Field Studies Council

Data Analysis Enhancement Module Handout

Learning Objectives:

1. Be introduced to descriptive and inferential statistical analysis techniques
2. Practice descriptive analysis skills in an unfamiliar context
3. Choose statistical tests appropriate to unfamiliar data
4. Explain how to draw statistical conclusions

Background information:

Descriptive analysis: Uses data to provide descriptions of the population measured either through numerical calculations or graphs and tables.

Inferential statistics: Makes inferences and predictions about a population based on a sample of data taken from that population as a whole.

Key term – Learning wall:

Mean	Interquartile range	Standard deviation	Critical values
Median	Percentages & ratios	Statistical significance	Probability
Mode	Normal distribution	Null hypothesis	Chance

Percentages and Ratios:

	At base of Tree	2m away from base of	At edge of tree canopy
Pin hits on Nettles	30	10	1
Pin hits on Grass spp	6	42	75
Total pins dropped	80		
% Frequency of Nettles			
% Frequency of Grass			

Area of shore	Height of Limpet / mm	Diameter of Limpet / mm	Height to Diameter ratio
Sheltered	10.5	12.1	
Semi-sheltered	6.9	9.5	
Exposed	4.5	8.2	

What does the ratio tell us?

Measures of central tendency:

Central tendencies are a single value used to express the clustering of data around a central point.

Mean:

Median:

Mode:

Chose and calculate the most appropriate central tendency for each set of data:

a) Number of woodlice in dark section of choice chamber in minute intervals:

Time / min.	1	2	3	4	5	6	7	
No. of Woodlice (dark)	15	16	12	30	18	2	17	

a) Leaf diameter in a random sample of a trampled area.

Sample no.	1	2	3	4	5	6	7	
Clover leaf diameter / mm	56.0	48.9	61.2	64.8	58.9	55.5	47.8	

b) Colour choice of pollinators in a survey watching flower visitation

Plant no.	1	2	3	4	5	6	7	
Colour of flower visited	Pink	Yellow	Pink	White	Pink	Red	Pink	

Spread of data:

n	1	2	3	4	5	6	7
Species diversity	0.165	0.486	0.490	0.555	0.649	0.688	0.694

Range: _____

Find the values which sit is the following positions:

Median (n+1) / 2 — The median is the _____th value = _____

Lower Quartile (n+1) / 4 — The Lower Q is the _____rd value = _____

Upper Quartile ((n+1) / 4) x 3 — The Upper Q is the _____th value = _____

Interquartile range: _____

Why might this be more useful than the full range?

Standard deviation:

N	x	(x-ẋ)	(x-ẋ)²
1	0.165		
2	0.486		
3	0.490		
4	0.555		
5	0.649		
6	0.688		
7	0.694		
ẋ		Σ(x-ẋ)²	

$$SD = \sqrt{\frac{\sum (x - \bar{x})^2}{N - 1}}$$

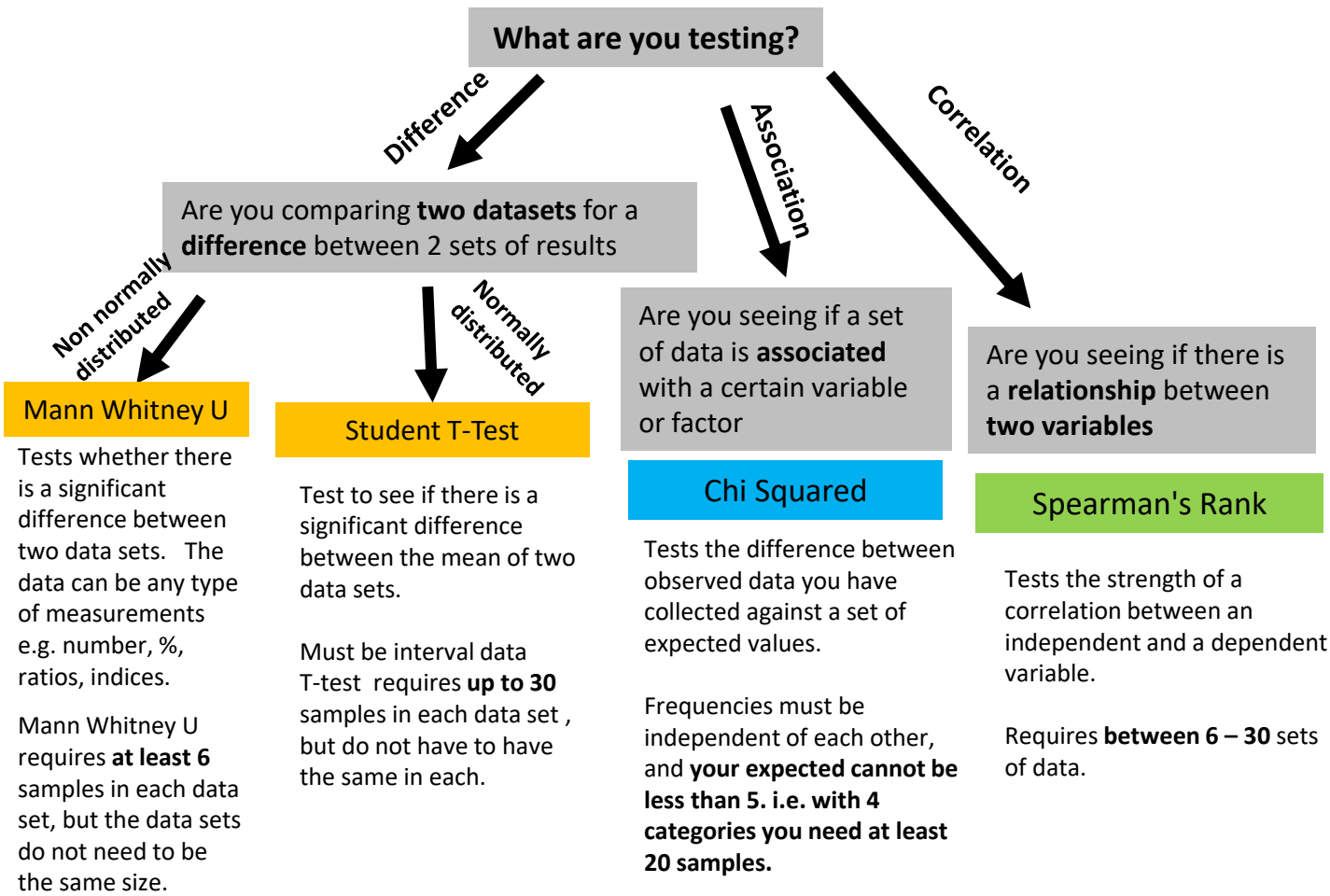
Standard deviation = _____

Statistics:

We use statistics to test whether data collected from a **sample** is **significant**, or could have arisen by **random chance** through sampling error.

Sampling Error:

Inferential statistics look at the **probability** that this has happened.
The chance that an observed pattern is the result of sampling error is called **statistical significance.**
If Probability (*p*)= 0.05, then there is a < **5%** probability your pattern is the result of a chance through sampling error. (i.e. you can be 95% confident in your result being statistically significant).
If there was a <1% probability of error what would the P value be? _____



Choosing statistics: For each of the following sets of data, choose and justify the most appropriate statistical analysis.

a) Freshwater pollution study - _____

	No. of Biological indicator species
1m above pollution source	2
At pollution source	110
1m below pollution source	150
10m below pollution source	45

b) Woodland plant adaptation study - _____

	Leaf surface area of Brambles / cm^2	
Area	Light	shade
n	10	12
Mean	37.5	49.6
Standard deviation	1.56	2.54

c) Lichen study - _____

Distance from Road	% cover Lichen on tree trunks
0	1
2.0	5
4.0	4
6.0	10
10.0	36
12.0	45
14.0	40
16.0	60

Statistical hypotheses: describe what we mean by:
Null hypothesis

Alternative hypothesis

How the statistics works:

	P value	
Number of pairs (n)	0.05	0.01
10	0.648	0.818
11	0.623	0.794
12	0.591	0.780
13	0.566	0.745
14	0.545	0.716
15	0.525	0.689

Critical values:

What to include in a statistical conclusion: