

TASK SUMMARY

This survey assesses support or lack of support in a community to an expensive project using public funds. The survey questionnaire was distributed to 5000 residents in Dunedin city, New Zealand. Responses were received from 2248 residents. The method of selecting the sample and constructing the questions is described in the video. A number of demographic variables were recorded for each respondent to allow post stratification, a feature often not reported in social surveys.

Threats to validity of the results are discussed in the second half of the video along with a discussion of the method of post stratification.

The full data set containing information for 2248 respondents produces the results presented in the video but the tasks suggested here should use the simplified data file with 1950 respondents at **stadiumreduced.xls**. The full data set was cleaned by removing a number of variables used in the data entry, removing cases where there were missing values and removing five cases under 20. Consequently the results from the following tasks may differ slightly from those in the video. If using GenStat for the analyses the Schools version can be used for graphs and simple analyses as it will operate with up to 2500 cases. If using the Undergraduate version where there is a limitation of 500 cases you can select a random sample of 500 cases for analysis.

Task 1 : Calculate the (crude) proportion for the “No” response category for all respondents and construct the 95% confidence interval for this proportion. Interpret this confidence interval.

Task 2 : Construct a cross tabulation of SEX against the outcome categories “No”, “Yes” and “Undecided”. Draw the bar graph summarising this table. Set up confidence intervals for the difference between the proportions of men and women in the three outcome categories.

Task 3 : Construct cross tabulations of AGE group against outcome; also for INCOME against outcome; also RATEPAYER against outcome. Draw multiple bar graphs as well. Report your conclusions. Set up appropriate confidence intervals for differences between proportions.

Task 4 : University students can select a random sample of 500 cases from this data file for analysis using the Undergraduate version

Task 5 : Take support for the stadium or no support (yes/no categories only) as the response for a logistic regression with sex, income and age as predictor variables. Report your conclusions.

Task 6 : With the same subset of 500 cases, carry out a multinomial logistic regression with “yes”, “no” and “undecided” as the categories of the outcome variable and sex, income and age as predictors or combinations of these. Report your conclusions.

As an optional further extension for university students investigate improved estimates using the idea of post-stratification. The procedure is described in detail in the file **post-stratification.pdf**

The demographic characteristics of Dunedin City (Statistics New Zealand Census 2006) are reported here for convenience (along with the estimated proportions used when carrying out the calculations for the full data set).

Strata	N_h	$\hat{\pi}_h$
SEX		
Female	64,398	23.51
Male	59,118	31.13
Sum	123,516	
AGE		
20-24	14376	41.18
25-29	7020	27.27
30-34	7434	35.96
35-39	7815	26.88
40-44	8511	30.60
45-49	8469	26.34
50-54	7584	28.81
55-59	7080	31.11
60-64	5222	26.77
65 and over	16542	23.37
Sum	90053	
INCOME		
20k or less	42202	14.05
20k - 30k	14064	30.69
30k - 40k	8916	24.24
40k - 50k	4836	21.92
50k - 70k	4179	29.91
70k-100k	1416	32.53
100k or more	1179	44.95
Sum	76791	

Task 7 : Using the reduce data file of 1950 respondents calculate the proportion supporting the building of the stadium after post-stratification by Sex. Compare with the value 27.16% reported in the video.

Task 8 : Find the proportions supporting the stadium after post-stratification first by Age and second by Income.

Note: It is possible to post stratify on these three variables together and this was reported in the video. This task will be left out here.