

Assignment 3 contained the following question.

Find an article in a newspaper or magazine published in 2009 that gives the results of a survey, observational study or experiment. Cut out the article, which should be no bigger than half of an A4 page, and staple it to the top half of an A4 page. Write down the name of the newspaper or magazine, the page where the article appeared, and the date of publication.

In the second half of the page, comment briefly on the article. For example, does it describe a survey, an observational study or an experiment? If a survey, how was the sample chosen? If an observational study, what conditions were compared? If an experiment, what were the treatments and the experimental units? What was measured?

Do you think this investigation was conducted well?

When we think about data we have to think about its

**collection** What question were the data-collectors trying to answer? How did they collect the data? What did they measure or ask?

**summary and display** This is mostly done in our Minitab practicals.

**analysis and interpretation** The rest of the course is about this.

The assignment question was about the *collection* phase. These notes contain a number of examples submitted in the assignment, to illustrate important points about data collection.

## **A survey**

**Source** *The Guardian*, 27/01/09, pages 1–2.

**Question of interest** What proportion of the UK electorate intends to vote for each party?

**Sample** “ICM interviewed a random sample of 1,003 adults aged 18+ by phone between 23 and 25 January 2009 ... across the country.”

**What exactly was asked?** *This doesn't appear to be stated.*

**Analysis** “... results have been weighted to the profile of all adults.”  
*What does this mean?*

**Display** A pie chart.  
*Suitable, because the data are on a nominal scale.*

## **A survey**

**Source** *Metro*, 26/01/09, page 21.

**Question of interest** How many children aged 11–16 in Britain have been injured trying to stop arguments between adults at home?

**Sample** The NSPCC asked 1075 children in this age-group.  
*How were they chosen?*

**What exactly was asked?** *This doesn't appear to be stated.*

**Analysis** “308,889 of them have been hurt in this way.”  
*Hmm... , that is rather precise, how do they know it is not 308,890?*

## An observational study

**Source** The student cited a report by the BMJ Group on 07/01/09 under [www.guardian.co.uk/lifeandstyle/besttreatments/](http://www.guardian.co.uk/lifeandstyle/besttreatments/)  
*Does this count as a newspaper? When I accessed the site on 6/02/09 I found that it did not contain the text reporting this study.*

**Question of interest** For elderly people, does a 'flu jab lower the risk of dying?  
*Define "elderly". Dying over what period of time?*

**Observational units** An unspecified large number of "elderly people".

**What was recorded?** Presumably, the numbers of "elderly" people in each of the four boxes below.

'Flu jab?	Yes	No
Died? Yes	<input type="checkbox"/>	<input type="checkbox"/>
No	<input type="checkbox"/>	<input type="checkbox"/>

**Analysis** "... fewer people who got the jab died."  
*This is a bit vague. How many fewer? And would it be more relevant to state the proportion?*

## An observational study

**Source** *The Economist*, 15/01/09.

**Question of interest** Can you predict the profit that a financial trader makes from "high-frequency trading" by the ratio

$$\frac{\text{length of ring finger}}{\text{length of index finger}}?$$

**Observational units** 44 men on the London trading floor, engaged in this type of trading, for 20 months.

**What was recorded?** For each man  $i$ , the ratio  $x_i$  and the profit  $y_i$ .

**Analysis**  $y$  tends to increase as  $x$  increases.

## **An observational study using surveys**

**Source** *Wall Street Journal*, 06/01/09, page D2.

**Question of interest** What are the best and worst jobs?

**Observational units** 200 job titles.

**What was recorded?** Numerical values for five variables—environment, outlook, income, stress and physical demands—, presumably based on surveys of people within each job.

**Analysis** The numbers were combined to produce a ranked list.

*HOW?*

The list begins

- “1. Mathematician
2. Actuary
3. Statistician
4. Biologist
- ...”

## **An observational study using a census**

**Source** [www.bbc.co.uk/1/hi/health/7852897.stm](http://www.bbc.co.uk/1/hi/health/7852897.stm), 27/01/09.

*Does this count as a newspaper?*

**Question of interest** Is the number of alcohol-related deaths per year changing? Is the trend different for men and women?

*Aren't the “proportions” more useful than the absolute “numbers”?*

**Data source** The Office for National Statistics.

**What was recorded?** The Home Office keeps records of *all* UK deaths, showing date, age, gender, cause of death, ....

**Analysis** For men, there were 18.1 such deaths per 100,000 in 2007; for women, 8.7 per 100,000. For both sexes, the number of such deaths per 100,000 people was slightly less in 2007 than in 2006, but much higher than in 1991.

*The detailed figures for 2006 and 1991 were not given.*

## **An experiment**

**Source** *The Sun*, 27/01/09, page 20.

**Question of interest** Can young men make themselves more attractive to young women by putting on scented deodorants?

**Treatments** { One particular fragrant deodorant;  
no deodorant.

**Experimental units** Some men aged 19–35.

*How many? Were they used once each? Or once with each treatment?*

**What was measured?** Attractiveness to “a test panel of women”.

*Exactly how was this measured?*

**Analysis** The fragrant deodorant made the men more attractive.

## **An experiment**

**Source** *The Epoch Times*, 21–27 January 2009, page 11.

**Question of interest** Can different diets change the blood-sugar levels in diabetic people?

**Treatments** { One low-carb diet;  
one low-GI diet.  
(GI = glycaemic index.)

**Experimental units** 49 obese people with type-2 diabetes.

**What was measured?** Change in the level of HBA1c after 24 weeks.

**Analysis** The level fell by 1.5% on the low-carb diet  
0.5% on the low-GI diet.

## **Forecasts**

Examples included the weather forecast, or predicted house prices.

Forecast numbers are predictions. I do not regard *predictions* as *data*.

## **Announcements**

Examples included announcements, by one to three firms, of their number of job losses this month, or an announcement by a single firm of its year-end share price.

I do not regard *announced* data as *collected* data. What about the number of job losses from the firms who made no announcement? Maybe there were no losses in most firms?

## **Collection of unrelated numbers**

One example, from *The London Paper*, 20/01/2009, page 6, gave a collection of numbers, on different scales and to different accuracies, with nothing in common except the inauguration of the current US president. These included:

“0 degrees Celsius is the maximum daytime temperature in Washington DC today”;

“1,000 security tips are received daily by the FBI”;

“800 pounds of bison stocked by the Marriott Wardman Park, Washington’s largest hotel, for meals”.

There is no possibility of comparing these numbers. They are too silly to be regarded as data.