Attendance Rate Survey

of Science Level 0-1 Modules

Survey Conducted by Survey Sampling Class Stat 30020/40220. 2008/09

Paul Behan
Aidan Boland
Ronan Daly
Simon Diver
Mark Forrest
Mark Grennan
Ciaran Horrigan
Samuel Ogwu Lekwadi
Donna Moynagh
Fabian Ofurum
Amanda Reilly
Caitríona Ryan

University College Dublin

26 / 11 / 2008

THE SURVEY SAMPLING CLASS OF 2008/2009 WHO CONDUCTED THE ATTENDANCE RATE SURVEY FOR THE TWO UCD SCIENCE COLLEGES.

Back Row (left to right): Amanda Reilly, Donna Moynagh, Fabian Ofurum, Samuel Lekwadi, Prof Stephen Gardiner, Ronan Daly, Aidan Boland, Caitriona Ryan.
Front Row (left to right): Dr Gabrielle Kelly (Module Coordinator), Mark Forrest, Mark Grennan, Simon Diver, Associate Dean Julian Menuge, Dean of Sciences Mark Rogers, Paul Behan.

Summary

The class of Survey Sampling, Stat 30020 and Stat 40220, carried out a survey to estimate the attendance rate at lectures in science modules in UCD.

The objective of the survey was to estimate the overall attendance rate and also to compare the attendance rate at different times and days during the week, to see how it varied. Our survey also aimed to give us practical experience in carrying out a survey, and of all the decisions that need to be made when designing a survey.

The survey focused on the attendance rate of modules 0 and 1 in the two science colleges: UCD College of Engineering, Mathematical and Physical Sciences, and UCD College of Life Sciences, as it is known that the majority of failures and drop-outs are among first year students, and addressing attendance rates is a first step in addressing the problem.

A stratified design was used where modules were subdivided into strata by time of day and day of the week and a simple random sample of modules selected from each stratum. Each of the 12 students participating in the survey was allocated to 3 - 4 modules to count the number of students at half past the hour attending the module.

The survey was conducted in the week of October 24, 32 classes were sampled out of a possible 84 classes; a ratio estimation method was used to compute the overall attendance rate.

The Overall attendance rate was 47.3% $\pm$ 4.4%. The lowest attendance rate was observed on Friday with a rate of 22.3% $\pm$ 4.7 %

1. Introduction

1.1 Background

The class of Survey Sampling, Stat 30020 and Stat 40220, carried out a survey to estimate the attendance rate at lectures in science modules in UCD. The objective of the survey was to estimate the overall attendance rate and also to compare the attendance rate at different times and days during the week, to see how it varied. Our survey also aimed to give us practical experience in carrying out a survey, and of all the decisions that need to be made when designing a survey.

We decided to survey the attendance rate of modules 0 and 1 in the two science colleges: UCD College of Engineering, Mathematical and Physical Sciences, and UCD College of life Sciences, with the exception of engineering modules. We felt that engineering modules were very practical-work orientated and so would have a higher attendance rate, and so these modules were omitted from our sampling frame. It is known that the majority of failures and drop-outs are among first year students, so the results of this survey are of interest as it is felt attendance may be a factor in the high failure rate. The results of the survey are, thus, of great interest to Professor Nick Quirke, the two College principles, the Dean of Science, and also to all module co-ordinators and students.

1.2 Data Assembly

<u>Labs/Tutorials</u>

We decided not to sample tutorials or labs as tutorial classes are often compulsory and there is sometimes more than one offering. It is also very difficult to obtain the list of all tutorial classes, their size and venues.

<u>Strata</u>

After careful consideration, we decided on a design that used modules as the unit of the sample and that divided the modules into strata. Our instructor provided us with a sampling frame listing all the relevant modules from the two colleges offered in semester 1. If a module was offered twice in the week it was listed twice, and if it was offered three times it was listed three times etc. We decided to choose our strata based on the day of the week and time of the day. We originally chose to have 6 strata which were as follows;

Strata	Day	Time
1.	Monday	9-11
2.	Monday	11 onwards
3.	Tuesday/Wednesday/Thursday	9-11
4.	Tuesday/Wednesday/Thursday	11 onwards
5.	Friday	9-11
6.	Friday	11 onwards

1.3 Selecting the sample

We chose our sample size in order to have an appropriate bound on the error of estimation and allowing for feasibility, as only twelve students participated in the survey. We decided on sampling 43 modules (3-4 modules each) out of a total of 86 modules. Dr. Gabrielle Kelly obtained the sampling frame of the total list of level 0 & level 1 modules for us. We used simple random sampling to select modules from each stratum by means of a random number generator. Students volunteered to survey particular modules that suited their timetable until every student had three or four modules to survey.

1.4 Preparation

Pilot

A pilot study allowed us to assess which time (quarter past the hour or half past the hour) was more appropriate to take attendance, and if it was feasible for only one student to count a large class size. We chose four modules to include in our pilot study, two large classes and two smaller classes. One student took attendance in one large class at a quarter past the hour and one student took attendance in the other large class at half past the hour. Likewise, one student took attendance in one small class at a quarter past the hour and one student took attendance in the other small class at half past the hour. We concluded from our pilot study that it was best to take attendance at half past the hour and that one student was enough to take attendance of a large class.

Letter

We sent a letter to the lecturers of all the modules in our sampling frame, asking their permission to use their module in our survey.

1.5 Timing of the Survey

We carried out our survey in week 7 of the semester. We note that this was the week before the bank holiday weekend, which may have affected stratum 5 and 6 attendance rate, and also that a student protest took place on Thursday afternoon in the city centre that week, which may have affected stratum 4's results.

1.6 Last year's results

Last year, the 2007 class of Survey Sampling carried out a survey similar to ours, which gave us an indication of what to expect. That class recorded attendance in the School of Mathematical Science only, and included modules at all levels in their sampling frame. However, their sample only included 12 modules at levels 0 and 1. Most of the level 0&1 modules surveyed had an attendance rate of between 30% and 50%. It is of interest to compare the results of our survey with last year's, and to obtain a more accurate estimate for the attendance rate of levels 0& 1 modules.

2.Methods

In our survey of the attendance rate, the following steps were undertaken: 1.Survey design. 2. Choosing the sample size. 3. Estimates and standard errors. 4. Plotting of the data.

2.1 Survey Design

In our survey, we used a combination of Stratified Random and Cluster sampling. The reasons why we used stratified sampling are:

1. Stratification may produce a smaller bound on the error of estimation, since the measurements within the strata are homogenous.
2. Separate estimates of population parameters in each stratum can be obtained.

By stratification we made sure that attendance rates are similar for classes in the same stratum.

Also, we used cluster sampling (clusters of unequal size) in each of the strata. Cluster sampling is a probability sample in which each sampling unit is a collection of elements. In our case each element was a student and the cluster is a class. We did one-stage cluster sampling, that is, all the student members in each selected cluster are to be used in overall sample. The cluster size is the enrolment for a particular course.

2.2 Choosing the sample sizes

There were six strata as listed in the introduction.
We used the method of proportional allocation to choose the overall sample size and the allocation for each stratum. This was because there were no costs involved and we assumed for simplicity that the variability was the same in each stratum.
The overall sample size *n* necessary to achieve a bound B on the error of estimation is:

$$n = \frac{N\sigma^2}{ND + \sigma^2}$$

Where n = overall sample size
N = total number of modules
σ^2 = Variance
B = 4%, 5%, 6%

$$D = \frac{B^2}{4}$$

We expect to see 0.2< attendance rate < 1,
That is range = 0.8.
Using the rule of thumb: 4σ = range, our estimate of σ was 0.2

From the above equation, when:
B = 4% n = 47
B = 5% n = 37
B = 6% n = 30

Our sample frame has N = 86 modules, and each stratum N_i has size:
$N_1 = 11$
$N_2 = 12$
$N_3 = 21$
$N_4 = 27$
$N_5 = 6$
$N_6 = 9$

We selected an overall sample size *n* of 43. This was between the values of n = 37 for B = 5% and n = 47 for B = 4%, which was done for practical considerations.

Sample sizes for the different strata (*i*) were got using:

$$n_i = n\left(\frac{N_i}{N}\right)$$

Where ni = allocated sample size to stratum *i*
N_i = size of i^{th} stratum
Stratum 1 = 5
Stratum 2 = 6
Stratum 3 = 11
Stratum 4 = 14

Stratum 5 = 3
Stratum 6 = 4

2.3 Estimates and standard errors

In order to get the individual estimates of the stratum means and 95% bound on the errors of estimation, we considered three different estimators for each stratum:

Case 1: $\bar{y}_j = \dfrac{\Sigma x_i y_i}{\Sigma x_i^2}$

Case 2: $\bar{y}_j = \dfrac{\Sigma y_i}{\Sigma x_i}$

Case 3: $\bar{y}_j = \dfrac{1}{n}\sum \dfrac{y_i}{x_i}$

Where y_i = number attending module *i*
x_i = number enrolled in module *i* .

The standard errors were got by:

95% Bound on error = $2\sqrt{\hat{V}\left(\bar{y}_j\right)}$

Where $\hat{V}(\bar{y}_j) = \dfrac{\sigma^2}{\sum w_i x_i^2}$

And σ^2 is estimated by $\hat{\sigma}^2 = \sum_{i=1}^{n} w_i \dfrac{(y_i - \bar{y}_j x_i)^2}{\sum w_i x_i^2}$

Where w_i = 1 for case 1

$w_i = \dfrac{1}{x_i}$ for case 2

$w_i = \dfrac{1}{x_i^2}$ for case 3

For the estimation of the entire population mean and 95% bound on the error of estimation, we used:

$$\bar{y}_{st} = \frac{1}{N}\sum N_j \bar{y}_j$$

95% Bound on error $= 2\sqrt{\hat{V}\left(\bar{y}_{st}\right)}$

$$= 2\sqrt{\frac{\sum N_j^{\,2}\hat{V}(\bar{y}_j)}{N^2}}$$

2.4 Plotting of the data.

The data from the survey will be plotted in order to see whether the variability increases as the enrolment (x) increases. This will help to determine which estimator 1, 2 or 3 is best.

3. Results

3.1 Sample Results

We selected a sample of 43 at random out of a total of 86 modules. 11 values were missing as discussed in Section 4, 2 of these not being available. This left 86 modules overall in our sampling frame out of which a sample of size 32 was achieved.

Table 3.1 displays the total number of modules available on a particular stratum (day(s) and times) together with enrolment figures and our counts for attendance.

Discussion on sample size, strata and module selection can be found in the Methods Section 2.

Table 3.1

Monday		**Monday**		**Friday**	
9am - 11am		**11am - 6pm**		**9am - 6pm**	
Stratum	1	**Stratum**	2	**Stratum**	5
Modules (N_i)	11	**Modules (N_i)**	12	**Modules (N_i)**	15
Sample (n_i)	4	**Sample (n_i)**	3	**Sample (n_i)**	4
Attendance	**Enrolment**	**Attendance**	**Enrolment**	**Attendance**	**Enrolment**
21	29	164	250	32	220
10	18	132	165	29	85
75	121	39	99	59	185
81	118			38	218
Tuesday, Wednesday, Thursday					
9am - 11am		**11am - 6pm**			
Stratum	3	**Stratum**	4		
Modules (N_i)	21	**Modules (N_i)**	27		
Sample (n_i)	8	**Sample (n_i)**	13		
Attendance	**Enrolment**	**Attendance**	**Enrolment**		
36	57	66	250		
22	29	127	248		
137	324	28	53		
54	92	44	85		
216	476	88	165		
77	262	55	114		
60	121	10	22		
146	307	68	91		
		64	91		
		267	476		
		147	476		
		92	167		
		53	85		

3.2 Estimates of Attendance Rate

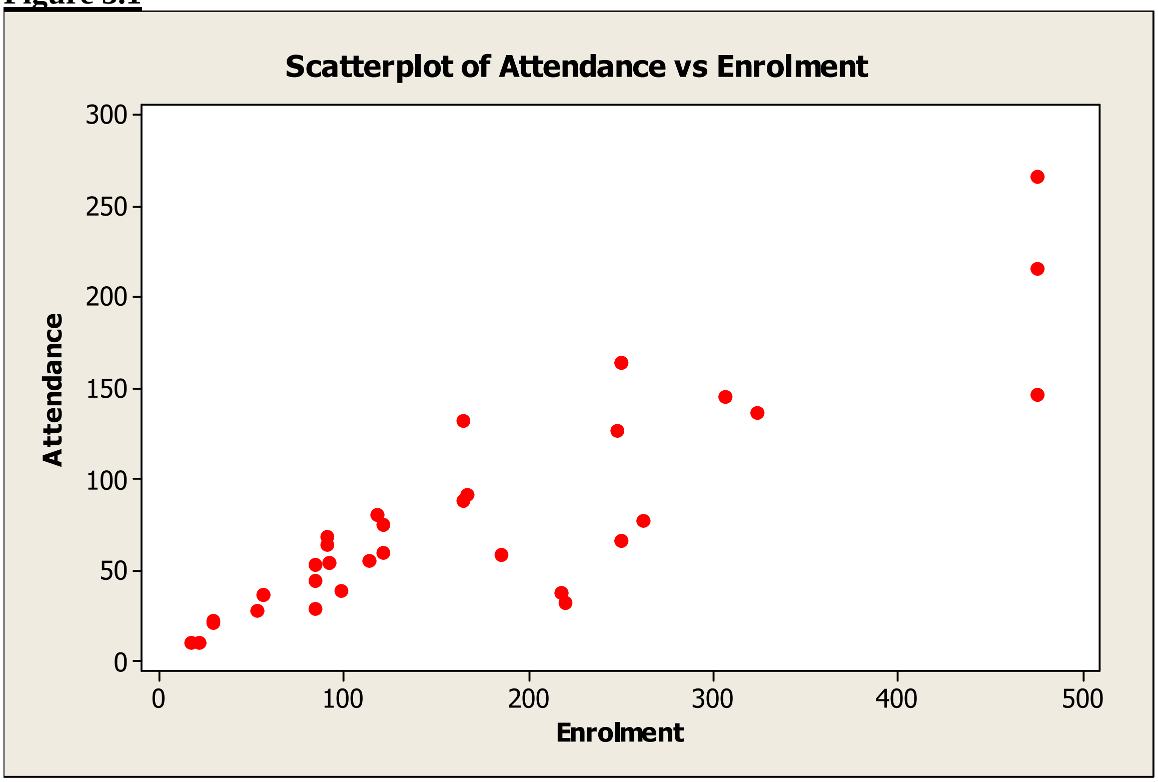
Table 3.2 displays attendance rates and bounds on the error for each stratum. These figures were calculated using a statistical package, SAS see **Appendix 1**. They are calculated for 3 cases with various weightings. This is discussed in Section 2, Methods.

Table 3.2

	Case I		**Case II**		**Case III**	
	Weighting = 1		**W = 1/Enrolment**		**W = 1/Enrolment^2**	
Stratum	**Rate**	**Error**	**Rate**	**Error**	**Rate**	**Error**
1	0.653	0.026	0.654	0.025	0.647	0.037
2	0.670	0.056	0.652	0.084	0.617	0.119
3	0.435	0.023	0.448	0.030	0.515	0.050
4	0.446	0.065	0.477	0.050	0.522	0.037
5	0.209	0.044	0.223	0.047	0.295	0.050
Overall Estimates of Attendance Rate						
	45.7 ± 4.7%		**47.4 ± 4.4%**		**50.6 ± 4.9%**	

3.3 Plots

Figure 3.1



A simple scatter plot of Attendance vs. Enrolment is displayed in **Figure 3.1**.

We use this plot to decide which weighting Case (I-III) in **Table 3.2** to use in our estimate of attendance rate. The choice of weighting is based on the variance or spread of the attendance figures as enrolment increases. Case I should be used if there is no increase in spread of the attendance figures with increase in enrolment. Case II should be used if there is an overall increase in spread of attendance with increase in enrolment. Case II should be used if there is a very large increase in spread of attendance with increase in enrolment.

In **Figure 3.1,** one can see an overall increase of attendance with increase in enrolment. This is to be expected as with larger class sizes a higher number of students would be present. We can see that the variance increases a large amount from small classes of size say 50 to 200. However the variance does not increase too much from class sizes of 200 to 500. This would indicate to me that Case II would be the most suitable estimator for our attendance rate i.e. a weighting of 1/enrolment which assumes that variance increases with increase in enrolment values. Our choice of estimator is therefore 47.4 ± 4%.

Figure 3.2

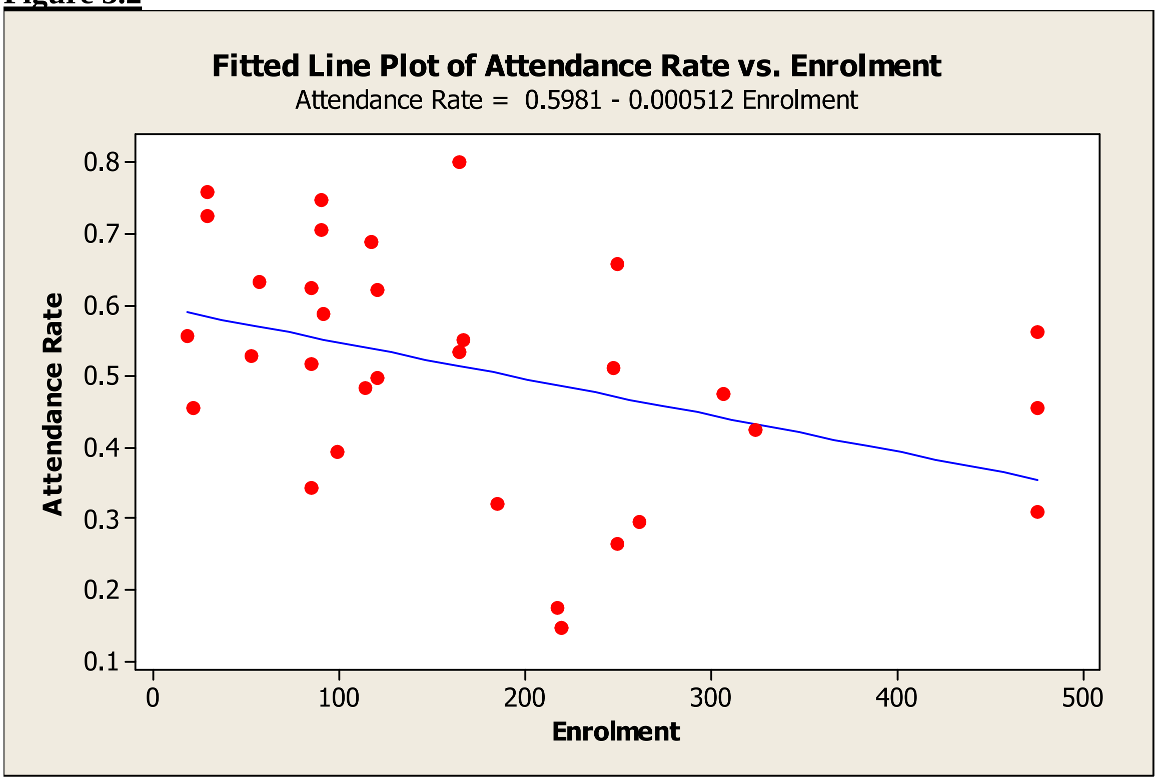


Figure 3.2 above is a plot of Attendance Rate vs. Enrolment which gives us a clearer picture. We can see that overall there is a decrease in attendance rate as class size increases. A straight line looks like a reasonable model for this relationship. The higher attendance rates between 70% and 80% occurred for smaller class sizes of between 30 and 160, whereas the lower rates of 15% to 35% occurred for larger class sizes of between 185 and 485.

This is verified by doing a statistical analysis and fitting a straight line through these data (the blue line shown on the plot) by the least squares method. The downward trend is shown to be statistically significant.

We propose the hypothesis that the downward trend is not significant. An analysis of variance (see **Table 3.3** below) calculates a p value of 0.025. This is less than our significance level of 5% which allows us to reject this hypothesis and conclude that the downward trend is significant. However we note the trend is very slight i.e a small slope, and the percentage of variation explained by the regression is only 15%.

The 2 lowest attendance rates in **Figure 3.2** occurred on Friday. These were rates of 15% and 17% with class sizes of 220 and 218 respectively. The highest attendance rate occurred on Monday afternoon with an attendance rate of 80% and class size of 165 students.

Table 3.3

This analysis of variance was calculated using a statistical package called Minitab.

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	0.134196	0.134196	5.59	0.025
Error	31	0.744580	0.024019		
Total	32	0.878775			

4 Discussion

4.1 Comments on the results

How our sample size was chosen

Our original sample size was 43 modules out of a possible 86. This meant that the twelve people available in our class sampled between 3 and 4 classes each. This choice of sample size was decided upon to give us a small error bound (±5%) and thus, hopefully increasing the accuracy of our findings.

Results of the stratification

The results of the data collection proved to be quite interesting. It was very unexpected to see that classes on Monday between 9.00am to 11.00am had the highest attendance rate of all strata - 65% attendance rate. Overall, there was very little difference in the attendance of classes in the morning and the afternoon and, no noteworthy difference between attendance rates on Tuesday, Wednesday and Thursday with all three days having a similar attendance rate between 45 and 50%. However, there was quite a substantial difference in attendance on Fridays compared to the rest of the week. There was a very poor attendance level in the Friday classes – 22% attendance. In light of these results Monday looks like the best day for scheduling lectures with Friday being by far the worst day for lecture attendance.

Should we have chosen different strata?

After the data was collected, we observed that we could have done a different stratification of modules. Of the classes we sampled, the majority of the classes were large lecture hall classes. We discovered that it may have been a good idea to stratify by class size: large classes, and small room classes. This could have given some interesting results regarding the difference in attendance rates between large and small classes. Our results point towards slightly poorer attendance in larger classes.

4.2 Problems encountered during the survey

Problems that arose when collecting the data

Our sample size got reduced from 43 to to 32 and there are unfortunately a large number of unavoidable missing values.

Firstly, one of our surveyors was sick during the week of the sampling and this went unnoticed until the end of the week. This meant his four classes went uncounted. The results from two other modules couldn't be taken as the lecturer involved did not fully understand what the surveyor was doing and therefore would not allow the observor to count his class. A further two modules had already been completed earlier in the term before the survey was undertaken. There were no results from these modules. One class was on a field trip on the day of counting. Finally one observor sampled two incorrect modules that could not be used.

This then gave us a final sample size of 32. However, we were also able to reduce N, the total size of population, from 86 to 84 as two of the modules had already been completed. In the end our sample size was still large enough as we ended up with an error bound of ±4.5% which was still lower than our aim of ±5%.

Problems Encountered with the strata

We encountered some problems in the strata of our survey. Because there were values missing in stratum 5, the entire stratum collapsed. This led us to actually omit this stratum entirely and instead, combine stratum 5 and stratum 6 into one stratum. So, instead of having a Friday morning stratum and a Friday afternoon stratum, we just had Friday as a whole. We thought this was the best decision as we could still salvage and use the data collected.

,

4.3 Comparison with last year's results

It is hard to directly compare our results from this year's survey to the results collected last year as they chose to survey the attendance rates of all Levels but only from the School of Mathematical Sciences, while we decided to restrict our survey to Level 0 and 1 but include all Science modules. Furthermore, they did not stratify their sample by the day of the week, as we did, but chose to stratify it by module level and by morning lectures, 9-11 and other lectures, 11-finish.

However, it is still possible to compare our results for modules 0 and 1 and, in doing so, we can see that the general pattern – that there is negligible difference between attendance in morning lectures and attendance in afternoon lectures – prevails. The attendance rate for early morning lectures in last years survey was 48% ± 2.6% and the attendance rate for afternoon lectures in last years survey was 51% ± 2.0%.

Overall there was a similar poor attendance rate of ~50% for Levels 0 and 1 modules this year and last year.

4.4 Conclusions

There was a disappointing overall attendance rate of 47.4% ±4.4%. We think that new measures, such as marks for attendance, might have to be put in place in order to encourage more students to turn up to lectures. Also, given the extremely poor attendance of Friday lectures (22%) we feel that, where possible, lectures should be scheduled earlier in the week.

Unexpectedly we found no noteworthy difference in attendance between early morning lectures (9-11) and afternoon lectures (12-finish). Therefore, we feel that the time of the day is of minor importance when scheduling lectures.

Finally, we also want to note that some modules may have included students repeating the module exam but not attending the lectures. This could have led to slightly lower results. However, in reality it was impossible for us to calculate the attendance of repeating students in each module.

4.5 Acknowledgements

Firstly we would like to thank the Dean of Science for allowing this unique module to be undertaken again this year. We want to thank all the lecturers for their co-operation and understanding when the survey was being held. And most importantly we want to thank

Dr. Gabrielle Kelly for her encouraging support and guidance throughout the course of this module.

Appendix

Our estimates of attendance rate and error bounds were calculated using a statistical package called SAS. The following is a portion of the code used in our analysis.

```
data a;
set dass.attendance;
wt1=1;
wt2=1/enrolment;
wt3 =1/(enrolment*enrolment);
run;

data a1;
set a;
if stratum=1;
run;

proc surveyreg data=a1;
model attendance=enrolment/noint;
weight wt3;
run;
```