
7.1. z test with unmatched data

EXAMPLE 7.1. The evolution of *Littorina littoralis* at Aberystwyth, 2002

BOX 7.2. How to carry out a z test for unmatched data

Step 1. **Parametric** statistical tests are accessed through the ‘Analysis Tool Pack’, which is activated via the add-in option under ‘Tools’ on the tool bar. Select ‘Tools’. On the drop-down menu, select ‘Add-ins’ and from the dialogue box check the square for ‘Analysis toolkit’. Returning to the ‘Tools’ option, you should now find that ‘Data Analysis’ is added to the drop-down menu. You are now ready to complete a parametric test.

Step 2. You first confirm that all criteria are met (e.g. BOX 7.1.) and then you may carry out the z test (BOX 7.2.). Select a new sheet in Excel or continue using that already prepared for the F test. The data must be entered in columns, one for each sample. Put a ‘label’ for each sample to identify it. Using ‘Tools’ from the menu bar and then ‘Data Analysis’, select ‘ z test: Two Sample for Means’.

In the dialogue box that opens, ensure that the cursor is flashing in the ‘Variable 1 Range: Input’ box. Input the cell references of the data by clicking on cell A4 (this includes the data or sample label) and dragging down the column to the last cell containing data. The area on the spreadsheet will now be highlighted and the cell references shown in the input box.

The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C
1	Periwinkle shell height data		
2			
3			
4	lower shore shell height (mm)	mid shore shell height (mm)	
5	5.3	11.7	
6	8.7	4.1	
7	5.3	8.8	
8	5.3	9.7	
9	5.9	5	
10	2.8	6.7	
11	8.7	6.6	
12	2	7.8	
13	5.3	7	
14	6.5	7.6	
15	4.3	9.3	
16	8	7.7	
17	10.2	6.6	
18	5.3	9.4	
19	7.8	8.8	
20	7	5.8	
21	6.1	5.8	
22	5	7.5	
23	5.4	6.3	
24	5.7	12.5	
25	6.5	11.3	

Now click in the box marked ‘Variable 2 Range:’ and repeat the process with the second column of data. The hypothesized mean difference under the null hypothesis is zero so enter 0 in the box below. Next the variances of both data sets need to be entered. The values for **variance** calculated for the F test should be entered into the appropriate boxes. The box marked ‘Labels’ should be clicked – this will put a tick in the box, which shows that the first cell for each data set contains a label and not data. If this box is not ticked, Excel will treat the material in the first cell as data and will not be able to complete the calculation. Note that it is useful to use the labels to identify your data. These labels are used by Excel to identify the output data. The default value for alpha (p) is 0.05; this can be changed if required, but you will not normally require to do so.

z-Test: Two Sample for Means

Input

Variable 1 Range: \$A\$4:\$A\$34

Variable 2 Range: \$B\$4:\$B\$34

Hypothesized Mean Difference: 0

Variable 1 Variance (known): 3.590161

Variable 2 Variance (known): 4.097345

Labels

Alpha: 0.05

Output options

Output Range: \$G\$5:\$I\$22

New Worksheet Ply:

New Workbook:

OK

Cancel

Help

Step 3. Next, select the output options. To return the output data below the input data, select 'Output Range:' and then click in the box (the cursor will now flash in the box). Drag over an area where you want the results to be displayed. Note that you could just select a couple of cells – Excel will determine the actual size that it requires for the results table. Note too, that it is essential to click in the box as well as selecting the 'Output Range' button. If this is not done, the location is entered into the 'Input Range' box and the analysis cannot be completed.

You can choose to have the results entered on to a 'New Worksheet Ply', in which case the results will be given on a fresh sheet, accessed by the tabs at the bottom of the current sheet. Alternatively a 'New Workbook' can be selected.

z-Test: Two Sample for Means

	<i>lower shore shell height (mm)</i>	<i>mid shore shell height (mm)</i>
Mean	5.546666667	7.47
Known Variance	3.590161	4.097345
Observations	30	30
Hypothesized Mean Difference	0	
z	-3.799465032	
P($Z \leq z$) one-tail	7.25288E-05	
z Critical one-tail	1.644853	
P($Z \leq z$) two-tail	0.000145058	
z Critical two-tail	1.959961082	

Step 4. The results table shows first summary data for the two sets of data; note that by including the 'labels' that the columns are appropriately titled.

The z statistic is shown. Note that the negative sign can be ignored, as this only indicates that the sample with the smaller mean was entered as data before the sample with the greater mean.

The critical value for the two-tailed test is 1.96 and the test statistic exceeds this value. The null hypothesis can be rejected and the actual p value for the calculated value of z is given as 0.000145. The result is thus highly statistically significant. Many scientific papers quote only the actual p value derived from the test statistic (and not the test statistic itself) to indicate the level of significance of the test result.