

## Identifying Experimental Units and Calculating Experimental Error

The evidence for conclusions about environmental effects on average body weights and measures of behavior in developing rats was assessed by an analysis of variance and associated *F*-tests in a report by Pearson *et al.* (1). In the experiment 12 litters were used. The *F*-tests indicated 106 degrees of freedom for experimental error. The tests almost surely ignored correlations among littermates and, consequently, underestimated experimental error and exaggerated the strength of evidence for inferences made in the study.

The experimental units were not identified correctly. Questions about the effects of litter composition, for example, require an experimental error derived from differences among litters of like composition. The litter is the experimental unit. There were only 12 litters in the study; there could not have been more than 9 degrees of freedom for an error used in testing the litter composition effects described.

D. F. Cox

Iowa State University,  
Ames 50011

### References and Notes

1. D. E. Pearson, M. H. Teicher, B. A. Shaywitz, D. J. Cohen, J. G. Young, G. M. Anderson, *Science* **209**, 715 (1980).

9 September 1980; revised 11 May 1981

Concern over littermate correlations was most clearly brought to the attention of researchers in our field by Abby and Howard (1). They found, using body weight as their sole parameter, that within-litter scores were tightly clustered and between-litter variability was comparatively great. Since individual weights in any one litter clustered within a restricted portion of the total range of weights, and differences between litters contributed largely to total variances, they concluded that *F* and degrees of freedom values would be overestimated in analysis of variance designs when individual subjects' scores were analyzed. Hence they recommended that litter means be used instead as the experimental units.

Unfortunately, there is a tendency to generalize these conclusions to all behavioral measures. The tight clustering of body weight within each litter is a unique consequence of the suckling situ-

ation (2), and comparable clustering is not observed in our behavioral measures. In activity and learning performance we have found that within-litter variability is comparatively great and between-litter differences are small and usually statistically insignificant. Furthermore, individual scores within each litter appear to be distributed randomly over the entire range of scores. Thus Abby and Howard's conclusions do not apply, and the individual subject's score remains the appropriate unit of analysis.

This situation is most readily appreciated by examining the weakest statistically significant effect we reported: the difference between T-Hom (treated homogeneous) and T-Het (treated heterogeneous) pups in the shuttle box [figure 3 in (3)]. Both groups of pups performed very poorly in this task, but the T-Het pups had somewhat steeper learning curves, and this was reflected in a group  $\times$  trial block interaction [ $F(4, 200) = 2.80, P = .0267$ ]. Since individual scores were well distributed across litter groupings (4), there were no significant effects of grouping or any group  $\times$  trial interactions (5). Analyzing these data in terms of litter group means greatly compressed the range of scores through regression to a common mean and diminished standard errors; and although the degrees of freedom in the denominator were reduced by 80 percent, the significance of the phenomenon was exaggerated by eliminating this individual variability [ $F(4, 40) = 3.44, P = .0164$ ].

In like fashion, analyzing all our data in terms of litter group means preserves the statistical significance of the behavioral effects we reported. Since there were no significant effects of litter grouping and no litter-group interactions, we felt justified in pooling litter groups and using individual subject scores.

The situation admittedly was quite different for body weight (6). Furthermore, whereas the correlation between individual scores in the shuttle box and litter group means across trial blocks was only .43, correlations between individual scores and litter group means across age averaged .91 for treated pups and .97 for controls when body weight was analyzed. Thus, while we did not exaggerate

the strength of our conclusions regarding activity and shuttle box performance, we did exaggerate body weight differences. However, since the main point we wished to make about body weight was the lack of difference between T-Hom and V-Hom (vehicle homogeneous) pups, we considered an approach that biased our results in the opposite direction acceptable. When analyzed in terms of litter group means, the overall difference in body weight between V-Het (vehicle heterogeneous) and T-Het pups was highly significant [ $F(1, 10) = 14.82, P < .005$ ], while the difference between V-Hom and T-Hom pups was not [ $F(1, 10) = 1.72$ , not significant].

The reported activity, body weight, and learning curves for T-Het and V-Het pups have been replicated by us many times in over 100 litters. Our activity and body weight results for T-Hom and V-Hom pups closely parallel data recently published by Erinhoff *et al.* (7), which we have now replicated in 12 additional litters.

MARTIN H. TEICHER\*

DAVID E. PEARSON†

BENNETT A. SHAYWITZ

DONALD J. COHEN

Laboratory of Developmental  
Neurobiology,  
Section of Pediatric Neurology,  
Yale University School of Medicine,  
New Haven, Connecticut 06510

### References and Notes

1. H. Abby and E. Howard, *Dev. Psychobiol.* **6**, 329 (1973).
  2. E. M. Blass and M. H. Teicher, *Science* **210**, 15 (1980); W. G. Hall and J. S. Rosenblatt, *J. Comp. Physiol. Psychol.* **91**, 1232 (1977); M. Leon, P. G. Crockery, G. K. Smith, *Physiol. Behav.* **21**, 793 (1978).
  3. D. E. Pearson, M. H. Teicher, B. A. Shaywitz, D. J. Cohen, J. G. Young, G. M. Anderson, *Science* **209**, 715 (1980).
  4. In our calculation of litter group means we obtained two data points per litter. For heterogeneous litters we calculated one mean for treated pups and one mean for controls. In the homogeneous litters we calculated separate means for males and females in order to create equal and acceptable cell sizes of six.
  5. No statistically significant effects of grouping were noted nor any group  $\times$  trial interactions.
  6. For body weight, statistically significant litter group and group  $\times$  age interactions were observed: T-Hom,  $F(5, 20) = 1.95$ , N.S., and  $F(25, 100) = 3.71$ ,  $P < .001$ ; T-Het,  $F(5, 20) = 3.70$ ,  $P < .025$ , and  $F(25, 100) = 5.78$ ,  $P < .001$ ; V-Hom,  $F(5, 23) = 3.83$ ,  $P < .025$ , and  $F(25, 115) = 3.51$ ,  $P < .001$ ; V-Het,  $F(5, 23) = 6.29$ ,  $P < .001$ , and  $F(25, 115) = 5.29$ ,  $P < .001$ , for group and group  $\times$  age interactions, respectively.
  7. L. Erinhoff, R. C. MacPhail, A. Heller, L. S. Seiden, *Brain Res.* **164**, 195 (1979).
- \* Present address: Department of Psychiatry, Harvard Medical School, McLean Hospital, Belmont, Mass. 02178.  
† Present address: Department of Sociology, Yale University, New Haven, Conn. 06520.

3 June 1981