Year 1 Knowledge Organiser **Statistics**

0. Large Data Set	
Trace (Tr)	Rainfall less than 0.05 mm
Beaufort Scale	Another measure for mean wind speed
Knot	A nautical mile per hour (1kn = 1.15mph)
Oktas	Measures cloud cover, eighths of the sky covered by cloud
Decametres (Dm)	Measures daily mean visibility, greatest horizontal distance at which an object can be seen in daylight
Hectopascals (hPa)	Measures mean pressure
1. Data Collection	
Population	The whole set of items that are of interest
Census	Observes or measures every member of the population
Sample	A selection of observations taken from a subset of the population
Sampling Unit	Individual units of a population
Sampling Frame	A list of sampling units which are named or numbered
Random Sample	Every member of the population or sample has an equal chance of being selected
Systematic Sampling	The population is placed in an ordered list and the sample is chosen at regular intervals, choosing the start position randomly
Stratified Sample	The population is proportionally divided into mutually exclusive groups and a random sample is taken from each group
Quota Sampling (non-random)	An interviewer or researcher selects a sample until each quota is filled that represents the characteristics of the whole population
Opportunity Sampling (non-random)	Sample is taken from people who are available at the time the study is carried out and who fit the criteria you are looking for

2. Measures of Location & Spread	
Mean (Population)	μ
Mean (Sample)	$\bar{x} = \frac{\Sigma x}{n}$
S_{xx} =	$\Sigma(x-\bar{x})^2 = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$
Variance σ^2 =	$\frac{s_{xx}}{n} = \frac{\Sigma x^2}{n} - \bar{x}^2$
Standard Deviation σ =	$\sqrt{Variance}$
General formula for coded data	$y = \frac{x-a}{b}$
Mean (coded)	$\bar{y} = \frac{\bar{x}-a}{b}$
Standard deviation (coded)	$\sigma_y = \frac{\sigma_x}{b}$

3. **Representations of Data**	
Common Definition for Outlier	Greater than $Q_3 + k(Q_3 - Q_1)$ Less than $Q_1 - k(Q_3 - Q_1)$
Cleaning Data	The process of removing anomalies from data set
4. **Correlation**	
Interpolation	Making an estimate of values within the range of the given data
Extrapolation	Making a prediction based on a value outside the range of the given data
5. **Probability**	
Event	A collection of one or more outcomes
Sample Space	The set of all possible outcomes
Mutually Exclusive	Two events that cannot occur at the same time
Independent events	Events that have no effect on one another
P(A')	Probability A does not occur
P(A | B)	Probability A occurs given that B has occurred
P(A ∪ B) for mutually exclusive events	P(A) + P(B)
P(A ∩ B) for independent events only	P(A) x P(B)
6. **Statistical Distributions**	
Criteria for Binomial Distribution:	• Fixed number of trials • Two possible outcomes (success or failure) • Fixed probability of success • Trials are independent of each other
Binomial Function $P(X = x)$	$\binom{n}{x} p^x (1-p)^{n-x}$ n = number of trials, p = probability of success
7. **Hypothesis Testing**	
H_0	The null hypothesis
H_1	The alternative hypothesis
Critical Region	A region of the probability distribution which, if the test statistic falls within it, would cause you to reject the null hypothesis
Critical Value	The first value to fall within the critical region
Significance Level	The probability of incorrectly rejecting the null hypothesis
Step 1	Define test statistic X, write its distribution and define the meaning of p in context
Step 2	Write null and alternative hypotheses
Step 3	Assume H_0 is true and determine probability of observed test statistic
Step 4	Two-part conclusion: 1. Do we reject H_0 or not? 2. Put in context of original problem.