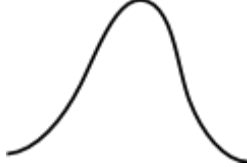

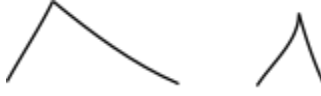


Discrete distributions	Select these distributions for counts - variables where the numbers are integers e.g. number of people, number of accidents etc.	The frequencies need to add to the total amount of data or the proportions need to add to 1 or 100% - data should be shown as a bar graph	Continuous distributions	Select these distributions for measurement data - variables with units e.g. distance, time, weight etc.	Area under the curve//model has to add to 1 or 100% - data should be shown as a histogram
binomial distribution	Looks like: <ul style="list-style-type: none"> lowest bar is 0 and highest bar is n the mean of the variable is the same as n times p the higher the value the p, the more the mean/centre of the distribution moves to the right the higher the value of n, the more smoother the shape (more bars!) 	Context conditions/clues: <ul style="list-style-type: none"> fixed number of trials (n) fixed probability of success (p) only two outcomes (e.g. yes or no, has it or doesn't have it, etc.) each event is independent (e.g. one person having blue eyes or not is independent from another person having blue eyes or not) 	normal distribution	Looks like /similar shape:  <ul style="list-style-type: none"> symmetric bell-shaped - high proportion of values close to mean 	Context conditions/clues: <ul style="list-style-type: none"> possible for values to be higher or lower than what has been given in the data could be originally discrete data, but the distribution looks normal need mean and standard deviation
uniform distribution	Looks like/similar shape: <ul style="list-style-type: none"> lowest bar is the minimum and the highest bar is the maximum each bar is similar in height 	Context conditions/clues: <ul style="list-style-type: none"> equally likely outcomes (spinners, dice, drawing names from hats with replacement) minimum and maximum values for the variable 	uniform distribution	Looks like/similar shape:  <ul style="list-style-type: none"> symmetric rectangular-shaped even distribution of values between min and max 	Context conditions/clues: <ul style="list-style-type: none"> minimum and maximum values for the variable each value between min and max has same probability of happening
Poisson distribution	Looks like: <ul style="list-style-type: none"> lowest bar is 0 shape becomes more symmetric as the value of lambda (the mean/average rate) increases shape typically skewed since "events are rare" so low mean mean similar to the variance (sd²) - as mean increases, so does the variation 	Context conditions/clues: <ul style="list-style-type: none"> average/mean rate of occurrence can't have simultaneous events events are rare events are random events are independent rate is proportional e.g. 2 accidents per day, 4 accidents per two days 	triangular distribution	Looks like /similar shape:  <ul style="list-style-type: none"> could be symmetric but also could be skewed triangular-shaped straight-ish edges 	Context conditions/clues: <ul style="list-style-type: none"> minimum and maximum values for the variable need mode or most likely value not sure of shape apart from min, max and mode

- We should be given with quite a large amount of data so we can be confident the features of the data are "real" and not just a by-product of sampling variation.
- With a small amount of data, we can expect there to be more "clunkiness" and will have less accurate estimates of the parameters for the model e.g. the mean for Poisson (remember how the MOE for a survey percentage gets smaller with the more data you have).
- Independence of events and randomness of events are always assumptions - independence so we can multiply probabilities to get combined probabilities and randomness so that they are actually chance events - there is uncertainty about what is happening and we can't control what will happen
- If you're not sure, you can make assumptions e.g. assume a maximum value if it is not clear whether there is one or not, and you want to use a uniform/triangular distribution model
- Sometimes use models if they match the data we have even if the conditions for the theoretical model are not all met (often the case for Poisson)