

Gerald's Column **by Gerald Fitton**

My column is shorter than usual this month since our editor decided to publish early! Although this month's column has a statistical flavour I think that even those of you who are not statisticians will find parts of it interesting.

Turbo Drivers etc

Let me tell you about the experience of Eddie Lord.

He wrote to Computer Concepts on 26th Feb 1995 complaining about his CC turbo driver. He was having a series of problems with both his A 540 and a new Risc PC. The problems with the Risc PC were much worse than those of the A 540. His 26th Feb 1995 letter to CC finishes "I have written about this before but not received a reply . . ."

The good news is that Eddie eventually received a reply in April offering upgrades to Impression and the turbo driver. The bad news is that these did not completely solve his problem. Eventually he asked for and obtained a free of charge replacement for his "old original" dongle. In his letter to me dated 6th July 1995 he says "Apparently my old original dongle was not bidirectional and could not cope with turbo driver version 4 . . . My version of Publisher is now V 4.05 and the turbo driver is 4.03". The new dongle, together with the upgrades has solved his problems.

Now to Peter Boxall and his problems – they were similar to those of Eddie.

After some correspondence with me he got in touch with CC and his last letter to me includes the following: "I have to purchase V 4.03 of turbo drivers (£10 + VAT) . . . As regards the bidirectional dongle . . . the advice (is) that the one I have should be OK". Peter concludes this part of his letter to me with "Once I have installed the new turbo driver I will experiment".

Printers V 1.28c

I referred to these new Acorn printer drivers in last month's column. Since recommending them I have had correspondence stating that things go wrong with V 1.28c. Upon pressing my unhappy correspondents to describe their problems in more detail they have often been similar to those reported by Eddie and Peter. My suggestion to these people was to upgrade their copy of Publisher, to upgrade their turbo driver and to try to get a replacement dongle from CC.

All those who have upgraded and obtained a replacement dongle are now happy people!

If you are having difficulty with Printers V 1.28c, use Publisher and use turbo drivers then please don't write to me for help until you've done the same as Eddie! By all means write to Colin Singleton since he keeps asking me what I have against CC and pointing out that their products give him no problems.

Let me say publicly to Colin that I have nothing against Impression, indeed I am on record as saying that as a DTP both Style and Publisher are very good, I recommend them to you and Impression is my DTP of choice. Some of you will know that I run a user group called ILine for users of Style and Publisher. It is the fact that the package has so many good features which leads to the viability of ILine. If nobody bought Impression then ILine wouldn't exist! My only complaint is that interaction between some of CC's products and other people's products does cause problems – and then people write to me (and not to Colin)!

Weighted Averages

As far as I am aware there is no spreadsheet in which there is a formula for the calculation of weighted averages with the added feature of ignoring blank slots.

I have to acknowledge that the inspiration for this next section of my column is due to James Ducat who writes: "Help! I am an English teacher trying to produce a mark sheet for my classes which is based around the idea of a weighted average for each unit of work".

James sent me a demonstration of what he has done towards solving this problem. With slight modification the file he sent me is shown in the screenshot below.

	A	B	C	D	E	F	G
1							
2	Assessment		Review	Talk	Story	Avg_1	
3			Weight	7	6	10	
4	Adams	Sophie	7	7	8	7.43	
5	Bennett	David	7	7	7	5.17	
6	Carpenter	Darren	5	7	7	6.39	
7	Dawkins	John	5	4	6	5.17	
8	Parks	Louise	7	7	7	3.04	

I guess that you'll have noticed that James is using PipeDream. If you have Fireworkz and not PipeDream, indeed if you have any spreadsheet other than PipeDream, I'm sure that you'll be able to make the appropriate adjustments to the formulae I quote so that it suits your spreadsheet.

Before taking the screen shot I placed the cursor in slot F7 so that you can see the formula used by James. Taking out the \$ signs, used only during the replication process, the formula reads:

((C4*C7)+(D4*D7)+(E4*E7))/sum(C4E4)

This is the classic definition of a weighted average.

Blank Slots

A further quote from James is: “. . . when a child has been absent or not handed in work . . . PipeDream is treating this as a zero mark whereas I really want it to ignore the blank cell”. In the screenshot you will see that the slot D7 is blank. What James wants is for the result in F7 to be 7.00 and not the 5.17 given by the formula.

If you have used PipeDream or Fireworkz for calculating ordinary (non weighted) averages of a range of slots some of which are blank, then you will have discovered that the function `avg(range)` does not treat the blank slots as zero but ignores them completely. Let me give you an example using the spreadsheet shown in the screenshot. The average of the range C7E7 calculated by `avg(C7E7)` will be 7.00; this is 14 divided by 2 and not 14 divided by 3.

What we want is a function which ignores blank slots in the way that `avg()` does, but which calculates the weighted average instead of the non weighted average.

Weighted Average with Blank Slots

Have a look at the screen shot below.

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2	Assessment		Review	Talk	Story	Avg_1	Review	Talk	Story	Weight	Avg_2	
3			Weight	7	6	10						
4	Adams	Sophie	7	7	8	7.43	49	42	80	23	7.43	
5	Bennett	David	7	7	7	5.17	49	42	70	17	7.00	
6	Carpenter	Darren	5	7	7	6.39	35	42	70	23	6.39	
7	Dawkins	John	5	4	6	5.17	35	24	60	23	5.17	
8	Parks	Louise			7	3.04			70	10	7.00	

Again I have placed the cursor in one of the crucial slots, I7, before taking the screenshot so that the formula is visible. It is:

```
if(D7= "a", "", (D4*D7))
```

In the ‘Application of Number’ skill courses I teach at College a distinction is made between formula expressed in symbols (as above) and the same formula expressed in words. Using symbols is a higher level skill. Indulge me whilst I practice converting the above formula into words as: “If D7 is a blank slot then make I7 a blank slot, otherwise make $I7 = D4 * D7$ ”.

Understanding the formula in K7 is important to following why this application works.

It is:

$\text{if}(C7="",0,C4)+\text{if}(D7="",0,D4)+\text{if}(E7="",0,E4)$

and gives the value 17. This 17 is made up of the 7 in C4 and the 10 in E4. The 6 in D4 is replaced with the value 0 because D7 is blank.

The average calculated in slot L7 ignores the blank slots. The formula it contains is:

$\text{sum}(H7J7)/K7$

and gives the value $(49+0+70)/17 = 7.00$.

This is the weighted average ignoring blank slots which James requires.

What is a Median

James also wants a formula for the median of each of the columns of data.

When dealing with highly skewed (lopsided) distributions such as the distribution of wages you can easily be misled if you concentrate only on the mean. One way of avoiding being misled is to think of the median instead. The median wage is a wage such that half the wage earners earn more and half less than the median wage. I notice that the media now refer to the median wage (much lower than the mean) and Labour politicians talk in terms of a proposed minimum wage which is some proportion not of the mean but of the median wage.

So that you can appreciate the advantage of using the median instead of the mean let me relate (in much shortened form) one of the many anecdotes I tell my statistics students and which, like all my best stories, is almost true!

Some years ago, when much younger, my son, David, talked his way into a salesman's job. He was taken on a three month trial under the following conditions. "Start at £3000 a year and, if you do well then I'll raise your salary to at least the same as the other sales staff." David asked what the average salary was and got told £10000 pa (it was a long time ago when £10000 was a lot).

He did well and 'won' a set of (cheap) cuff links as 'Salesman of the Month' in his third month. He was looking forward to at least the monthly equivalent of £10000 pa and was amazed to receive only half the amount he expected.

After an argument he was told that the £10000 'average' for the company consisted of 9 salesmen (including David) on £5000 pa and the 'boss' on £55000 pa! So much for arithmetic means! The median wage for that company was £5000. David should have asked not about the mean salary, £10000, but about the median salary, £5000.

A second even shorter example: About ten years ago when inflation was in double figures nurses had a pay rise which averaged 18% (true) but 80% of nurses received less than an 8% rise (also true)! The median wage rise was around 6%. A few on high wages received wage rises of 25% or so pulling up the mean to 18%!

Finding the Median

PipeDream does not have a built in function to calculate the median of a set of data. No problem! Just write a custom function called “median”! One of the lovely things about well written custom functions is that you don’t have to understand their inner workings to use them. Over the years I have accumulated a wide range of custom functions provided by correspondents. I haven’t looked at the inner workings of some of those, like “median” which I use the most!

Back in early 1992, a contributor who shall remain anonymous (at least so far as this column is concerned) sent me a custom function which I have used on many occasions to calculate the median of a range of values. Inspired by James, I studied its inner workings for the first time last week – now I’ve improved it so that it ignores blank slots.

The modified function is on the Archive monthly disc. It is:

```
....function(“median”,“list:array”)
\ return the median value from an unsorted list
\ sort the list
...[c_sort]sort(@list)
\ define and initialise ‘local’ variables
...set_name(“arow”,B9)
...set_value(arow,rows(A6))
\ delete blank slots
...repeat
...if(index(A6,1,arow)=“”,set_value(arow,arow-1),)
...until(index(A6,1,arow)>“”)
\ if odd number of items, find the middle one
...if(mod(arow,2)=1,result(index(A6,1,(arow+1)/2),))
\ if even number of items, find middle two and split the difference
...index(A6,1,arow/2)
...index(A6,1,arow/2+1)
...result((A20+A21)/2)
```

The word “median” on the first row is the name of the function.

The “list:array” is a parameter passed to the custom function and is the unsorted array.

In the custom function the slot A6 is the slot containing[c_sort]sort(@list). After “sort” has done its work the slot A6 contains the sorted version of the array. I am not going to include a textual version of “sort” in this column; it is on the monthly disc and consists of a simple sort routine containing a couple of nested for – next loops.

At the bottom of the sorted array you will find all the blank slots – if there are any.

I have defined a local variable 'arow'. Initially the value of 'arow' is the number of rows in the sorted array. The repeat – until loop which follows the heading “\ delete blank slots” reduces the value of 'arow' until 'arow' is the number of items in the sorted array which are not blank.

The median of a sorted list is the middle one. The remaining rows of the custom function find the value of the middle item. If the number of items in the list is even then the median is considered to lie half way between the two middle values.

James would like a custom function which calculates the two quartiles. The upper quartile is a value such that 25% of the set of data is larger than the upper quartile. The lower quartile is defined in a similar way. Those of you in teaching contemporary uses of statistics will know that there is a great deal of interest at the moment in the 10% of households that are either at the top or bottom of the sorted list! Perhaps you'll be able to modify the custom function so that it calculates these percentiles. Anyway, if you do improve the custom function then please let me know; send me a disc copy so that I can make it more easily available to others such as James.

The Lottery

I knew that this item would generate much interest. I have had many letters about the statistics, philosophy and morality of gambling together with many amusing and some distressing anecdotes. One piece of philosophy which is repeated in many of my letters is “Be careful what you wish for – it may come true!”. I'm sorry but most will have to wait for another day – but please keep writing. I've selected just one piece of correspondence; files relating to it are on the Archive monthly disc.

Donald C Bindon has sent me 32 weeks of lottery results together with an analysis of frequencies using the Poisson distribution. I reckon that the true distribution must be Binomial (with $n = 32$) but a Poisson distribution is a good approximation; the approximation improves as the weeks go by and n gets larger. His submission includes an interesting custom function which, almost as a throw away item, includes an insertion sort routine which I would like to include in my 'median' calculations! His custom function also uses the deref() function in conjunction with index(,,) in a way which I find most interesting. I'm sorry I don't have time to go into detail, perhaps another day – anyway you'll find it on the monthly disc.

Donald reports that:

“An article in The Times newspaper of Saturday 28th Jan 1995, reported that a chi-square analysis of 12 years of the results of the Australian National Lotto had concluded that they were far from truly random. Anyone wanting to improve their chances of winning could do so by following the trends”.

I can't believe it! I would prefer to believe that something has gone wrong with the statistical analysis. Over the years I have had many contributors who have sent me chi square custom functions both for PipeDream and for Fireworkz. I am always interested in receiving alternative versions. If you have one which analyses Donald's data (or, if you're one of my many antipodean correspondents, the Australian National Lotto results) then let's have it!

Finally

Thanks for your many letters. Please continue to send them to the Abacus Training address (not to Archive) which you'll find at the back of Archive.s