xf. BIRTH ORDER.

(2) JUL 14 1965

Anthony
Edwards

7/98 priceless!

A CAUTIONARY TALE

Demographer: Good morning, Mr. Statistician, you are just the person I wanted to see.

Statistician: Good morning; what can I do for you?

Dem Well, I need to know how to measure association in a contingency table. I am interested in seeing whether first sons follow in their fathers' footsteps more often than second and later sons, so I have classified both fathers and sons as having either 'professional' or 'non-professional' occupations, and then recorded whether father and son were in the same occupation or not, not forgetting to separate first sons from the others.

Stat I see. And now you have a 2×2 contingency table to interpret?

Dem Yes, here it is:

	Occupation		
	Same	Different	
1st sons	600	700	1300
later sons	700	600	1300
	1300	1300	2600

ANTHONY EDWARDS

I'm puzzled, because I should have thought that first sons would tend to follow their fathers' occupation more often than later sons, but they don't seem to. How should I measure the strength of the effect?

Stat Chi-squared comes to 200/13, if my ability to do sums in my head is not impaired at this early hour, which is 15.4, a very significant value for one degree of freedom. So you certainly need a measure of association. But let's skip that for a moment; why do you think you have found such an unexpected result?

Dem I don't know at all. I know about biassed samples, so I was careful about that.

Stat Do you know about stratification?

Dem Well, I've heard of it, but isn't that to do with correlation and heterogeneity?

Stat Yes, it is. What happens if you look at the contingency tables for the fathers' occupations separately? A professional one, and a non-professional one.

Dem I don't know; I am not interested in which occupation, but only if father and son had the same occupation. But we can look at the original figures:

Father's Occupation Professional

Son's Occupation

	same	different	
1st sons	300	100	400
later sons	600	300	900
	900	400	1300

Father's Occupation Non-Professional

Son's Occupation

	same	different	
1st sons	300	600	900
later sons	100	300	400
	400	900	1300

Stat Now they are very interesting figures. Don't you see that the proportion of first sons with their fathers' occupation is greater than the proportion of later sons with their fathers' occupation, in both cases, just as you had expected?

Dem Why, that's extraordinary! No, it's absurd! If heterogeneity can cause that, I'll eat my hat. Let's look at the strength of the association in both cases. The sample sizes are the same, so we can use chi-squared can't we?

Stat We can indeed, and in both cases we find a value of 9.02, very significant, and the association is in the same sense in each case.

Dem But if the values are the same, that means there's no heterogeneity. Oh, this is more than I can stand! I have two contingency tables in full agreement about the association, but when I ask them the same question together I get a silly answer.

Stat Is it so silly?

Dem Of course it is. I don't know whether I'm standing on my head or my heels.

Stat Exactly what was your question?

Dem I just wanted to know whether first sons followed in their fathers' footsteps more often than the others.

Stat I see. Well, the answer, according to your data, is no, they don't.

Dem (sarcastically) Very funny. What's true of the two halves of the population isn't true of the whole. You statisticians should go and..... .

Stat D'you know, I think you're upset about the answer to your question because you really didn't mean to ask that question at all. What you really wanted to know was whether first sons followed in their fathers' footsteps more often than later sons, other things being equal.

Dem What do you mean, 'other things being equal'?

Stat Let's look at the figures again. First, forget about the son's occupation. Then we have:

Father's Occupation

	Professional	Non-Professional	
number of 1st sons	400	900	1300
number of later sons	900	400	1300
	1300	1300	2600

You see, the distribution of family sizes must be quite different according to the father's occupation. In the professional class larger families seem to be the rule, whilst amongst the non-professional class there were at least five hundred only sons. As a result of this there is a pretty high association between being a first son and having a father in a non-professional occupation.

Dem Go on.

Stat Secondly, forget about the son's birth-order. Then:

	Father's Occupation		
	Professional	Non-Professional	
Sons Occupation			
same	900	400	1300
different	400	900	1300
	1300	1300	2600

We find that sons of fathers in the professions tend to become professional people, but that sons of fathers not in the professions tend to break away and become professional also. This, of course, is due to the well-known fact that a larger and larger proportion of people is employed in the professions.

Dem Go and teach your grandmother to suck eggs.

Stat Patience, please! Now if there is a high association between being a first son and having a father in a non-professional occupation, and at the same time a high association between having a father in a non-professional occupation and a son in a different occupation, then it's not really surprising to find a high association between being a first son and having a different occupation from your father.

Dem No ... I mean yes, well, I suppose so.

Stat And that's just what you found.

Dem Was it?

Stat In spite of the fact that first sons do tend to follow in their father's footsteps in both the professional and non-professional classes.

Dem Yes ... I mean no. What time is it?

Stat So when I say 'other things being equal' I mean that, so far as the data will permit, the effects of subsidiary correlations should be elimintated. So if I were you I would ask a slightly different question, one which is, I suspect, more meaningful from a demographic point of view.

Dem (after long pause) Ah, now I see! You mean that I should ask whether the first sons of bus conductors are bus conductors more frequently than the later sons, and whether the first sons of bus drivers are bus drivers more frequently than the later sons, but on no account must I ask whether the first sons of the employees of a bus company more often follow their father's occupation than the later sons?

Stat No, that's not what I mean at all. What I want you to do is simply to ask the question to which you want an answer, and not some other question. You see, if we further subdivide the bus drivers into tall ones and short ones, we will get yet another answer, won't we?

Dem I see. Ask a silly question and you'll get a silly answer. I wasn't brought up to think this clearly.

Stat So your father wasn't a statistician? Tell me, are you an only son?

Analysis of First Versus Later Sons' Tendency to Follow their Fathers' Occupation.

Raw data. Eleven 2×2 tables (one for each occupational class) each organized as follows:

	Occupations of Father and Son	
	same	different
1st sons	a	b
later sons	c	d

Object of Analysis. To determine in which (if any) of the 2×2 tables the association should be judged significant.

Measure of Association. log ab/cd, with large sample variance $\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$.

Father's Occupation	Association	S.D.	X_1^2
Professional	.573	.151	14.376***
Farm	.305	.092	10.979***
Managers	.027	.108	.063
Clerical	.157	.294	.285
Sales	.124	.229	.292
Crafts	-.035	.086	.164
Operatives	-.242	.102	5.592*
Service	.403	.266	2.298
Farm Labor	.152	.429	.125
Laborers	-.069	.285	.058
Weighted Mean	.084	.043	3.826

Total	X_{10}^2	34.232***
Weighted Mean	X_1^2	3.826
Heterogeneity	X_9^2	30.406***

Notes.

(1) The occupational class 'household' has been omitted because of insufficient data.

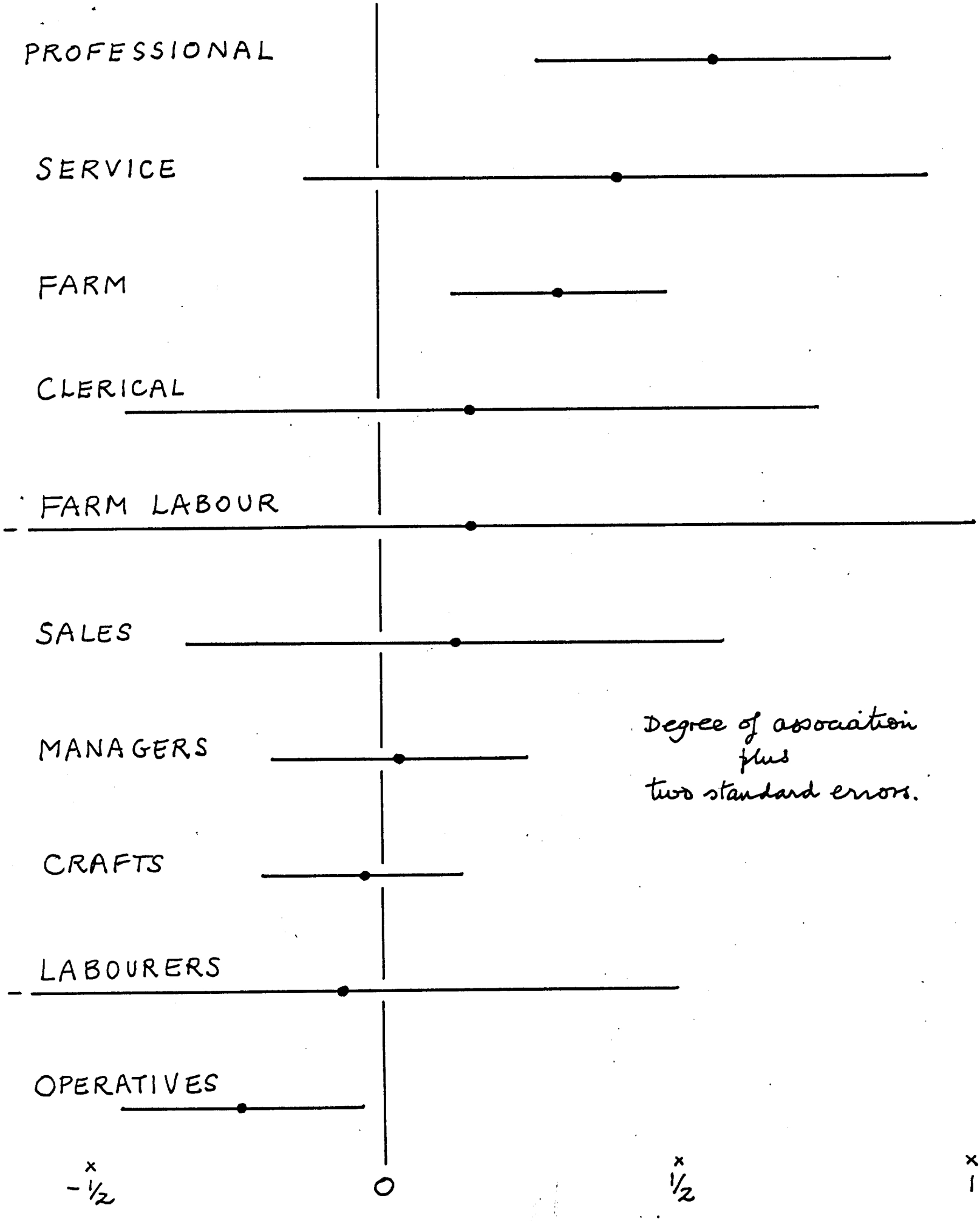
(2) The X_1^2 values have been calculated from the associations and their variances, but should not differ sensibly from the contingency X_1^2 values.

(3) The weighted mean association has not been calculated from the 2×2 table formed by summing the separate tables, but by forming the weighted mean of the separate associations. The former technique would have been incorrect.

(4) The calculations have not been checked.

Conclusion.

There is too much heterogeneity in this matter between occupational classes for an overall measure of association to be valid. However, two classes are so outstanding that there is little doubt that when the father is professional or a farmer, the first son tends to follow in the same occupation move frequently than later sons. The remaining classes can probably be regarded as homogeneous and showing no such effect. The difference in this tendency between the professions and farmers is not significant.


PROFESSIONAL
SERVICE
FARM
CLERICAL
FARM LABOUR
SALES
MANAGERS
CRAFTS
LABOURERS
OPERATIVES
-½
0
½
1
Degree of association
plus
two standard errors.