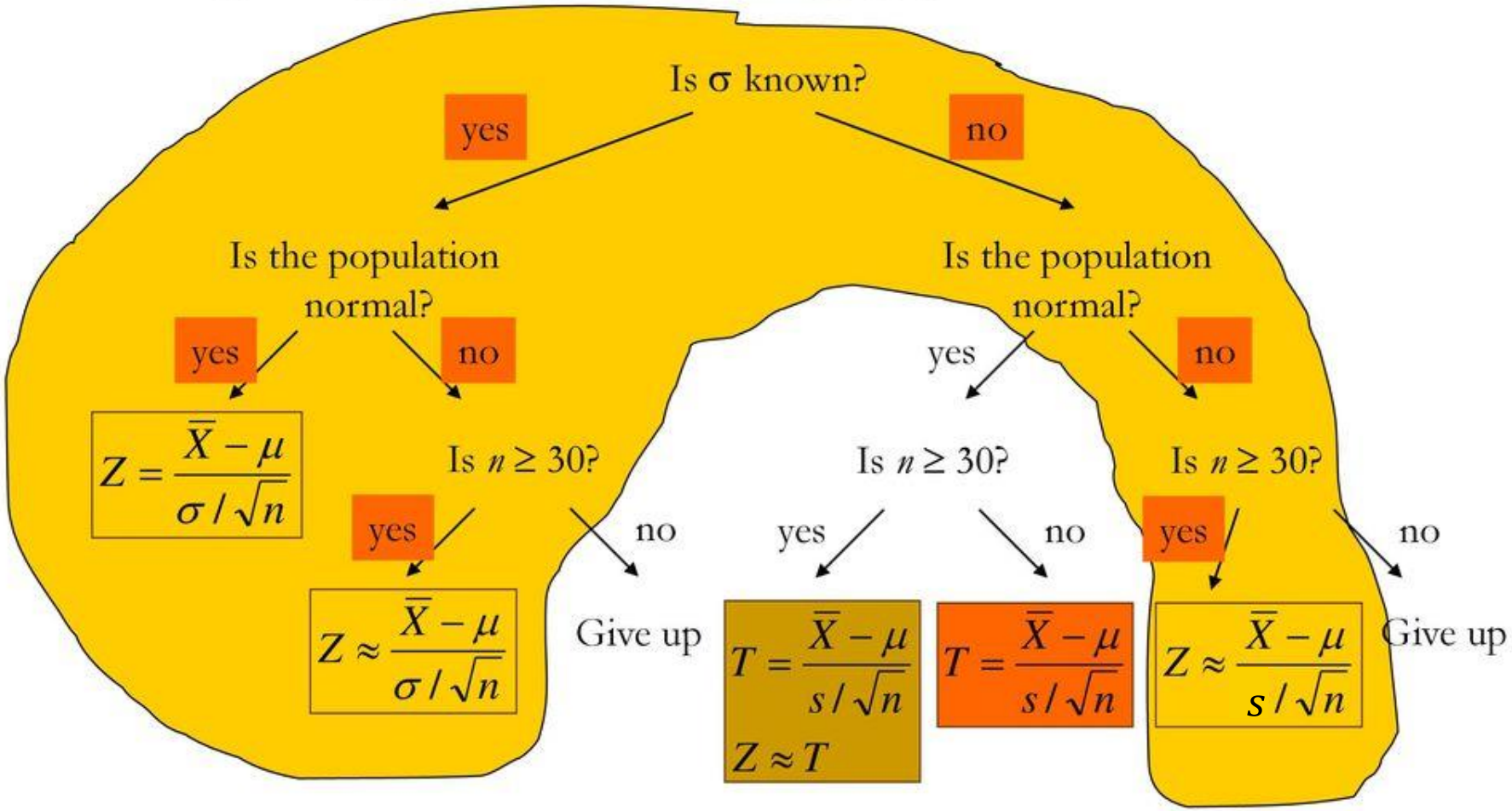


The Decision Tree

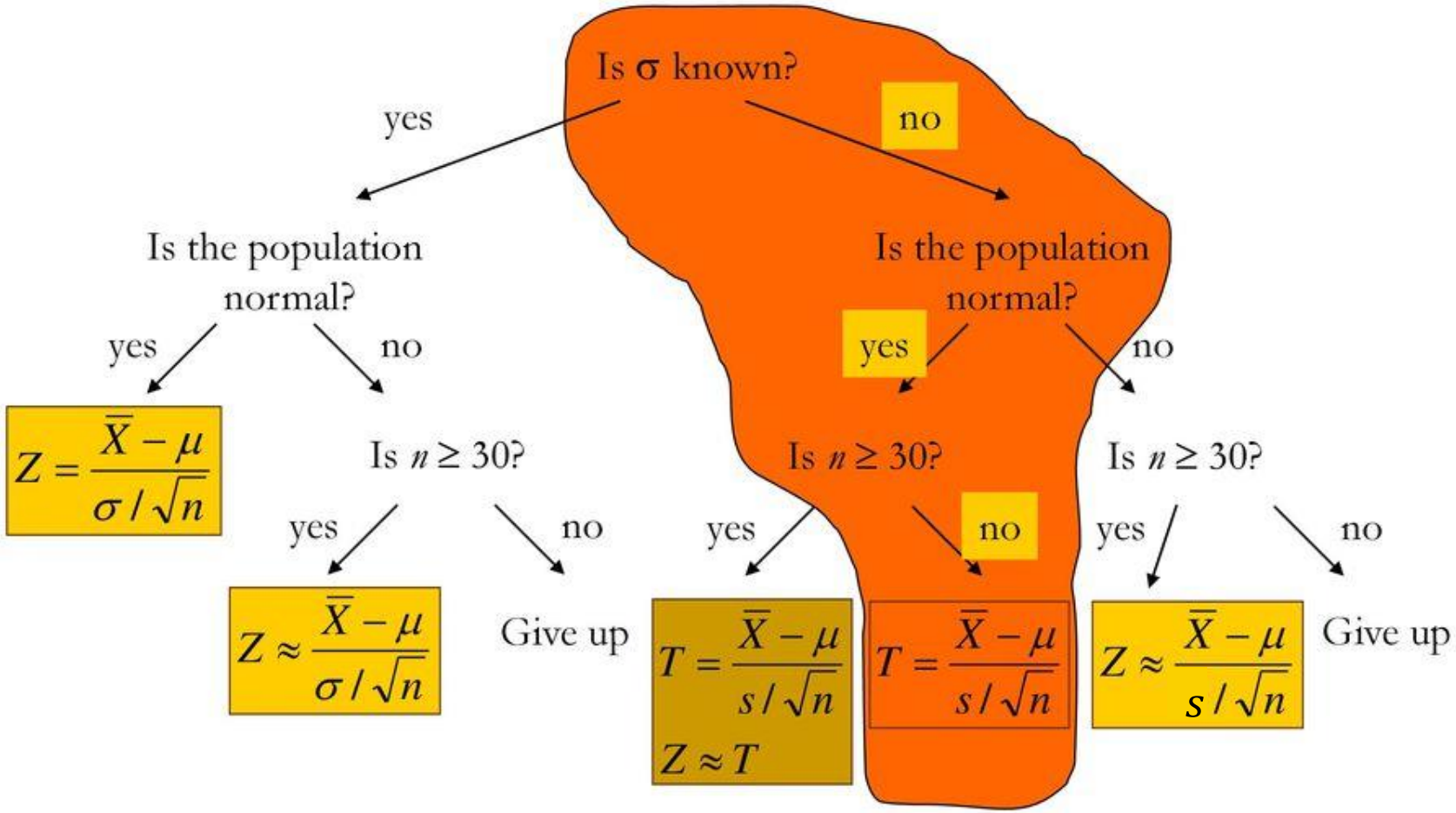


out of scope

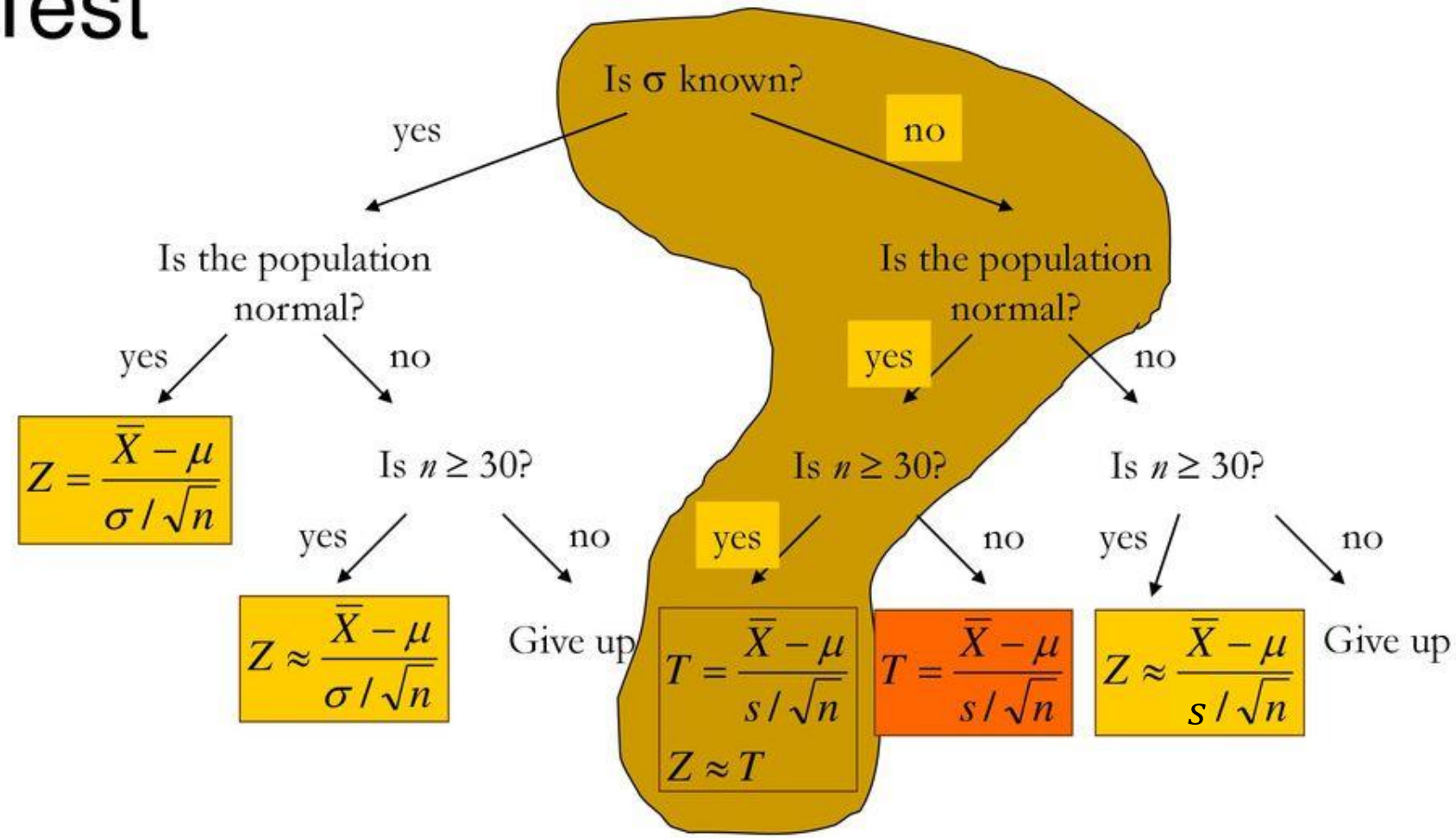
When to Use Z-Test



When to Use T-Test

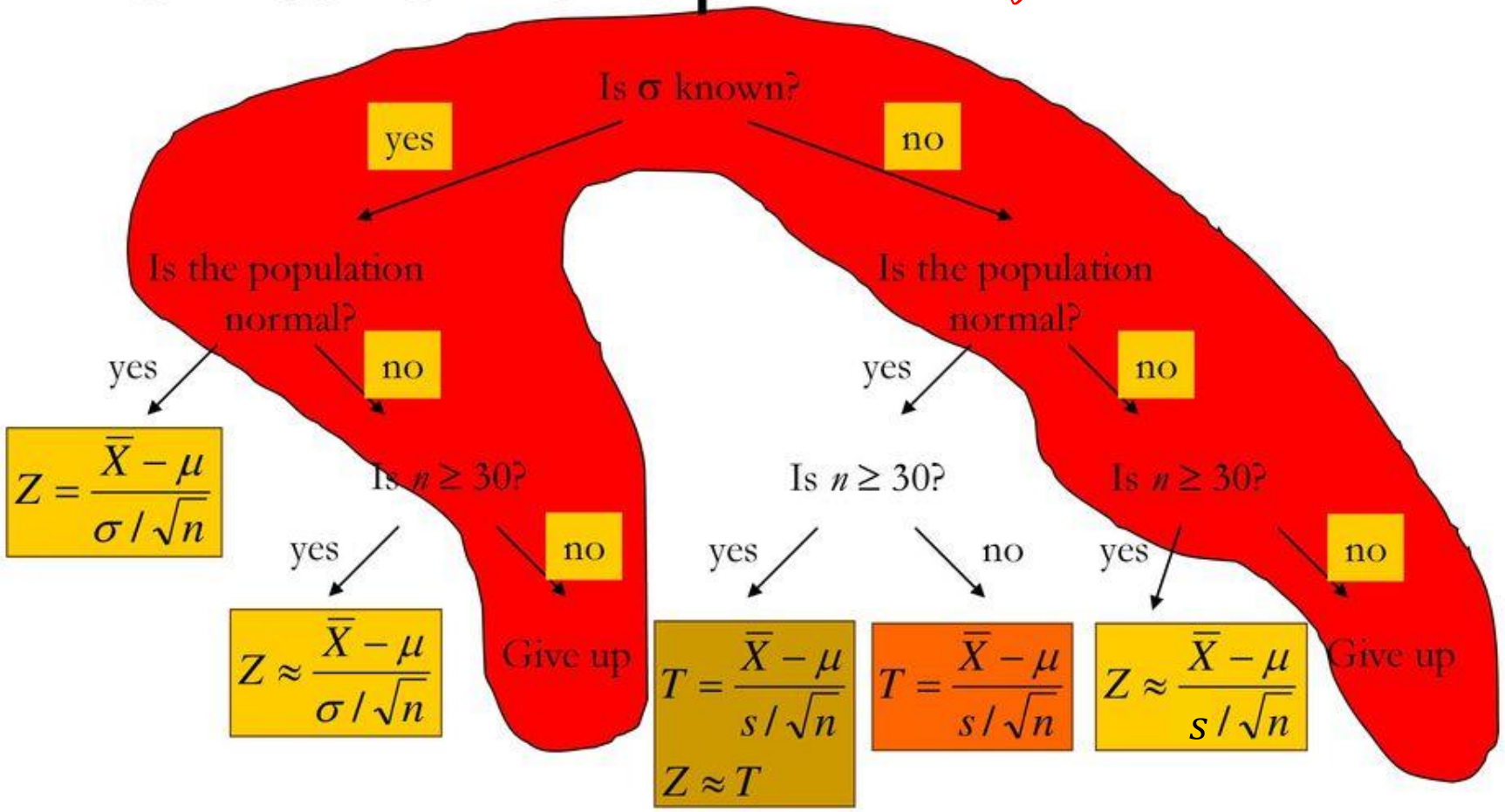


When to Use Either Z-Test or T-Test



When to Give Up

out of scope



3. The F-distribution

F - Distribution

- $V \sim \chi^2_{k_1}$ $W \sim \chi^2_{k_2}$ V, W independent
- $F = \frac{V/k_1}{W/k_2}$ → $\boldsymbol{F}$ distribution with ($\boldsymbol{k_1}$, $\boldsymbol{k_2}$) degrees of freedom
- Application ($\{X_i\}$ and $\{Y_j\}$ independent):

$$X_1, \ldots X_{n_1} \sim iid\ N(\mu_1, \sigma_1^2) \quad Y_1, \ldots Y_{n_2} \sim iid\ N(\mu_2, \sigma_2^2)$$

- $V = \frac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2_{n_1-1}$ $W = \frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2_{n_2-1}$ V, W independent
- $F = \frac{V/k_1}{W/k_2} = \frac{\frac{(n_1-1)S_1^2}{\sigma_1^2}/(n_1-1)}{\frac{(n_2-1)S_2^2}{\sigma_2^2}/n_2-1} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$

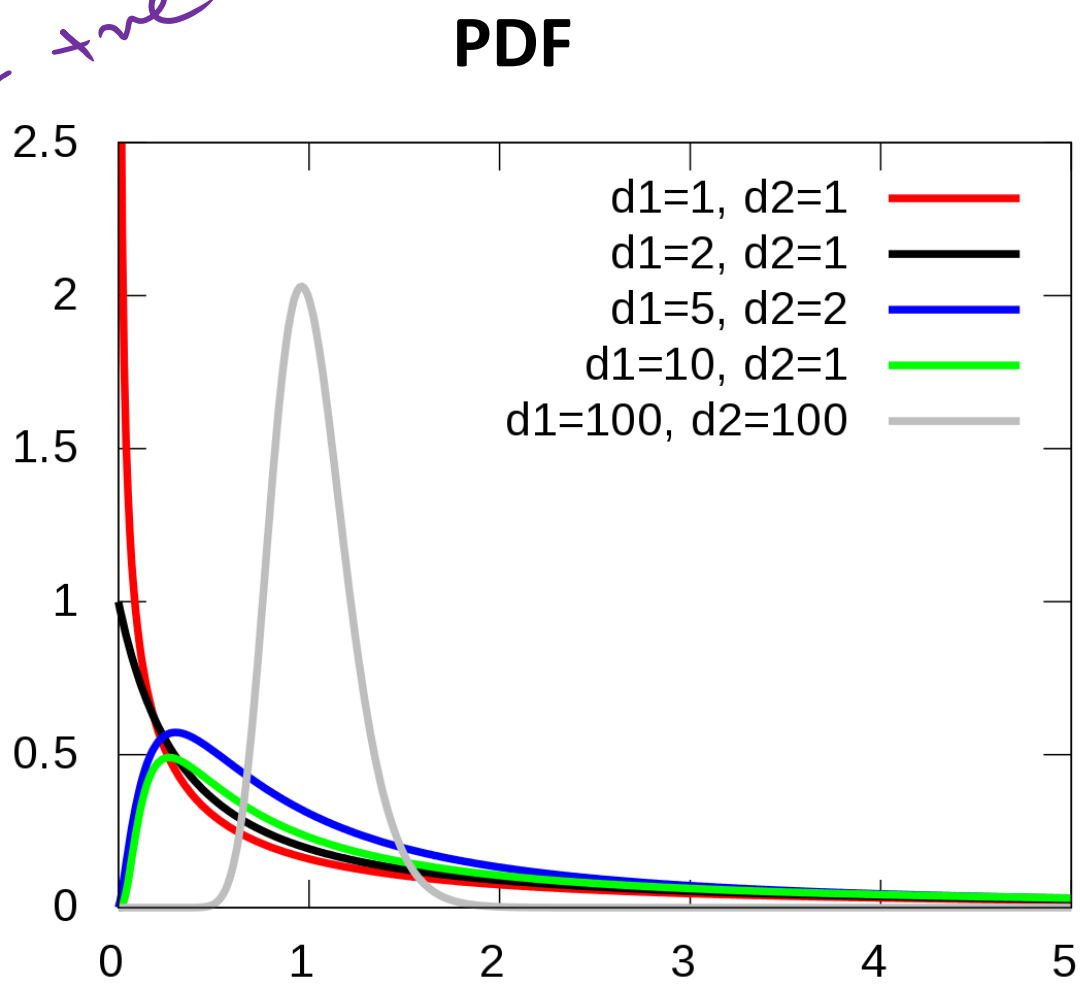
$\boldsymbol{F}$ distribution with $(\boldsymbol{n_1 - 1}, \boldsymbol{n_2 - 1})$ degrees of freedom

Compare Variance of two Population

F - Distribution

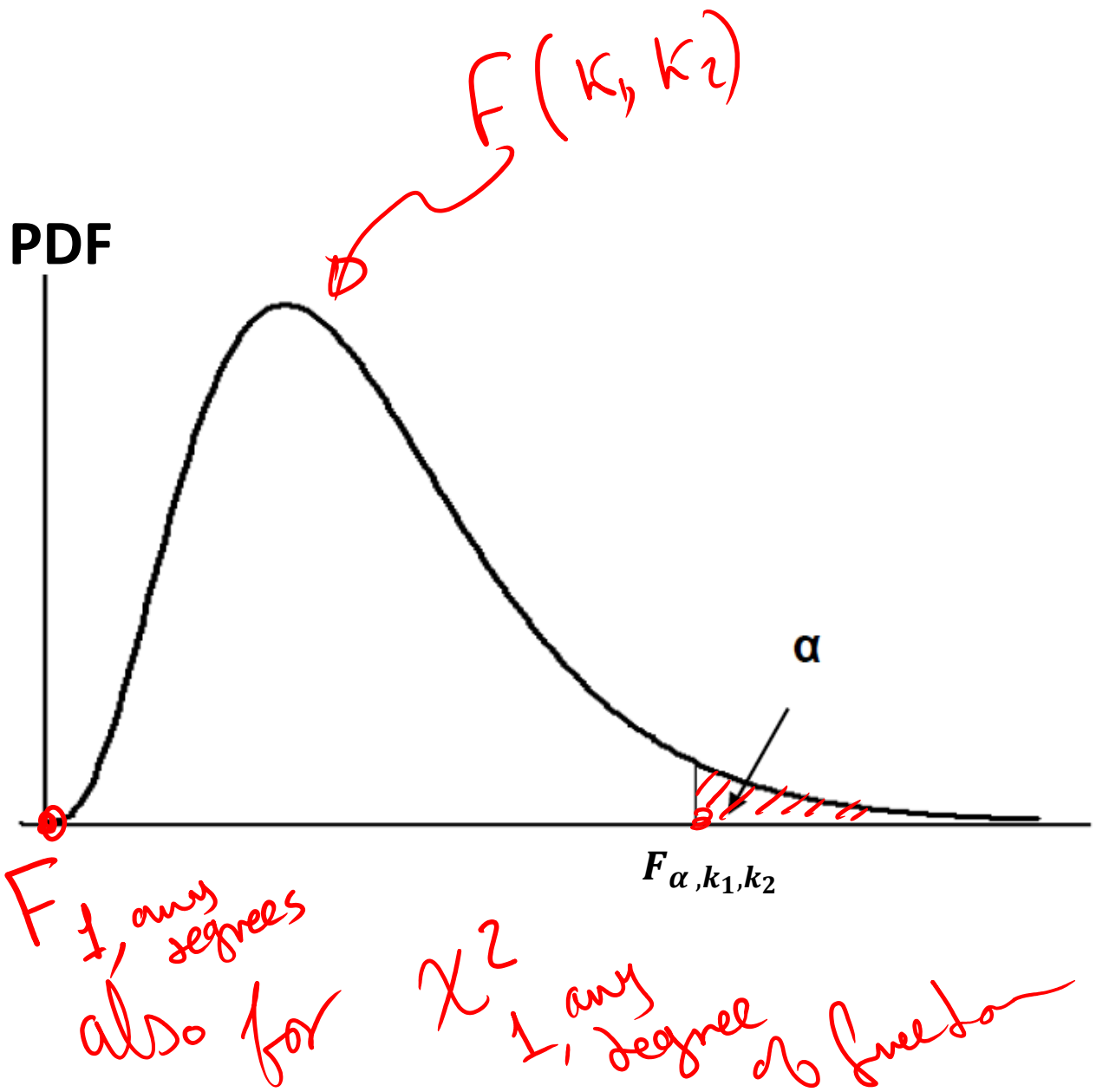
- No –ve value
- Not symmetric

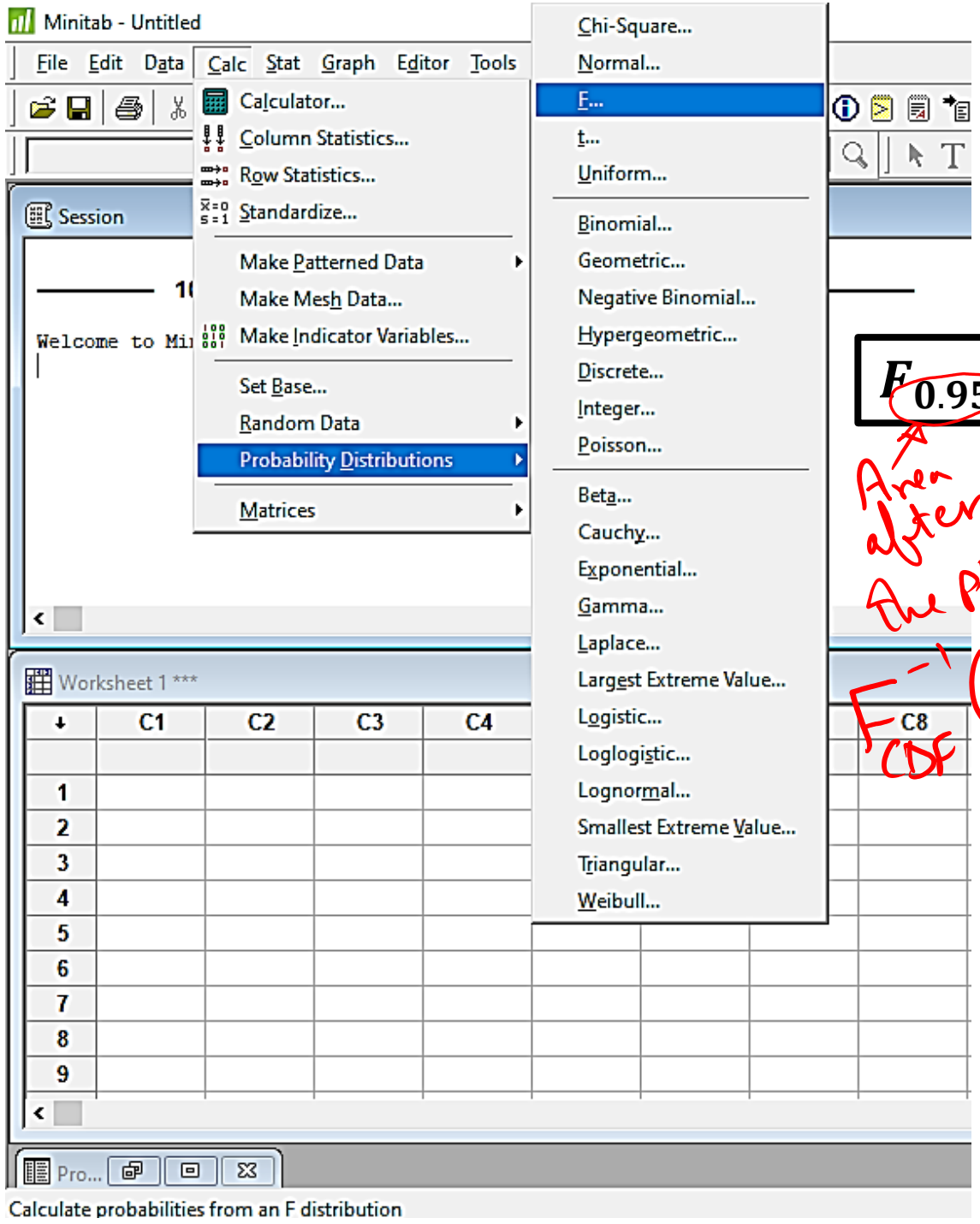
$\frac{\chi^2_{k_1}}{\chi^2_{k_2}}$ +ve +ve



F - Distribution

- F_{α,k_1,k_2}
- $P\left(U > F_{\alpha,k_1,k_2}\right) = \alpha$
- $F_{CDF}\left(F_{\alpha,k_1,k_2}\right) = 1 - \alpha$
- $F_{\alpha,k_1,k_2} = F_{CDF}^{-1}(1-\alpha)$
- $F_{\alpha,k_1,k_2} = ?$
- Using MINITAB or R



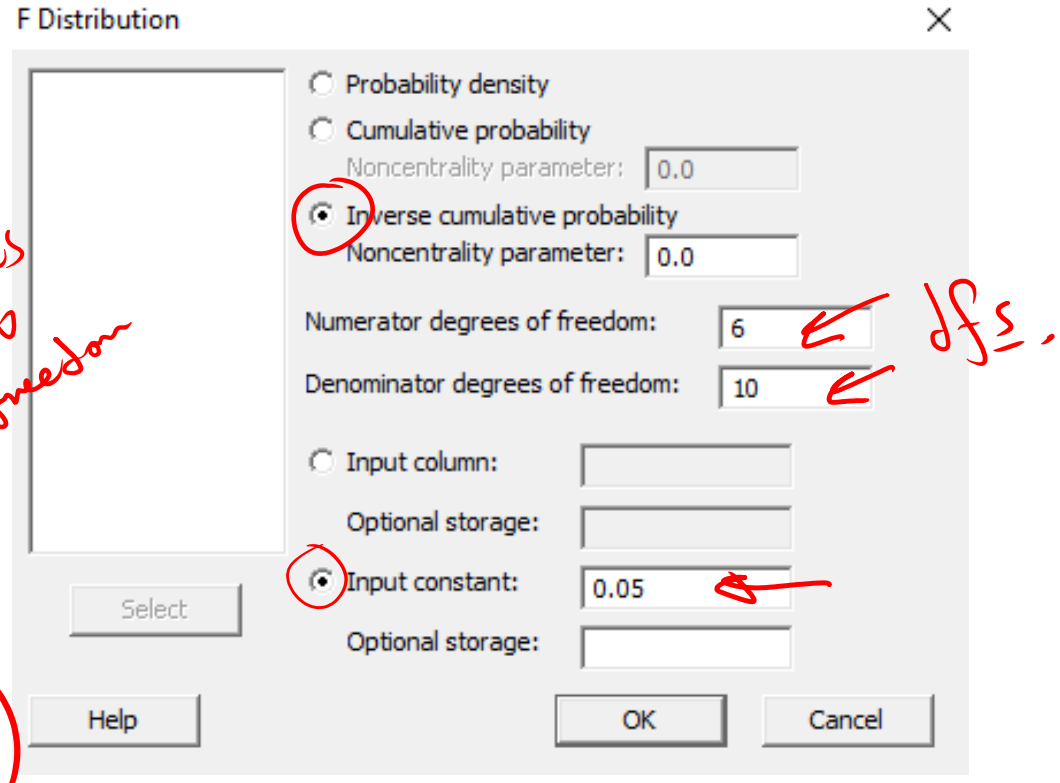


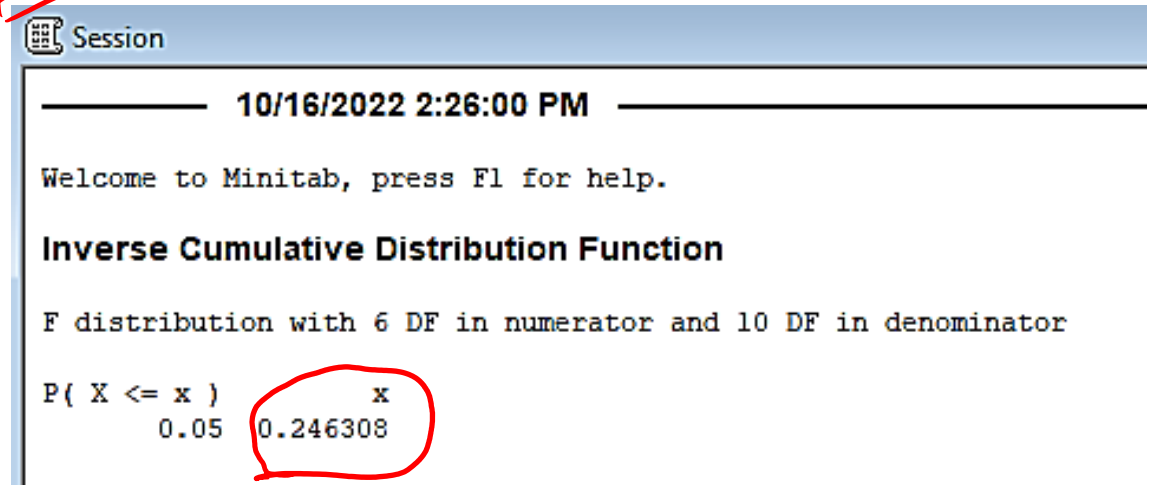
$F_{0.95,\,6,\,10}$

degrees of freedom

Area after the pt.

$F^{-1}_{CDF}(0.05)$ $1-0.95$





Ex: Consider the four independent random variables X, Y, U and V such that X ~ N(0,16), Y ~ N(5,4) , U ~ $\chi^2(4)$ and V ~ $\chi^2(16)$.
State the distribution of each of the following variables

X+2Y

$E(X+2Y) = E(X) + 2E(Y) = 0 + 2*5 = 10$

$V(X+2Y) = V(X) + 4V(Y) = 16 + 4(4) = 32$

$\sim N(10, 32)$

2X−Y

$E(2X-Y) = 2E(X) - E(Y) = 2(0) - (5) = -5$

$V(2X-Y) = 4V(X) + V(Y) = 4(16) + (4) = 68$

$\sim N(-5, 68)$

A Linear Combination of Normally Distributed Random Variables

Let $X_1, X_2, \ldots X_n$ be independent, normally distributed random variables with expected values $\mu_1, \mu_2 \ldots \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \ldots \sigma_n^2$ respectively and let $a_1, a_2, \ldots a_n$ be constants.

$$\sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

$E(aX+bY+C) = aE(X) + bE(Y) + C$

$V(aX+bY+C) = a^2 V(X) + b^2 V(Y)$

Ex: Consider the four independent random variables X, Y, U and V such that X ~ N(0,16), Y ~ N(5,4) , U ~ $\chi^2(4)$ and V ~ $\chi^2(16)$.
State the distribution of each of the following variables

Sum of the squares of indep. Std normal

$$\frac{X^2}{16} + \frac{(Y-5)^2}{4}$$

$\frac{X-0}{4}$ ~ Std. Normal

$Z_1^2 + Z_2^2$

$\frac{Y-5}{2}$

$\sim \chi^2(2)$

$$\frac{X}{\sqrt{V}}$$

should be Std. normal

should be divided by its degree of freedom

$$\frac{\frac{X-0}{4}}{\sqrt{V/16}} = \frac{X/4}{\sqrt{V}/4} = \frac{X}{\sqrt{V}}$$

$\sim t(16)$ (تقريبا)

$$\frac{4U}{V}$$

$$\frac{U/4}{V/16} \sim F(4,16)$$

$$\frac{4U}{V} \sim F(4,16)$$