Don’t discard last week’s TV Guide!

Alan Graham
Centre for Mathematics Education,
The Open University, UK

◆INTRODUCTION ◆

AN **OLD** TV Guide has long been regarded as the ultimate symbol of uselessness. Here are some ideas for purposeful investigations based on data from this under-rated magazine which can involve simulation using a spreadsheet.

There seem to be two problems that confront a teacher who tries to bring the real world into the teaching of statistics.

- How can I find topics that will interest the students?
- Even if I find an interesting topic, where will I/ they get suitable data from?

This article offers some practical suggestions for tackling these questions by exploiting that much maligned and greatly under-rated of statistical sources – last week’s TV Guide.

It is useful for the teacher to begin any statistical investigation with a class topic or theme. Having agreed a context, each student or each small working group of students, can pose a clear question within the class theme. This question should be one to which they are personally interested in finding an answer and which will drive their statistical investigation through its main stages to a reasoned conclusion.

Contained in this article are some suggestions for statistical investigations which arose out of a casual remark by a student concerning films shown on TV. She claimed that: ‘The Guardian TV film reviewer always goes for the oldies. If the film isn’t in black-and-white, he won’t recommend it!’

This remark generated interest in a number of statistical investigations on the theme of TV films and last week’s TV Guide proved an admirable source of information. The investigation below is presented using the following four stages, known as the PCAI stages, described in more detail in Graham (1990, 1994):

Stage P	Pose the question
Stage C	Collect the data
Stage A	Analyse the data
Stage I	Interpret the results

◆INVESTIGATION-FIRST DRAFT◆

Stage P pose the question

‘Are more films shown on terrestrial TV (i.e. BBC 1 and 2, ITV and Channel 4) at weekends than on weekdays?’

Stage C collect the data

Last week’s TV Guide was the source used to generate the following data.

Films shown in the four terrestrial channels during Oct 9-15th 1993

Saturday	8
Sunday	7
Monday	4
Tuesday	6
Wednesday	6
Thursday	3
Friday	5
Total	39

Stage A analyse the data

This stage of the investigation usually involves performing a calculation or representing the data in a useful graph or diagram. Access to a suitable spreadsheet or statistical package will greatly aid the analysis, allowing the student to try out a range of possibilities. In this example, several different forms of analysis are explored, all aimed at clarifying the patterns and trends in the data. We start with a graphical representation in the form of a simple barchart.

(a) Graph the data using a barchart (Figure 1)

(b) Calculate the mean number of films Mean = 39/7 = 5.6 films per day

(c) Compare the barchart in Figure 1 with the daily average of 5.6 (Figure 2). Note that in this second version the students also reordered the bars to have weekend films on the right.

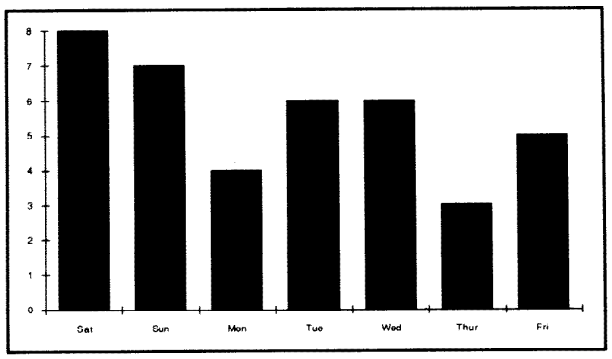

Fig. 1 Barchart showing the data in Table 1.

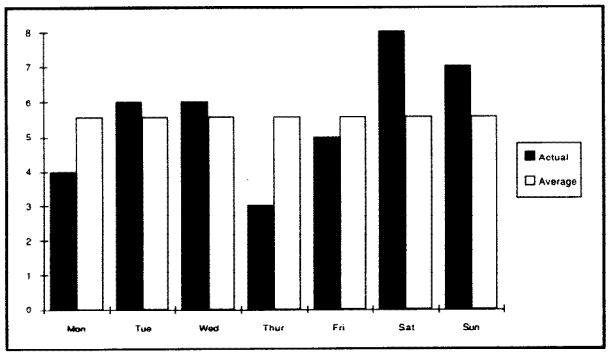

Fig. 2 Compound barchart comparing actual data with the daily average.

	A	B	C	D	E	F
1	Sim	Bins-array	Day	Freq		
2	3	1	Mon	8	Max-->	8
3	4	2	Tues	6	Min-->	3
4	5	3	Wed	4	Range-->	5
5	1	4	Thurs	6		
6	6	5	Fri	7		
7	5	6	Sat	3		
8	5	7	Sun	5		
9	3		TOTAL	39		
10	1					
...	...					
40	7					

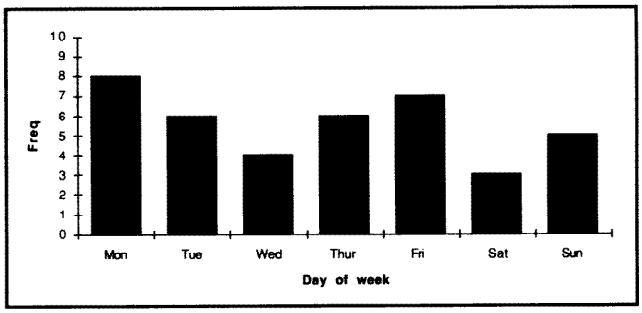

Fig. 3 Spreadsheet and graph showing a simulation of the number of films over a week.

Stage I interpret the results

The 'I' stage involves looking closely at the different forms of graphical representations and calculations and seeking some sensible conclusion about the original question. Clearly, just by glancing at the table or the barcharts, we can see that the number of films on the Saturday and Sunday is greater than any other day. However, it is hard to say whether this difference is significant. We need to investigate whether such a difference could have occurred by chance alone. So, an important interpretative question here is:

> How unlikely would it be to get this large a difference from the average just by chance alone?

This question is now tackled by a computer simulation on a spreadsheet. The particular spreadsheet used here is Excel4 (on a Macintosh) and the display values and corresponding graph are shown in Figure 3.

Column F gives the Max and the Min values and also the Range of the values contained in column D. The commands to generate these values, = Max(), =Min(), and =Range(), are available within the library of functions resident in Excel.

Finally, the graph shows the current state of column D. It is a good idea to fix the scale on the vertical axis to prevent it from automatically re-scaling, otherwise the comparison from one simulation to the next is less easy to make.

The principle of the simulation is to generate 39 random numbers in the range 1 to 7, corresponding to the days of the week. These are shown in column A and are generated by the command =INT(RAND()*7+1). Column B shows the 'bins-array' into which Excel4 will sort these 39 numbers. The particular command used in cell D2, namely,

{=FREQUENCY($A$2:$A$40,B2:B8)}

counts the frequencies of each of the seven values. This command must be copied down column D.

To run the simulation, press 'Command =' repeatedly. Each press of 'Command =' causes the random values in column A to be recalculated and the graph will redraw appropriately to match the new values. Figure 3 shows the results of just one typical run. Running a series of simulations takes only a few seconds. The benefit is that it shows, both graphically and numerically, just how much variation one can expect if the 39 films were allocated randomly throughout the seven days of the week. You will probably find that spreads are generally wider than you might expect and may reach the conclusion, therefore, that the observed values of a maximum of 8, minimum of 3 films on particular days does not seem to be highly unusual.

◆ INVESTIGATION ◆ - SECOND DRAFT

At this stage in any statistical investigation it is common to decide that a different form of analysis might be more helpful. For instance, in this case, the group of students came to the view that it was unhelpful to separate off the number of films for each day of the week. Given the wording of the question, they were really only interested in two measures – the number of films shown on weekdays and the number shown on weekends. They collapsed the data into these two headings as follows:

Films shown in the four terrestrial channels during 9-15 Oct 1993

Weekends	15
Weekdays	24
TOTAL	39

They now needed to make a second loop around the A and the I stages of the PCAI cycle, as follows: Using the same thinking as before, the 'expected' numbers of films shown at weekends and weekdays would be 2/7 and 5/7 of 39 respectively. In other words:

	Observed	Expected	
Weekends	15	2/7 x 39	= 11.1
Weekdays	24	5/7 x 39	= 27.9
TOTAL	39		= 39.0

So, as before, there are rather more films at the weekend than we might expect, but is the difference (between 15 and 11.1) significant? Figure 4 shows a computer simulation to investigate this key statistical question of just how much variation we might expect from chance alone.

	A	B	C	D	E
1	Sim		Simulated	Expected	Actual
2	7	Weekends	13	11.1	15
3	3				
4	3				
5	3				
6	5				
...	...				
40	4				

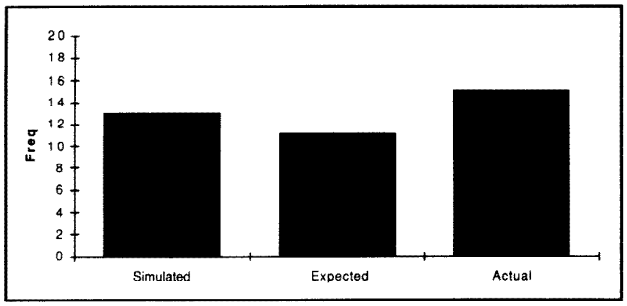

Fig. 4 Second computer simulation and graph showing the number of films shown at the weekend.

As before, the 39 random numbers are generated in column A.

Cell C2 again uses the =FREQUENCY() command and here it gives the count of how many of the 39 values are less than 3 (an outcome of 1 or 2 in column A is deemed to be a weekend film).

This time the students were rather more systematic in their simulation. (As before, the simulation was generated by pressing 'Command ='.) They ran the simulation 50 times and kept a tally of how often the cell value, C2, was 15 (the observed value) or greater. With each run, it was quite easy to see from the graph whether the simulated number of films at the weekend equalled or exceeded the observed value of 15.

In the event, this proved to be the case exactly seven times out of the *50* trials. The results seemed to confirm their earlier scepticism about whether there was a significant pattern in their data. From this last simulation they reached the conclusion that 'you could expect a result at least as extreme at the observed one something like 14% of the time by chance alone'.

◆IDEAS FOR FURTHER ◆ INVESTIGATIONS

Below are some of the other statistical investigations that arose from this theme of TV films and which you might like to pursue with your own class. And, just in case your departmental budget doesn't run to a copy of last week's TV Guide, I have also included the data which the students collected to get you started (Table 1).

- Does the Guardian TV film reviewer tend to select old films?
- Do satellite TV films tend to be more recent than those shown on terrestrial TV?
- Are films shown on Sky more recent than those on the Movie channel?
- Do modem films have a shorter viewing time than older films? (The data in Table I will not support this investigation, but suitable relevant data should be available in a TV Guide).

Note, by the way that this data set contains gaps (shown as 'data missing'). It is worth stressing to students that this is a common feature of real data and they will need to be prepared to use their common sense to deal with such situations.

References

Alan Graham, 1990, *Investigating Statistics — a beginner's guide*, Hodder and Stoughton.

Alan Graham, 1994, *Teach Yourself Statistics*, Hodder and Stoughton.

	Sky	Movie Channel	All terrest	Reviewed terrest
SAT	66 69 78 77 91 87 92 91 90 91	83 90 64 69 90 91 91 91 91 82 90	84 74 50 43 63 83 59 31	50 43 59 63 31
SUN	DATA	MISSING	52 85 55 37 38 90 85	37 38 85 90 85 55
MON	91 69 68 80 91 91 91 92 69 90	53 90 34 91 73 90 91 91 91 91 90 81 91	47 87 87 40	40
TUES	91 79 66 80 91 91 90 91 90 91 90	57 89 66 57 56 89 91 91 91 90 91 90	85 36 44 79 41 41	36

	Sky	Movie Channel	All terrest	Reviewed terrest
WED	90 64 80 51 90 91 91 70 91 90	62 87 38 82 65 87 92 92 90 91 73 88	89 75 32 88 36 87	36 32 75
THURS	92 65 71 65 92 91 91 90 68 73	36 58 74 90 90 91 91 91 89 84	43 56 92	56 92
FRI	DATA	MISSING	85 71 58 88 69	85 69 88

Table 1. Year of film (taken from the Guardian Guide, Oct 9th 1993) 66 = 1966.