Human Communication
Lecture 15
Tools for Analysing Data

1. Visualisation Techniques

Visualisation Techniques

Visualisation techniques - used for exploratory data
- make patterns in data apparent to human analyst,
- display visually relationships between different data variables

Tools for this include MATLAB, matrix manipulation system with excellent graphical display abilities.

Apparent effects can be confirmed by simple statistical techniques
- allows us to determine extent to which anticipated effect is present in data from experiment

Visualisation: less significant here, though useful for summaries of data for reporting results

Histograms (bar charts)

Shows how many data fall into each of a number of classes

Record temperature at noon each day for a year, then count how many days between 16 and 17 Celsius, 17 and 18, 18 and 19, and so on.
- plot of ***the number of days*** (vertical axis) against the various ***temperature categories*** (horizontal axis)
- shows the ***distribution*** of the data

Multiple peaks indicate something going on

Split set of data into clusters associated with peaks

Investigate whether members of the clusters differ from each other in consistent ways.

e.g. peaks around 25 and 16 with trough in between;
days in **25** cluster **'bright'**, but in **16** cluster **'cloudy'.**
Infer bright days are hotter than cloudy ones

Evaluating Usability Example

We have developed an interface for a variety of users to communicate with

We ask users to rate the usability of the interface as:
1. easy to use
2. average
3. difficult to use

We test it on different groups of users, recording how many users select each rating, for each of:
a. children (under 12 years)
b. teenagers (13 to 18 years)
c. adults (over 18 years)

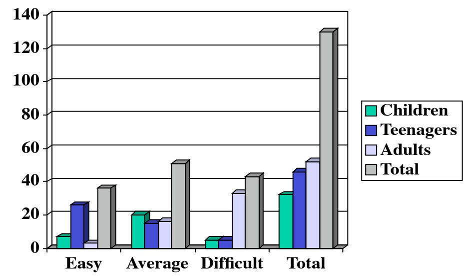
If there is no consistency of usability then the ratings should be equally spread across 1 to 3 ratings.

Is there a difference between different users?

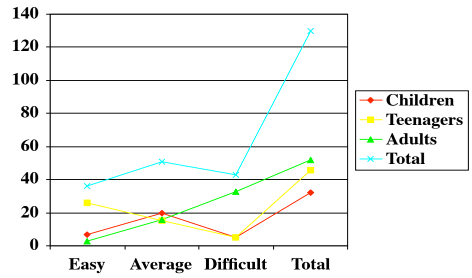
Usability: by age group and ease of use

Ratings:	easy	average	difficult	Totals
Children	7	20	5	32
Teenagers	26	15	5	46
Adults	3	16	33	52
Total	36	51	43	130

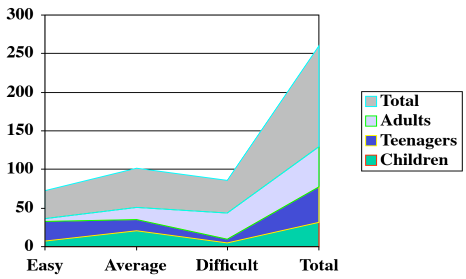
Age group v ease of use: Bar chart



Age group v ease of use: graph



Age group v ease of use: area chart



Is sample sufficient?

1. Get initial data for a small sample
2. Do a larger study focussing on selected variables that seem interesting in the smaller study
3. Categories may be independent for one variable but not for another - this will affect the results

2. Using statistical tests

Confirmation using statistical tests

Experimental Hypothesis: ***teaching child using an Intelligent Tutor increases language test score***
Measure: ***performance on language test***

Compare distributions of:

1. Score distribution of children not using ITS
2. ITS taught students score distribution

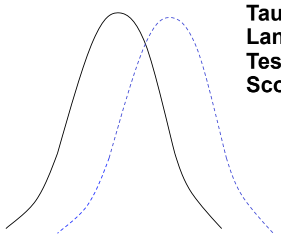
Assume:

- Language test score **normally distributed**
- Test is designed to ensure normal distribution

Null Hypothesis: ***ITS taught score distribution will be the same as the normal score distribution***

Score Distributions

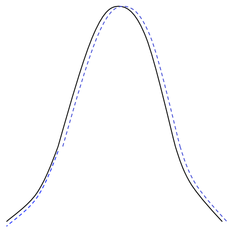
Usual Language Test Score
μ= 100
s.d.=15



ITS Taught Language Test Score

Null Hypothesis

Usual Language Test Score
μ= 100
s.d.=15



ITS Taught Language Test Score

Variance

Variance is the mean deviation from the centre:

$$\text{Variance} = \frac{\sum \{(x1 - \mu)^2, \ldots (xn - \mu)^2\}}{N}$$

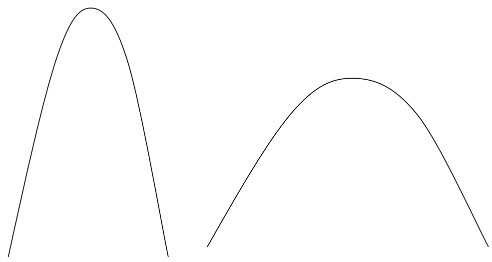
$$= (\sum (X - \mu)^2)/N$$

The Standard deviation σ is the square root of the variance:

$$\sigma = \sqrt{(\sum (X - \mu)^2)/N}$$

Distributions

Left curve shows less variance than right curve



Example Student: Peter

Peter: Taught by ITS
Does his test score come from the normal distribution?
1. Test him = 120
2. Find position in the distribution:
3. Find Z score = x - μ(mean)/σ (standard deviation)
   = 120-100/15 = 1.33
4. Look this up in tables: 1.33 > probability of 0.0918

So 9.18% of the usual-score population score higher than Peter - not very convincing….

What is he scored 145 instead?
Z = 3, p = 0.0013 of the score higher…
So, 0.13% (1 in 769) would score higher
-> so rare that it seems likely this comes from a different population
-> Reject the null hypothesis

Rejecting the Null Hypothesis

What do we set as a cut-off for this?
- Reject if the probability is less than or equal to 5% or 0.05
- Write as **"p < 0.05" or "significant at p = 0.05"**

Means that:
when the score from the unknown distribution could only arise from the known one
(i.e have the same distribution)
with a less then 5 from 100 chance or less,
we reject the null hypothesis,
and say the score is from a different distribution.
p of 0.05 = significance level

Significance Level of 0.05

This means that there is a greater than 95% chance we are correct in accepting a different distribution
[in only 5% of extreme cases would this score arise by chance]

To be more certain, we may take the 0.01 level (1% chance)
[99% confident in claim of differences]

All statistical tests follow this basic logic:
- Research hypothesis predicts a difference in distribution
- Null hypothesis predicts no difference

Looking for effects in data

We look for suspicious data:
- we assume that nothing is going on,
- that no effects are present,
- that our data are independent of the factors that might influence them,

and we search for evidence that we are wrong.

What signs are there of independence or otherwise in our data?

3. Scatter plots and correlation

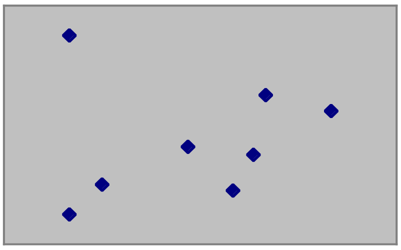
Scatter Plots

Look at data in more than one dimension at once, e.g. age, height and weight.
- a large set of triples of numbers, plot these numbers as points in space
- three axes at right angles to each other, triples specify co-ordinates of a point

If there is no relationship between the individual measurements, the points ought to be scattered randomly

If there is an effect, the points will be clustered more densely

Consider scatter plots...

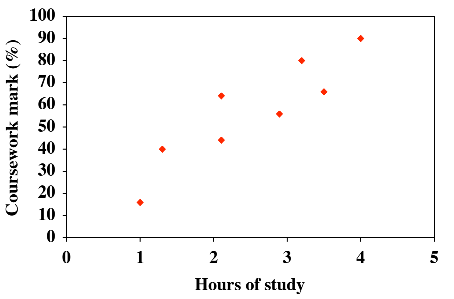


Where would you draw a line?

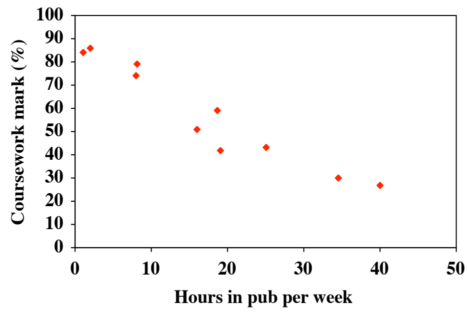
Hours spent v coursework mark

STUDENT								
	a	b	c	d	e	f	g	h
Hours spent	1	1.3	2.1	2.1	3.2	2.8	3.5	4
% on cwork	16	40	44	64	80	56	66	90

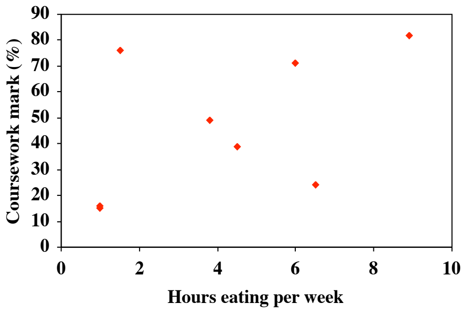
Positive Correlation



Negative Correlation



No correlation



Linear correlation

Linear correlation measures ***how well the data fit the model of a straight line*** relationship.

1. **Compute the means** of the x and y data from the scatter plot separately.
2. For each point in the scatter plot (pair of data) **calculate the deviation** of each datum from its mean and multiply, that is:
   compute (x - mean(x))*(y - mean(y))
3. **Sum these products** for all the data pairs **and divide by N-1** for N data.
4. **Work out the standard deviation of x and y separately**, and **divide the sum from step 3. by the product of these** standard deviations.

Pearson's Correlation Coefficient

Measures how well the data fit the straight line model it assumes:

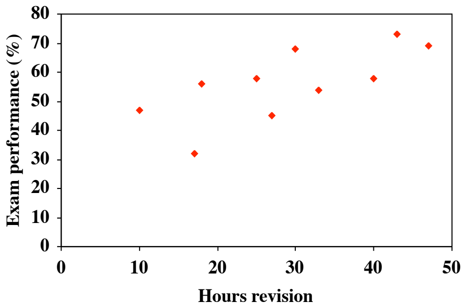
$$\text{correlation} = \frac{\sum \{(x - \mu x)(y - \mu y)\}}{(N-1)\,\sigma x\,\sigma y}$$

Lies between -1 (low X means high Y) and +1 (high X means high Y) with 0 meaning no correlation

Revision v exam performance
(example from Hinton, 1995)

STUDENT										
	a	b	c	d	e	f	g	h	i	j
Hours studied	40	43	18	10	25	33	27	17	30	47
% on exam	58	73	56	47	58	54	45	32	68	69

Revision v Exam performance



Using Pearson's Correlation Coefficient

$$\text{correlation} = \frac{\sum\{(x - \mu x)(y - \mu y)\}}{(N-1)\, \sigma x\, \sigma y} = 0.72$$

To see if this might be due to chance, we need to know the **degrees of freedom** = n-2 = 8

One-tailed test - is correlation +ve or -ve?

Two-tailed test - is there a significant correlation?

Here, +ve correlation predicted, so one-tailed

From tables of probability for one tailed = 0.05, for 8 d.f. r = 0.5494

0.72 is greater than that, **so significant correlation with less than 5% probability it is due to chance**

Interpretation of the size of a correlation

Cohen (1988) suggests guidelines interpreting correlation coefficient :

Correlation	Negative	Positive
Small	-0.29 to -0.10	0.10 to 0.29
Medium	-0.49 to -0.30	0.30 to 0.49
Large	-1.00 to -0.50	0.50 to 1.00

Criteria are somewhat arbitrary, should not be observed too strictly. Interpretation depends on the context and purposes.

Correlation of 0.9:
- may be very low if verifying physical law with high-quality instruments
- may be very high in social sciences where there may be a greater contribution from complicating factors.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.) Hillsdale, NJ: Lawrence Erlbaum Associates.

Comments on Correlation…

A **high positive correlation** between two variables **doesn't mean that one causes the other**……

Say we get a correlation of 0.8 between exam performance and hours of study:

Does this mean that the longer you study the better your exam results will be?
- or the better the exam results the more you will study?
- or some other variable influencing both (you are conscientious and bright)

Or ***time spent watching television and incidence of lung cancer are correlated***, but neither causes the other:
- both are caused by economic factors providing people with leisure time and money to buy cigarettes…

Statistical dependence is not the same thing as causal dependence.

Methods for collecting data

Methods for collecting data

Task analysis	**Observation**
Cognitive Walkthrough	Mock-ups
Protocol analysis	Wizard of Oz
Video Recording	**Interview**
Questionnaire	Focus groups
Sensitivity Analysis	Expert evaluation
Post-hoc analysis	**Logging use**
Manipulation experiment	**Self Report**
Dialogue mark-up and analysis	Sentient analysis

Direct Observation

Commonly used in **early stages of hypothesis formation** (or system design)
Identify potential **interactions between parameters** that might otherwise be missed

To help focus and record observations:
- **use tools** e.g. *event counters, checklists, structured behavioural annotation sheets*
- **restrict bandwidth** e.g. via *chat interface*

Very useful when used with other methods

Observation issues

Disadvantage: ***presence of the observer may affect behaviour being observed***

To reduce observer effects:
- **repeated sessions** enable participants to become accustomed to the observer's presence
- **careful placing of the observer** to avoid intrusion
- **train the observer** to resist interceding
- **explaining the role of the observer** to the participants

Mock-ups and paper prototypes

Goal: for system design, to get feedback on early design ideas before any commitment is made, mock-ups or prototypes of the system are used

1. **electronic prototypes** can be developed and presented on computer screen
2. **paper-based interface designs** can be used to represent different screen shots

Elicits ***responses to actual interfaces*** *and not other issues surrounding the operational access of technology*

Facilitates ***more imaginative feedback****, actively encourages* ***"hands on" interaction***

Video recording

Videoing user and system (or user and expert in WOZ studies) interaction **enables all visible user behaviour** (verbal and non-verbal) **to be used as data**

Video can be used for:
- **detailed behavioural analysis of participant**
- in less detail, **for reference**, to determine interesting episodes in the interaction
- **to transcribe verbal interactions** between expert/tutor and user/student in WOZ studies

Video recording of screen interactions also enables data capture of keyboard use and mouse movement

Tools that permit replay of the interaction are becoming more common and reliable.

Interviews

Elicit knowledge from a user by direct verbal questioning, and can be:

1. **very structured**: pre-determined questions in specified order, little room for elaboration in responses
2. **semi-structured**: variation in order of coverage of questions, open-endedness in responses, flexibility in question selection, potential generation of new questions
3. **open-ended:** few specific pre-determined questions, further questions determined by the previous response

Generally easy to administer and to respond to…

Commonly used:

1. for feedback on **interface design** and **usability**
2. to determine **users feelings** and **attitudes**
3. to determine **appropriate variables**
4. post-session **to confirm other data** collected

Questionnaires

Present questions to be answered in **written form** and are **usually structured**

To determine:
- **user characteristics** e.g. demographic, goals, attitudes, preferences, traits
- **users task knowledge**

Interviews versus questionnaires:
- conducted ***verbally*** rather than in ***written*** form
- suitable for ***eliciting*** a wider range of ***data*** which ***users may find difficult to elucidate*** in writing and without prompting
- interviews ***more objective*** than open-ended, unstructured feedback

Risk of respondent being influenced by questioner

Think Aloud/Protocol Analysis

User recorded while talking through what he is doing
- ***what he believes is happening***
- ***why he takes an action***
- ***what he is trying to do***

Advantages:
1. Simple, requires little expertise, provide useful insights
2. Encourages criticism and analysis
3. Points of confusion can be clarified at time

Disadvantages:
1. But process itself can alter task
2. Analysis can be difficult
3. Possible Cognitive Overload

Logging Use

Automatic recording of user actions can be built into software for later analysis
– Enables replay of full interaction
– Keystroke and mouse movement
– Errors
– Timing and duration of tasks and sub-tasks

Advantages:
1. Objective data
2. Can identify frequent use of features
3. Automatic, and unobtrusive

Disadvantages:
1. Actions logged need to be interpreted
2. Technical problem and file storage
3. Privacy issues

Cognitive Walkthrough

User asked to reflect on actions and decisions taken in performing a task, **post-task**
1. Re-enact task, replay session or use session transcript
2. User is asked questions at particular points of interest

Timing:
- immediately post-task (easier for user to remember)
- ***later*** (more time for evaluator to identify points of interest)

May help understand more about what going on in an interaction - could be used to better understand any communication - with or without system

Useful when talk aloud would be too intrusive

Physiological Responses:Eye Tracking

Reflects amount of cognitive processing required for tasks
- patterns of movement may suggest areas of difficulty

Useful for assessing relationship between verbal and non-verbal behaviour

Can measure:
1. Number of fixations
2. Fixation duration
3. Scan path

Need more work on how to interpret*, e.g. if looking at text is user reading it?*
Becoming standard equipment - now less invasive

Reading

Dix, A., Finlay, J., Abowd, R. and Beale, R. (2004) *Human-Computer Interaction.* Prentice Hall Chapter 9: Evaluation Techniques pp 318 - 364

Hinton, P. (1995) *Statistics Explained,* Routledge, London, UK