

# CACTUS

Harley Hyde

## In search of beauty



This statue of the goddess Aphrodite was originally carved by Menides of Antioch somewhere between 130 BCE and 90 BCE. A peasant found her in the ruins of the ancient city of Milos, in 1820.

Originally she would have stood in a niche in the wall of a gymnasium where she would have been colourfully painted and wearing jewellery. She was gazing at an apple that she held high in her left hand. Her right arm clutched her garment. The statue was

carved from seven blocks of marble and although the arms were recovered, an “expert” at that time did not think they quite matched.

She was purchased by a naval officer, who took her home to Paris. She is now displayed in the Musee du Louvre and she has been given the Latin name Venus de Milo.

In 1854 the German scientist Zeising claimed that the ratio of a person’s height to the height of their navel is in the same ratio as the Golden Ratio ( $\phi = 1.62$ ). When I printed a full length picture of the Venus de Milo, I measured the height of her image at about 26 cm and the height of her navel at about 16 cm. The ratio of 26 to 16 is about 1.625. If you measure some other great works of art you will get similar answers. Is this just a case of artists using the Golden Ratio as a “rule of thumb” or did their models actually have those measurements?

There is so much hype about the Golden Ratio that it is worth reading an article at [http://www.maa.org/devlin/devlin\\_06\\_04.html](http://www.maa.org/devlin/devlin_06_04.html). It explains why some of the more far-fetched ideas are actually true and why some sugges-

tions, such as Zeising’s, gain credibility.

One answer has been provided by the Australian Bureau of Statistics. The ABS wishes to see its statistics widely used as it contributes to its mission of “assisting and encouraging informed decision making, research and discussion.” It therefore takes quite seriously the task of educating, and assisting those who educate, about statistics. We are grateful for their cooperation when preparing this article as the ABS data mentioned here is used with permission from the ABS. The entire contents of the ABS website is now accessible free of charge from their website at [www.abs.gov.au](http://www.abs.gov.au).

Of particular interest to teachers is the CensusAtSchool section of their website. Teachers can ask students from Years 5 to 12 to go online and answer a series of questions about themselves. The ABS keeps a database of the responses and anyone can go online and download a random sample of those responses. I chose a random sample of 200 from Year 10 students of both genders and from all postcodes and school sizes.

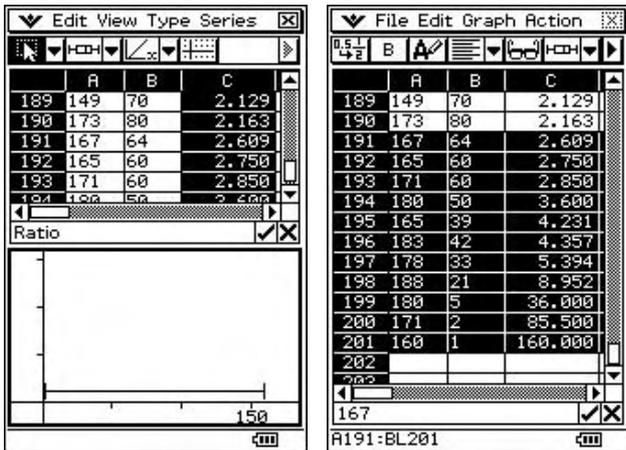
Teachers can also download very useful materials such as question sheets that help their students work through and understand the data they download.

The download of a random sample contains two files. The first is an Excel spreadsheet that lists and explains all of the variables and the various codes used. The other file is a .csv file that contains the actual data. Modern versions of Excel will recognise the file type and open it as a regular Excel file but students using older versions of Excel may need some help when first reading the file because they have to check the box that specifies that a comma is a valid delimiter.

The data available varies from time to time. At the time of writing this article, the data in Column Z describes each student’s height and the data in Column AB describes the height of their “belly button” above the floor. I copied these two columns to a new file and from there

I was able to drag the data into a ClassPad Emulator Spreadsheet.

In column A we have data describing each student's height and in column B we have data describing the height of the same student's navel. In cell C2 we create the formula =A2/B2 and copy it down to row 201. When we tap OK, the column fills with numbers that look excitingly close to 1.6.



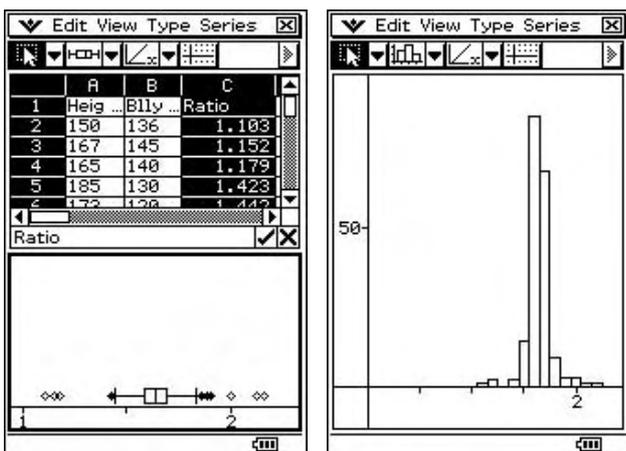
Now select column C by tapping the column heading and from the graph menu select Box Whisker and you probably get... nothing! And where did that 150 spring from?

Closer examination shows some quite stupid answers for the belly button height; that is what happens when people do not understand why a question is asked.

Select all the data in the range A2:C201.

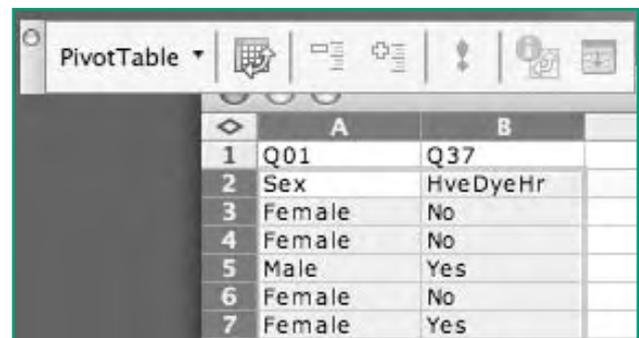
From the Edit Menu choose Sort. Fill in the Sort Dialogue Box giving the Range as A2:C201 and the Key column as C. When you press OK, the stupid answers go to the top or bottom of the heap. I deleted the data below row 190. There were still plenty of outliers.

Both the box plot and the histogram show that the ratio of the heights does cluster narrowly about a value near 1.6.

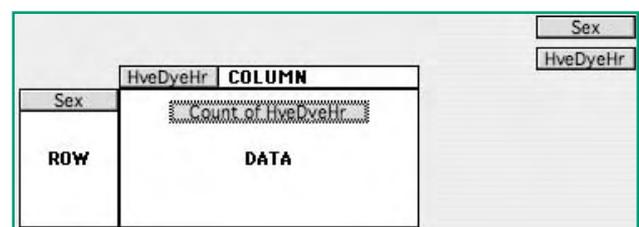


When I downloaded data supplied by Year 12 girls I discovered a higher proportion of stupid answers. It would appear that senior girls have reservations about answering questions about their belly button. I also noticed a relatively high spike of data close to 1.62. Twenty-five girls had all given answers that resulted in a ratio amazingly close to  $\phi$ . Then I noticed the data from a girl whose navel floated in a divine proportion above her head: she had multiplied her height by 1.62 instead of divided. I suspect that at least one teacher had to explain the purpose of the question to get the girls to cooperate.

Returning to the original Year 10 data download, you will find in column B the variable HveDyeHr which is the code for "Have dyed hair". The data consist of the words "Yes" or "No" which are nominal values. Other examples in question 37 include "has a ring," "has some body piercing" and "has a tattoo." The answers to these questions are of most interest if we compare them in a cross tabulation with age, gender or even music preference. I copied columns for Sex and HveDyeHr into a new spreadsheet and from there I was able to drag the data into a ClassPad Emulator Spreadsheet.



While in the Excel spreadsheet, generate a cross tabulation of the data using a Pivot Table. With Sex in column A and HveDyeHr in column B select the data as shown below. Then click on the Pivot Table icon. The second step of the Pivot Table Wizard will ask you where the data is. Correct the range to read \$A\$2:\$B\$202. At the third step you need to push the Sex icon into the Row Box and the HveDyeHr icon into the Column Box and into the Data Box like this:



When you click finish you have a Pivot Table formatted as a 2x2 Contingency Table.

	D	E	F	G
1	Count of HveDyeHr	HveDyeHr		
2	Sex	No	Yes	Total
3	Female	40	68	108
4	Male	79	13	92
5	Grand Total	119	81	200
6				
7				
8	Expected	No	Yes	Total
9	Female	64.26	43.74	108
10	Male	54.74	37.26	92
11	Total	119	81	200
12				

Students can make their own expected value table using these formulae.

At cell E9 type =G3\*E5/G5.

At cell E10 type =G4\*E5/G5.

At cell F9 type =G3\*F5/G5.

At cell F10 type =G4\*F5/G5.

At cell G9 type =E9+F9 and copy down.

At cell E11 type =E9+E10 and copy across.

They can now compare the observed values with the expected values. This much analysis suggests that girls are more likely than boys to dye their hair. In later years a complete  $\chi^2$  analysis is not then a large step.

A contingency table is simply a way of building a frequency table for more than one variable at a time. Many spreadsheets do not provide a function to do this. The code in the next panel describes how you can build a 2x2 contingency table using the spreadsheet software that is included with a ClassPad. It can be used with other spreadsheets using the formula “=if(D\$2 ..” instead of “=Cellif(D\$2 ..”

You will find that much of the data in the ABS data download can be analysed using this simple 2x2 contingency table.

ABS staff must be congratulated for the wide variety of educational material that they provide at their well designed and user friendly website. Do visit and enjoy.

## Build your own contingency table

With data cells at A1:B200 of a Spreadsheet, we will form a contingency table at E2:F3  
At D2 type “Female” and at D3 type “Male”.  
At E1 type “Yes” and at F1 type “No”.

At I1 type =Cellif(D\$2=A1,1,0)

At J1 type =Cellif(D\$3=A1,1,0)

At K1 type =Cellif(E\$1=B1,1,0)

At L1 type =Cellif(F\$1=B1,1,0)

Copy these formulae down to row 200.

These rows generate a block of boolean values that show if the data match the contingency table labels. Now we need to discover how the data are paired.

At N1 type =I1\*K1 and at O1 type = J1\*K1.

At P1 type = I1\*L1 and at Q1 type = J1\*L1.

Copy these formulae down to row 200.

Then we count how many of each pairing.

All we need to do is add up the columns.

At E2 type =sum(N1:N200)

At E3 type =sum(O1:O200)

At F2 type =sum(P1:P200)

At F3 type =sum(Q1:Q200)

Then we must find the row and column totals

At cell E4 type =E2+E3

At cell F4 type =F2+F3

At cell G2 type =E2+F2

At cell G3 type =E3+F3

At cell G4 type =G2+G3

At D1 type “Actual”

This completes the contingency table.

Copy D1:G4 to D7:G10

At cell E8 type: =E4\*G2/G4

At cell E9 type: =E4\*G3/G4

At cell F8 type: =F4\*G2/G4

At cell F9 type: =F4\*G3/G4

Change D7 to “Expected”

At D8 type = D2 and at D9 type = D3

At E7 type = E1 and at F7 type = F1

Once this spreadsheet has been saved, you can use it for any other 2x2 analysis. All you need to do is drop a new data set into A1:B200 and then change the table labels at D2, D3, E1 and F1 to match the data values in columns A and B.

D	E	F	G
Actual	Yes	No	
Female	68	40	108
Male	13	79	92
	81	119	200
Exptd	Yes	No	
Female	43.74	64.26	108
Male	37.26	54.74	92
	81	119	200

