Available online at www.sciencedirect.com
ScienceDirect
Journal of Theoretical Biology 245 (2007) 597–599
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

Letter to Editor

Letter to the editors regarding some papers of Dr. Satoshi Kanazawa

Dear Editors,

Dr. Satoshi Kanazawa has published several papers in your journal recently finding evidence for differential sex ratios, with big and tall parents, engineers, violent men, and less attractive parents being disproportionately more likely to have sons than daughters. As a statistician, not a biologist, I cannot speak to the theoretical content of these papers, but I believe the statistical arguments therein to be seriously flawed. This is not to say that the results are not scientifically correct, just that they have not been convincingly demonstrated by the statistical evidence.

In this letter, I discuss two serious statistical problems, along with one mistake in interpretation of a statistical analysis. I hope that these comments will be useful in obtaining more reliable results in the future.

1. Potential for selection bias in controlling for total number of children

First, in the papers on big and tall parents, engineers, and violent men (Kanazawa, 2005, 2006; Kanazawa and Vandermassen, 2005), the statistical analysis is a regression of number of boys or girls born, controlling for income, the total number of children of the other sex, and other predictors. I will discuss in the context of the paper on engineers and nurses but the same issues arise in the other papers.

The predictors of interest in Kanazawa and Vandermassen (2005) are indicators for whether the parent's occupation is ''systematizing'' (e.g., engineering) or ''empathizing'' (e.g., nursing). The coefficients of ''systematizers'' and ''empathizers'' are statistically significant, and the authors conclude that systematizers are much more likely to have boys and empathizers are much more likely to have girls, and then back this up with some serious biodeterministic theorizing (for example, referring to occupations as ''brain types'').

These regression coefficients cannot be interpreted so cleanly, however. One of the predictors in the regression is the total number of children of the other sex. A problem arises because different people may very well try for another child or not after having one son, or one daughter, or other pattern—and the regression analysis does not account for this. The number of children in the family is an intermediate outcome (in econometric terminology, an ''endogenous variable''; see Woolridge, 2001) because it can be itself affected by the predictor of interest (the parent's occupation).

To see how this could distort the regression analysis, consider a simplified scenario in which all families have either one or two children, with the following rules: (a) engineers will stop at one child if first child is a boy, but if the first child is a girl, they will have another child; (b) nurses will stop at one child with probability 30% and continue on to a second child with probability 70%, regardless of the sex of the first child. In this model, we shall also suppose that the probability of a boy is exactly 50% for all births; thus we are supposing that the true effect—the difference in sex ratios between engineers and nurses—is actually zero.

Under this model, engineers will have the following distribution of family types: 50% b, 25% gb, 25% gg. Nurses will have the distribution: 15% b, 15% g, 17.5% bb, 17.5% bg, 17.5% gb, 17.5% gg. We then simulate 800 families (400 from engineers, 400 from nurses) under this model and then regress the number of boys on parental occupation and number of girls, to obtain the following result:

lm (formula = n.boys ~ engineer + n.girls)

	Coef.est	Coef.se
(Intercept)	1.18	0.02
engineer	−0.14	0.02
n.girls	−0.56	0.02

The coefficient for ''engineer'' is negative and statistically significant, even though under our model the probability of a boy birth is 50% in all circumstances. The nonzero coefficient is an artifact arising from controlling for the number of girls, a variable that is influenced by the other regression predictor.

In addition, the coefficient for the number of girls is negative. Kanazawa and Vandermassen (2005) observe and misinterpret a similar pattern in their data, writing,

> when we control for all the variables included in our equations, those who have more biological daughters have fewer biological sons, and those who have more biological sons have fewer biological daughters. This seems to suggest that parents specialize in producing children of one sex or the other, some producing mostly

0022-5193/$ - see front matter
doi:10.1016/j.jtbi.2006.11.005

or exclusively boys, and others producing mostly or exclusively girls.

In fact, however, this pattern can be explained as a pure statistical artifact, as can be seen in this pattern occurring in our simulation in which births are completely random.

The model as fit by Kanazawa and Vandermassen has additional, related, problems in that it controls for income, which is also an intermediate outcome (that is, it is influenced by parental occupation) so it does not make sense to compare two people with different occupation-classes and the same income. Also, occupation itself is intermediate, in that one might, for example, decide to become a social worker after having a girl.

The usual way that would be recommended to analyze such data would be: (1) treat the proportion of boy births (that is, number of boys divided by total number of children) as an outcome; (2) *not* control for income, total number of children, or other variables that could be influenced by parental occupation; (3) and use a measure of occupation choice that occurs before any children are born. If the latter measure is not possible, the results will necessarily be more speculative, but steps (1) and (2) can certainly be done and would instill much more confidence in the results.

2. Apparent statistical significance in the context of multiple comparisons

A second, unrelated, statistical error came in the paper on attractiveness and the sex ratio (Kanazawa, in press). Physical attractiveness in the survey used by this paper was measured on a five-point scale, from "very unattractive" to "very attractive". The key result was that 44% of the children of surveyed parents in category 5 ("very attractive") are boys, as compared to 52% of children born to parents from the other four attractiveness categories. With a sample size of about 3000, this difference is statistically significant (2.44 standard errors away from zero).

In interpreting this statement of statistical significance, however, we should consider the arbitrariness of picking out category 5 and comparing it to 1–4. Why not compare 4 and 5 ("attractive" or "very attractive") to 1–3? Given the many comparisons that could be done, it is not such a surprise that one of them is statistically significant at the 5% level.

Perhaps the most natural analysis of these data would be a regression of the proportion of boys on the numerical attractiveness measure. Using the data in Fig. 1 of the paper, the estimated regression coefficient is -1.5 with a standard error of 1.4—thus, not statistically significant. (Weighting by the approximate number of parents in each category does not appreciably change this result).

I have little to say about the difficulties of measuring attractiveness except that, according to the paper, interviewers in the survey seem to have assessed the attractiveness of each participant three times over a period of several years. I would recommend using the average of these three judgments as a combined attractiveness measure. General advice is that if there is an effect, it should show up more clearly if the x-variable is measured more precisely. I do not see a good reason to use just one of the three measures.

One way to summarize the multiple comparisons criticism is to consider the number of possible analyses that could have been considered by Kanazawa in comparing different levels of attractiveness. In addition to the linear regression, there is the comparison of category 1 to categories 2–5, the comparison of 1–2 to 3–5, the comparison of 1–3 to 4–5, and the comparison of 1–4 to 5. Any of these, if statistically significant, could have been chosen to be reported. In addition, with three waves of data, the research could report the results from wave 1, wave 2, wave 3, or the average of all three waves. This comes to $5 \times 4 = 20$ possible comparisons. A simple Bonferroni correction multiplies the significance level (p-value) by the number of potential comparisons, so that to achieve statistical significance at the 5% level, one would need an individual comparison with p-value of $0.05/20 = 0.0025$. In comparison, Kanazawa's reported result is 2.44 standard errors away from zero, which corresponds to a p-value of 0.015 (that is, 1.5%), which would be statistically significant on its own but not as one of 20 possible comparisons. In short, the observed result in this study could easily occur by chance, given the large number of potential comparisons that could be made with these data. (In fact, this p-value is not even statistically significant as one of five comparisons, if we were to ignore the possibility of using data from either of the three waves).

3. Misinterpretation of a logistic regression coefficient

Finally, Kanazawa (in press) includes a mistake in interpreting a logistic regression coefficient. The difference reported in this study was 44% compared to 52%—the most attractive parents in the study had an 8% higher rate of girls. One could also say that the proportion of girls was $0.08/0.52 = 15\%$ higher among the most attractive parents. But the paper reports that "very attractive respondents are about 26% less likely to have a son as the first child". This appears to be based on an incorrect interpretation of a logistic regression of sex of child on an indicator for whether the parent was judged to be very attractive. The logistic regression coefficient was -0.31. Since the probabilities are near 0.5, the correct way to quickly interpret the coefficient is to divide it by 4: $-0.31/4 = -0.08$, thus a difference of 8 percentage points (which is what we saw above). For some reason, Kanazawa exponentiated the coefficient: $\exp(-0.31) = 0.74$, then took $0.74-1 = -0.26$ to get a result of 26%, which cannot be interpreted in the way suggested in the paper. 26% can be interpreted in terms of the odds ratio (i.e., $p/(1-p)$, where p is the probability), but the statement of "26% less likely" is an incorrect summary of the regression (setting aside the multiple comparisons problems discussed in point 2 of this letter). This is particularly unfortunate since 26% was the number reported in the press.

4. Summary

Dr. Kanazawa has looked for some interesting patterns, and it is certainly possible that the effects he is finding are real (in the sense of generalizing to the larger population). But the results could also be reasonably explained by chance and by selection effects. I think a proper reporting of Kanazawa's findings would be that they are interesting, and compatible with his biological theories, but not statistically confirmed.

It is admirable that Dr. Kanazawa pursues these open research questions. However, I think that the data should be analyzed so as to minimize concerns of statistical errors, and that such problems should be clearly identified in the abstract and in the body of the article so that the readership, as well as the popular press, does not over-interpret speculative research.

We thank the referees for helpful suggestions and the National Science Foundation for financial support.

References

Kanazawa, S., 2005. Big and tall parents have more sons: further generalizations of the Trivers–Willard hypothesis. J. Theor. Biol. 233 (4), 583–590.

Kanazawa, S., 2006. Violent men have more sons: further evidence for the generalized of the Trivers–Willard hypothesis. J. Theor. Biol. 239 (4), 450–459.

Kanazawa, S., in press. Beautiful parents have more daughters: a further implication of the generalized Trivers–Willard hypothesis. J. Theor. Biol. Available online 24 July 2006. doi:10.1016/j.jtbi.2006.07.017.

Kanazawa, S., Vandermassen, G., 2005. Engineers have more sons, nurses have more daughters: an evolutionary psychological extension of Baron–Cohen's extreme male brain theory of autism. J. Theor. Biol. 233 (4), 589–599.

Woolridge, J.M., 2001. Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge, MA.

Andrew Gelman
Department of Statistics, Columbia University, New York, NY 10027, USA

Department of Political Science, Columbia University, New York, NY 10027, USA
E-mail address: gelman@stat.columbia.edu