



Mediated Models for the Analysis of Confounded Variables and Self-Selected Samples

Michael H. Birnbaum; Barbara A. Mellers

Journal of Educational Statistics, Vol. 14, No. 2. (Summer, 1989), pp. 146-158.

Stable URL:

<http://links.jstor.org/sici?sici=0362-9791%28198922%2914%3A2%3C146%3AMMFTAO%3E2.0.CO%3B2-J>

Journal of Educational Statistics is currently published by American Educational Research Association.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/aera.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.

Mediated Models for the Analysis of Confounded Variables and Self-Selected Samples

Michael H. Birnbaum

California State University, Fullerton

and

Barbara A. Mellers

University of California, Berkeley

Key words: correlation and causation, dropouts from treatment, mediated models, missing data, self-selected samples, selection bias

This paper addresses the interpretation of data that are contaminated by self-selected samples and/or lack of experimental control. Wainer (1989) reviewed different methods for treating self-selected samples and concluded that the most defensible approach is to model the process that caused the data to be missing. In concert with Wainer, the present paper emphasizes the value of specifying models; however, the theses of the present paper are that (a) any analysis should be interpreted in the context of a set of rival theoretical models, (b) these models should allow latent mediating variables that are imperfectly measured by observed variables, and (c) modeling brings clarity to the conclusion that confounded data can be misleading. Simple models clarify the limitations on conclusions one might otherwise attempt to draw from tainted data.

Wainer (1989) cited Tukey's comment that statisticians are like lawyers (Wainer, 1986). Some lawyers tell you, "Don't do it!" and others—in our opinion—tell you how to get away with murder. Wainer gave cogent arguments against ignoring bias in samples or attempting to "correct" the results for the proportion of missing data by means of covariance adjustments. He went on to describe the "state of the art," for those who hope to find a legal loophole, of methods that might allow one to draw conclusions in the face of missing data. Wainer gave a sympathetic discussion of modeling the process in order to draw inferences from self-selected samples of confounded data. The "bullet-hole approach" (model the process) seems more justified than ignoring the bias or attempting to partial out the participation rate. It makes sense to put armor where returning planes have shown little damage, because we believe the theory that bullets are shot uniformly at planes, and the missing data are the aircraft that were shot

down. This theory is more plausible than the theory that returning aircraft reveal a random sample of where bullets strike planes.

The purpose of this article is to argue that scientists should examine any statistical analysis in the context of a set of rival theories. Our support of modeling is in partial agreement with Wainer's (1989) paper; however, we want to push modeling a step further and emphasize that theories often warn us to side with the type of lawyer who explains why certain actions are improper and should not be done. Simple theories help us understand why better schools might produce lower average Scholastic Aptitude Test (SAT) scores, why a pesticide might be a good idea even if crop yields are equal in treated and untreated fields when the number of eelworms is partialled out, and why—in the absence of bias—men might have higher salaries than women with the same merit and have higher merit than women who receive the same salaries.

In the following examples, we employ mediated causal models in which the observed variables are assumed to be imperfect measures of the latent constructs. Because it is unrealistic to suppose that dropping out of a study would be perfectly correlated with any observed variable, we will also use mediated models to represent selection processes that result in missing data. The models discussed by Wainer postulate functional relationships between observed variables and probability of participation. To represent probability of participation as a correlated consequence of a mediator complicates the analysis, but it also shows why one cannot correct for participation rates by simple regression/covariance adjustments.

Covariance Adjustments: Another Can of Eelworms

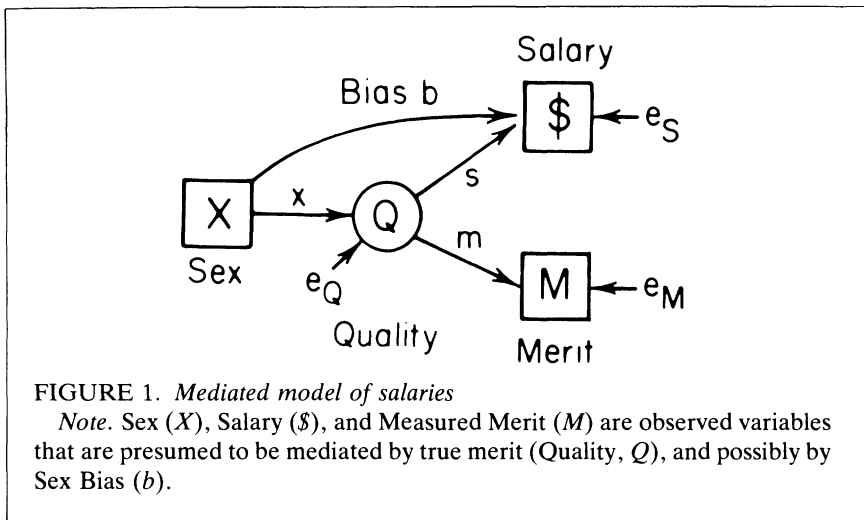
One of the most persistent misconceptions in the use of regression/covariance adjustment is the belief that when there remains a significant relationship between each of two variables and the dependent variable when the other has been partialled out, then there are at least two independent, causal influences on the dependent variable. This single misconception has led to the misinterpretation of results in several fields of research, including subception (Eriksen, 1960), learning without awareness (Dulany, 1968), unconscious affect (Birnbaum, 1981; Birnbaum & Mellers, 1979a, 1979b), and others (Brewer, Campbell, & Crano, 1970). In fact, the null hypothesis (that one latent variable underlies all three observed variables) implies that analysis of covariance will lead to the "discovery" of two partial effects (Birnbaum, 1979b).

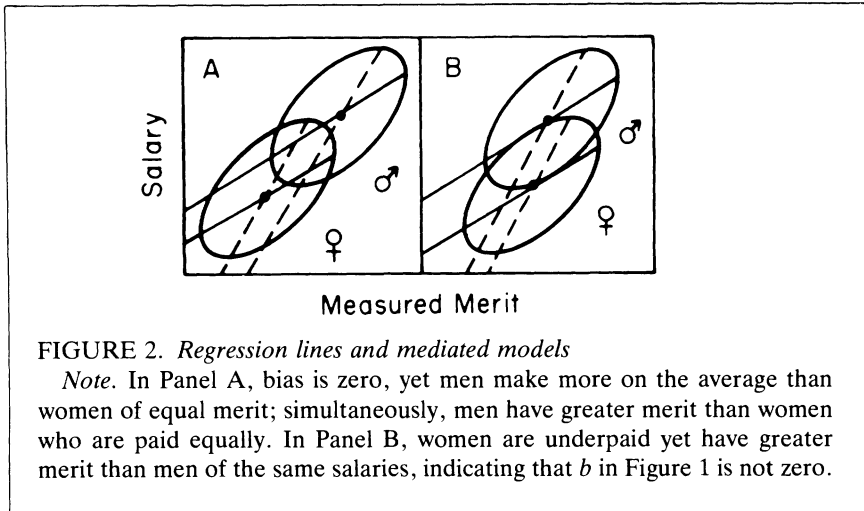
The analysis of sex bias in salaries provides a good illustration of how regression is misused and misinterpreted. It has been found that women earn less on the average than men with the same merit (where merit is a single measure or a composite that includes measures of experience as well as of performance). Because there is a sex difference in salary with merit

partialed out, some investigators have concluded that this difference indicates sex bias. However, for the same sample in which women are paid less than men of equal merit, it can also be the case that women are lower in merit than men with the same salaries. This paradoxical pattern is consistent with a single mediator theory that assumes no sex bias (Birnbaum, 1979a, 1979b, 1981, 1983, 1985; Birnbaum & Hynan, 1986).

A mediated model of salary is illustrated in Figure 1. This model allows that men and women may differ in true merit, denoted by Q , for quality. The correlations between the mediator, Q , and the observed variables, Sex (X), Salary ($\$$), and Measured Merit (M), are x , s , and m , respectively, in the absence of bias (when $b = 0$). The model of Figure 1 explains why standard regression analysis leads to paradoxical conclusions.

Panel A of Figure 2 shows a pattern of data that is consistent with the assumption that there is no sex bias ($b = 0$). When statisticians examine graphs like Figure 2, they conclude that there is no bias; however, regression analysis leads to different conclusions (Birnbaum & Hynan, 1986). The solid lines depict the regression lines predicting salary from merit for each sex. The vertical gap between the solid lines shows that women are paid less on the average than men with the same merit. The dashed lines show the regression of merit on salary. The horizontal gap between two dashed lines shows that men have more merit on the average than women of the same salaries. Neither comparison of regression lines alone tells the whole story. Panel B of Figure 2 shows a situation in which women are underpaid but have higher merit than men of the same salaries, unlike Panel A, and the theory that $b = 0$ can be rejected. As noted by Birnbaum





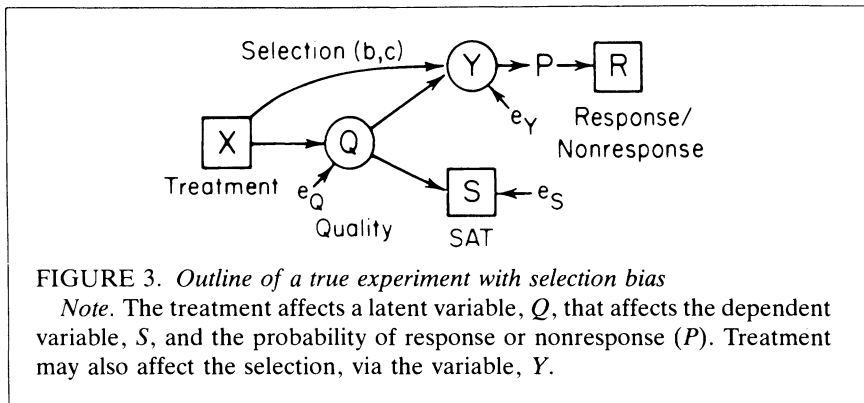
(1985), the higher the correlation between merit and salary within groups, the easier it would be to detect true sex bias in this model. However, regression analysis has the property that the *lower* the correlation between merit and salary, the easier it is to falsely conclude that there is sex bias when $b = 0$, if regression adjustments are used.

Wainer (1989) reviewed the eelworm study as a classic example of misuse of statistical adjustment. The mediated model of eelworms (substituting measured eelworms for measured merit, fumigation for sex, and crop yield for salary) shows why the effect of treatment with eelworms partialled out would not be an estimate of the treatment effect. It also shows, however, that covariance analysis does have an interpretation. If fumigation were to produce an increase in crop yield and yet the treatment effect on crop yield were negative, with eelworms partialled out, the results would indicate that fumigation has harmful side effects. The main point is that the explicit statement of a plausible theory (as in Figure 1) clarifies the meaning of the analysis. In the next section, a mediated model will be developed to illustrate *why* the proportion of missing data cannot be simply partialled out, nor can it be interpreted as a simple indicator of the mediating variable.

Missing Data in True Experiments

Even when treatments are manipulated experimentally, systematic loss of data (dropouts) can invalidate the interpretation of observed treatment effects. Simple models show how the observed effect of a variable can go in the opposite direction of the true causal effect, even when the data are obtained in a proper experiment.

Figure 3 shows an outline of the model that will be used to create missing



data in simulated experiments. The experimental treatment is labeled X . The treatment, X , is presumed to affect a latent variable, Quality, Q , which in turn is correlated with the observed dependent variable, S (SAT). The probability (P) of completing the experiment or not (R) depends on both the latent variable (Q) and the treatment, according to a selection process. For example, X may refer to a program of extra educational expenditures, intended to improve performance on the SAT, S . Students of higher quality (Q) may be more likely to take the SAT (i.e., have a higher value of P). In addition, the treatment itself may have a publicity component that encourages students to take the SAT, thereby increasing the value of P for the treatment group.

Although Figure 3 resembles a path diagram for a linear model, some nonlinear and interactive relationships are built into the model. For example, the treatment may affect *selectivity* as well as the overall likelihood of participating. Suppose an educational intervention included taking alternate forms of the SAT as practice. Such a treatment may not increase the overall probability of taking the SAT, but instead might *increase the relationship between Q and probability of taking the test*. That is, those who perform well or poorly on the practice tests might be encouraged to take or discouraged from taking the SAT, respectively. It will be shown below that such an effect can make neutral or even harmful treatments appear beneficial.

Mediated Model of Missing Data

To illustrate the difficulty of reaching proper conclusions in the presence of missing data, calculations were made using a model outlined in Figure 3. The equations are as follows:

$$Q = aX + e_Q \tag{1}$$

$$S = 200(Q + e_S) + 1000 \tag{2}$$

$$Y = (Q + bX + e_Y)(1 + cX) \quad (3)$$

$$P = 1/(1 + \exp(-Y)) \quad (4)$$

where X represents treatment or no treatment (1 or 0 respectively); a represents the treatment effect on Q ; Q is the mediator, quality; S is the dependent variable of the experiment (SAT). The terms, e_Q , e_S and e_Y are mutually uncorrelated errors that cause the observed variables to be imperfectly correlated.

The parameter b determines whether the treatment group is more or less likely to drop out of the study; the parameter c determines the selectivity—that is, the extent to which probability of participation depends on the mediating variable, Q . Y determines the probability, P , that the subject will participate. Figure 4 shows the relationship between Quality (plus error) and Probability of participation (P) for different Selection models (b and c). Panel A illustrates the effect of parameter b on participation rate for the control group ($a = b = c = 0$) compared with a treatment that is ineffective but encourages participation ($a = c = 0$; $b = .5$). Panel B illustrates the effect of parameter c by comparing the selection models for the control group and an ineffective treatment that increases selectivity ($a = b = 0$; $c = 1.6$).

Twenty-seven different hypothetical experiments were calculated, combining three assumptions concerning the true treatment effect ($a = -.4, 0$,

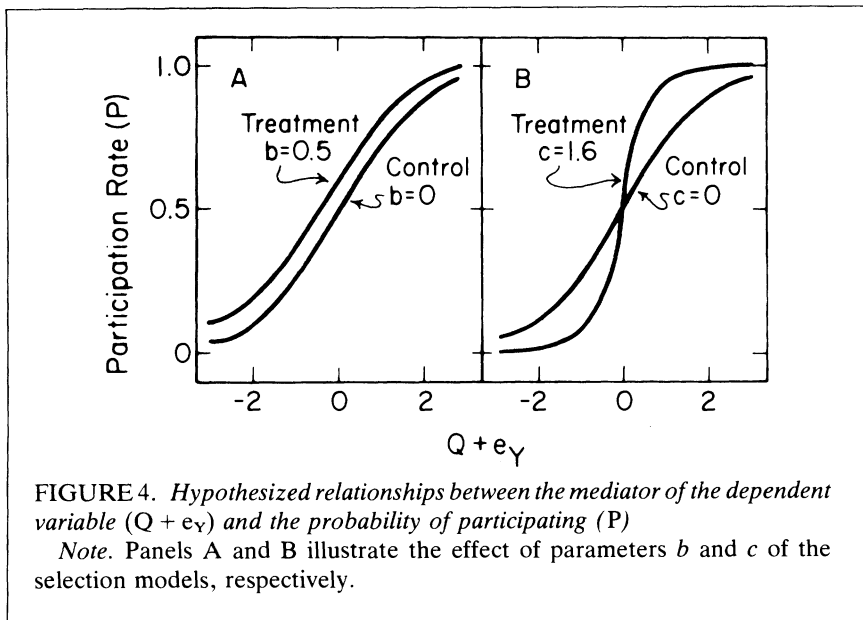


TABLE 1
Expected participation rates for Treatment group

Selection parameters		True effect (200a)		
<i>b</i>	<i>c</i>	-80	0	80
-.5	-.8	46	48	50
-.5	0	36	42	48
-.5	1.6	38	45	49
0	-.8	48	50	52
0	0	44	50	56
0	1.6	46	50	54
.5	-.8	50	52	54
.5	0	51	58	64
.5	1.6	51	55	62

Note. Each entry is the percentage of participation for the Treatment group (rounded to nearest percent). Control group participation rate = 50%.

or .4) with nine different selection models ($b = -.5, 0, \text{ or } .5$; and $c = -.8, 0, \text{ or } 1.6$). Within each "experiment," the following variables took on two levels each: X (1 or 0), e_Q , e_S , and e_Y ; they were factorially combined to create $2 \times 2 \times 2 \times 2 = 16$ basic conditions, with uncorrelated errors in each "experiment." The values of e_Q , e_S , and e_Y were -1.5 or 1.5 ; $-.5$ or $.5$; and $-.5$ or $.5$, respectively.

Results: It's Easy to Be Fooled by Mother Nature

The results are shown in Tables 1 and 2. In Table 1, columns represent values of the true mean difference between Treatment and Control groups (in the entire population), including harmful (-80), ineffective (0), or beneficial (80) treatment effects ($a = -.4, 0, \text{ or } .4$, respectively). Each row shows the participation rate for a different variation of the selection model (determined by parameters, b and c). Table 1 shows that the participation rates for the Treatment group vary from 36% to 64%, depending on the selection model for each condition. The Control group has a participation rate of 50%.

Table 2 shows the expected sample mean difference for each true effect size. Table entries show the expected mean difference in SAT scores for each group. Positive entries indicate a difference favoring the Treatment group. In each column of Table 2, the true effect is the same; however, depending on the selection model (rows), the observed effect can go in the opposite direction of the true effect, or it can go in the same direction as the true effect. The problem for the investigator would be to try to infer the true effect from a single value of the observed effect. For example, a

TABLE 2
Expected Treatment – Control differences

Selection parameters		True effect (200a)		
<i>b</i>	<i>c</i>	–80	0	80
–.5	–.8	–215	–137	–59
–.5	0	–39	25	85
–.5	1.6	33	109	179
0	–.8	–217	–139	–61
0	0	–59	0	57
0	1.6	27	94	144
.5	–.8	–219	–141	–63
.5	0	–86	–30	25
.5	1.6	8	53	78

Note. Each entry is the expected mean difference in the selected sample (Treatment – Control) as a function of the mean difference in the population (True effect), and the selection model (Rows). Positive scores indicate higher means for Treatment group.

treatment that actually *increases* mean SAT by 80 points appears to *decrease* SAT by an average of 59 points, despite having the same participation rate as the Control group (see first row of third column of Tables 1 and 2).

Consider the first column of the tables, where the true effect is harmful (–80 SAT points). In this case, the program that is supposed to help people score higher on the SAT is actually lowering SAT scores by an average of 80 points. If, however, the program includes a selection effect ($c = 1.6$), which might be accomplished by giving feedback on practice tests to the students, then the observed treatment effect will appear to be beneficial, by 8, 27, or 33 points, depending on whether the program encourages or discourages taking the SAT. The participation rates for the treatment group for these three cases varied from 38% to 51%, compared to 50% in the control group. Note that a change in participation rate is not necessary to reach the wrong conclusion. This result also reveals *why* one cannot simply partial out the participation rate: because higher participation rates can co-occur with either higher or lower scores, depending on the true effect and the type of selection bias.

From Data to Theory and Back

The problem can be stated as follows: From a given result, under what conditions can one determine the true state of nature that led to it? Unfortunately, some people believe that data can prove theories true and that

theoretical conclusions can be calculated directly from data. Instead, the connection between theories and data can be stated thus: "If the theory is true *then* the data will have a certain form." Should the data satisfy the predicted form, the theory might or might not be true. Should the data violate the predictions, then something is wrong with either the theory or the data. Because many different true states of nature can generate the same observed effect sizes and because the connection between a set of theories and data is itself a theory, the inference from data to theory is far more difficult than the calculations of predictions from theory.

When You Lose Control, All Hell Breaks Loose: Correlation and Causation Can Be Opposites

The above examples illustrate how, even in controlled experiments, missing data and covariance adjustments can lead to wrong conclusions. However, when the so-called "independent" or "exogenous" variable is merely observed rather than manipulated, the situation becomes even worse. One reason not to infer causation from correlation is that true causation can be of the opposite sign as real-world correlation.

Suppose that, instead of performing an experiment on the benefits of modern medicine, it were decided to conduct a correlational study. Because temporal order is considered important to causal inference, medical treatments in one year might be correlated with the health of the person in the following year. Such a study would reach the following conclusion: People who received medical treatments one year are in *worse* health the next year than people who received no medical treatments at all. On the other hand, when people are randomly assigned to conditions, one might reach the opposite conclusion: If modern medical practice is beneficial, then people who received medical treatment last year would be in better health this year than people from whom the medical treatments were randomly withheld. It is important to realize that such "wrong" conclusions of correlational research would persist for the analysis of any particular treatment or diagnosis.

Thus, correlational and causal relations can be opposite in sign. Perhaps correlations and causal relationships are often opposites. Until causal research is done, it is impossible to know whether a correlational study has found a wrong conclusion. Some authors use the word "spurious" for some correlations, as though correlations can be identified as "spurious" or "genuine." However, all correlations should be regarded as "spurious" when offered as evidence of causation. Nevertheless, both causal and correlational relations of medicine would be of value to an insurance company, but they have different meanings. When evaluating potential clients, insurers should *not* sell life insurance to people who received extensive medical treatments last year—they are poor risks. On the other hand, when

clients *already have policies* and get sick, the insurer should advise them to receive treatment.

Now, suppose there is a true *negative* correlation between educational expenditures and SAT scores. It may be that the causal link is *positive*. Perhaps educational expenditures are like medical ones, and although higher expenditures are causally positively beneficial, they correlate negatively with the performance of the recipients. As Wainer noted, it would be very difficult to get the control necessary to experimentally change educational expenditures, so investigators are tempted to look at correlational surveys as the "best one can do." It would be better, however, for investigators to admit that correlations are not relevant to causal questions and to admit that these difficult questions cannot be addressed until experimentation becomes possible.

As Time Goes By, Things Get Worse

It has long been known that many variables that fluctuate freely over time are found to be highly correlated. For example, economic conditions have been found to correlate with the lengths of women's skirts. When skirt lengths are long, economic indicators are down, but when skirts are short, things are looking up. The recent decline and rise of SAT scores adds time to the other problems of missing data, confounded variables, aggregation of mismatched groups, and covariance adjustments for participation rates. Zajonc (1986) tried to predict the mean SATs for the years 1973 through 1985 from Average Family Size in each year's cohort (which he labeled Birth Order) and Proportion Who Take the SAT in a given year. Zajonc reported a multiple R^2 of .67, which he interpreted as evidence supporting the theory that birth order has a causal influence on SAT scores.

Zajonc's (1986) prediction was not as accurate as that of Nihm (1976). As predicted by the satirical Sue Doe Nihm, mean SAT scores can be perfectly fit by a polynomial of degree 12 for the 13-year period from 1973 to 1985. Even a simple quadratic equation produces a squared correlation of .961, as follows:

$$SAT = a_0 + a_1t + a_2t^2 \quad (5)$$

where $a_0 = 1503130.4$, $a_1 = -1517.503$, $a_2 = .383117$, and t is measured in years (A.D.). A graph of the fit is shown in Figure 5. Because Equation 5 provides a better fit than the theory of Zajonc (1986), using the same number of parameters, the SAT data could be cited in satire as evidence favoring Nihm's law.

In confounded data, there is a tiny correlation between birth order and IQ, with family size partialled out (Zajonc & Markus, 1975). There is a difference of less than one half of an IQ point per birth order position. This correlation may be due to the child's low IQ causing the child to be last,

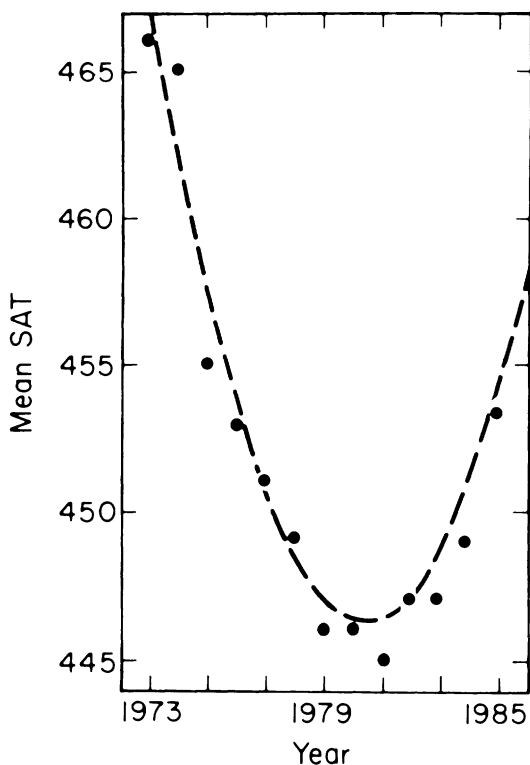


FIGURE 5. Mean SAT scores plotted against years

Note. Dashed line shows fit of quadratic polynomial ($R^2 = .961$). Zajonc (1986) reported an R^2 of only .67 when he attempted to predict the same data from two variables, Average Family Size for the cohorts, and Proportion Taking the SAT.

rather than birth order causing the last child to have a low IQ. Perhaps people are more likely to quit having children when a child is lower in IQ than the other children in the family. Alternatively, a third factor may explain the tiny effect. As parents get older, they are more likely to have retarded children. Therefore, because later children are born to older parents than the first child, and older parents tend to have more retarded children, it is possible that the age of the parents causes the trend. Until experiments are done, birth order data will remain an excellent source of erroneous hypotheses and improper reasoning.

Environmental theories could be tested by an experiment in which people who volunteer to adopt four children are randomly chosen to receive

one, two, three, or four children to adopt at birth. Such an experiment would allow one to ascertain whether biological or adoptive family size makes a difference for IQ or SAT scores. Furthermore, in families of any fixed size (two or four) one could examine whether the first adopted child was higher or lower in IQ than subsequent adopted children. In this way, one could unconfound the variables of adoptive family size, adoptive family order, biological family size, and biological birth order. According to environmental theory, adoptive “birth” order should be correlated with IQ or SAT, even when it is unconfounded from other variables by experiment.

Summary and Conclusions

There is no substitute for properly designed experiments. No fancy statistical analysis can work magic on a badly confounded set of data, and the best message to convey is that many interesting and important questions must remain doubtful until experimenters obtain the funding, control, and authority to carry out the proper research. Sometimes it is best to remember that correlation is the tool of the devil (Birnbaum, 1973, 1974), and that no matter how tempted we may be, we should conclude that we can't reach a conclusion.

In the meantime, we will be seeing many puzzling results on important problems that are difficult to study experimentally. The best we can do under such circumstances is to devise alternative models that permit us to clarify how uncertain any conclusion would be based on such findings. These models make us feel better about concluding that, despite the expense and efforts that were invested in data collection and analysis, such data do not permit definitive answers.

References

- Birnbaum, M. H. (1973). The devil rides again: Correlation as an index of fit. *Psychological Bulletin*, *79*, 239–242.
- Birnbaum, M. H. (1974). Reply to the devil's advocates: Don't confound model testing and measurement. *Psychological Bulletin*, *81*, 854–859.
- Birnbaum, M. H. (1979a). Is there sex bias in salaries of psychologists? *American Psychologist*, *34*, 719–720.
- Birnbaum, M. H. (1979b). Procedures for the detection and correction of salary inequities. In T. R. Pezzullo and B. E. Brittingham (Eds.), *Salary equity: Detecting sex bias in salaries among college and university professors* (pp. 121–144). Lexington, MA: Lexington Books.
- Birnbaum, M. H. (1981). Reply to McLaughlin: Proper path models for theoretical partialling. *American Psychologist*, *36*, 1193–1195.
- Birnbaum, M. H. (1983). Perceived equity of salary policies. *Journal of Applied Psychology*, *68*, 49–59.
- Birnbaum, M. H. (1985). Relationships among models of salary bias. *American Psychologist*, *40*, 862–866.

- Birnbaum, M. H., & Hynan, L. G. (1986). Judgments of salary bias and test bias from statistical evidence. *Organizational Behavior and Human Decision Processes*, 37, 266-278.
- Birnbaum, M. H., & Mellers, B. A. (1979a). Stimulus recognition may mediate exposure effects. *Journal of Personality and Social Psychology*, 37, 391-394.
- Birnbaum, M. H., & Mellers, B. A. (1979b). One-mediator model of exposure effects is still viable. *Journal of Personality and Social Psychology*, 37, 1090-1096.
- Brewer, M. B., Campbell, D. T., & Crano, W. D. (1970). Testing a single-factor model as an alternative to the misuse of partial correlations in hypothesis-testing research. *Sociometry*, 33, 1-11.
- Dulany, D. (1968). Awareness, rules, and propositional control: A confrontation with S-R behavior theory. In T. Dixon & D. Horton (Eds.), *Verbal behavior and general behavior theory* (pp. 340-387). Englewood Cliffs, NJ: Prentice-Hall.
- Eriksen, C. W. (1960). Discrimination and learning without awareness: A methodological survey and evaluation. *Psychological Review*, 67, 279-300.
- Nihm, S. D. (1976). Polynomial law of sensation. *American Psychologist*, 31, 808-809.
- Wainer, H. (Ed.) (1986). *Drawing inferences from self-selected samples*. New York: Springer.
- Wainer, H. (1989). Eelworms, bullet holes, and Geraldine Ferraro: Some problems with statistical adjustment and some solutions. *Journal of Educational Statistics*, 14, 121-140.
- Zajonc, R. B. (1986). The decline and rise of Scholastic Aptitude scores. *American Psychologist*, 41, 862-867.
- Zajonc, R. B., & Markus, G. B. (1975). Birth order and intellectual development. *Psychological Review*, 82, 74-88.

Authors

- MICHAEL H. BIRNBAUM, Professor, Dept. of Psychology, California State University, Fullerton, CA 92634. *Specializations*: psychophysics, judgment and decision making, salary equity, source credibility, methodology.
- BARBARA A. MELLERS, Associate Professor, Dept. of Psychology, University of California, Berkeley, CA 94720. *Specializations*: mathematical psychology, human judgment, decision making.

LINKED CITATIONS

- Page 1 of 1 -



You have printed the following article:

Mediated Models for the Analysis of Confounded Variables and Self-Selected Samples

Michael H. Birnbaum; Barbara A. Mellers

Journal of Educational Statistics, Vol. 14, No. 2. (Summer, 1989), pp. 146-158.

Stable URL:

<http://links.jstor.org/sici?sici=0362-9791%28198922%2914%3A2%3C146%3AMMFTAO%3E2.0.CO%3B2-J>

This article references the following linked citations. If you are trying to access articles from an off-campus location, you may be required to first logon via your library web site to access JSTOR. Please visit your library's website or contact a librarian to learn about options for remote access to JSTOR.

References

Testing a Single-Factor Model as an Alternative to the Misuse of Partial Correlations in Hypothesis-Testing Research

Marilynn B. Brewer; Donald T. Campbell; William D. Crano

Sociometry, Vol. 33, No. 1. (Mar., 1970), pp. 1-11.

Stable URL:

<http://links.jstor.org/sici?sici=0038-0431%28197003%2933%3A1%3C1%3ATASMAA%3E2.0.CO%3B2-X>

Eelworms, Bullet Holes, and Geraldine Ferraro: Some Problems with Statistical Adjustment and Some Solutions

Howard Wainer

Journal of Educational Statistics, Vol. 14, No. 2. (Summer, 1989), pp. 121-140.

Stable URL:

<http://links.jstor.org/sici?sici=0362-9791%28198922%2914%3A2%3C121%3AEBHAGF%3E2.0.CO%3B2-%23>