1. Which statement(s) about the standard deviation of a data set is/are false (circle all that apply): The standard deviation ...

   (a) is never negative.

   (b) **must be greater than zero.**

   (c) **is the mean of the differences from the mean.**

   (d) is the (positive) square root of the variance.

   (e) has the same dimensions as the data.
   Explain why you rejected one of the alternatives in a sentence: *The s.d. cannot be negative as it is the square root of a mean of squares. By definition the s.d. is the square root of the variance. The variance is in the squares of the units of the data, so its square root is in the same units as the data.*

2. In a company, suppose employees can be categorised as either male or female as defined by their birth certificates. Suppose in the company the average salary of all male workers is \$68,000, while the average salary of all female workers is \$64,000. Which of the following must be true? (Select all that apply):

   (a) The mean salary within the company must be \$66,000.

   (b) **The mean salary within the company must be at least \$64,000.**

   (c) The mean salary within the company is less than the median.

   (d) **The mean salary within the company is between \$64,000 and \$68,000.**

   (e) It is impossible to say anything about the mean salary within the company from the information given.

3. A 2002 survey included the question "Would you say that you are very happy, pretty happy, or not too happy?". Each respondent was recorded as being in one of three possible categories of family income.

The counts for the responses are provided in the table below:

		Happiness level		
		Not Too Happy	Pretty Happy	Very Happy
Income	Above average	17	90	51
	Average	45	265	143
	Below average	31	139	71

What percentage of individuals from families of below average income responded as being "Very Happy"?

(a) **29.5%**

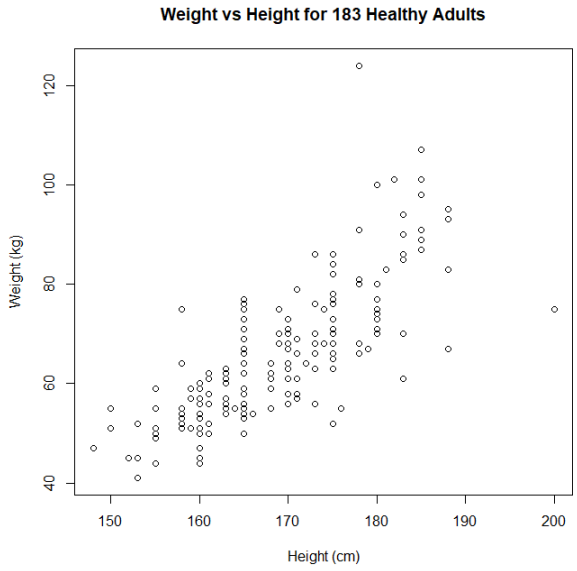
(b) 28.2%

(c) 26.8%

(d) 12.9%

(e) 8.3%

4. The heights (in cm) and weights (in kg) were self–reported by 183 healthy adults who regularly exercise. A scatterplot of the data is below:



The correlation between weight and height for these adults is:

(a) – 0.42

(b) 0.06

(c) 0.31

(d) **0.76**

(e) 1

Justify your choice of answer in a sentence or two: *The data show an upward trend so the correlation must be positive; the points are roughly, but not exactly, following a line so a correlation of about 0.8 looks reasonable. (2 marks: 1 for written answer)*

5. Sam is interested in bird nest construction, and finds a correlation of 0.82 between the depth of a bird nest (in inches) and the width of the bird nest (in inches) at its widest point. Sue, a classmate of Sam, is also interested in looking at bird nest construction, and obtains the same data on the same bird nests as Sam, except she takes her measurements in centimeters, instead of inches. We know that 1 inch is 1.64 centimeters. Which of the following is correct?

   (a) Sue's correlation should be 1, because it will match Sam's exactly.

   (b) Sue's correlation would be $1.64 \times 0.82 = 1.34$, because you need to change the units from inches to centimeters.

   (c) Sue's correlation would be $0.82/1.64 = 0.50$, because you need to change the units from inches to centimeters.

   (d) **Sue's correlation would be 0.82, the same as Sam's.**

   (e) It is impossible to determine Sue's correlation from the information given.

   Explain your answer clearly: *The correlation does not depend on the units of measurement used for either variable. Hence changing from inches to centimetres will not affect a correlation between length data.*

6. A study published in *Neurobiology of Aging* investigated how diet and exercise can affect mental agility in elderly dogs. Beagles aged 7–11 years were assigned at random to either a control group or to a special diet and exercise regime. After some time, the dogs were tested on problem–solving tasks involving finding treats under different coloured blocks. Suppose of the twelve dogs on the diet and exercise regime, eleven were able to solve the entire task, but of the eight dogs in the control group, only two could do so.

   (a) This study is an example of:

      i. **An experiment.**
      ii. An observational study.
      iii. A census.

   (b) Display the data from this study in a contingency table, clearly labelling your rows and columns and providing marginal totals.

	Special diet & exercise	*Control*	
Solved task	*11*	*2*	*13*
Did not solve task	*1*	*6*	*7*
	12	*8*	*20*

7. It has been suggested that people may be able to predict the appearance of a randomly selected image if the content of the image is emotive to the person, such as being erotic or frightening. Savva *et al.* (2004) report a study involving 50 volunteers, 25 of whom were deemed to be "spider phobes" (either arachnaphobic or disliking spiders). The volunteers were shown images of two types, either containing a spider or a "calm" (or *low–affect)* image, such as a coastline. Subjects were shown image pairs of the same type on a computer and asked to indicate which image they preferred; one of the two images was then selected at random and flashed quickly and repeatedly on the screen (i.e., so–called *subliminal exposure*). There were 12 such trials for each of the spider images per subject and 36 trials per low–affect image. The subject's preference at each trial either matched or did not match the randomly selected image to follow.

(a) Identify the response variable in this study.
The response variable is the image selected by the subject at a trial, this either matching or not matching the image to be shown subliminally after.

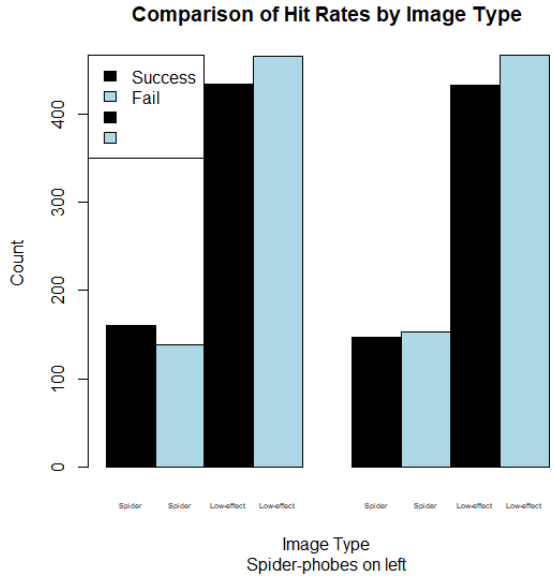
(b) Why might the researchers have thought the "spider phobes" would have a higher success rate for the spider images than for the low–affect images?
The spider images would presumably be more emotive to the spider phobes than the neutral images, and the research hypothesis suggested that predictive success rate would be higher for images the subject finds more emotive.

(c) The table below summarises the data from the study, where "S" indicates a preference matching the image following, "F" indicating a non–match.

Image pair type:	Spider		Low– affect	
	S	F	S	F
Spider-phobe	161	139	434	466
Non Spider-phobe	147	153	433	467

Sketch a graphic using the data from the study that would be suitable to show whether the spider–phobes and non spider-phobes differed in their success rates for the two types of images.

Not perfect, but a grouped bar chart like the following works best:



Note it is preferable to group by volunteer type, as we are interested in comparing hit rates both within and across the two groups.

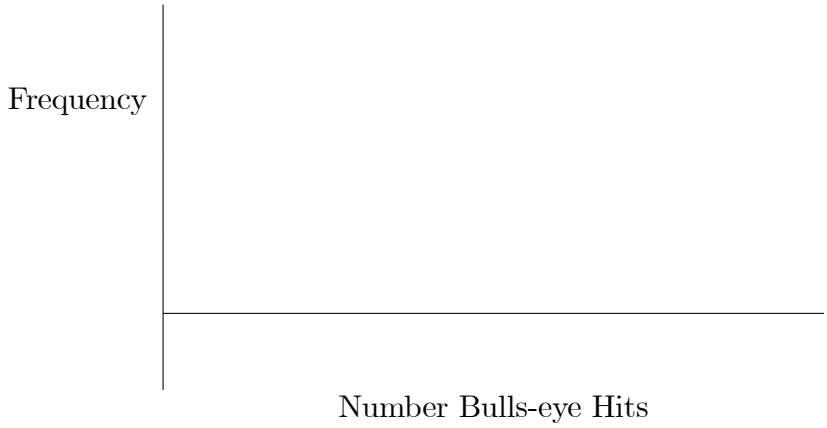
8. Jane is a keen archer, and practices shooting at a circular target 20 m away. Over ten shots, she records the number of times she lands her arrow in the bulls-eye (the centre of the target).

   (a) To model the number of successful bulls-eye hits Jane scores in her ten attempts, give two reasons why a Binomial model could be appropriate. Ensure you relate your criteria to the context in question.
   The response is binary (either hits bulls-eye or does not), the number of trials is fixed at ten. It might also be assumed that Jane's skill level does not vary across ten trials and so the chance of success is fixed at each attempt.

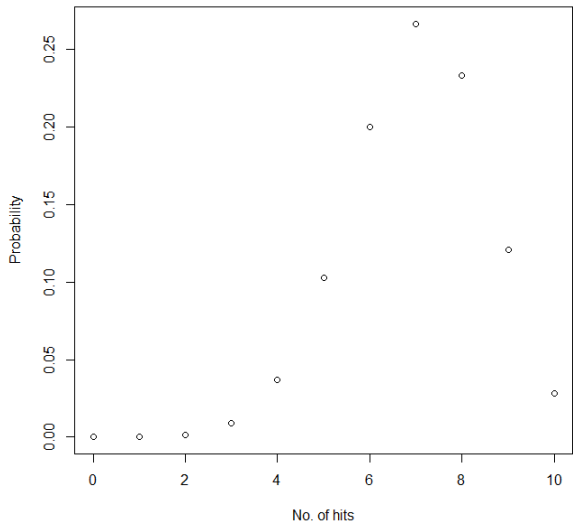
   (b) You observe Jane having ten attempts to hit the bulls-eye and record she is successful on seven attempts. Assuming a Binomial model is appropriate, use this information to assign the parameters you consider most suitable based on what you have observed.

We would take $n = 10$ and the probability of success estimated to be 0.7.

(c) Using your parameters from (b), sketch the Binomial distribution you would use to model the number of bulls-eye hits Jane records over her next ten attempts.



Something with the following shape (note values on the vertical axis not required):



(d) Which score would you expect to be the most likely, based on the

model?
Most probability will be on seven.

(e) Suppose instead you observed the number of times Jane hits the bulls-eye over 100 shots. Assuming her skill level does not change and is the same as for the original set of shots you observed, explain how a Binomial distribution modelling the number of bulls-eye hits would differ from that described in (c) in regards to its

- Location:
  The mode will move up to around seventy.
- Shape:
  The distribution will be less skewed, i.e., more symmetric.