Statistics 1 Knowledge Organiser

Probability

P(A) + P(A') = 1

Mutually exclusive events:
P(A) **or** P(B) = P(A) + P(B)

Conditional Probability
P(A and B) = P(A) × P(B|A)

Independent events:
P(A) **and** P(B) = P(A)× P(B)

Probability distribution:
A listing of the possible values of the variable and the corresponding probabilities

$\sum P(X = x) = 1$

Correlation

Product moment correlation = r
(in the formula booklet)

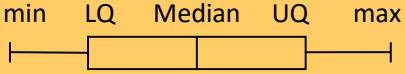
If r = -1 perfect negative correlation
If r = 0 no correlation
If r = 1 perfect positive correlation

Spearman's rank = r_s
(in the formula booklet)

If r_s = -1 rank orders are in exact reverse order.
If r_s = 1 rank orders are identical

Measures of Spread

Box plots:



Variance: $\sigma^2 = \frac{1}{n}\sum x_i^2 - \bar{x}^2$

From a frequency table:

$\sigma^2 = \frac{\sum x_i^2 f_i}{\sum f_i} - \bar{x}^2$

Standard deviation = σ
$(the\ square\ root\ of\ the\ variance)$

Adding the same value to each number will not alter the variance as the spread will remain the same.

Binomial distribution

$$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$$

If you want less than don't forget you can use the tables in the formula booklet.

Geometric distribution

$$P(X=x) = p\,(1-p)^{x-1}$$

If you want more than, remember it must have had that many fails.

Expectation and Variance

	Random Variable	Binomial Distribution	Geometric Distribution
E(X) = μ	$\sum x_i p_i$	np	$\frac{1}{p}$
Var (X) = σ^2	$\sum x_i^2 p_i - \mu^2$	npq	

Measures of location

Median: middle if in order
Mode/Modal: Most common (2 modes = bimodal)
Mean: $\bar{x} = \frac{\sum x_i}{n}$, if from frequency table $\bar{x} = \frac{\sum x_i f_i}{f_i}$
Adding the same value to each number will alter the mean by that amount.

Permutations (order important)

The number of permutations of n distinct objects is n!

The number of different permutations of r objects from n distinct objects is ${}^nP_r$

If objects aren't distinct:
$\frac{n!}{p! \times q \times r! \times \cdots}$

where p + q + r + … = n

Combinations (order unimportant)

The number of different combinations of r objects from n distinct objects is ${}^nC_r$

Histograms

f.d. = Frequency / Class width

Frequency polygons

Plot the midpoints

Cumulative frequency

Plot the upper bounds

If discrete data don't forget to alter the bounds with .5

Regression

(in the formula booklet)

y (depends) on x:
y = a + bx

x (depends) on y:
x = a' + b'y

These regression lines are only reliable if the value you use to estimate with is within the data range and there is strong correlation.

@pbrucemaths