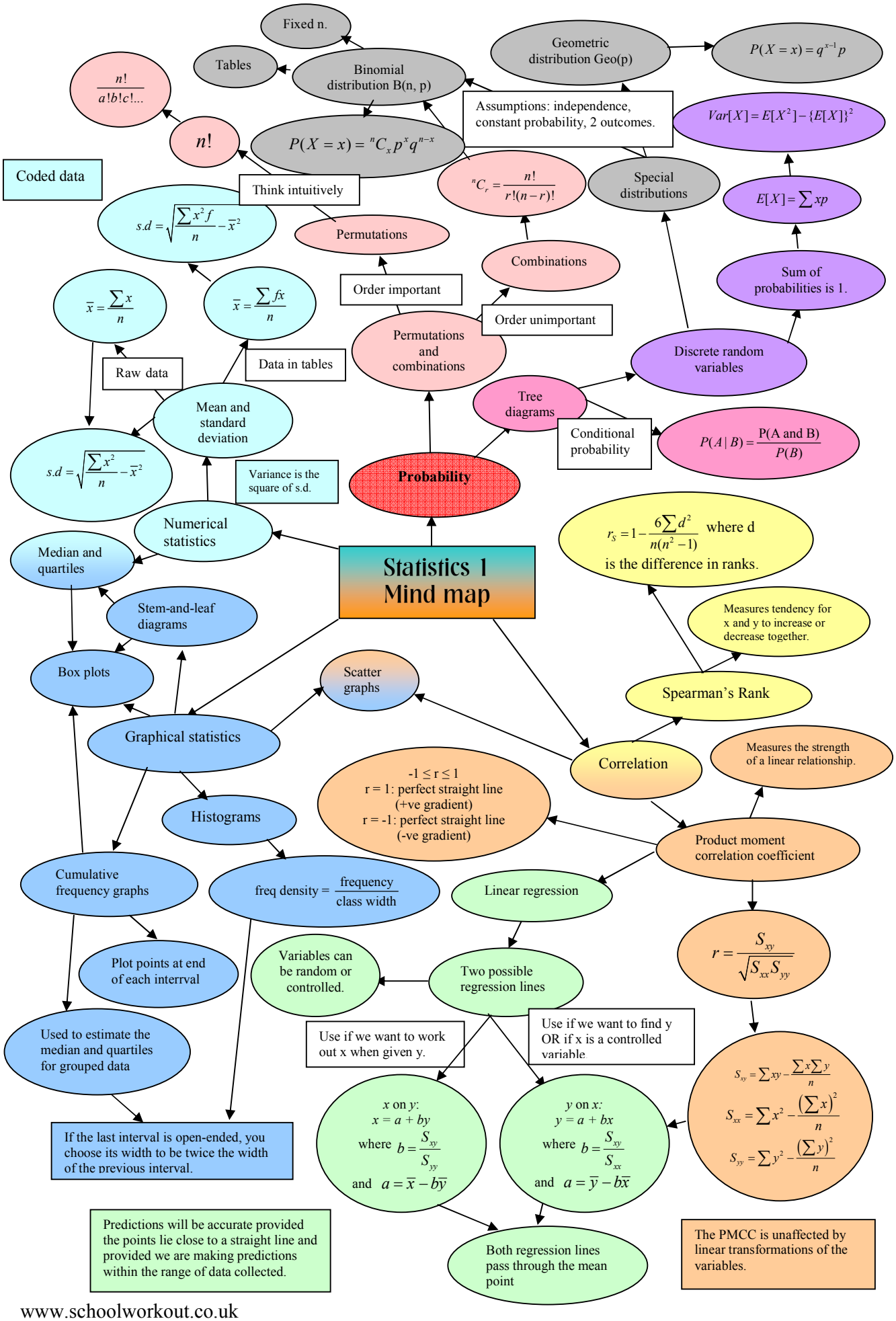
Statistics 1 Mind map
Probability
Fixed n.
Tables
Binomial distribution B(n, p)
Geometric distribution Geo(p)
$P(X = x) = q^{x-1}p$
$\frac{n!}{a!b!c!...}$
$n!$
$P(X = x) = {}^nC_x p^x q^{n-x}$
Assumptions: independence, constant probability, 2 outcomes.
$Var[X] = E[X^2] - \{E[X]\}^2$
Coded data
Think intuitively
${}^nC_r = \frac{n!}{r!(n-r)!}$
Special distributions
$E[X] = \sum xp$
$s.d = \sqrt{\frac{\sum x^2 f}{n} - \bar{x}^2}$
Permutations
Combinations
Sum of probabilities is 1.
$\bar{x} = \frac{\sum x}{n}$
$\bar{x} = \frac{\sum fx}{n}$
Order important
Order unimportant
Permutations and combinations
Discrete random variables
Raw data
Data in tables
Tree diagrams
Mean and standard deviation
Conditional probability
$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$
$s.d = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$
Variance is the square of s.d.
Numerical statistics
$r_S = 1 - \frac{6\sum d^2}{n(n^2-1)}$ where d is the difference in ranks.
Median and quartiles
Stem-and-leaf diagrams
Measures tendency for x and y to increase or decrease together.
Box plots
Scatter graphs
Spearman's Rank
Graphical statistics
Correlation
Measures the strength of a linear relationship.
-1 ≤ r ≤ 1
r = 1: perfect straight line (+ve gradient)
r = -1: perfect straight line (-ve gradient)
Histograms
Product moment correlation coefficient
Cumulative frequency graphs
freq density = $\frac{\text{frequency}}{\text{class width}}$
Linear regression
$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$
Plot points at end of each interrval
Variables can be random or controlled.
Two possible regression lines
Used to estimate the median and quartiles for grouped data
Use if we want to work out x when given y.
Use if we want to find y OR if x is a controlled variable
$S_{xy} = \sum xy - \frac{\sum x \sum y}{n}$
$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}$
$S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n}$
x on y: $x = a + by$ where $b = \frac{S_{xy}}{S_{yy}}$ and $a = \bar{x} - b\bar{y}$
y on x: $y = a + bx$ where $b = \frac{S_{xy}}{S_{xx}}$ and $a = \bar{y} - b\bar{x}$
If the last interval is open-ended, you choose its width to be twice the width of the previous interval.
Predictions will be accurate provided the points lie close to a straight line and provided we are making predictions within the range of data collected.
Both regression lines pass through the mean point
The PMCC is unaffected by linear transformations of the variables.