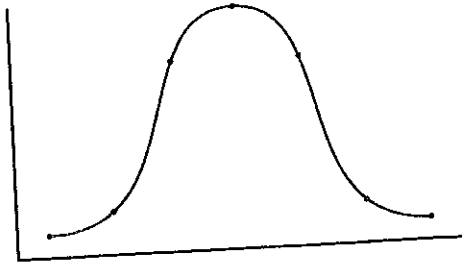


# 26

## Statistics

### A Basic statistical terms

Notice the key vocabulary in these three short texts about statistics.



A **normal distribution** of data means that most of the examples in a **set of data** are close to the average, while relatively few examples tend to one extreme or the other. Normally distributed data shown on a chart will typically show a **bell curve**. It will often be necessary to work out the extent to which individuals **deviate**<sup>1</sup> from the **norm**<sup>2</sup> and to calculate the figure that represents **standard deviation**<sup>3</sup>.

Six children are 7, 8, 8, 8, 11 and 12 years old. Their **average** age is 9 years old (the **sum** of their ages divided by six). The **mode** (the most frequent value) is 8. The **median** is 9.5 (the **halfway point** between the two **extremes** of the range).

Statisticians are often concerned with working out **correlations**<sup>4</sup> – the extent to which, say, left-handedness **correlates with** intelligence. They must ensure that any data they collect is **valid**, i.e. that it is measuring what it claims to measure – all the subjects in the **sample**<sup>5</sup> must be appropriately and accurately assessed as left- or right-handed, for example. The figures must also be **reliable**, i.e. they would be **consistent**<sup>6</sup> if the measurements were repeated. Usually, statisticians hope that their calculations will **show/indicate a tendency**, e.g. that left-handed people will be shown to be **significantly**<sup>7</sup> more intelligent than right-handed people.

<sup>1</sup> differ <sup>2</sup> the average <sup>3</sup> average difference from the norm <sup>4</sup> connections, often as cause and effect <sup>5</sup> the subjects of the experiment or group representing the total population measured <sup>6</sup> the same <sup>7</sup> noticeably

### B A probability<sup>1</sup> problem

Notice the vocabulary in this problem from a statistics textbook.

Sue picks a card **at random**<sup>2</sup> from an ordinary pack of 52 cards. If the card is a king, she stops. If not, she continues to pick cards at random, without replacing them, until either a king is picked or six cards have been picked. The random **variable**<sup>3</sup>, C, is the total number of cards picked. Construct a **diagram** to illustrate the possible **outcomes**<sup>4</sup> of the experiment, and use it to calculate the **probability distribution**<sup>5</sup> of C.

<sup>1</sup> likelihood of something happening <sup>2</sup> by chance <sup>3</sup> number or element of a situation that can change <sup>4</sup> results <sup>5</sup> assessment of probabilities for each possible value of C

### C Other useful nouns for talking about statistics

In a class of 8 women and 4 men, what **proportion**<sup>1</sup> are male? Answer: one third

In the same class what is the female to male **ratio**<sup>2</sup>? Answer: 2:1

The figures show a **trend**<sup>3</sup> towards healthier eating habits.

The study investigates the increase in the **volume**<sup>4</sup> of traffic on the roads.

<sup>1</sup> number compared with another number <sup>2</sup> relationship between two numbers showing how much bigger one is <sup>3</sup> change in a particular direction <sup>4</sup> amount, quantity



We say **10 per cent** (NOT the 10 per cent or 10 percentage) of students got an A for their exam but the **percentage** of students achieving an A has increased.

## Exercises

### 26.1 Complete the sentences.

- 1 The six subjects who took the test scored 24, 22, 16, 16, 16, and 14 points out of 30. The \_\_\_\_\_ was 16. The \_\_\_\_\_ score was 19 and the \_\_\_\_\_ score was 18.
- 2 The \_\_\_\_\_ of all donations to the charity in 2003 was \$3,938. The smallest donation was \$10 and the largest was \$130. Most were around the \_\_\_\_\_ point of \$60.
- 3 Each questionnaire item asked respondents to choose one of a \_\_\_\_\_ of six options, with the two \_\_\_\_\_ being 'very dissatisfied indeed' and 'completely satisfied'.

### 26.2 Use the correct form of the words in the box to complete this text.

distribute    trend    significant    probable    random    correlation    outcome    vary

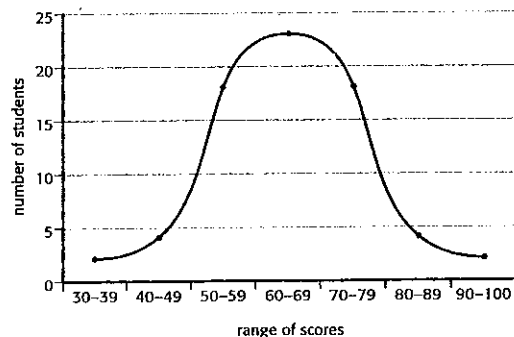
Life insurance companies base their calculations on the laws of \_\_\_\_\_, that is they assess the likely \_\_\_\_\_ given the different \_\_\_\_\_ such as age, sex, lifestyle and medical history of their clients. The premiums are therefore not chosen at \_\_\_\_\_ but are carefully calculated. The \_\_\_\_\_ of ages at which death occurs and causes of death are studied to see if they \_\_\_\_\_ with other factors to be taken into account in setting the premiums. Naturally, the companies also monitor social \_\_\_\_\_ and react to any changes which might \_\_\_\_\_ affect mortality rates.

### 26.3 Answer the questions.

- 1 There are 12 male students and 6 female students in the class. What is the ratio of males to females? And what proportion of the class is male?
- 2 If I am collecting data on course choices among second-year undergraduates and my sample is too small, what exactly do I need to do?
- 3 If my data show that students have a tendency to choose the type of clothing their friends choose, does it mean that they always, often or rarely choose similar clothes?
- 4 If I repeat the same experiment three times and the results are not consistent, is my method reliable?
- 5 If 20 out of 200 students fail an exam, what proportion, in percentage terms, failed?
- 6 If the average score in a test is 56, and Barbara scores 38, by how many points has she deviated from the norm?
- 7 If the volume of court cases increases, what changes: the type of case, the size of each case or the total number of cases?
- 8 What does standard deviation tell us? (a) What the standard of something is, (b) what the norm is, or (c) what the average difference from the norm is?
- 9 If a general survey of teenage eating habits asks questions about what teenagers eat for breakfast and lunch, is the survey likely to be valid?
- 10 Here is a graph showing how many students got scores within each 10-mark band in a biology test. Are the scores normally distributed? What is the shape of the graph called?

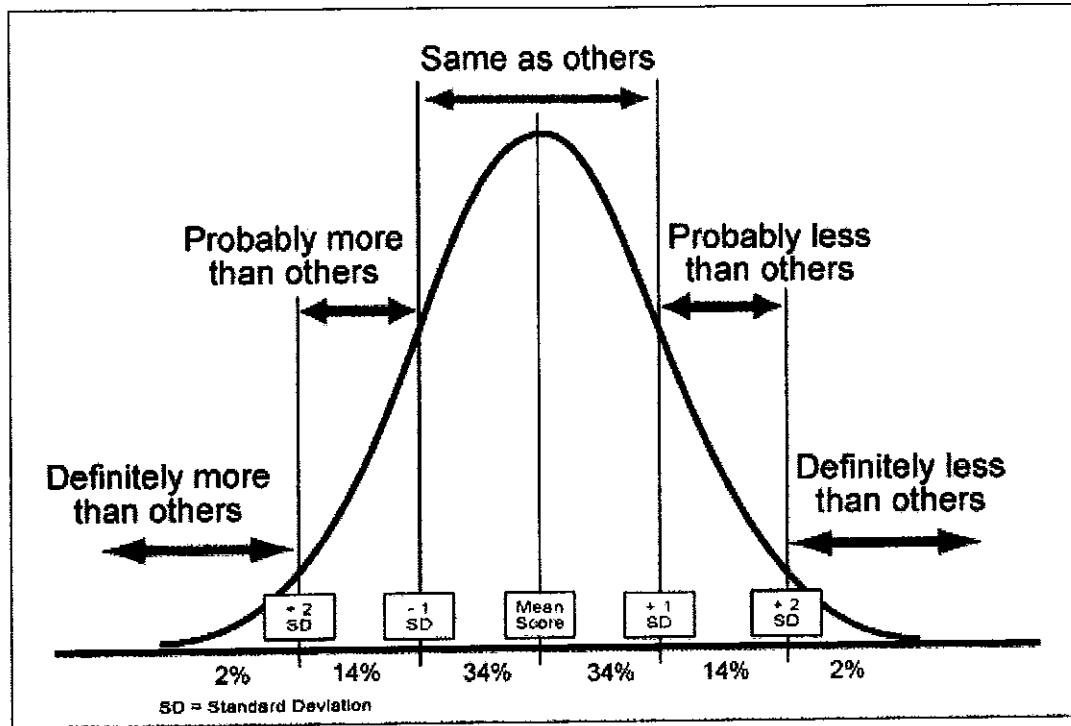
**FOLLOW UP**

What kinds of statistical data are likely to be discussed in your discipline? Find a relevant chart, graph or table and write about it using some terms from this unit.



# Normal distribution (1)

[http://www.youtube.com/watch?v=ed-vkd46\\_m4&feature=relmfu](http://www.youtube.com/watch?v=ed-vkd46_m4&feature=relmfu)



Listen to the recording and answer questions.

- 1) What does the speaker want to show? .....
- 2) How is the mean denoted? .....
- 3) What is a random variable in the example? .....
- 4) How is the standard deviation denoted? .....
- 5) What does the number 62 denote? .....
- 6) What does the shaded area show? .....
- 7) What does z represent? .....
- 8) What does x denote? .....
- 9) What do we need tables for? .....
- 10) What does the function  $\phi(z)$  denote, .....
- 11) Why should we round the  $\phi(z)$  value? .....
- 12) How do you read the notation

$$Z = \frac{X - \mu}{\sigma}, \quad \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

# The Greek Alphabet

Letters		Name	Pron.	Letters		Name	Pron.
Cap-ital	Small			Cap-ital	Small		
A	α	alpha	/'ælfə/	N	ν	nu	/nju:/
B	β	beta	/'bi:tə/	Ξ	ξ	xi	/'ksaɪ/
Γ	γ	gamma	/'gæmə/	Ο	ο	omicron	/'əʊmɪkrən/
Δ	δ	delta	/'deltə/	Π	π	pi	/'paɪ/
E	ε	epsilon	/'epsɪlən/	Ρ	ρ	rho	/'rəʊ/
Z	ζ	zeta	/'zi:tə/	Σ	σ, ς	sigma	/'sɪgmə/
H	η	eta	/'i:tə/	Τ	τ	tau	/'taʊ/
Θ	θ	theta	/'θi:tə/	Υ	υ	upsilon	/'jʊpsɪlən/
I	ι	iota	/'aɪ'əʊtə/	Φ	φ	phi	/'faɪ/
K	κ	kappa	/'kæpə/	X	χ	chi	/'kaɪ/
Λ	λ	lambda	/'læmdə/	Ψ	ψ	psi	/'psaɪ/
M	μ	mu	/'mju:/	Ω	ω	omega	/'əʊmɪgə/

Which Greek letters will you use to represent:

- a) angles in a triangle.....
- b) the first infinite ordinal .....
- c) the ratio of a circle's circumference to its diameter.....
- d) the summation operator .....
- e) set membership .....
- f) the golden ration 1.618... in mathematics, art and literature .....
- g) a general eigenvalue in linear algebra .....
- h) the population mean or expected value in probability and statistics .....
- i) a risk management measure in mathematical finance .....
- j) a finite difference or difference operator .....