JUMPING TO CONCLUSIONS:
DATA INTERPRETATION BY YOUNG ADULTS

Helen L. Chick
University of Melbourne
<h.chick@edfac.unimelb.edu.au>

The increased emphasis on the Chance and Data strand of the mathematics curriculum means students should be approaching adulthood better prepared to interpret data and recognize relationships between variables and between a sample and its corresponding population. This study examines the statistical appropriateness of conclusions drawn by young adults from data in a small sample. While many of the analyses took into account the data's limitations, there were still some erroneous assumptions and omissions.

INTRODUCTION AND BACKGROUND

Statistical study has acquired a more prominent place in the school curriculum in recent years. As a consequence, Australian students approach adulthood with at least some experience in data handling and interpretation, varying in complexity from unsophisticated descriptive statistics and data depiction to hypothesis testing. Thus the depth of statistical understanding exhibited by individuals can vary hugely. Shaughnessy (1992, p.485) categorizes the sophistication of statistical conceptions into four levels, of which the first three are relevant here. The lowest of these is *non-statistical*, in which responses are based on beliefs or causality, and there is no attention to or awareness of chance or random events. In the second, called *naïve-statistical*, the individual has some understanding of chance and random events, can use simplistic judgemental heuristics and can make decisions based on experiences which may or may not be relevant to the data at hand. Individuals at the *emergent-statistical* stage have some training in probability and statistics and can apply normative models in some statistical settings. Australian students at the end of compulsory schooling will exhibit abilities in one (or more) of these levels.

One of the key ideas of statistical understanding, which may receive only intuitive coverage as part of the Chance and Data strand of the mathematics curriculum, is that of the relationship between a sample and a population. Tversky and Kahneman (1971/1982) write that many people view samples as highly representative of the populations from which they are drawn, and that small samples will show no more variability than larger samples. Such views may be held even by those at the emergent-statistical stage.

In studies which have considered students' statistical skills students rarely have been called upon to make their *own* decisions about how to investigate relationships in data. In 1987 Curcio urged that students should be encouraged to verbalize their observations of the relationships and patterns in data, but to date very few researchers have examined the extent to which students can do this. One such study by Lehrer and Romberg (1996) described the work of a group of ten children collecting and analysing data in various ways. Other studies (e.g. Chick & Watson, 1998; Lidster, Chick & Watson, 1997; Watson, Collis, Callingham & Moritz, 1995; Watson & Callingham, 1997) have been conducted using the *data cards protocol*. These studies mainly considered upper primary students and examined their abilities to represent and interpret data. The data cards protocol used there was adapted for the current study and is described in the next section.

Furthermore, while many studies have considered adult students undertaking statistics courses, little work has been done with high school students or those just leaving school on the threshold of adulthood (Batanero, Godino, Vallecillos, Green, & Holmes, 1994). The purpose of this study, then, is to examine the extent to which young adults — with an above average background of high school Chance and Data — can recognize relationships in a set of data, and consider the legitimacy of their conclusions about the data.

METHOD

The 32 students (21 males and 11 females) involved in this study were enrolled in a first year university basic mathematics course intended as a service subject for students needing some mathematics as part of a science or arts degree. It had no statistical content. One of two pre-tertiary mathematics subjects was an acceptable pre-requisite for the university course; each of these subjects had a statistics component. In particular, one of the pre-requisite subjects had, as one of its aims, that students develop skills in collecting, organising, displaying and interpreting numerical data as well as calculating summary statistics from this data. The other subject, used as a pre-requisite for mainstream tertiary calculus courses, incorporated the development of certain specific statistical skills associated mainly with normally distributed data; more descriptive and intuitive data interpretation was a component of mathematics subjects studied prior to this.

The *data cards protocol* used for this study supplied students with details about a set of 16 fictitious Tasmanian juveniles; giving name, age, weight, weekly fast food consumption, favourite activity and eye colour information for each one. The Appendix gives the complete set of data. In previous studies the students were supplied the data on 16 separate cards (hence the name of the protocol); in the current study the information was supplied in a table like that in the Appendix. Students were then asked to investigate the data, in such a way as to allow students to choose their own areas of focus. This made the task exceptionally open-ended, with no discernible end-point indicating when students had achieved "the answer." In this study the protocol was set as an assignment question:

> "Examine the data and produce a report which highlights any aspects of the data which you think are interesting. You are free to choose your approach to this problem and what to include in your report. If you make any claims about the data be sure to back them up with suitable evidence."

In response to this question, students examined various aspects of the data, producing written reports which occasionally included graphs and summary statistics, and which varied in complexity from simple frequency information to investigating associations between variables and discussing causes and effects. For the purpose of this study, the focus is on the conclusions drawn, not on the type of representations used nor the particular relationships examined. Legitimate conclusions which can be drawn about the data include (a) a very strong correlation between age and weight; (b) boys appear to be marginally heavier than girls; (c) boys, and older boys in particular, tend to consume more fast food meals per week; and (d) there is a strong correlation between fast food consumption and the activity or passivity of favourite activity, with high consumption associated with more passive activities, such as TV, reading, and board games. It should be noted that since the data was supplied to students the tasks associated with the protocol involve descriptive statistics rather than hypothesis testing.

It must also be said that the protocol's deliberate avoidance of specific directions means that we have to be cautious about the conclusions that this study might attempt to draw. For example, the students were *not* asked to discuss whether their observations of the sample might apply to a broader population. Consequently, a student's failure to address this does not mean that the concept of the relationship between a sample and the population from which it is drawn is not understood. If, however, we are interested in developing statistical skills in students to the extent that they can make valid judgements about data in various contexts, it is to be hoped that they will realise for themselves that this is an issue that needs to be considered, and will consider it. The extent to which students spontaneously do this, then, is one focus of this study.

RESULTS AND DISCUSSION

In looking at the nature of students' conclusions this study will examine the extent to which students were able to recognize that (i) the small sample size of the supplied data limits the applicability of conclusions to a broader population; (ii) the small sample size of the supplied data

may invalidate conclusions drawn about the data itself due to the higher variability of data in small samples; and (iii) relationships between two variables may be influenced by a third or fourth variable, that is, correlation does not imply that one variable is the sole effect on the other. Table 1 indicates what kind of conclusions and clarifications the students made, based on the data in the protocol. Some students' responses were classified as being in more than one category.

Table 1
The nature of students' conclusions on analysing the data cards problem

Conclusion type		% (N=32)
a.	Over-generalized or made a claim not reasonably supported by the data	34
b.	Specifically recognized that conclusions only apply to the given data	47*
c.	Observed that sample is too small to apply to the real world	31*
d.	Observed that unbiased selection of the sample is required	19*
e.	Some use of external reference or knowledge outside the data itself, either cited or asserted as fact	56
f.	Gave a qualified hypothesis to explain an observed phenomenon (e.g. "this *could* be because of ...")	66
g.	Did *both* f and a (i.e. some hypotheses qualified, other claims were too strong)	25
h.	Mainly listed some summative statistics and observations from the data only	38
i.	Did h but *not* b (i.e. focussed only on data but made no reference to data's limitations)	25
j.	Made an unwarranted assumption about the data (as opposed to an unwarranted conclusion)	9
k.	Made and tested an hypothesis (and may have accepted refutation of the hypothesis)	6
l.	Overlooked an important third component where a two variable correlation exists	41
m.	Recognized a third component influencing a seeming correlation	31
n.	Did *both* l and m	9

* Note that the set of students doing d is a subset of those doing c, which is a subset of those doing b, except for one student in d and b but not c.

The Sample and its Relationship to the Population

As can be seen from Table 1, 47% of the students made a comment to the effect that their conclusions only apply to the given data; of these, 66% went further and stated that the sample was too small to apply to a larger population. For example, in acknowledging that her observations apply to the data only, one student wrote "This shows that *of the students surveyed*, males had the higher number of fast food meals consumed per week [emphasis added]." [In these quotations of student work, spelling and more obvious punctuation errors have been corrected.]

Another student was more specific about the fact that the sample size is too small for conclusions to be drawn. She also expressed the need for a representative sample.

> A sample group of 16 may be a little small for an accurate conclusion to be made on Tasmanian school children; the larger the sample the more accurate a picture will be, and also a variety of children from the North, South, east and western schools including public and private schools would draw more accurate conclusions.

One student did not acknowledge that sample size was an issue, but was also concerned about how representative a sample it was, noting "that the table of results may have invalid data, i.e is the data a true representation of the cross-section of Tasmanian school children?" This point was raised by 19% of the students, with all but this particular student recognizing the issue of sample size as well.

In contrast to these expressions of caution, there were some unjustifiable conclusions and generalizations among the students' responses. In fact, over one-third of the students made some sort of statistically inappropriate interpretation of the data. In some instances students' difficulties were a consequence of trying to reconcile the data supplied with their own knowledge and expectations. One student, for example, wrote that "with 50% of the female population playing netball it is easy to understand how it has become one of Australia's largest followed sports." Interestingly, he was only talking about the younger students at this time. Another noted that males eat more fast foods per week and then concluded "Clearly by this information, it is possible to make the assumption that females take more care in their appearance, and their health." A third student, on observing that swimming was the favourite activity of only one person in the sample, claimed "Swimming is becoming an obsolete favourite activity."

One student drew on some external references to make some sweeping claims, based on her observation that no boys chose reading as a favourite activity. There is an implicit assumption on the part of the student that the data is representative of a larger population, that the small sample really does illustrate the claimed relationship, and that the observations really are a consequence of the claimed causes.

> These findings clearly show the relationship between gender and literacy skills [web site cited]. It can be shown that males are much slower at picking up reading and writing skills than females, especially at primary school level [another citation] ... Thus lower literacy levels in males, and their reliance on television for information, makes them more susceptible to advertising, especially food advertising, and because they are not as good at analysing the information they will be at a higher risk of not understanding the concept of a "well-balanced diet" and the importance of eating healthy food regularly ... Therefore literacy in children affects not just information gathering and analysis, but health and physical shape.

In contrast, one student had a particularly good understanding of the limitations of the data. The final two sentences in the quote below are her report's concluding remarks.

> In such a small sample it is hard to define what is correlated and what is coincidence. For instance, all [those] who enjoy television are males and as it happens, three out of those four children have blue eyes, yet it is likely that with a larger sample size, enjoying television wouldn't be dominated by blue-eyed males (especially the blue-eyed part) ... This type of data possibly demonstrates if you want to find relationships, you probably will, but with little evidence to support it. Without substantial evidence you cannot draw conclusions.

Students' Conclusions About the Sample Itself

Students had no difficulty making observations about various aspects of the data itself. As will be seen later in this report, many attempted to give explanations for the trends that they noticed. Very few students, however, were aware that some of their speculated conclusions could be due to random fluctuations. One student recognized that "The number of students is way too small to make general statements or assumptions from as there could exist ... coincidences within the data or the sample could be unrepresentative of the whole population."

While making general statements about the data, one student did not make it clear that in stating that "the majority of board game players are female," there are only three children in the set of board game players. This, of course, is the minimum-sized set required to have a non-unanimous majority! While his statement is true as a factual observation of the data, there is no indication that he is aware that (a) his statement is very strong in the light of so little evidence, and (b) it only takes a change of mind by one person in the sample to alter his conclusion.

One student made what could be regarded as a valid observation about the data, but backed it up with the example of two individuals rather than attempting a more legitimate statistical analysis. He wrote that "fast food consumers tend to be heavier than lower fast food consumers over the age range for both males and females. i.e. Mary Minski is 13 years old and weighs 55 kg whereas Dorothy Myers is 15 years old and weighs 50 kg." (It is conceded that the student may have confused "i.e." with "e.g." but more substantial evidence should still have been supplied.)

Three students made unwarranted assumptions about the actual data itself. Two asserted that two of the boys in the data were brothers (it *is* true that there is some evidence in favour of this possibility, and an additional student made this distinction), while another student claimed that a variety of children had been selected, with "ethnic background" a factor in that variety, based on the occurrence of some non-Anglo-Saxon surnames.

As a final example of students' difficulties with the sample itself, one student attempted to assign children of particular ages to particular school grades, allowing only one age per grade grouping (thus precluding the possibility that, for example, a grade 6 class could have both 11 and 12 year olds in it). He tabulated his results and then said that the "table seems to suggest there is some data missing. It seems logical that if in some grades one boy and girl were interviewed then there should be a girl and a boy from all grades." His identification here could, in fact, be an implicit expectation that the data be "representative" in some sense.

Influences on Relationships Between Variables

Qualified Hypotheses

Of the 12 students (38%) whose main approach in writing their reports was to list some summative observations from the data, only four were among the 21 students (66%) who proposed *possible* reasons for their observations. The remaining eight of the 12 made no attempt to give reasons for the results at all, just noting their observations without comment. The three students who neither hypothesised tentative reasons nor merely listed factual observations were among those making unjustified sweeping claims. The existence of a cautiously proposed hypothesis indicates that students *are* seeing the data in the broader context of a population, but that they appreciate that there is insufficient evidence to be sure of the claim. As will be discussed later, many students used "outside" knowledge or experience as a source of explanation for identified trends.

After observing that males ate more fast food than females and yet the average weights differed by only 4.4kg, one student wrote "This could mainly be accounted for by the fact females mature faster than males;" his use of the word "could" permits the possibility of an alternative explanation or recognizes the need for appropriate evidence. Another used three qualified statements in attempting to explain the connection between fast food consumption and television watching. It should be noted that very few students, and not this one, recognized that "favourite activity" did not mean "most frequent activity."

> One of the ... assumptions [sic: observations] that can be made from the data is that people who enjoy watching TV, also eat at least 7, and up to 12, fast food meals a week. This could be because of television advertising fast food, such as McDonalds ... The people who watch TV the most, would most likely have no sports commitments and so would not have strict diets to cling by. This could be one reason for being the heaviest in their age group.

After making the connection between the same variables, one student wrote that "it could possibly be stated that television causes children to eat an excess of fast food." Here the conditional "could" again indicates statistical caution, but no further explanation is offered.

Awareness of Factors Which Influence Relationships Between Two Variables

Many students were able to identify some two-variable correlations in the data. In many cases, however, these correlations were influenced by a third variable; this is best exemplified by the possible interconnections between age, weight, fast food consumption and the passivity of the favourite activity. In identifying a particular two-variable correlation, less than one-third of the students were able to recognize the existence of a third variable which would almost certainly influence the relationship, while 41% of students failed to identify such a factor despite noting the two-variable correlation. There were three students who did both: identifying a third variable on some occasions and failing to do so on others, sometimes even in adjacent paragraphs.

One student, who *was* able to recognize a third (or even fourth) correlating variable, observed that higher fast food consumption indicated greater weight, then noted "This could also be attributed to the gender difference (males tend to be heavier than females) and the age difference (the older people get the heavier they get (as they grow up))." Two students considered three variables successfully by considering the age to weight ratio, and then compared that with fast food consumption (in one case) and passivity of favourite activity (in the other).

In contrast, several students failed to take age into account when comparing weight with other factors. One student simply wrote "Those that play sport have a lower weight than those who prefer other activities such as TV or board games." Another actually calculated the average weights of the "active" and "inactive" children, before writing "It can be compared that the average weights of children that are more likely to have an outdoor activity are less than the rest, 34.3 kg [compared with] 52.6 kg," again failing to take age into account (the older, and thus heavier, children prefer the passive activities). Others made similar connections between weight and favourite activity while omitting the age link, yet after making this observation one student's very next paragraph stated that "it becomes apparent that age is also a major factor. It [is] noted that all students of age 13 and over (adolescents) have a weight that is above average, regardless of the amount of fast food meals eaten per week." This student, like one or two others, failed to make any further connection between these observations.

As a final example, one student identifies a possible third factor not included in the data in order to explain a two-variable relationship but overlooks a possible third variable (fast food consumption) which *is* present in the data.

> If we graph the weight to age then as the age rises so does weight as expected, but, in the 12 to 15 age group there seems to be a significant rise above the normal trend. This could be because of puberty and the accelerated growth rate associated with it.

The Use of External Knowledge and Assumptions

As indicated earlier and in some of the quotes already seen, some students interpreted the data in light of knowledge and assumptions based on their experience or other research. In fact, over half the students (56%) made some use of external information, either explicitly citing it as such or incorporating it without comment. The correlation between age and weight, for example, was regarded as unsurprising common knowledge (when it was considered at all); some students also regarded the connection between weight and fast food consumption in a similar light. One student's response illustrates both aspects:

> To think about it logically, the more fast food consumed, the heavier the child is going to weigh. But also the child's age can come into the equation. The older the child the heavier he is more likely to weigh, so it is going to be hard to determine that weight is a major factor involved.

At least three students indicated that eye colour should be irrelevant, although two of them observed that a possible correlation existed. One of these students was also one of two who actually hypothesised relationships in advance before testing them with the data. He wrote:

> Now we can look to see if eye colour has any relation to the amount of fast food consumed per week. My prediction is that it won't have any effect on the result, simply because it is a physical feature of each student. But, in fact, it seems as though it does play a major part in determining the number of meals consumed. [Shows table supporting his claim.] Well, there you have it. Blue eyes seem as though they play a major part, but as pointed out before they are only a physical feature, and it doesn't make sense that it would be possible to make a direct effect on the outcome. In my opinion it is only a coincidence that it appears to show to have any direct effect.

CONCLUSION

The statistical sophistication of the students in this study seems to range from non-statistical to emergent-statistical. There are examples of unsupported conclusions and tendencies to over-

generalize, contrasted with examples of deeper awareness of statistically appropriate analysis. Most of the analysis carried out by the students was fairly simplistic, but this is hardly surprising given the size of the sample and the vagueness of the protocol. As mentioned earlier, the deliberate non-specificity of the task has implications for what might be expected from students in their responses. There was no specific request for them to make generalizations from the data nor to discuss implications nor to propose explanations for observed aspects of the data. Consequently it should not be surprising that some students did not address these issues nor that they took a fairly simplistic approach to interpreting the data. Nevertheless, if one of the objectives of the mathematics curriculum is to improve statistical literacy, then the results of this study suggest that students should be given more opportunities to deal with drawing valid conclusions from raw data, to identify relationships between variables, and to address the extent to which claims about a population can be made based on a sample.

REFERENCES

Batanero, C., Godino, J.D., Vallecillos, A., Green, D. R., & Holmes, P. (1994). Errors and difficulties in understanding elementary statistical concepts. *International Journal of Mathematical Education in Science and Technology*, 25, 527-547.

Chick, H.L., & Watson, J.M. (1997). *The ups and downs of collaboration in mathematics*. Manuscript submitted for publication.

Chick, H.L., & Watson, J.M. (1998). Showing and telling: Primary students' outcomes in data representation and interpretation. In: Clive Kanes, Merrilyn Goos, Elizabeth Warren (Eds.) *Teaching mathematics in new times* (Proceedings of the Twenty-first annual conference of the Mathematics Education Research Group of Australasia Incorporated, pp.153-160). Gold Coast, QLD: MERGA.

Curcio, F.R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education*, 18, 382–393.

Lehrer, R., & Romberg, T. (1996). Exploring children's data modelling. *Cognition and Instruction, 14*, 69–108.

Lidster, S.T., Chick, H.L., & Watson, J.M. (1997). Developing cognition in interpreting data. In: N. Scott and H. Hollingsworth (Eds.) *Mathematics — Creating the future* (Proceedings of the 16th Biennial Conference of the Australian Association of Mathematics Teachers, pp. 202–209). Adelaide: Australian Association of Mathematics Teachers.

Shaughnessy, J.M. (1992). Research in probability and statistics: Reflections and directions. In D.A. Grouws (Ed.) *Handbook on Research in Mathematics Education* (pp. 465–494). New York: Macmillan.

Tversky, A., & Kahneman, D. (1982). Belief in the law of small numbers. In D. Kahneman, P.Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp.24–31). Cambridge, U.K.: Cambridge University Press. (Reprinted from 1971 issue of *Psychological Bulletin, 76*, 105–110).

Watson, J.M., & Callingham, R.A. (1997). Data Cards: An introduction to higher order processes in data handling. *Teaching Statistics, 19*, 12-16.

Watson, J.M., & Chick, H.L. (1997). *Collaboration in mathematical problem solving*. Paper presented at the Australian Association for Research in Education Conference, Brisbane, Australia.

Watson, J.M., Collis, K.F., Callingham, R.A., & Moritz, J.B. (1995). A model for assessing higher order thinking in statistics. *Educational Research and Evaluation, 1*, 247-275.

APPENDIX — DATA SUPPLIED TO STUDENTS

Table 2
Table of data supplied to students for interpretation

Name	Gender	Age	Favourite Activity	Eye Colour	Weight (kg)	Fast food meals consumed per week
Adam Henderson	M	12	Football	Blue	45	5
Andrew Williams	M	14	TV	Blue	60	10
Anna Smith	F	11	Boardgames	Brown	32	1
Brian Wong	M	9	Football	Green	26	1
David Jones	M	8	TV	Blue	30	7
Dorothy Myers	F	15	Swimming	Blue	50	2
Janelle MacDonald	F	18	Reading	Blue	66	4
Jennifer Rado	F	9	Boardgames	Green	33	4
John Smith	M	10	Football	Green	29	0
Kathy Roberts	F	12	Netball	Brown	32	0
Mary Minski	F	13	Reading	Green	55	3
Peter Cooper	M	16	Boardgames	Green	54	2
Rosemary Black	F	8	Netball	Brown	24	0
Sally Moore	F	17	Reading	Brown	56	1
Scott Williams	M	17	TV	Blue	66	8
Simon Khan	M	18	TV	Brown	74	12