

Unit 7 – Statistics

I Basic statistical terms

1 a) Match the terms to their explanations.

(arithmetical) average	the most frequent value
mode	the mean (calculated by dividing the sum of the values in the set by their number)
median	without method or conscious decision
correlation	noticeable
deviation	the amount by which a single measurement differs from a fixed value
significant	a quantity measuring the extent of the interdependence of variable quantities
at random	the halfway point between two extremes of the range

b) Six children are 8, 7, 8, 11, 12 and 8 years old. Work out their average age, the mode and the median.

- the average age:
- the mode:
- the median:

2 Correlation

a) Give examples of correlations, e.g. the hot weather and the consumption of ice-cream.

b) Read the text about correlations and fill in the gaps with the correct forms of words in the brackets.

Statisticians are often concerned with working out correlations - the extent to which, e.g. left-handedness correlates with (INTELLIGENT). They must (SURE) that any data they collect are valid, i.e. that it is measuring what it claims to measure: all the subjects in the sample must be (APPROPRIATE) and (ACCURACY) assessed as left- or right-handed, for example. The figures must also be (RELY), i.e. they would be consistent if the (MEASURE) were repeated. Usually, statisticians hope that their calculations will show or indicate a (TEND), e.g. that left-handed people will be shown to be (SIGNIFICANCE) more intelligent than right-handed people.

The text based on English Vocabulary in Use

c) When discussing correlation, we need to be aware of one fact. Please fill in the vowels in the statement and think about some examples.

C - R R - L - T - - N D - - S N - T I M P L - C - - S - T - - N

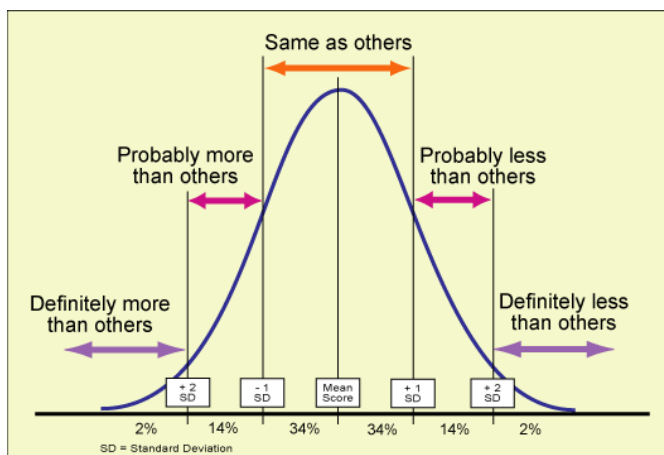
3 Read the text about life insurance and fill in the gaps. Use the words given, two will be left.

distribution trends significantly deviation probability
 random outcomes indicate correlate variables

Life insurance companies base their calculations on the law of, i.e. 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 they 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.

4 Normal Distribution

Look at the chart and explain what a normal distribution is. Then check with some information in the text on the right side.



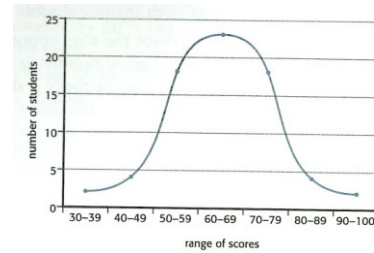
http://classes.kumc.edu/sah/resources/sensory_processing/learning_opportunities/sensory_profile/bell_curve.htm

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 plotted on a chart will typically show a **bell curve** (Gaussian curve). It will often be necessary to work out the extent to which individuals **deviate from the norm** and to calculate the figure that represents **standard deviation**.

5 Answer the questions:

- a) 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?
- b) If I am collecting data on course choices among second-year undergraduates and my sample is too small, what do I need to do?
- c) 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?
- d) If I repeat the same experiment three times and the results are not consistent, is my method reliable?
- e) If 20 out of 200 students fail an exam, what proportion, in percentage terms, failed?

- f) If the average score in a test is 56 and John scores 38, by how many points has he deviated from the norm?
- g) If the volume of court cases increases, what changes: the type of case, the size of each case or the total number of cases?
- h) What does standard deviation tell us: (i) what the standard of something is, (ii) what the norm is or (iii) what the average difference from the norm is?
- i) 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?
- j) Here is a graph showing how many students got scores within each 10-mark band in a biology test. What can you say about the distribution of the scores?



6 Listening

a) Pre-listening: Do you know how to solve the following problem?

The battery has a lifetime which is normally distributed with a mean of 62 hours and a standard deviation of 3 hours. What is the probability of a battery lasting less than 68 hours?

b) Watch a tutorial discussing this problem and answer questions below.

http://www.youtube.com/watch?v=ed-vkd46_m4&feature=relmfu

- What does the speaker want to show?
- How is the mean denoted?
- What is a random variable X?
- How is the standard deviation denoted?
- What does the number 62 denote?
- What does the shaded area show?
- What does z represent?
- What does x denote?
- What do we need tables for?
- What does the function $\Phi(z)$ denote?
- Why should we round the $\Phi(z)$ value?

c) Read the following notation

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

II 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	/'bɪ:tə/	Ξ	ξ	xi	/'ksaɪ/
Γ	γ	gamma	/'gæmə/	O	\omicron	omicron	/'əʊmɪkrən/
Δ	δ	delta	/'delta/	Π	π	pi	/'paɪ/
E	ϵ	epsilon	/'epsɪlən/	P	ρ	rho	/'rəʊ/
Z	ζ	zeta	/'zɪ:tə/	Σ	σ, ς	sigma	/'sɪgmə/
H	η	eta	/'i:tə/	T	τ	tau	/'taʊ/
Θ	θ	theta	/'θɪ: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 letter will you use to represent

- angles in a triangle
- the first infinite ordinal
- the ratio of a circle's circumference to its diameter
- the summation operator
- set membership
- the golden ratio
- the population mean or expected value in probability and statistics
- a risk management measure in mathematical finance
- a finite difference or difference operator