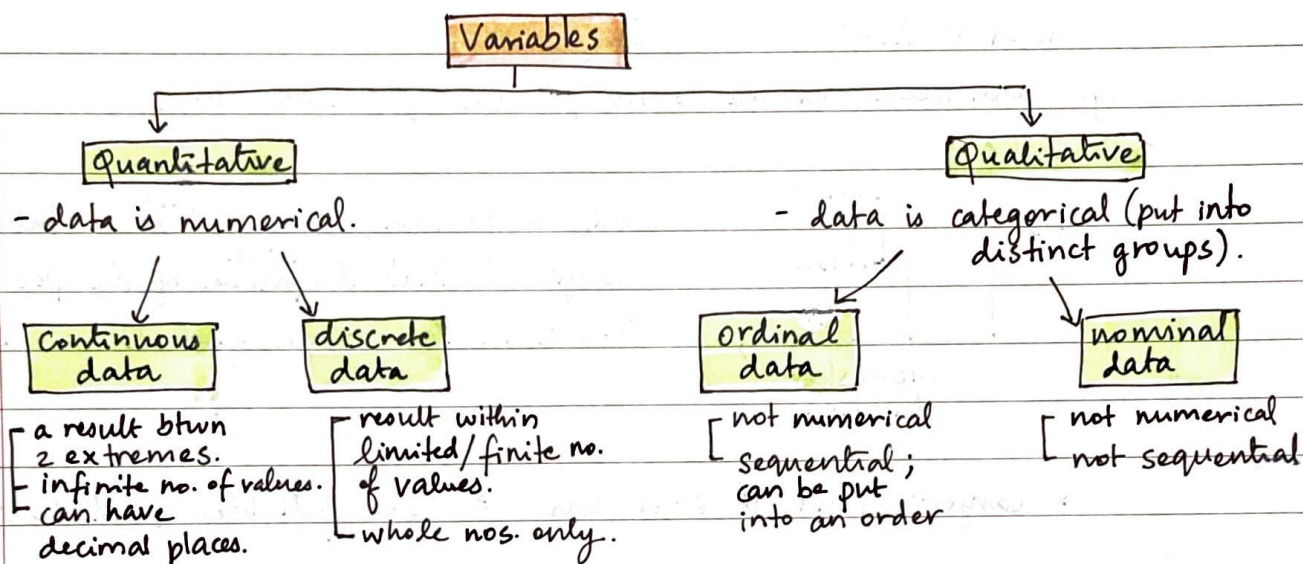


BIOLOGY - STATISTICS



Standard Deviation

- Tells us how data is spread out from its mean value.
- Involves data with normal distribution (a graph with a symmetrical, bell shape)

$$\rightarrow s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$$

- answers will be given in this form: mean $\pm$ S.D.
- tells us which data is more reliable.

/ \

more consistent less errors when calculating data
↳ less skewing of results ↳ better controlled/constant variables

Standard Error

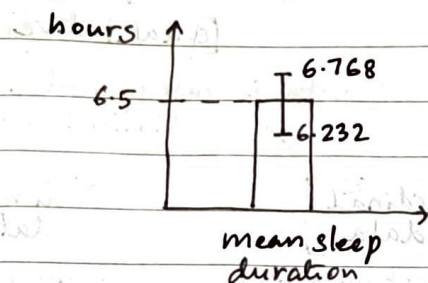
- Calculates how close the sample mean is to the true mean. (entire population's mean).

$$\rightarrow S.E = \frac{s}{\sqrt{n}}$$

→ calculate $2 \times S.E$

$$\text{mean} \pm 2S.E$$

eg. mean sleep duration (hours) = 6.5 ; $2S.E = 0.268$



This means that, we are 95% certain/
confident that the ^{true} mean of the population
is within this range of 6.232 - 6.768

* larger standard error bar $\Rightarrow$ less reliable (and vice versa)

What the error bars tell us

- 95% chance of the true mean being within that range.
- can be used to compare different means:
 - $\hookrightarrow$ if error bars overlap (for 2 diff means) $\Rightarrow$ the means are not significantly different. (and vice versa)

T-test

→ Statistical test to compare the means of 2 sets of data to see if they are significantly different.

→ when can it be done?

- $\hookrightarrow$ there are 2 sets of data
- $\hookrightarrow$ the data are continuous
- $\hookrightarrow$ the data are quantitative
- $\hookrightarrow$ the ^{data} population shows a normal distribution
- $\hookrightarrow$ 10 ~~more~~ or more data for each set of the data being compared

1) Construct null hypothesis

↳ No significant difference in mean ~~for~~ btwn the 2 data sets.

2) Calculate t value

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

* they can never be -ve.

3) Calculate the degree of freedom

$$(n_1 - 1) + (n_2 - 1)$$

4) Consult probability table given and find critical value

d.o.f	probability value			

0.10 prob $\Rightarrow$ 90% confident that the 2 means are sig. different,

0.001 prob $\Rightarrow$ 99.9% confident that the 2 means are sig. different, etc.

↓
- it is preferred to be atleast 95% confident, to conclude that the 2 means are sig. diff; cut-off point

- anything < 95% $\Rightarrow$ not significantly diff.

→ find the intersection of your calculated ~~diff~~ d.o.f and 0.05 (95%).
this is the critical value.

5) Compare t-value to critical value

→ if t-value < critical value : ~~we are not 95% confident that~~ the 2 means are not significantly different $\therefore$ accept H_0 .

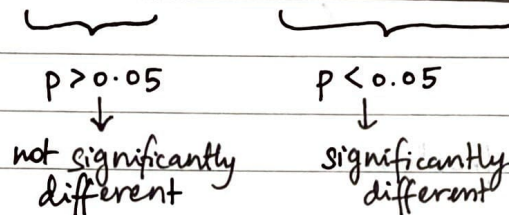
→ if t-value > critical value : more than 95% ~~are~~ confident that the 2 means are significantly different.

NOTE: if t-value > critical values at 99% or 99.9% as well, then more than 99.9% confident that the 2 means are significantly different.

the 2 means are significantly different $\therefore$ reject H_0

If only p value is given, and you can't find a t-value/critical:

probability value			
0.10	0.05	0.01	0.001



Chi-square test (χ^2)

→ statistical test to measure if the observed and expected values are significantly different.

→ when can it be done?

- you want to compare expected (predicted) & observed (actual) data.
- data are quantitative
- data are discrete

1) Construct null hypothesis

↳ observed and expected values are not significantly different.

2) χ^2 value

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

3) degree of freedom

$$n-1$$

4) Consult the table and find critical value

0.05 ⇒ 95% confident that expected and observed values are significantly different.

find critical value = intersection of d.o.f and 0.05

5) compare χ^2 value to critical value

- if $\chi^2 < \text{critical value}$

observed and expected not significantly diff. $\therefore$ accept H_0 .

- if $\chi^2 > \text{critical value}$

observed and expected significantly diff. $\therefore$ reject H_0 .

not sig diff

- difference due to chance.
- genes unlinked.

sig. diff

- difference not due to chance.
- genes linked; autosomal linkage.
- genes not on diff chromosomes and may be on the same one.
- they do not assort independently during meiosis.

Pearson's Linear Correlation

- To measure if there's a relationship btwn 2 variables.

- When it is used :

- want to see correlation btwn x & y .

- data are quantitative

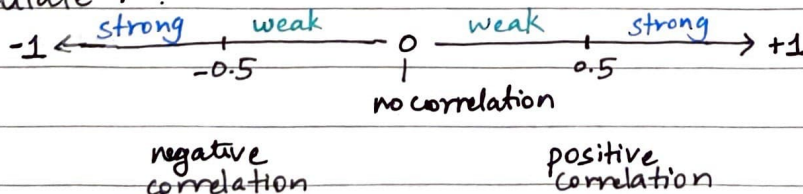
- data are continuous

- ↳ data shows normal distribution

- scatter graph shows a (+ve or) -ve linear correlation.

- at least 5 paired data.

- calculate 'r'.



- critical value: intersection of probability level and no. of data pairs

- ↳ if $r >$ critical value, more than 95% confident that X and Y are correlated. * when comparing r to c.v take $|r|$.

- ⇒ Significant correlation btwn X & Y.

Spearman's Rank Correlation

(same purpose as Pearson)

→ when it is used :

- the 2 data do not show a linear correlation on scatter graph.
- data do not show normal distribution.
- at least 5 paired data.
- data are quantitative and can be ranked (highest → lowest and vice versa)

→ when ranking, if 2 data are same, :

~~rank~~ eg. 15 and 15, ranks 5 and 6 :

both will be ranked as $\frac{5+6}{2} = 5.5$.

the next rank will be 7.

→ calculate 'r_s'. (correlation & strength same as for Pearson)

→ same process for critical value.