# Lesson: Support for Public Funding for the Otago Stadium

Support (or lack of support) from the Otago community to use public funding to construct the new Otago stadium, Forsyth Barr Stadium, has been widely discussed and debated for some time. A survey questionnaire about this topic was distributed to 5000 residents in Dunedin city, New Zealand. Responses were received from 2248 residents.

This lesson investigates the factors that influence residents support for public funding of the stadium.

1. To open the data we click on **File>Example Data Sets:**
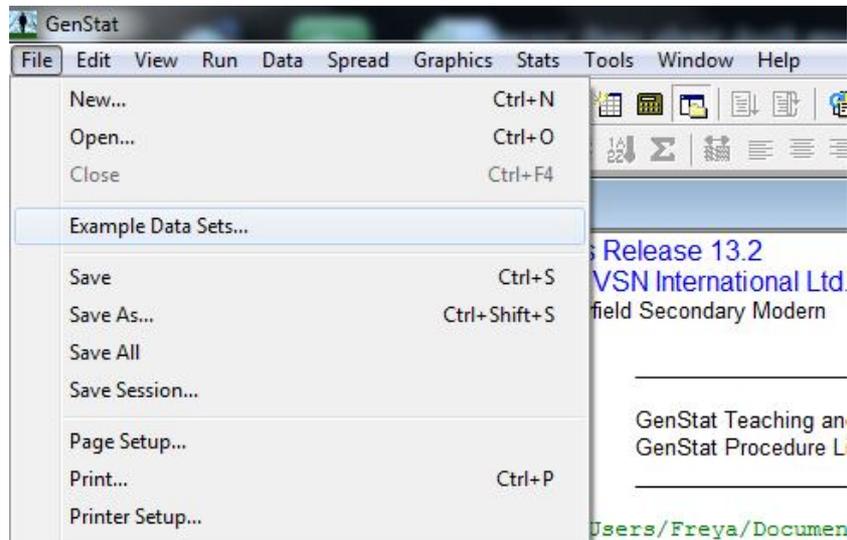
This brings up the Example Data Sets dialog shown below. Click on the **Filter by topic** drop down menu and select the **NZ Schools Example Data sets** option. Choose the file **Stadium.gsh** and click on **Open data:**
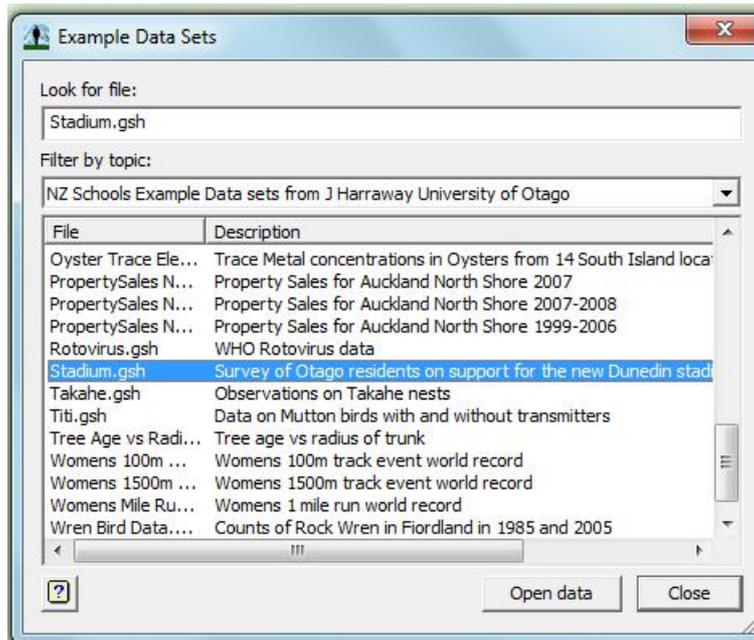
This opens a spreadsheet (displayed below) containing data from a study investigating what proportion of residents support using public funds to construct the new Otago Stadium (note that this means support *public funding* of the stadium, not support the stadium). The study was carried out by John Williams (Department of Marketing, University of Otago):


In the spreadsheet the data recorded are **Support** (Do you support public funding of the Otago Stadium? Yes/No)**, No** (Support Yes/No recorded as 0/1)**, Age** (age of respondent)**, Employment** (employment situation)**, Education, AgeGroup** (categorical age group of respondent)**, HomeOwn** (living situation)**, Ratepayer** (pays rates Yes/No)**, Income** and **Sex** (Male/Female)**.**

2. The first step of the analysis is to calculate the (crude) proportion of respondents that do not support public funding for the stadium (the "No" response category) and construct the 95% confidence interval for this proportion. Click on **Stats>Binomial Tests:**
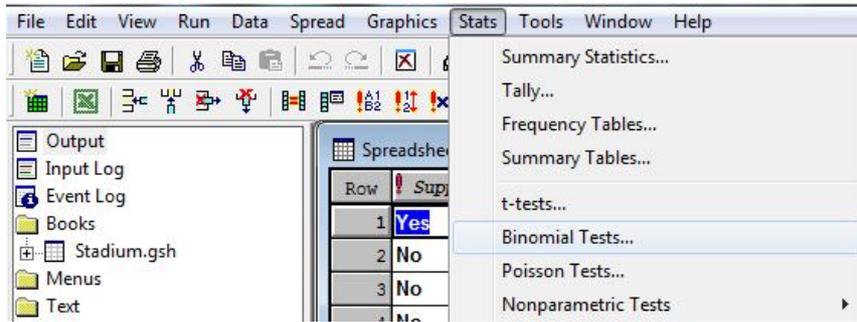
In the **Sample set** box enter the variable **No** and in the **Proportion of Success** box enter **0.5**. Click on **Options** and tick **Summary** and **Confidence Interval.** Press **OK** then press **Run:**
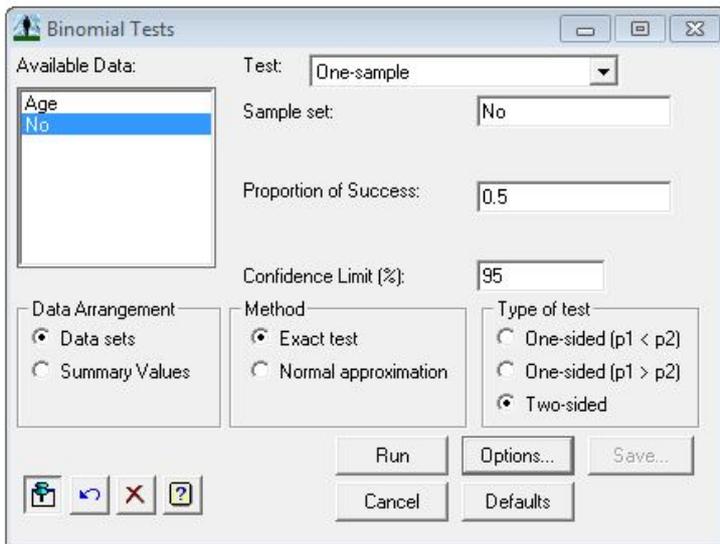
Switch to the GenStat Output Window, where you should have results as shown below:

One-sample binomial test
Summary

| Sample Size	Successes	Proportion
1950	1290	0.662

95% confidence interval for proportion: (0.6400, 0.6825)

Interpret this confidence interval in terms of the number of respondents that do not support public funding of the Otago Stadium.

3. Construct a cross tabulation of **Sex** against each of the categories in **Support** (No, Yes and Undecided), and draw the bar graph summarising this table. Click on **Stats>Frequency Tables:**
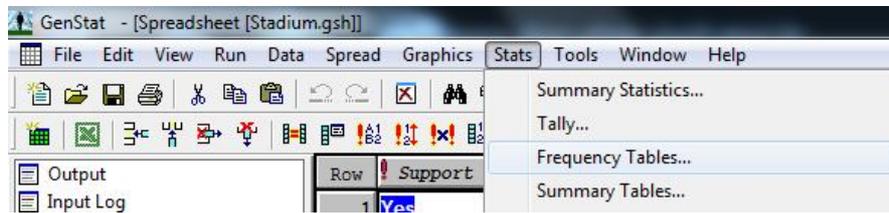
In the **Groups** box enter the variables **Sex** and **Support**, and in the **Options** box tick **Use Tabbed-table** (this makes copying results from the output into a new document more user-friendly) and **Set Margin**. In the **Margin** drop-down menu select **Overall Margin**. Click on **Graphics**, then in the **Factor for X-axis** box enter **Support** and in the **Groups** box enter **Sex**. In the **Method** box select **Bar chart** then click **OK:**
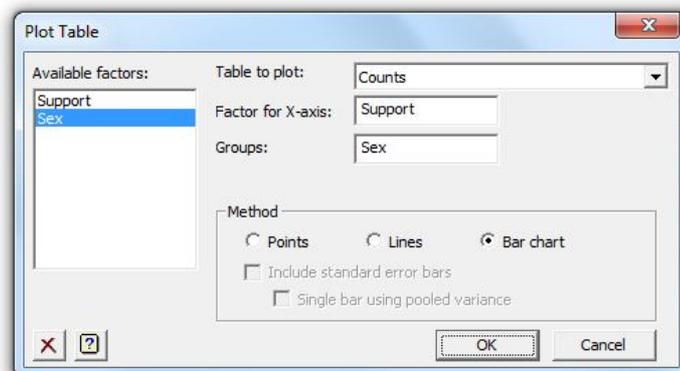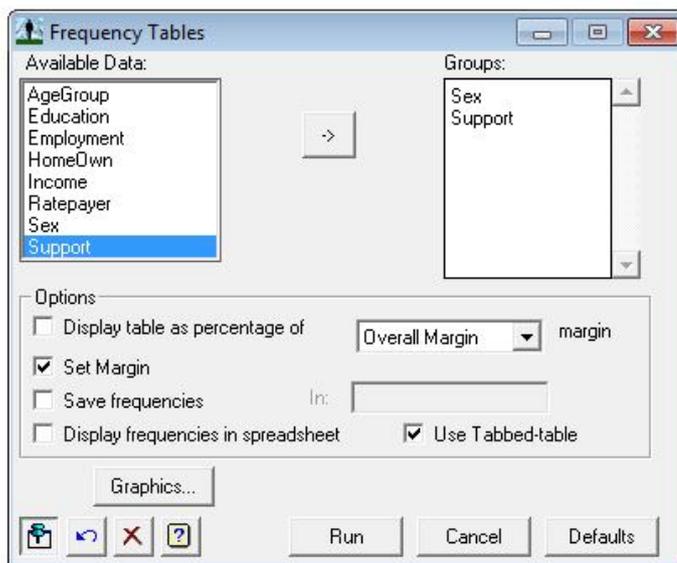
A bar chart will automatically appear in the GenStat Graphics Viewer, similar to the one pictured below:

Support for Public Funding of the Stadium
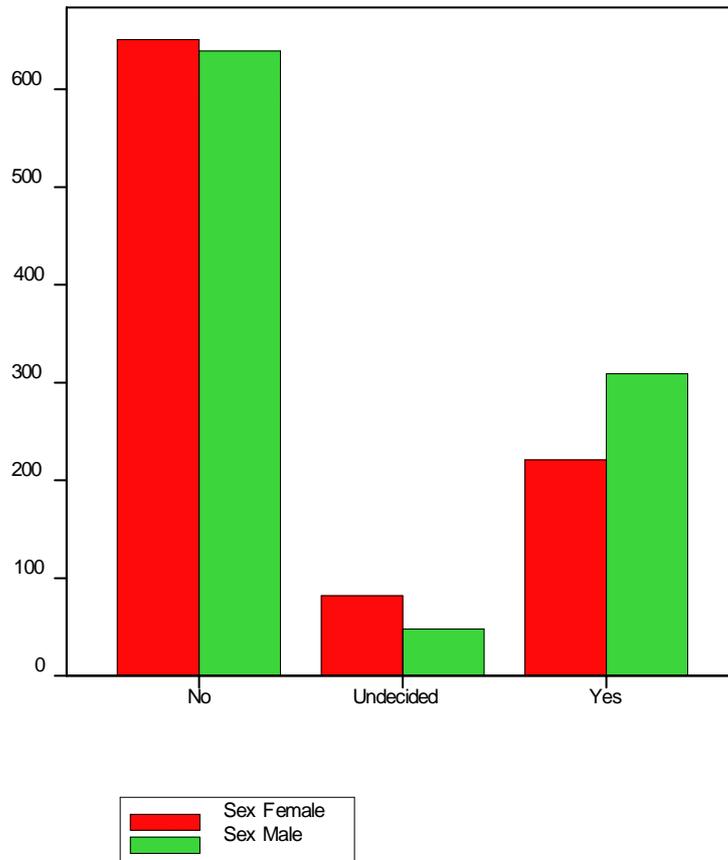
What tentative conclusions can you make from this bar chart?

Switch back to the **Frequency Tables** menu, and press **Run:**

GenStat Output Window:

Support	No	Undecided	Yes	Count
Sex				
Female	651	82	221	954
Male	639	48	309	996
Count	1290	130	530	1950

4. Set up confidence intervals for the difference between the proportions of men and women in the three outcome categories (No, Undecided and Yes). Click on **Stats>Binomial Tests** as was done in part 2.

For this example we will calculate the difference between proportions of men and women who do not support public funding of the stadium (answered "No").

In the drop-down menu for **Test** select **Two-sample** and in the **Data Arrangement** box select **Summary Values.** Here we will take sample 1 (y1 and n1) as Female and sample 2 (y2 and n2) as Male. In **Number of Successes (y1)** enter **651** (the number of females that answered "No"), and in **Sample Size (n1)** enter **954** (the total number of female respondents). In **Number of Successes (y2)** enter **639** and in **Sample Size (n2)** enter **996.** Click **Run:**
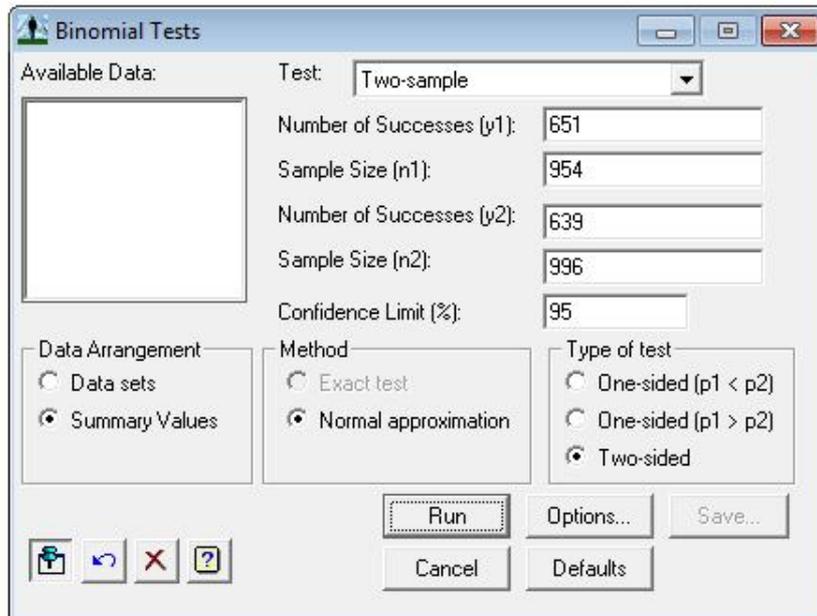
GenStat Output Window:

Two-sample binomial test
Summary

Sample	Size	Successes	Proportion
1	954	651	0.682
2	996	639	0.642

Approx s.e. of difference between proportions: 0.0214

95% confidence interval for difference between proportions: (-0.001124, 0.08277)

Interpret this confidence interval and state any conclusions you can make.

Repeat the above procedure for the difference in the proportions of men and women who selected a) "undecided" and b) "Yes" for their support of public funding for the stadium and interpret your confidence intervals.

5.
a. Construct a cross tabulation of **AgeGroup** against **Support**, and create a bar graph showing this information, as was done in part 3.
   Calculate the confidence interval for difference between the proportion of respondents aged "**25-34**" and "**55-64**" that answered "**No**" in support for public funding for the stadium and interpret your result, as was done in part 4.

b. Construct a cross tabulation of **Income** against **Support**, and create a bar graph showing this information (with "Income" as the factor on the x-axis and "Support" as the grouping variable).
   Calculate the confidence interval for the difference between the proportion of respondents with income "**<$30,000**" and "**>$100,000**" that answered "**No**" in support for public funding for the stadium. Interpret your confidence interval.

c. Construct a cross tabulation of **Ratepayer** against **Support**, and create a bar graph showing this information (with "Ratepayer" as the factor for the x-axis and "Support" as the grouping variable).
   Calculate the confidence interval for the difference between the proportion of respondents who were **ratepayers** and those who were **not ratepayers** that answered "**No**" in support of public funding for the stadium. Interpret your confidence interval.

   Calculate any other appropriate confidence intervals for difference between proportions. What conclusions can be drawn from the confidence intervals found above?

   Note: the question below is designed for students undertaking a first year statistics course at university.
   To change from the GenStat School Edition to the GenStat Undergraduate Edition click on **Tools>Options** and in the **GenStat Edition** box select **Undergraduate Edition**. You will have to restart GenStat for this change to take effect.

6. Now we want to select a random sample of 500 respondents from the stadium data set for the following questions. Click on **Spread>Calculate>Random Subset:**
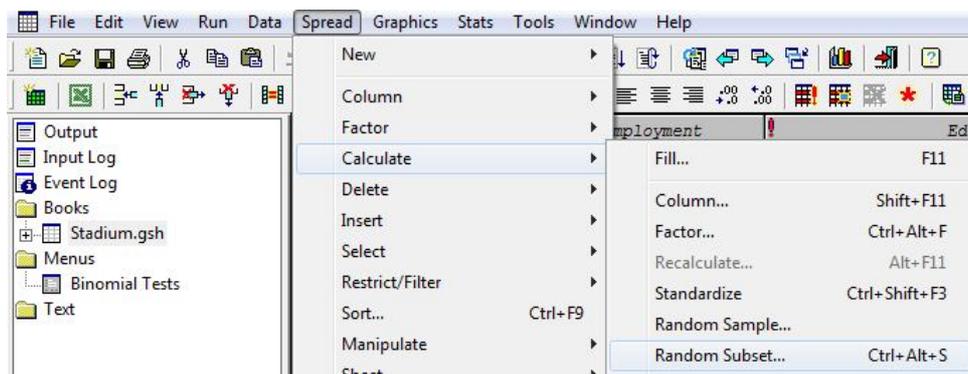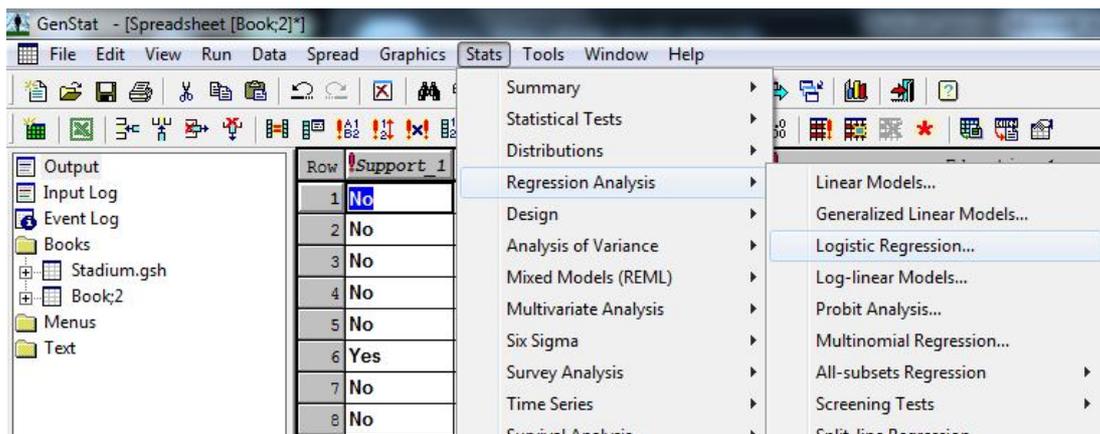
In the **Number of Samples** box enter **500** and press **OK:**
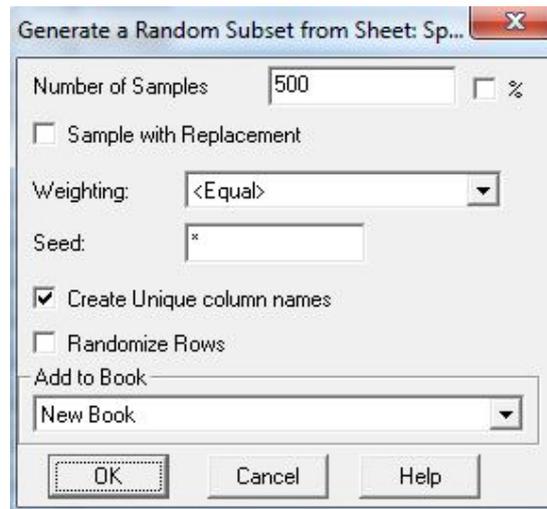
This will create a new dataset with the name "Book;2" or equivalent, with 500 respondents that were randomly selected from the original data set (Stadium.gsh). The variables in the new data set hold the same names but with _1 added to the end to differentiate them from the original variables.

7. Carry out a logistic regression with **No_1**) as the response (support for the stadium yes/no recoded as 0/1, note that 'undecided' is coded as 0 for illustrative purposes only) and **Sex_1**, **Income_1** and **Age_1** as predictor variables. Click on **Stats>Regression>Logistic Regression:**

In the **Number(s) of Subjects** box enter **500,** and in the **Number of Successes** box enter **No_1**. In the **Model to be Fitted** box enter the model **Sex_1 + Income_1 + Age_1** then click **Run:**
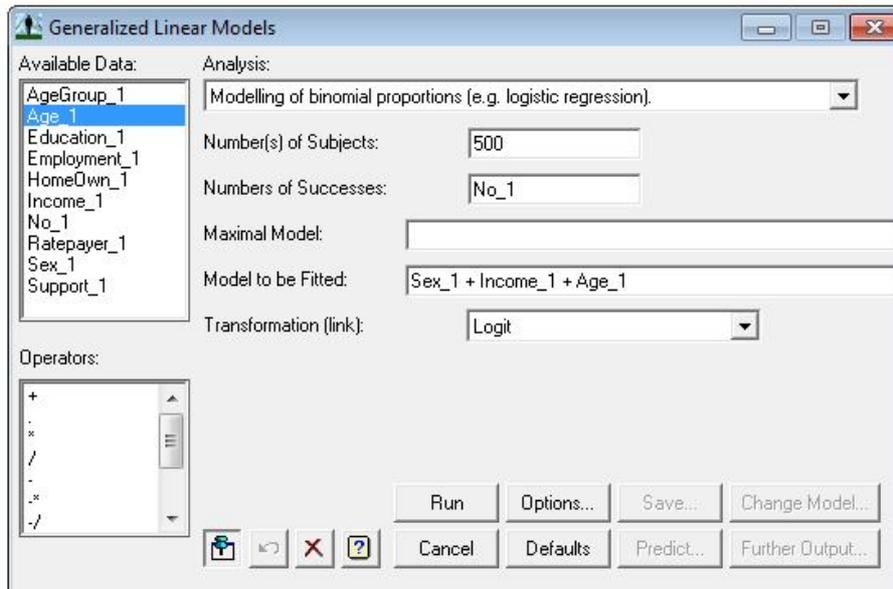
GenStat Output Window:

Regression analysis

Source	d.f.	deviance	mean deviance	deviance ratio	approx chi pr
Regression	6	9.6	1.6043	1.60	0.141
Residual	477	263.5	0.5523		
Total	483	273.1	0.5654		

Estimates of parameters

Parameter	estimate	s.e.	t(*)	t pr.	antilog of estimate
Constant	-6.524	0.256	-25.53	<.001	0.001468
Sex_1 Male	-0.006	0.117	-0.05	0.957	0.9938
Income_1 $30000-$50000	-0.037	0.154	-0.24	0.808	0.9634
Income_1 $50000-$70000	-0.231	0.195	-1.18	0.237	0.7937
Income_1 $70000-$100000	-0.293	0.189	-1.55	0.120	0.7460
Income_1 >$100000	-0.475	0.202	-2.36	0.018	0.6220
Age_1	0.00072	0.00376	0.19	0.848	1.001

Parameters for factors are differences compared with the reference level:

Factor	Reference level
Sex_1	Female
Income_1	<$30000

Write out the logistic regression model and report any conclusions you have from the results above.

8. Extension Question: With the same subset of 500 cases, carry out a multinomial logistic regression with **Support_1** (Yes, No and Undecided) as the outcome variable and **Sex_1**, **Income_1** and **Age_1** as predictors. Report your conclusions.