Storks Deliver Babies ($p = 0.008$)

KEYWORDS:
Teaching;
Correlation;
Significance;
p-values.

Robert Matthews
Aston University, Birmingham, England.
e-mail: rajm@compuserve.com

Summary
This article shows that a highly statistically significant correlation exists between stork populations and human birth rates across Europe. While storks may not deliver babies, unthinking interpretation of correlation and *p*-values can certainly deliver unreliable conclusions.

◆ INTRODUCTION ◆

Introductory statistics textbooks routinely warn of the dangers of confusing correlation with causation, pointing out that while a high correlation coefficient is indicative of (linear) association, it cannot be taken as a measure of causation. Such warnings are typically accompanied by illustrative examples, such as the correlation between the reading skills of children and their shoe size, or the apparent relationship between educational level and unemployment (see e.g. Freedman *et al.* 1998). However, such examples are often either trivially explained via an obvious confounder (e.g. age, in the case of reading age and shoe size) or are not obviously cases of mere association (e.g. educational level may indeed be at least partly responsible for time spent unemployed). In what follows, I give an example based on genuine data of an association which is clearly ludicrous, but which cannot be so easily dismissed as non-causal via an obvious confounder.

My starting point is the familiar folk tale that babies are delivered by storks. The origins of this connection are believed to lie partly in the association between storks and the concept of women as bringers of life, and also in the bird's feeding habits, which were once regarded as a search for embryonic life in water (Cooper 1992). The legend lives on to this day, with neonate-bearing storks being a regular feature of greetings cards celebrating births.

While it is (I trust) obvious that the legend is complete nonsense, it is legitimate to ask precisely how one might set about refuting it scientifically. If one were approaching the question in the same way that many other links are investigated (e.g. suspected links between diet and cancer risk), one may well decide to carry out a correlational study, to see if the number of storks in a country bears a simple relationship to the number of human births in that country. Although the presence of a statistically significant degree of correlation cannot be taken to imply causation, its absence would certainly constitute evidence against a simple relationship. This possibility can quickly be investigated in the present case using standard hypothesis testing, with the null hypothesis being the absence of any correlation between the number of storks and the number of live births in a particular country. This I now proceed to do.

◆ TESTING THE STORK-BIRTH ◆ RELATIONSHIP

The white stork (*Ciconia ciconia*) is a surprisingly common bird in many parts of Europe, and data on the number of breeding pairs are available for 17 European countries (Harbard 1999, pers. comm.); the latest figures, covering the period from 1980 to 1990, are given in table 1, along with demographic data taken from Britannica Yearbook for 1990.

Plotting the number of stork pairs against the number of births in each of the 17 countries, one can discern signs of a possible correlation between the two (see figure 1).

The existence of this correlation is confirmed by performing a linear regression of the annual number of births in each country (the final column in table 1) against the number of breeding pairs of white storks (column 3). This leads to a correlation coefficient of $r = 0.62$, whose statistical significance can be gauged using the standard t-test, where $t = r \cdot \sqrt{[(n-2)/(1-r^2)]}$ and n is the sample size. In our case, $n = 17$ so that $t = 3.06$, which for $(n-2) = 15$ degrees of freedom leads to a p-value of 0.008.

Country	Area (km^2)	Storks (pairs)	Humans (10^6)	Birth rate (10^3/yr)
Albania	28,750	100	3.2	83
Austria	83,860	300	7.6	87
Belgium	30,520	1	9.9	118
Bulgaria	111,000	5000	9.0	117
Denmark	43,100	9	5.1	59
France	544,000	140	56	774
Germany	357,000	3300	78	901
Greece	132,000	2500	10	106
Holland	41,900	4	15	188
Hungary	93,000	5000	11	124
Italy	301,280	5	57	551
Poland	312,680	30,000	38	610
Portugal	92,390	1500	10	120
Romania	237,500	5000	23	367
Spain	504,750	8000	39	439
Switzerland	41,290	150	6.7	82
Turkey	779,450	25,000	56	1576

Table 1. Geographic, human and stork data for 17 European countries

◆ ANALYSIS ◆

What are we to make of this result, which points to a highly statistically significant degree of correlation between stork populations and birth rates? The correlation coefficient is not particularly high, but according to its p-value, there is only a 1 in 125 chance of obtaining at least as impressive a value *assuming* the null hypothesis of no correlation were true. Yet as with any p-value (and contrary to what unwary users of them believe),

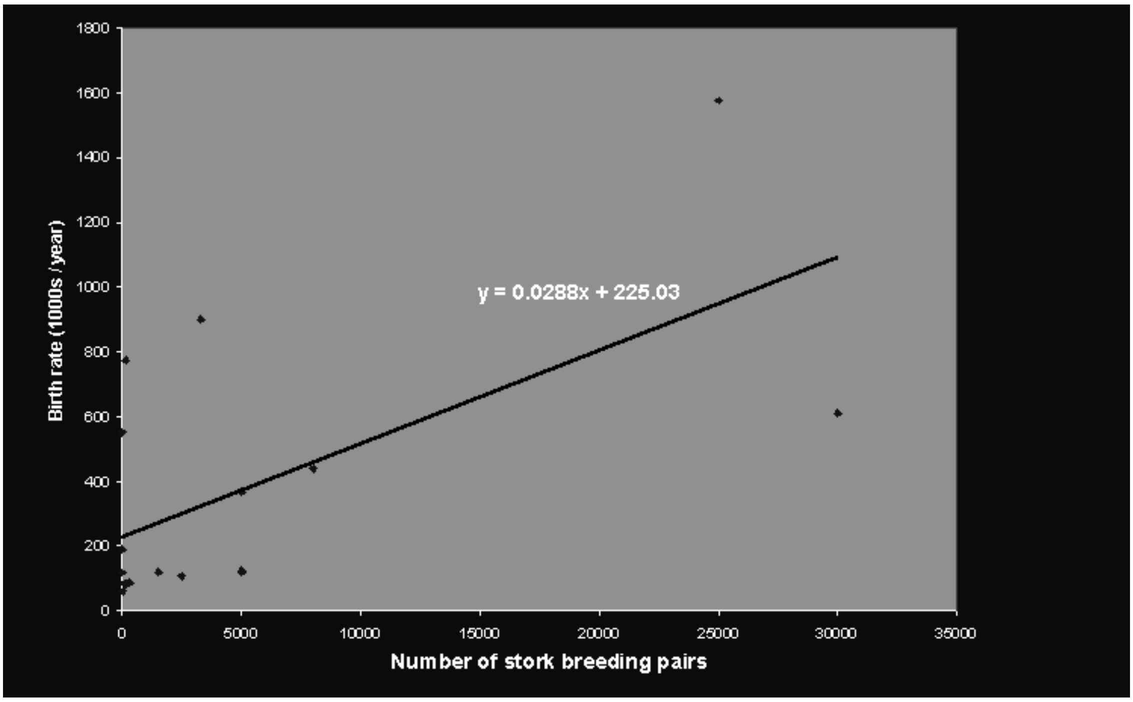

Fig 1. How the number of human births varies with stork populations in 17 European countries.

this does *not* imply that the probability that mere fluke *really is* the correct explanation is just 1 in 125; still less does it imply a $124/125 = 99.2\%$ probability that storks *really do* deliver babies.

Such apparent nit-picking distinctions are frequently overlooked by consumers of p-values. In the case of the correlation between storks and human births, however, they no longer seem so pedantic: indeed, they provide the very welcome 'escape route' by which to avoid a patently ludicrous inference. The most plausible explanation of the observed correlation is, of course, the existence of a confounding variable: some factor common to both birth rates and the number of breeding pairs of storks which – like age in the reading skill/shoe-size correlation – can lead to a statistical correlation between two variables which are not directly linked themselves. One candidate for a potential confounding variable is land area: readers are invited to investigate this possibility using the data in table 1.

◆ CONCLUSION ◆

Standard statistical texts routinely warn of the fallacy of mistaking correlation for causation, but the examples they provide are usually either trivial, with obvious confounders, or lack clear non-causality. The empirical relationship between the number of stork breeding pairs and human birth rates in 17 European countries provides a non-trivial example of a correlation which is highly statistically significant, not immediately explicable and yet causally nonsensical. Indeed, its sheer absurdity has pedagogic value beyond the correlation/causation fallacy alone, as it compels greater attention to be paid to the precise meaning of p-values, and promotes greater recognition of the fact that rejection of the null hypothesis does not imply the correctness of the substantive hypothesis.

Acknowledgements

The author is very grateful to Chris Harbard of the Royal Society for the Protection of Birds for supplying the stork data, and to Professor Dennis Lindley for valuable discussions.

References

Cooper, J.C. (ed.) (1992). *Brewer's Myth and Legend*. London: Cassell.

Freedman, D., Pisani, R. and Purves, R. (1998). *Statistics* (3rd edn). New York: W.W. Norton.