Quantitative Methods in Computing Education Research

(A brief overview – tips and techniques)

Dr Judy Sheard
Senior Lecturer
Co-Director, Computing Education Research Group
Monash University

judy.sheard@monash.edu

Overview

- Quantitative research in computing education
- Quantitative research design
  - Variance
  - Sampling
- Data collection
  - Questionnaire design
- Variables
- Measurement
- Data preparation
- Descriptive statistics
- Inferential statistics

A research study of computing education research

- An investigation of research into the teaching and learning of programming.
- 164 programming research papers from six key computing education conferences from 2005-2008.
- Research questions:
  - What is the nature of the research?
  - What are the main areas of research?

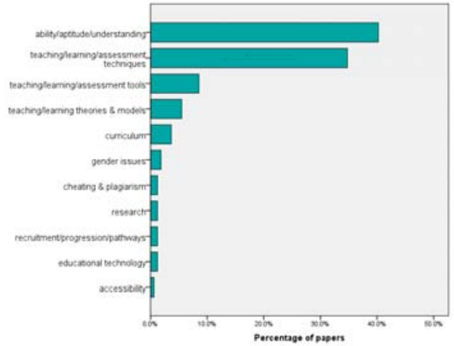
Sheard, J., Simon, Hamilton, M., & Lönnberg, J. (2009). *Analysis of research into the teaching and learning of programming. Proceedings of the International Workshop on Computing Education (ICER 2008), Berkeley, San Francisco, USA.*

Findings: nature of research

- 37% quantitative (describes phenomena in numbers and measures)
- 21% qualitative (describes phenomena in words)
- 42% mixed (research that is both quantitative and qualitative)

Therefore ... 79% of studies involved some form of quantitative research.

Findings: research themes



Research that is quantitative in nature

Experimental research – at least one variable is deliberately manipulated or varied by the researcher to determine the effect of that variable.

Quasi-experimental research – similar to experimental research but researcher has less control over the conduct of the experiment.

Survey (non-experimental research) – attempts to portray some aspect of a population at a point in time by gathering data from a sample of the population.

Research questions

Experimental – What was the effect on the experimental variable(s)? Does pair-programming help students to learn programming?

Quasi experimental – As for experimental but with naturally occurring rather than randomly assigned groups. Does pair-programming help students to learn programming?

Survey – What are the characteristics of the variables? What are the relationships and possible effects among the variables? Is there a relationship between learning style and use of online resources?

Research design for quantitative research

Quantitative research tends to be structured and prescriptive in nature – an assumption of an objective reality.

The outcomes are expressed as numbers.

These numbers must be interpreted by the researcher to produce valid and usable results – these results are used to answer research questions or problems.

This requires careful research design.

Variance

In a quantitative educational research study, there will be differences in measurements taken for variables – individuals are not all the same!

It is important to explain or control *variance* (difference) to get a clear view of variables of interest while eliminating, limiting, or explaining influences of other variables.

Some techniques used:

- Randomisation
- Hold variables constant or include in the design
- Statistical adjustment

Controlling variance – an example

A study of the effect of problem-based learning (PBL) on the performance of introductory engineering students. Sixty students are to be taught using PBL or a traditional method. The effect will be measured by performance on an end of semester exam.

Controlling variance – an example

Variable	**Method of controlling variance**

Generalisability

Quantitative research is often concerned with studying a *sample* from a population and generalising results back to that population.

An important consideration is obtaining an unbiased representative sample of the population.

Sampling techniques include:

- Random sampling
- Stratified sampling
- Systematic sampling

Biased samples

A sample is *biased* when some individuals are more likely to be sampled than others – also called a *non-random* sample.

Consider the following:

A lecturer wants to determine how much time his students spend preparing for exams. He collects data from students in one lecture towards the end of semester.

Could this be a biased sample?

Sample size

Determining an appropriate sample size can be complex. The researcher needs to consider:

- Cost, time and resources.
- Access to, and availability of, participants.
- Precision of statistical analysis.
- Variability within the target population.
- The chosen sampling scheme.

Generally, the larger the sample the better.

Some "rules of thumb"

- Statistical analysis of samples less than 10 not recommended for many tests.
- Samples of 30 or more recommended for many tests.
- When samples are divided for analyses then the "rules of thumb" apply to the sub-samples.
- In *regression*, *factor analysis* and other multivariate research, a minimum sample size of 10 for each variable is recommended.

Data collection methods

Many data collection methods are used in quantitative research. For example:

- Questionnaires
- Observations
- Structured interviews
- Ability tests
- Attitude inventories
- Computer log files

Questionnaires

Many studies in educational research involve administering a questionnaire.

In designing a questionnaire the researcher must consider:

- information required
- profile of the respondents, e.g. reading level, sensitivities
- style of questions
- data entry
- data analysis

Designing a questionnaire is a demanding task!

Why pilot your questionnaire?

Please indicate the total time in hours you spent in preparation for the exam.

0-10 ☐ *11-20* ☐ *21-30* ☐ *more than 30* ☐

Why pilot your questionnaire?

Please indicate the total time in hours you spent in preparation for the exam.

0-2 ☐ *3-5* ☐ *6-10* ☐ *more than 10* ☐

MONASH University 19

Variables

A *variable* is any attribute or property that differs between people or varies across time.

In research we often compare the values and examine the relationship between variables.

Educational research encompasses:

- A diversity of possible variables from concrete to abstract, e.g. *exam results, age, learning style, course satisfaction, resource usage*
- A variety of measurement instruments – *tests, computer software, questionnaires, inventories, observations*

MONASH University 20

Types of variables

Independent variables – assumed to produce an effect on, or be related to, a phenomena of interest. In experimental research they are manipulated by the researcher. May be used for classification.

Dependent variables – what is measured but not manipulated or controlled.

An example:

A study of cheating behaviour of IT students in different disciplines.

MONASH University 21

Data characteristics

Two broad distinctions:

- *Discrete* – data has specific values, e.g. number of students in a class, grade achieved, gender
- *Continuous* – measurement from a continuous interval, e.g. age, time on task

The type of data largely determines how the data is measured, described and analysed.

MONASH University 22

Types of measurement scales

Nominal (or categorical) – measures without order; allows classification or grouping, e.g. course, gender.

Ordinal – measures with order; allows ranking, e.g. grade, attitude towards course.

Interval – measures with order and equal intervals on a scale, e.g. test score, IQ.

Ratio – measures with order, equal intervals on a scale and a *true zero* point, e.g. age, length.

Note that in SPSS no distinction is made between interval and ratio.

MONASH University 23

Levels of measurement

	Ordered	Equal interval	True zero
Nominal			
Ordinal	√		
Interval	√	√	
Ratio	√	√	√

Spring 2010 24

What about Likert scales?

strongly agree				strongly disagree
1	2	3	4	5

What type of measurement scale is this?

… and these Likert scales?

strongly agree	agree	neutral	disagree	strongly disagree
1	2	3	4	5

What type of measurement scale is this?

Statistics

The appropriate data analysis approach for numerical data is *statistical analysis*.

Statistics — "bits of information", "theory, procedures, and methods by which data are analyzed in a quantitative manner" (Wiersma, 2005, p.492)

Statistical analysis has been greatly facilitated by computers.

Many statistical analysis software packages are available. E.g., SPSS, SAS, R.

Descriptive vs. inferential statistics

Two broad areas:

- *Descriptive statistics* — used to describe sets of quantitative data. This involves descriptions of distributions of data and relationships between variables.
- *Inferential statistics* — used to make inferences about populations from analysis of subsets (samples) of the population.

Sample data set from: Anxiety and Exams study

We will explore a set of data collected from a class of 25 students. The following data was collected from the students at the end of semester, prior to their exam and after submission of an assignment.

- Gender
- Number of hours spent in exam preparation.
- Anxiety rating (Likert scale where 1 = low anxiety to 5 = high anxiety). The students were asked to rate their level of anxiety about the forthcoming exam.

After marking, the exam and assignment marks were added to the data set and a final result computed as follows:

- final mark = exam mark * 7 + assignment mark * 3

Data preparation

Enter data into a spreadsheet with one *case* (data for one person or one observation) per row and one *variable* per column. Data from Excel can be imported into SPSS.

Need to decide how to handle:

- missing data
- *Don't Know*, *Not Applicable* and *Other* responses
- Extended text responses

May need to recode or create variables:

- allocate a numeric code to *nominal* or *ordinal* variable values (e.g. 1=male, 2=female)
- allocate grouping values e.g. age ranges, grades (0-49 → 'N')

Data preparation (contd.)

Cleaning – almost all data is dirty!

- duplication or missing cases
- out of valid range, e.g. 999 for an exam result
- unreasonable value, e.g. 1000 hours spent on homework
- anomalies, e.g. a student enrolled in a course that they have already passed
- missing values – note that missing values are not zero

Exploratory graphs are useful.

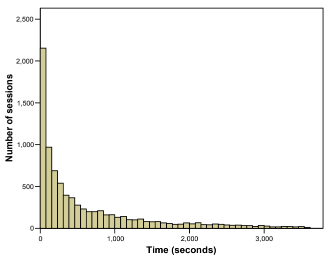
Running **Descriptives** and **Frequencies** in SPSS is useful.

Using SPSS in data preparation

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std.	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Gender	25	1	2	1.36	.490	.621	.464	-1.762	.902
Assignment Mark	25	1	10	6.32	2.688	-.301	.464	-.801	.902
Exam Mark	25	1	11	5.88	2.386	-.226	.464	.047	.902
Hours of Exam Preparation	25	2	100	13.28	18.596	4.550	.464	21.929	.902
Level of Anxiety	25	0	6	3.12	1.424	-.134	.464	-.119	.902
Final Mark	25	20	93	57.32	20.591	-.260	.464	-.505	.902
Valid N (listwise)	25								

Sometimes outliers are not obvious



Descriptive vs. inferential statistics

Two broad areas:

- *Descriptive statistics* — used to describe sets of quantitative data. This involves descriptions of distributions of data and relationships between variables.
- *Inferential statistics* — used to make inferences about populations from analysis of subsets (samples) of the population.

Descriptive statistics

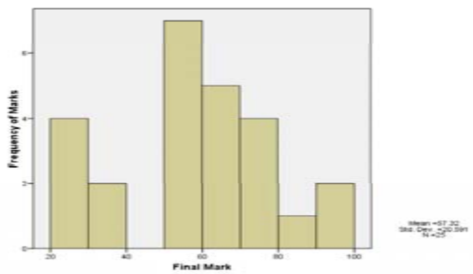
Ways of describing data include:

- *frequency distributions* – bar graphs, histograms, stem-and-leaf plots, boxplots;
- *measures of central tendency* – means, medians, modes;
- *measures of dispersion* – standard deviations, interquartile ranges
- crosstabulations ; and
- linear relationships between variables – correlation coefficients, scatterplots.

Also

- factor analysis
- clustering

Histogram: numerical data



Stem-and-Leaf Plot

```
Final Mark Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     3.00 Extremes    (=<24)
     1.00        2 .  7
     2.00        3 .  47
      .00        4 .
     7.00        5 .  2446779
     5.00        6 .  00344
     4.00        7 .  0379
     1.00        8 .  6
     2.00        9 .  03

 Stem width:        10
 Each leaf:      1 case(s)
```

Indicates 5 leaf values from 60-69 on this stem

Leaf values: 60, 60 63, 64 and 64

(Output from SPSS)

Describing distributions

A group of scores on a variable is known as a *distribution*.

Ways of describing a distribution:

- The *location* of the distribution on the scale of measurement – measure of *central tendency*;
- The *dispersion* of the distribution – how the scores are spread out – measure of *variability*; and
- The *shape* of the distribution.

Measures of central tendency

Mean – average of the scores. A population mean is represented by the symbol 'μ'.

Median – point on the scale of measurement below which one-half of the scores lie.

Mode – score with the greatest frequency (not that useful).

Note:

- The mean is used for symmetrical data.
- The median is used for skewed data.

Measures of variability

Range – the difference between the maximum and the minimum values on the scale of measurement.

Inter quartile range (IQR) – data set divided four equal sections. Gives the range of the middle half of the data set.

Variance – average of the squared deviations from the mean.

Standard deviation (SD) – positive square root of the variance. A population standard deviation is represented by the symbol 'σ'.

Descriptive statistics – so far

Level of Measurement	Central Tendency	Dispersion
Nominal	Mode	
Ordinal	Median Mode	Range
Interval/Ratio	Mean Median Mode	Standard deviation Variance Range

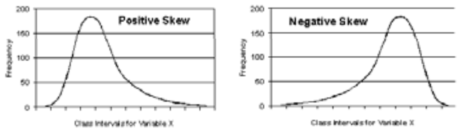
Shape of a distribution

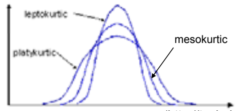
Distributions may take on a variety of shapes.

We typically focus on these aspects:

- *Skewness* – extent to which the distribution is asymmetrical.
- *Kurtosis* – extent to which the distribution is flat or peaked.
- *Modality* – number of peaks in the distribution.

Skewness and kurtosis





(http://techniques.geog.ox.ac.uk/)

A special distribution – the normal distribution

A distribution that occurs frequently in educational research is the *normal distribution*.

This is a smooth symmetrical distribution with a bell-shaped curve.

A researcher is usually more concerned with the theoretical shape than the shape of the observed distribution. It is important to know:

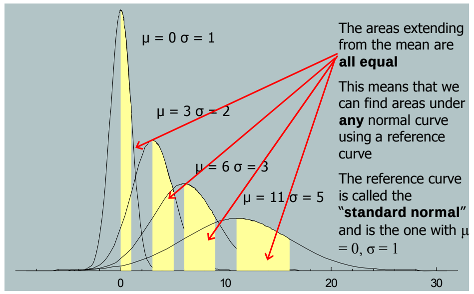
- What shape the data from the variable should take.
- What assumed distribution is required for the procedures applied to the data.

Normal curves

The normal distribution is represented by a curve with the following properties:

- Symmetrical about the mean.
- Unimodal – one peak
- Mean, median and mode are equal.
- Asymptotic to the horizontal axis – approaches but does not touch the axis
- Within the same number of standard deviations from the mean, the proportion of scores under the curve is the same as for the *standard normal curve*

The common property



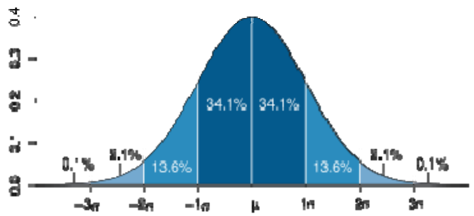
How we use the normal curve

The normal curve provides information about the percentage of cases that lie between different scores for any normally distributed data with a known mean and standard deviation.

Since the area under the normal curve is proportional to the frequency of scores in the distribution, we can determine the proportion of scores between any two points under the normal curve,

It is not trivial to find normal curve areas but computers (Excel, SPSS) can do it or we can use tables and methods described in most statistics texts.

The "68.3-95.5-99.7" rule



Descriptive vs. inferential statistics

Two broad areas:

- *Descriptive statistics* — used to describe sets of quantitative data. This involves descriptions of distributions of data and relationships between variables.
- *Inferential statistics* — used to make inferences about populations from analysis of subsets (samples) of the population.

Hypothesis testing

In inferential statistics, a *hypothesis* is used to determine whether an observation has an underlying cause or whether it was due to some random fluctuation or error in a sample.

The researcher will test to see if the hypothesis is consistent with the sample data – if not the hypothesis is rejected.

Two different ways of stating a hypothesis:

- Looking for a difference between groups;
- Looking for relationships between groups.

Hypothesis testing

On what basis do we accept or reject a hypothesis? Consider this example:

A set of exercises to encourage reflection on design was hypothesized to improve students' mechanical engineering design skills. This method was used on a class of 30 students. In a test on engineering design, the class scored a mean of 60% with a standard deviation of 10. The same test on another class that had not used these exercises, resulted in a mean score of 55% with a standard deviation of 12.

Does the hypothesis seem reasonable? What if the experimental class mean was 70%? What about 57%?

Level of significance

The *level of significance* is a probability used in testing hypotheses. It is a criterion used in making a decision about the hypothesis.

The common level used in educational research is 0.05. Occasionally other levels are used: 0.01, 0.001 and 0.1.

A level of 0.05 means that when the probability is lower than 0.05, the null hypothesis is rejected.

It then follows that if the null hypothesis is true it will only be rejected 5% of the time.

Median vs. Mean

Mean

- Most statistical tests are based on the mean – called *parametric* tests.
- The mean is badly affected by asymmetry and outliers in the data.
- To use tests based on the mean the data should have a symmetric distribution and have no outliers.
- It is also helpful, but not essential, if the data have a normal distribution
- Tests based on the mean are often the most *powerful* tests if the assumptions are met

Median

- Most non-parametric tests are based on the median
- Almost no assumptions about the data are needed in order to use *non-parametric* tests.
- If a non-parametric test exists it is almost always appropriate to use it.
- Non-parametric tests may not be very *powerful.*

Excellent health statistics - smokers are less likely to die of age related illnesses.'

Parametric analysis - assumptions

- Measurement of the data on the *interval* or *ratio* scale.
- Scores are *independent*.
- Scores are selected from a *normally distributed* population. (If the sample is large this is not important.)
- *Homogeneity of variance* – if two or more groups are studied, they must come from populations with similar dispersions in their distribution.

Common parametric statistical tests

Statistical test	Hypothesis tested
t-test (independent-samples)	There is no difference in the mean scores from two samples.
t-test (paired-samples)	There is no difference in the means of two related measures on a sample or a sample of matched pairs of subjects.
ANOVA (one-way)	There is no difference in the means of scores of two or more samples. Single independent variable.
ANOVA (two-way)	There is no difference in the means of scores of two or more samples. Two independent variables are included and a hypothesis for their interaction

Sample research questions for parametric statistical tests

Statistical test	Sample research question
t-test (independent-samples)	Do male and female students spend different amounts of time on exam preparation?
t-test (paired-samples)	Is there a difference between the scores students achieve on assignment work and exams?
ANOVA (one-way)	Is there any difference in learning with interactive visual tools, as measured by a test score, between students with visual, verbal or kinesthetic learning styles?
ANOVA (two-way)	Is there any difference in learning with interactive visual tools between novice and experienced students and with visual, verbal or kinesthetic learning styles?

t-test

- One of the most commonly used statistical tests.
- Used to determine whether the means of two groups differ to a statistically significant degree.
- There are different versions of the t-test.

Which t-test to use?

Independent samples t-test – the mean score from a group is tested against the mean score from another group. The hypothesis tested is:

$H_0 : \mu_1 = \mu_2$

$H_A : \mu_1 \neq \mu_2$

Paired or dependent samples t-test – the mean score of a group on one measure is tested against the mean score on another measure.

$H_0 : \mu_d = 0$

$H_A : \mu_d \neq 0$

One sample t-test – a mean score is tested against a particular value.

$H_0 : \mu_1 = \text{value}$

$H_A : \mu_1 \neq \text{value}$

Assumptions

Assumptions for an independent samples t-test:

- Interval scale.
- Independence of observations and groups.
- Normality – can test this in SPSS.
- Homogeneity of variance – test this in SPSS.

Independent-samples t-test example

Anxiety and Exams study

We would like to see if there are any differences between the male and female students for:

- assignment marks
- exam marks
- number of hours spent on exam preparation.

To do this we will compare the mean values of these variables for the males and females using *independent-samples t-tests*. In each case we are testing the null hypothesis of no difference between the groups.

$H_0 : \mu_1 = \mu_2$

$H_A : \mu_1 \neq \mu_2$

Independent-samples t-test – test output

Independent Samples Test

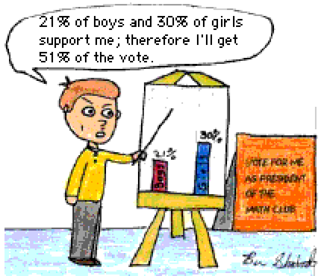
		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Assignment Mark	Equal variances assumed	1.258	.274	-.476	23	.639	-.542	1.139	-2.897	1.814
	Equal variances not assumed			-.515	20.793	.612	-.542	1.051	-2.729	1.645
Exam Mark	Equal variances assumed	9.447	.005	-2.765	23	.011	-2.375	.859	-4.152	-.598
	Equal variances not assumed			-3.410	21.654	.003	-2.375	.696	-3.821	-.929
Hours of Exam Preparation	Equal variances assumed	4.798	.039	-1.878	23	.073	-3.278	1.745	-6.888	.333
	Equal variances not assumed			-2.100	22.130	.047	-3.278	1.561	-6.513	-.042

Independent-samples t-test results

We now report the results of our tests:

Independent-samples t-tests (n=25) were conducted to evaluate the hypotheses that males and females differ in their marks for assignments and exams and in their preparation time for exams.

No difference was found in the assignment marks between the male and female students. However, the female students scored higher marks (M=7.0, SD=1.0) than males (M=4.6,SD=2.5) in the exam (t(21.7)= -3.41, p<0.05) and the female students spent longer hours (M=11.8,SD=3.1) than the males (M=8.5,SD=4.7) on their work (t(22.1) = -2.10, p<0.05). These differences were significant.



Nonparametric analysis

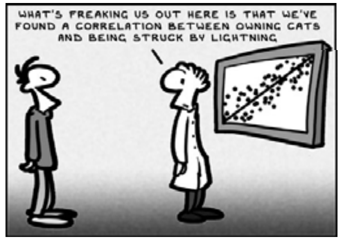
- Measurement of the dependent variable on the *nominal* or *ordinal* scale (sometimes also interval or ratio scale).
- Usually no assumptions about the shape of the distribution.
- Scores are independent.
- Nonparametric tests are less powerful than parametric tests.

Common nonparametric statistical tests

Statistical test	Hypothesis tested
χ^2 test for: ▪goodness-of-fit ▪independence	A population has a hypothesized shape. Two variables are independent in the population.
Mann-Whitney U test	There is no difference in the scores from two samples.
Wilcoxon matched-pairs signed rank test	There is no difference in scores when subjects are tested on two occasions or no difference in scores of matched pairs of subjects.
Kruskal-Wallis	There is no difference in the scores from three or more populations.

Common nonparametric statistical tests

Statistical test	Hypothesis tested
χ^2 test for: ▪goodness-of-fit ▪independence	Is the population from which this sample was drawn normally distributed? Does the occurrence of visual, verbal and kinesthetic learning styles vary according to gender?
Mann-Whitney U test	Is there any difference in the level of anxiety (as measured by a 5-point Likert scale) experienced by male and female students?
Wilcoxon matched-pairs signed rank test	Does the rating of self-efficacy of students (5-point Likert scale) change pre and post exam?
Kruskal-Wallis	Is there a difference in the grades achieved by students across four different data structure classes?



Relationship (association)

- In educational research we are often interested in the relationship between variables. E.g. *Will an increase in class size be associated with a decrease in class performance?*
- These relationships can be shown using *scatterplots*.

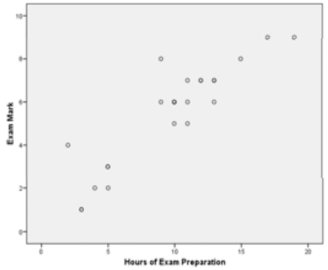
Measure of relationship

- The degree of relationship is called *correlation*.
- The measure of relationship is a *correlation coefficient*. This is represented by 'r'.
- Correlation coefficients can take in values from -1.00 to +1.00.
  - the absolute value indicates the strength of the relationship
  - the sign indicates a negative or positive relationship
  - zero indicates no *linear* relationship.
- For most statistical tests we are interested in linear relationships.

Correlation of exam marks and hours spent on study

This shows a strong positive relationship between exam marks and the number of hours spent on study.

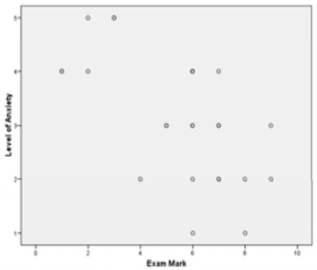
(r = 0.901)



Correlation of anxiety and exam marks

This shows a moderately negative relationship between anxiety and exam marks

(r = -0.608)



MONASH University

73

Measures of relationship

Level of Measurement	Correlation coefficient
Ordinal	Spearman rank order
Interval/Ratio	Pearson product-moment

Note that if one variable is interval and the other is ordinal then Spearman rank order should be used

74

Measures of relationship

Level of Measurement	Correlation coefficient
Ordinal	What is the correlation between rank in the final year of high school and the first year of university?
Interval/Ratio	What is the relationship between hours of study and exam performance?

Note that if one variable is interval and the other is ordinal then Spearman rank order should be used.

75

Questions ?