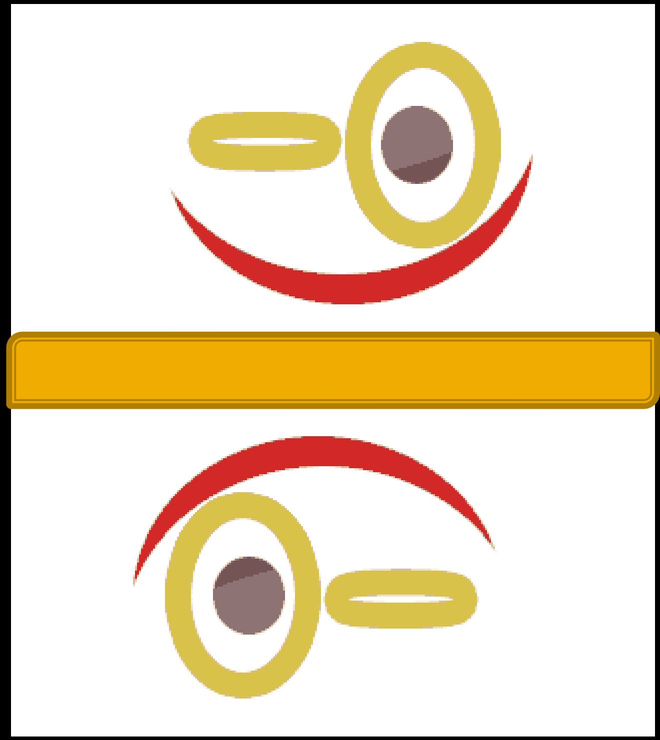


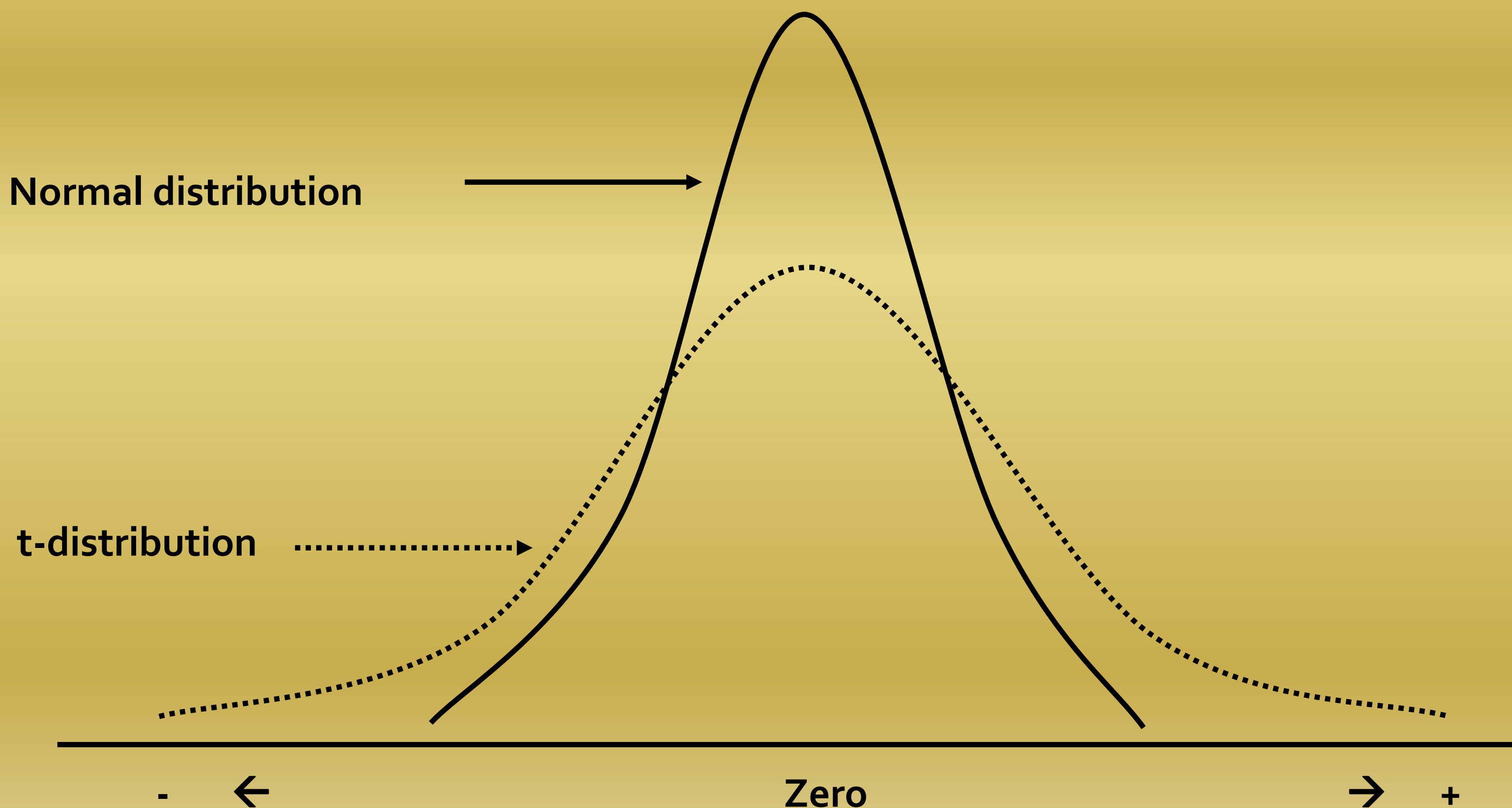
Test of significance:
t-test

t-test;

- **It is one of the commonly used tests for testing hypothesis for testing the significance of difference for the quantitative data. It depends on a distribution called the t-distribution with (n-1) degree of freedom. This distribution was introduced by William S Gosset, who used the pen-name "student" and is often called students't-test in his honor (or t-test)**

t-test;

- **t-distribution is like the normal distribution of a symmetrical bell-shaped distribution with a mean of zero but it is of lower peak, higher tails, and more spread out (more probability in the tails and less in the center), having two tails**


Normal distribution
t-distribution
-
Zero
+

t-distribution

The exact shape of the t-distribution depends on the degree of freedom (d.f.= n-1), and on the sample SD, the fewer the degrees of freedom, the more the t-distribution is spread out

t-test use is restricted to the small sample size (less than 30)

$$t = \frac{\text{Difference between two means}}{\text{Standard error of difference}}$$

t-test represent the measurement of the significancy of difference between two means;

$$t = \frac{\text{Difference between two means}}{\text{Standard error of difference}}$$

Applications of t-test;

1. For calculation of population mean.
2. For calculation of significance of difference between sample mean and population mean.
3. For calculation of significance of difference between two independent means.
4. For calculation of significance of difference between two dependent means (paired observations).

1-Calculation of population mean:

In general, a confidence interval "C.I." (Population mean; μ) is calculated using t distribution through appropriate significance level ($\alpha=0.05$ for 95% C.I., $\alpha=0.01$ for 99% C.I.) with (n-1) degrees of freedom

This is applied for small sample size (n <30) because for large degrees of freedom, the t-distribution is almost the same as the standard normal distribution

e.g.

The followings are the numbers of hours of relief obtained by 6 patients after receiving a new drug;

2.2, 2.4, 4.9, 2.5, 3.7, & 4.3 hours

Mean = 3.3 hours
SD= ±1.3 hours
n= 6
Calculate population mean? (Using α=0.05).

Calculation of μ

Tabulated t for α=0.05, for d.f. (n-1) is;

$$t^{\alpha=0.05}_{df.=n-1=6-1=5} = 2.571$$

$$\mu = \text{mean} \pm t^{\alpha}_{d.f.} \times SE$$

$$\mu = 3.3 \pm 2.57 \times (1.13/\sqrt{6})$$

$$\mu = 3.3 \pm 2.57 \times 0.46 = 3.3 \pm 1.2$$

Lower limit = 3.3-1.2→2.1 hours

Upper limit = 3.3+1.2→4.5 hours

$$\mu \rightarrow (\ 2.1 \ \text{---}\ 4.5\ )\ \text{hours}$$

$$\mu = \text{mean} \pm t^{\alpha}_{d.f.} \times SE$$

2-Difference between sample mean and population mean;

The t-test is used to test the significance of the difference of sample mean from a population mean or a standard mean or a standard value

e.g.

The followings are the heights in cm of 24 two-years-old boys with sickle cell disease

84.4	89.9	89.0	81.9	87.0	78.5	84.1	86.3
80.6	80.0	81.3	86.8	83.4	89.8	85.4	80.6
85.0	82.5	80.7	84.3	85.4	85.0	85.5	81.9

Height standard for United Kingdom (U.K.) give a reference height for two-years-old boys of 86.5 cm (represent μ).

Does the above sample suggest that the two-years-old boys with sickle cell (SC) disease differ in height from the standards? (use α=0.05).

=Data;

Data represent the heights of 24 two-years-old boys with SC disease, with;

Mean = 84.1 cm

SD= ±3.11 cm

SE=SD/√n = 3.11/√24

=3.11/4.9=

±0.63 cm

$$SE = \frac{SD}{\sqrt{n}} = \frac{3.11}{\sqrt{24}} = \frac{3.11}{4.9} = \pm 0.63 \text{ cm}$$

2=Assumptions;

We assume that the sample of 24 two-years-old boys with SC disease was selected randomly from a normally distributed population of boys with SC disease

Null hypothesis (H_o);

H_o: There is no significant difference between mean height of boys with SC disease from the normal standard height
($m1=\mu$; $m1-\mu=0$)

OR There is no significant influence (effect) of SC disease on the height of children

Alternative hypothesis (H_A);

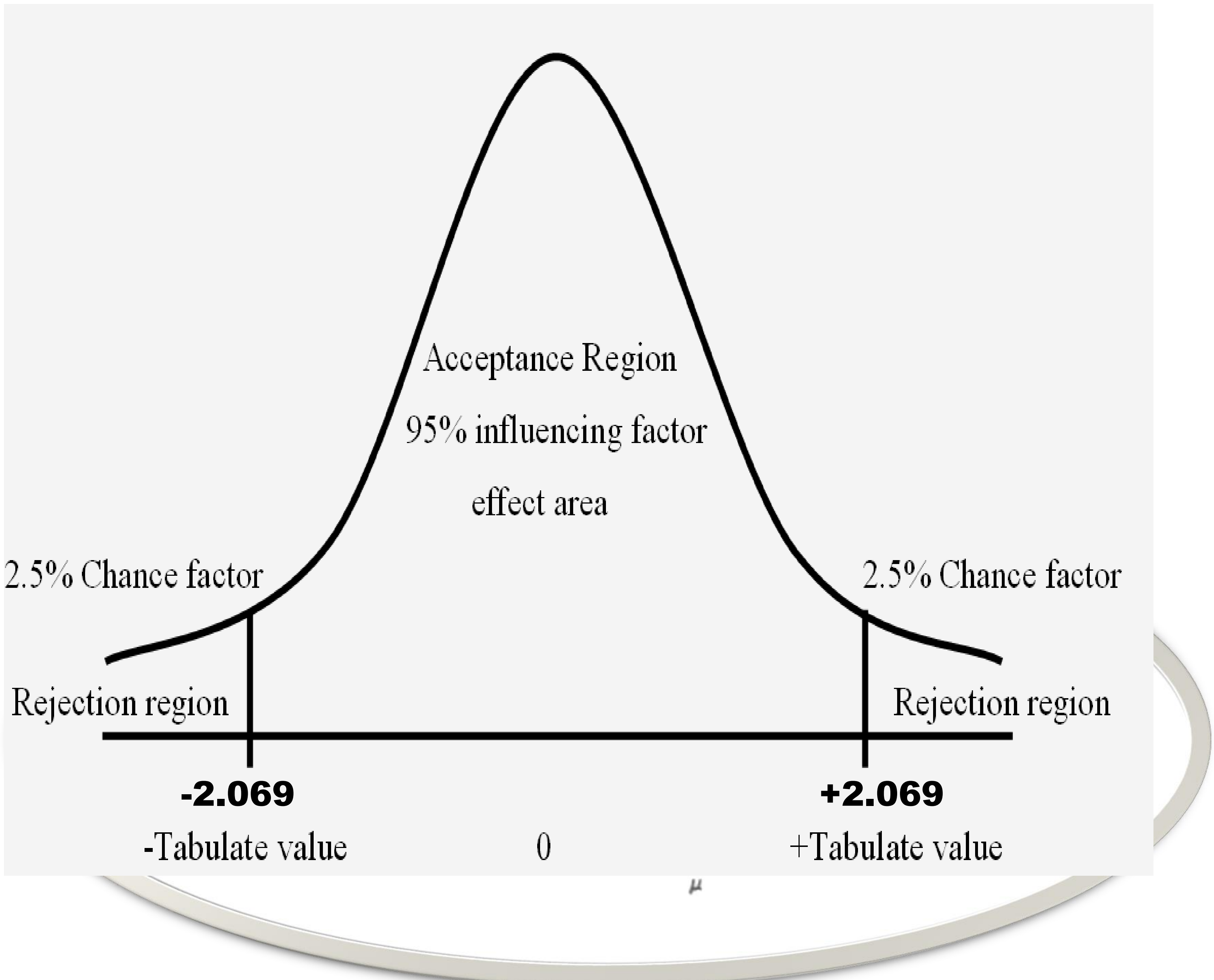
H_A: There is significant difference between mean height of boys with SC disease from the normal standard height
($m1 \neq \mu$; $m1 - \mu \neq 0$)

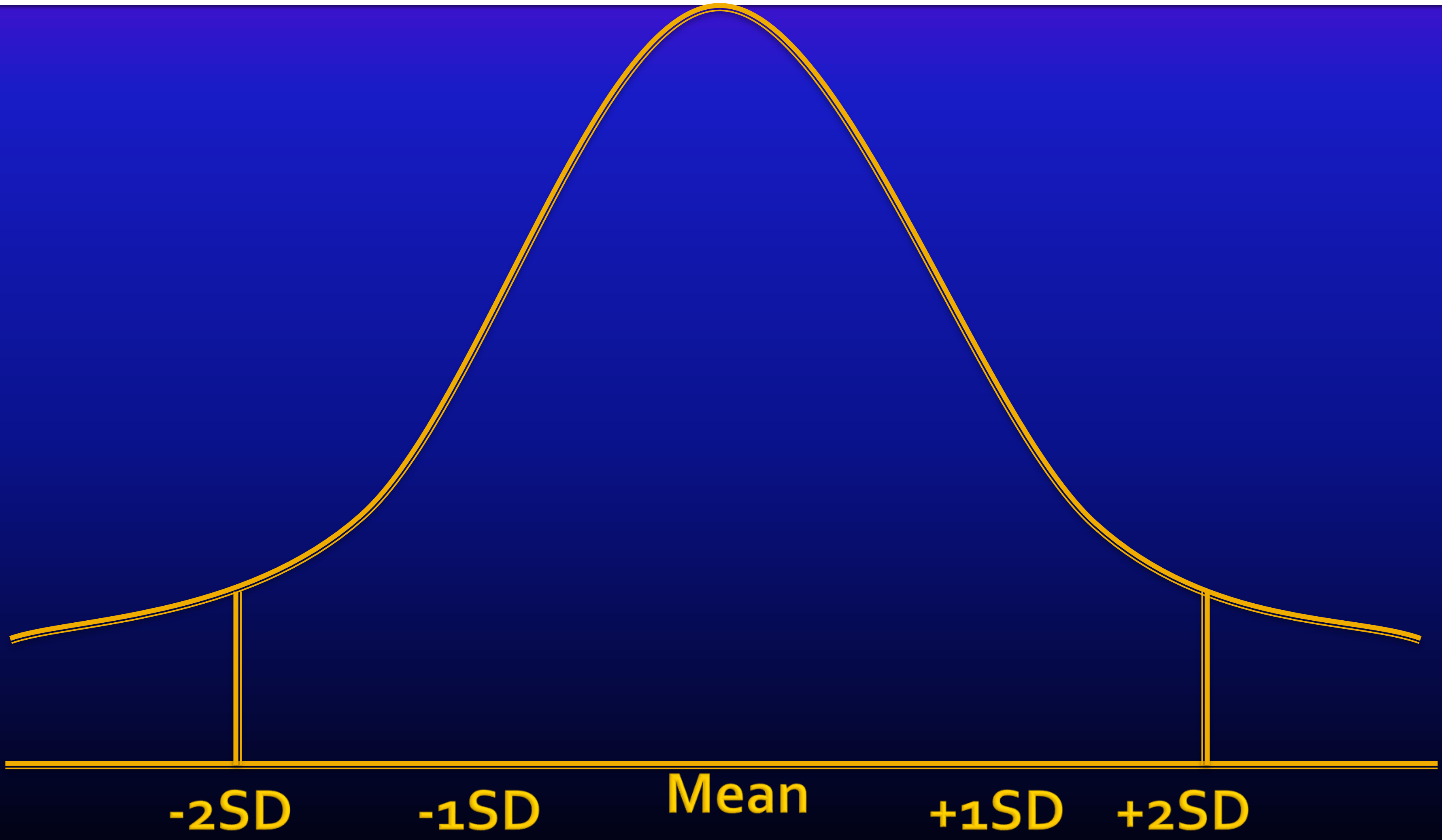
OR There is significant influence (effect) of SC disease on the height of children

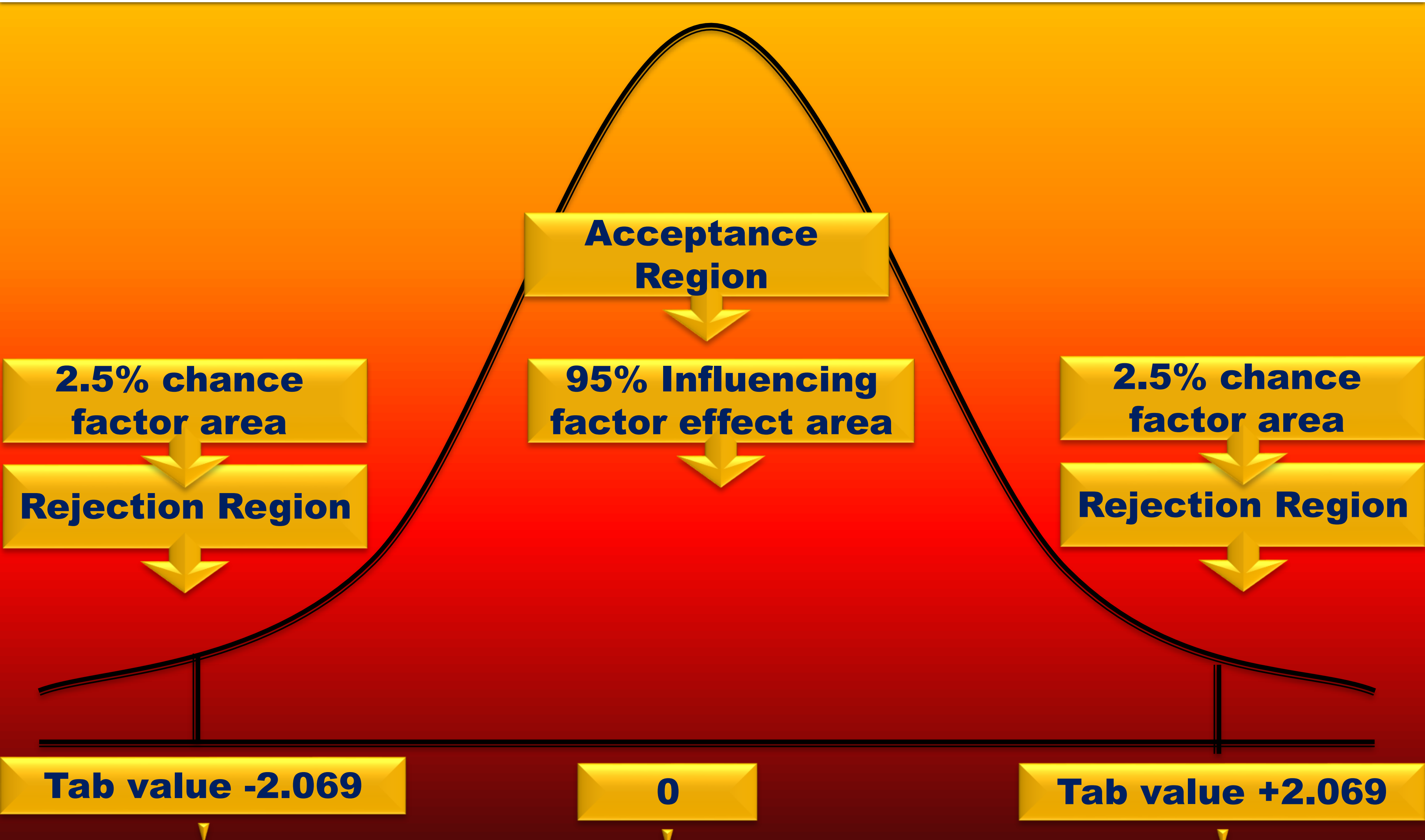
=Level of significance;

Level of significance; (α = 0.05);
5% Chance factor effect area
95% Influencing factor effect area (SC disease)

d.f.=n-1;
tabulated t for d.f {(n-l)=24-1=23} at α 0.05 equal to 2.069


Acceptance Region
95% influencing factor
effect area
2.5% Chance factor
2.5% Chance factor
Rejection region
Rejection region
-2.069
+2.069
-Tabulate value
0
+Tabulate value
μ


-2SD
-1SD
Mean
+1SD
+2SD



Acceptance Region
2.5% chance factor area
Rejection Region
95% Influencing factor effect area
2.5% chance factor area
Rejection Region
Tab value -2.069
0
Tab value +2.069

=Apply the proper test of significance;

Apply the proper test of significance

$$t = \frac{\text{Difference between two means}}{\text{Standard error of difference}}$$

$$t = \frac{m - \mu}{\text{standard error of sample}}$$

$$= \frac{84.1 - 86.5}{0.63} = -3.81 \text{ (calculated t value)}$$

3.81 > 2.069

=Statistical decision;

Since Calculated t > Tabulated t

So $P<0.05$

Then reject Ho and accept HA ….

There is significant difference between mean height of boys with SC disease from the normal standard height

There is significant influence (effect) of SC disease on the height of children

Significantly SC disease lowering height of children

→ There is stunting of height due to the effect of SC disease.

=Statistical decision;

$\alpha=0.05$	$\alpha=0.02$	$\alpha=0.01$	$\alpha=0.001$
$t = 2.069$	$t = 2.50$	$t = 2.81$	$t = 3.77$
df.=23	df.=23	df.=23	df.=23

P value < 0.05 <0.02 <0.01 <0.001

Highly significant effect

3-Difference between two independent means

1. **Assumption; we assume that we have two independent samples randomly chosen each one from a normally distributed population with equal variances of populations.**
2. **Equation… as the SE of difference is calculated as; Standard deviation of population (pooled SD) → SP**

$$SP = \sqrt{\frac{S_1^2\,(n1 - 1) + S_2^2\,(n2 - 1)}{n_1 + n_2 - 2}}$$

S_1^2 → variance of group 1
S_2^2 → variance of group 2

$$\text{SE of difference} = SP\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

3. **The d.f. = $n_1 + n_2 - 2$ or $(n_1 - 1) + (n_2 - 1)$**

A study of birth weight of infants born to 15 non-smoker and 14 heavy smoker mothers (during pregnancy)

Birth weight (Kg) of infants born to non-smoker mothers (n=15)	Birth weight (Kg) of infants born to heavy smoker mothers (n-14)
3.99	3.18
3.79	2.84
3.60	2.90
3.73	3.27
3.21	3.85
3.60	3.52
4.08	3.23
3.61	2.76
3.83	3.60
3.31	3.75
4.13	3.59
3.26	3.63
3.54	2.38
3.51	2.34
2.71	

1=Data;

Data represent the birth weight in kilograms of two independent groups of smoker and non-smoker mothers infants with mean birth weight of non-smoker mother's infants of 3.5933, and of heavy smoker mother's infants of 3.2029 kilograms

2=Assumptions;

We assume that the two independent groups (of infants born to non-smoker mothers and those born to heavy smokers mothers) were randomly drawn each one form a normally distributed population with equal variances of populations

Null hypothesis (H_o);

There is no significant difference between mean birth weight of infants of non-smoker mothers and birth weight of infants of heavy smoker mothers ($m_1=m_2$; $m_1 - m_2=0$)

OR There is no significant influence (effect) of mother smoking during pregnancy on birth weight of their infants

Alternative hypothesis (H_A);

There is significant difference between mean birth weight of infants of non-smoker mothers and birth weight of infants of heavy smoker mothers ($m_1 \neq m_2$; $m_1 - m_2 \neq 0$)

OR There is significant influence (effect) of mother smoking during pregnancy on birth weight of their infants

▪Level of significance;

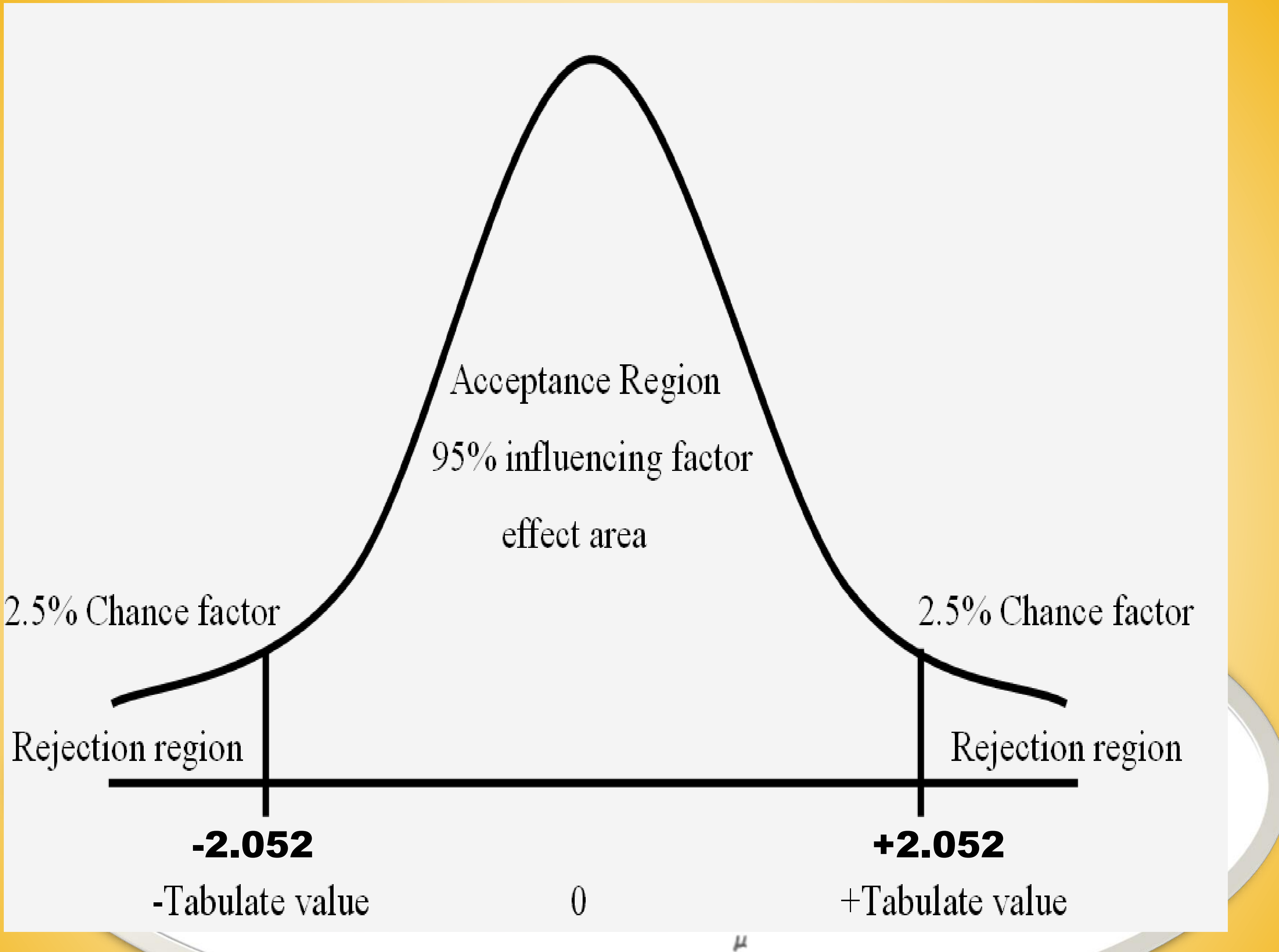
Level of significance; ($\alpha = 0.05$);

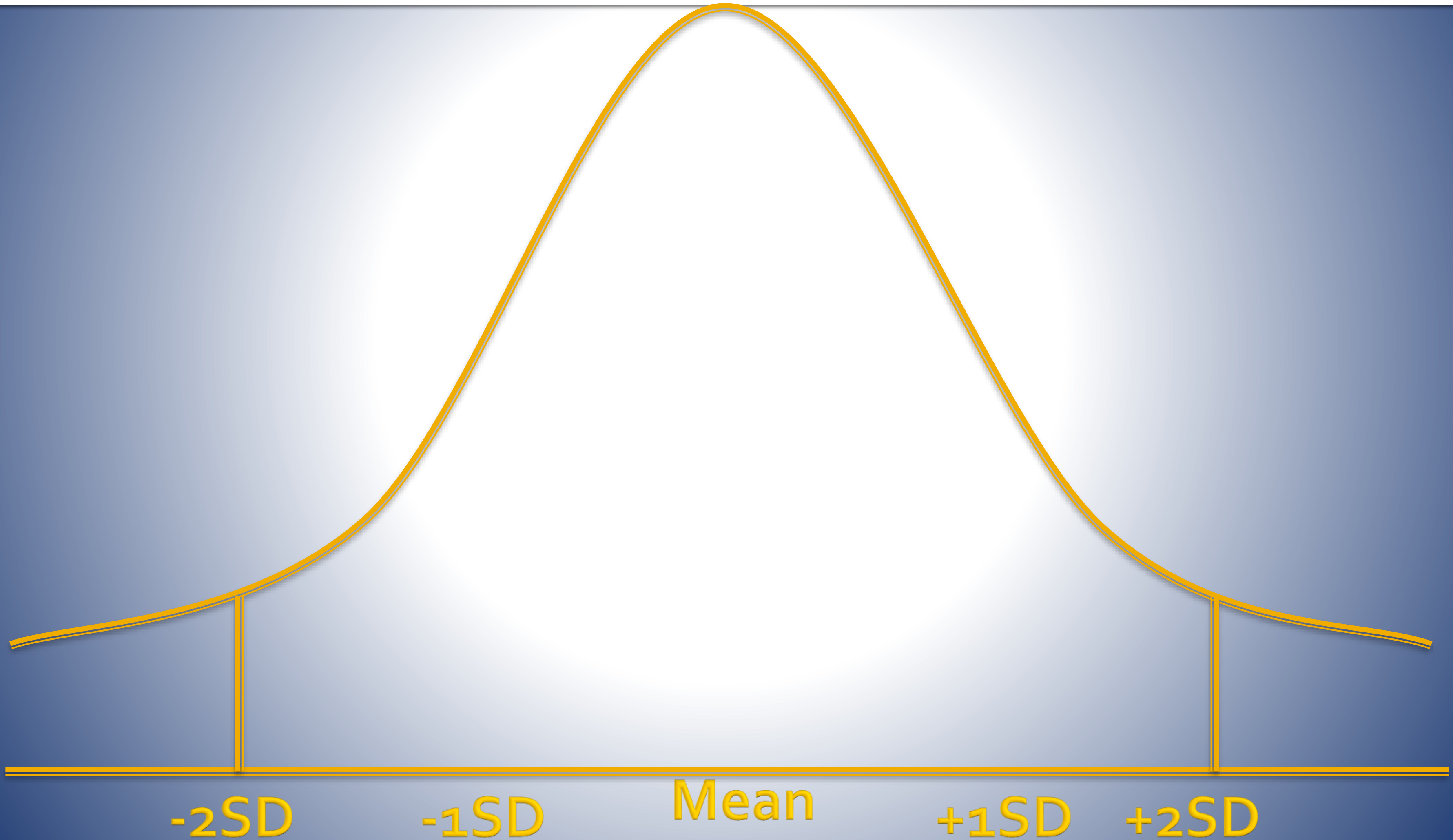
5% Chance factor effect area

95% Influencing factor effect area (Smoking during pregnancy)

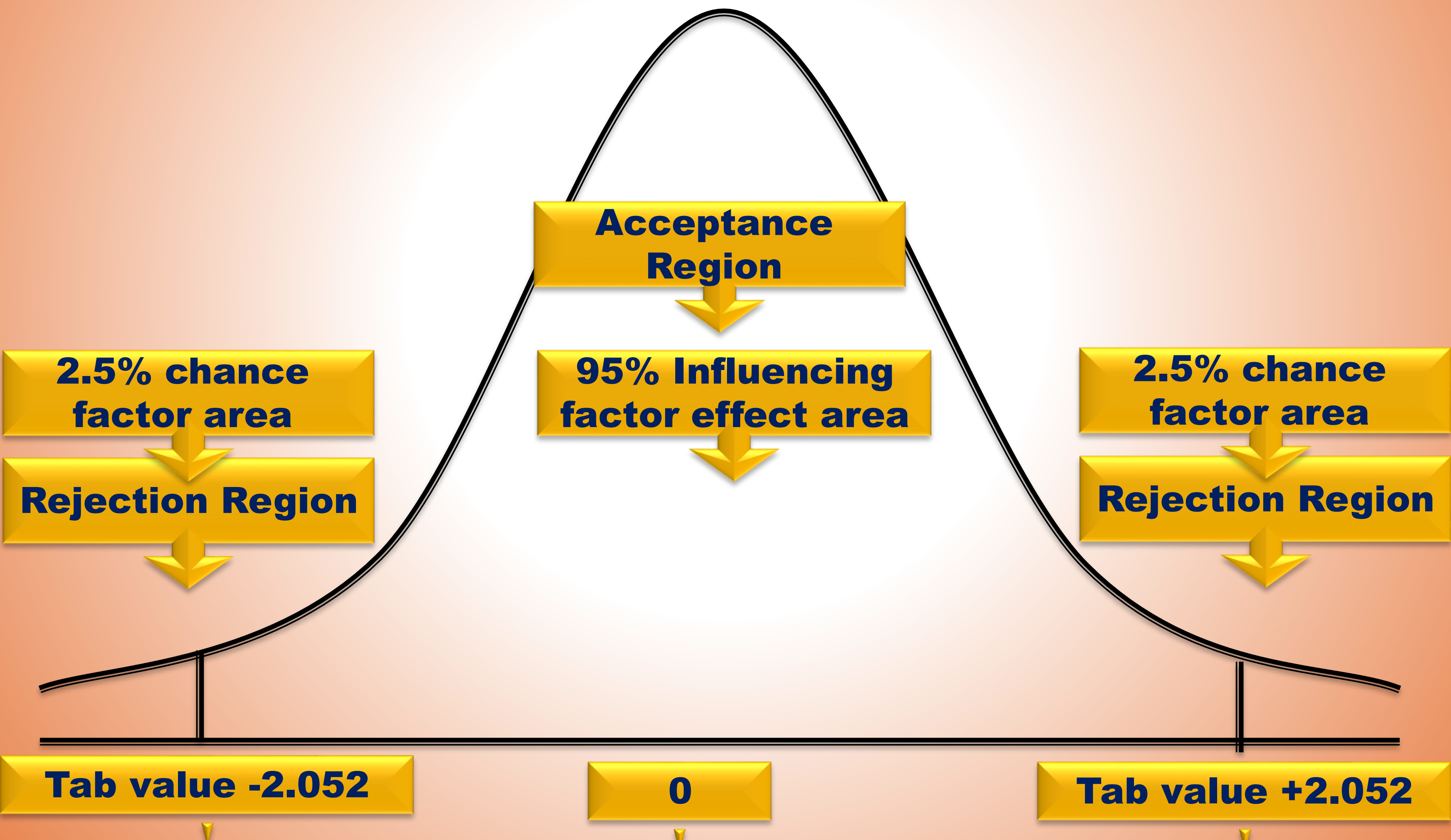
d.f.= $n_1 + n_2 - 2$ or $(n_1-1) + (n_2-1)$;

tabulated t for d.f ($n_1 + n_2 - 2$ = 15+14-2=27) equal to 2.052


Acceptance Region
95% influencing factor
effect area
2.5% Chance factor
2.5% Chance factor
Rejection region
Rejection region
-2.052
+2.052
-Tabulate value
0
+Tabulate value
μ



Mean
-2SD
-1SD
+1SD
+2SD



Acceptance Region
2.5% chance factor area
Rejection Region
95% Influencing factor effect area
2.5% chance factor area
Rejection Region
Tab value -2.052
0
Tab value +2.052

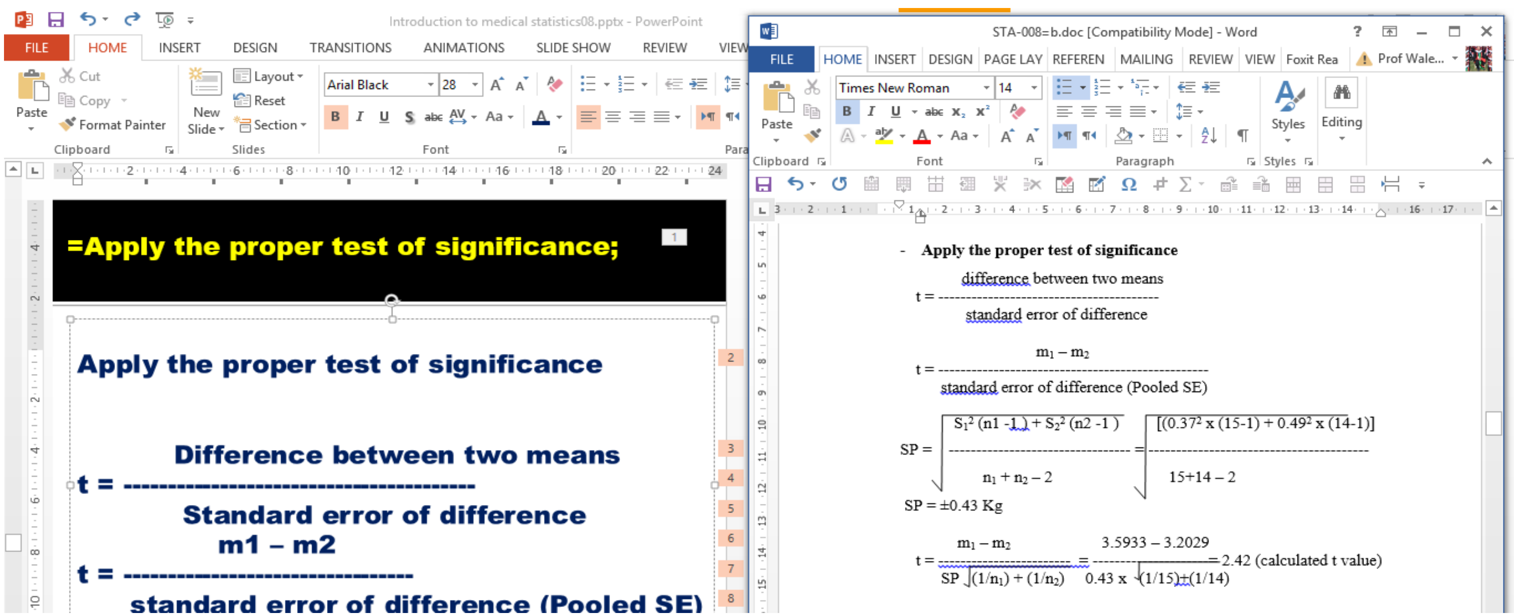
=Apply the proper test of significance;

Apply the proper test of significance

$$t = \frac{\text{Difference between two means}}{\text{Standard error of difference}}$$

$$t = \frac{m1 - m2}{\text{standard error of difference (Pooled SE)}}$$

=Apply the proper test of significance;



2.42 > 2.052

=Statistical decision;

Since Calculated t > Tabulated t

So $P<0.05$

Then reject Ho and accept HA ....

There is significant difference between mean birth weight of infants of non-smoker mothers and birth weight of infants of heavy smoker mothers ($m_1 \neq m_2$; $m_1 - m_2 \neq 0$)

OR There is significant influence (effect) of mother smoking during pregnancy on birth weight of their infants

Significantly smoking during pregnancy lowering the birth weight of infants

→ There is decreasing of weight due to the effect of smoking during pregnancy.

=Statistical decision;

$\alpha=0.05$	$\alpha=0.02$	$\alpha=0.01$	$\alpha=0.001$
$t = 2.052$	$t = 2.47$	$t = 2.77$	$t = 3.69$
df.=27	df.=27	df.=27	df.=27

P value < 0.05

Significant effect