The Hospital Problem Revisited. Tertiary Student's Perceptions of a Problem Involving the Binomial Distribution

Robyn Reaburn
University of Tasmania
<rreaburn@utas.edu.au>

This paper considers the intuitive solutions of 26 first year tertiary students to a binomial probability problem on entry to a statistics unit. For this problem a successful solution requires consideration of the sample size. On the basis of a Rasch analysis, students were classified into three groups according to their ability, and the reasoning they used compared. The problem was again posed at the end of the unit and the answers and reasoning used by the students were compared with their earlier responses.

Half of all newborns are girls and half are boys. Hospital A records an average of 50 births per day. Hospital B records an average of 10 births per day. On a particular day, which hospital is more likely to record 80% or more of female births?

A variation of this problem was first reported by Tversky and Kahneman (1982) in a study of the heuristic rules used by undergraduate students to judge events that are uncertain in outcome. Since then this same problem, or variations of it, have been used in studies of school students (Fischbein & Schnarch, 1997, Watson & Moritz, 2000), and pre-service teachers (Watson, 2000). This study describes the types of reasoning used by tertiary students to answer this question on entry to a statistics unit, and after the unit was completed. This question was used as part of a wider study to investigate students' intuitive reasoning in statistical inference.

To answer the Hospital Problem successfully, it is necessary to look beyond the proportions and to appreciate the effect of sample size. In the studies cited, the most common answer given was that the likelihood of recording more than 80% of female births was equal for both hospitals. Tversky and Kahneman (1982) refer to this as an example of the representativeness heuristic, where samples are assumed to be more like the overall population than sampling theory suggests. In the hospital problem, this heuristic leads to the conclusion that sample size is not relevant, that as the two events are described by the same statistic they will be equally representative of the general population. Sampling theory, however, suggests that the smaller sample is more likely to deviate from the 50% rate of births for each gender (Tversky & Kahneman, 1982). As a sample increases in size, the sampling statistic (here the proportion of girls born) is more likely to approach the theoretical value for the entire population (Fischbein & Schnarch, 1997).

In Tversky and Kahneman's study, 53 out of the 95 undergraduate students answered that the two hospitals were equally likely to record an uneven proportion of births. Fischbein and Schnarch gave a similar question to students in grades 5, 7, 9, 11 and to college students who were prospective teachers specialising in mathematics, none of whom had previously studied probability. The students in the lower grades had a high number of non-responses. When the question was answered by the younger students, the most common response was that of the largest hospital. As the age of the students increased, the number of responses also increased, and whereas the likelihood of choosing the larger hospital decreased, the likelihood of answering that the events were equally likely increased. Out of the 18 college students 16 gave the answer of equal likelihood (the other two did not respond). Fischbein and Schnarch suggested that as the understanding of ratio improved with age, this understanding became dominant at the expense of an understanding of the effect of sample size. Out of the whole study only one grade 9 student gave the smaller hospital as the answer.

Watson and Moritz (2000) interviewed 62 students from grades 3, 6 and 9 from a variety of school regions, including suburban and rural schools in Tasmania. There were equal numbers of males and females. During these interviews students were asked about the size of a sample needed to study the weights of grade 5 children, and for the grade 6 and 9 students were then asked a variation of the hospital problem. The question was:

> The researchers went to two schools: One school in the centre of the city and one school in the country. Each school had about half girls and half boys. The researchers took a random sample from each school: 50 children from the city school, 20 children from the country school. One of these samples was unusual; it had more than 80% boys. Is it more likely to have come from:
>
> - The large sample of 50 from the city school, or
> - The small sample of 20 from the country school, or
> - Are both samples equally likely to have been the unusual sample?
>
> Please explain your answer.

Out of the 41 respondents, only 8 chose the small sample, with only 6 of these being able to give adequate reasons. Those students who picked the larger sample suggested that as there were more children to pick from, there were more children to get the higher number of boys. The most common response was that of equal likelihood (61%), and the proportion of this response did not vary between the grade 6 and grade 9 students. The reasons given were either that the process was random or because each school population from which the samples were taken had a 50% occurrence of each gender. It is apparent from this study that the context of the question is of importance. In an earlier question students had been asked about the number of students needed to study the weights of grade 5 children. Eighty percent of the students who had stated that larger samples were needed to study the children's weights did not recognize that a smaller sample was more likely to give extreme results in this question.

Watson (2000) gave the hospital problem to 33 preservice students who were all in a post-graduate teaching program. There was wide variation in the mathematics background of these students; 23 had at least studied mathematics up to the second year as part of their previous university courses (one of these was on leave from a PhD enrolment in mathematics), and 10 had less than this. They were given the hospital problem to work on overnight and asked to complete it on their own.

This study recorded the reasoning used by the students. The students were divided into those who used intuitive reasoning only, mathematical reasoning only, or a combination of the two. The mathematical reasoning was divided into whether the binomial distribution was used (formal), or more elementary mathematics such as percentages were used (basic). The results are summarised in Table 1.

It is apparent that the students who used mathematical arguments alone were more successful than those who used intuition alone. It is also apparent that mathematics alone is not entirely successful, as those who made errors in their formal mathematical calculations were unaware of their error. However those who used both mathematical and intuitive reasoning were only 50% successful.

Table 1

Responses to the Hospital Problem (Watson, 2000)

	Strategy		
Correctness of conclusion	*Intuition*	*Mathematics*	*Intuition and Mathematics*
Correct (n = 18)	7	Formal: 2 Basic maths: 6	Formal: 3
Incorrect (n = 15)	8	Formal: 4	Basic maths: 3
Total	15	12	6

Method

The study of the Hospital Problem described in this paper is being carried out at as part of a wider study of students' intuitive statistical reasoning and inference at an Australian university. On entry to the unit the students were given a questionnaire where they were required to interpret probability statements, to recognise independence and sampling variation, and to make simple inferences. The Hospital Problem, as described in the opening paragraph, was part of this questionnaire. At the end of the unit the students were given another questionnaire that required them to make statistical inferences and explain their reasoning. The Hospital Problem was one of three questions that were repeated from the first questionnaire.

Participants

The participants were volunteers who were enrolled in a first year statistics unit. This unit is a service course for students who are studying Biomedical Science, Aquaculture and Environmental Science. The unit is also taken as an elective by students studying Health Science, Computing, and Education. The initial questionnaire was given to 26 students. Of these, one had studied mathematics at year 11, 20 had studied mathematics at year 12, one at TAFE, and four at University. Nineteen of these 26 reported that they had studied some form of statistics in their last mathematics course. Due to circumstances beyond the researcher's control, the second questionnaire was completed by nine of these students.

Data Collection and Analysis

Answers to the questionnaires were rated according to the SOLO taxonomy. With this taxonomy the answers were scored so that answers that showed more sophisticated levels of statistical thinking were given higher scores. A Rasch analysis (Bond & Fox, 2007) which simultaneously gives a score for both the items and individuals, was used to rank the students and items on all the items on the first questionnaire. Based on this analysis, the students were divided into three groups, above average, average and below average. These groups were then examined to see if there was any pattern in the type of response according to ability. After the second questionnaire, the answers to the Hospital Problem were then examined to see how these answers may or may not have changed.

Results and Discussion

With the Rasch analysis the Hospital Item showed misfit ($z = 5.4$), suggesting that the students were using a different form of reasoning for this question than for the other items in the questionnaire. Using the Rasch rankings, the students were divided into three groups, above average (0.9 logits or above), average (-0.13 to 0.75 logits), and below average (-0.27 logits and below). The overall responses to the Hospital Problem in the first questionnaire are summarised in Table 2. There was one non-response to this question.

Table 2

Pattern of Response to the Hospital Problem

	Previous study of statistics	
	Yes	No
Hospital A (incorrect), n = 3	3	0
Hospital B (correct), n = 12	5	7
Equally likely (Incorrect -main inappropriate conception), n = 10	9	1
Total	17	8

When asked to explain their answers, all the students who chose Hospital B used a form of reasoning that showed that they recognized that it was more likely, or as some stated, "easier", for Hospital B to deviate from the 1:1 ratio. For example: It is more likely that 8/10 will be female than 40/50, as it only requires that 3 births are female instead of the probable male births, instead of 15 that is needed to be female.

All the students who chose Hospital A used a form of reasoning that suggested that as there are more births to choose from, it is more likely that there will be more female births. These answers are similar to those noted by Watson and Moritz (2000) for the grade 6 and 9 students. For example: Because Hospital A has more births each day than Hospital B, it is likely that there will be more female births too.

Two of the students who chose the equally likely option used reasoning that involved proportions or ratios. For example: Percentage is independent of the total number of births, it is a proportion.

The other students who chose this option used reasoning that involved the constant probability for each individual outcome. For example: The likelihood of gender is individual to the delivery not on the hospital and the number the hospitals deliver.

The responses according to student ability and the form of reasoning used are summarised in Table 3.

None of the above ability students chose Hospital A or used proportional reasoning. The least sophisticated reasoning, that is there are more births in Hospital A therefore more girls would be born, was evenly spread between the average and below average groups. All of these students who chose Hospital A had stated that they had been exposed to statistics in their school mathematics. Of all the students who chose Hospital B, only one, who was in the average group, specifically mentioned the effect of increasing sample size: A large amount of births will allow the average of boys to girls to even out. Hospital B has a lower amount of births and has a higher chance of reaching 80% female.

Table 3

Types of Responses Used in the Hospital Problem Grouped by Students' Ability

		Type of reasoning used			
Grouping	Answer	More likely for Hospital B	Independence of each single birth	Proportional reasoning	More births in hospital A
Above average	Hospital A				
(n = 9, 1 no answer)	Hospital B	4			
	Equally likely		4		
Average	Hospital A				2
(n = 12, 1 no answer)	Hospital B	6			
	Equally likely		2	1	
Below average	Hospital A				1
(n = 5)	Hospital B	2			
	Equally likely		1	1	
	Total	12	7	2	3

In the last week of semester the students were then given a second questionnaire that also included the Hospital Problem. Responses were available for nine students. The responses and types of reasoning used by these students in both questionnaires are described in Table 4.

Four out of the five students who had initially chosen the equally likely option had now chosen Hospital B so now all except one student now gave the answer of Hospital B. Of interest is that the ability of the student who gave this exceptional response was rated as above average. The student quoted earlier, who acknowledged that the expected statistic will be approached with a higher sample size, stated a similar argument in the second questionnaire, whereas one student (in the above average group) also now acknowledged the effect of an increasing sample size.

There was no specific intervention to address the inappropriate reasoning displayed in this question, but during the statistics unit the students did study probability. During this module the students were introduced to the definition of probability in terms of long term frequencies, and the length of run it might take for coin tosses to reach a 1:1 ratio was discussed. The students had also used the binomial distribution for the calculation of probabilities. While no definite conclusions can be drawn from this small number of students, it is encouraging that some students were able to make the transition from purely proportional reasoning to consideration of sample size.

Table 4

Comparison of Reasoning Used in the Hospital Problem in Questionnaires 1 and 2

		Questionnaire 1		Questionnaire 2	
Person	Grouping	Answer	Reasoning	Answer	Reasoning
1	A. average	Equal	Constant probability	Equal	Constant probability
2	A. average	B	More likely for B	B	Effect of larger sample size
3	Average	A	More births, more girls	B	More likely for B
4	Average	Equal	Constant probability	B	More likely for B
5	Average	Equal	Constant probability	B	More likely for B
6	Average	B	More likely for B – effect of larger sample size	B	More likely for B – effect of larger sample size
7	Average	A	More births, more girls	B	More likely for B
8	Average	B	More likely for B	B	More likely for B
9	Average	B	More likely for B	B	More likely for B

Conclusions and Recommendations

It is apparent that with the right experience, students can move from proportional reasoning only to reasoning that allows for the effect of sample size. It is of interest that in the initial questionnaire all three students who gave the larger hospital as their response, and nine out of the ten who said the two hospitals are equally likely, had all studied statistics in a previous course. In contrast, seven out of the 12 correct responses came from students who had not studied statistics previously. It could be inferred that the previous studies in statistics had, at the least, not helped in their reasoning. It would be of interest to determine the details of the students' previous mathematical experience to see if the dominance of proportional reasoning comes about by the students misconstruing course content, or whether this dominance has come about purely by lack of experience with sampling.

The unit completed by the participants in this study was an applied statistics unit, and required a lower level of mathematical ability than statistics units that are theoretically based. It would also be of interest to see how students in theoretical statistics courses which require this higher level' of mathematical ability, can apply their theory to practical problems such as this.

References

Bond, T. & Fox, C. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences.* Mahwah, NJ: Lawrence Erlbaum Associates.

Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education, 28*, 96-105.

Tversky, A., & Kahneman, D. (1982). Judgement under uncertainty: Heuristics and biases. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases.* New York: Cambridge University Press.

Watson, J. (2000). Preservice mathematics teachers' understanding of sampling: Intuition or mathematics. *Mathematics Teacher Education and Development,* 2, 121-135.

Watson, J., & Moritz, J. (2000). Developing concepts of sampling. *Journal for Research in Mathematics Education,* 31, 44-70.