APPLICATION OF W. HENDRICKS' METHOD OF
ADJUSTING FOR BIAS IN MAIL SURVEYS *

By D. S. Robson BU-23-M August, 1951

Method:

1) Questionnaires are mailed to a sample of k individuals, and n_1 persons respond to this mailing

2) Questionnaires are again mailed to the nonrespondents of the first mailing and n_2 persons respond to this second mailing

.....

i) Questionnaires are again mailed to the nonrespondents of the i-1 mailing and n_i persons respond to this i'th mailing.

.....

It is then argued that the number of mailings required to extract a response from an individual is a measure of that person's resistance to the questionnaire. Denote this measure by X, i.e.,

X = 1 for those individuals who respond to first mailing

.....

X = i ------------------"----------------- i'th --"---

.....

Let $\mu_x = E(X)$ and assume that $\ln \frac{X}{\mu}$ is $N(0,\sigma)$

Let $D = \frac{\ln X - \ln \mu_x}{\sigma}$, then D is N(0,1) and

$$\ln X = \ln \mu_x + \sigma D \qquad (1)$$

---

*See "Agricultural Economics Research" Vol. 1, No. 2, April 1949, pp. 52-56.

Note that $D = D(X)$, hence $Pr[\underline{X} \leq 1] = Pr[\underline{D} \leq D(1)]$, and the sample produces the estimate $Pr[\underline{X} \leq 1] \simeq \frac{n_1}{k} = Pr[\underline{D} \leq \hat{D}(1)]$ and,

similarly, $$Pr[1 < \underline{X} < 2] \simeq \frac{n_2}{k} = Pr[\hat{D}(1) < \underline{D} < \hat{D}(2)]$$, or

$$Pr[\underline{X} < 2] \simeq \frac{n_1+n_2}{k} = Pr[\underline{D} < \hat{D}(2)]$$, etc.

The estimate $\hat{D}(X)$ of $D(X)$ is thus computed from a table of normal deviates.
Let Y be the characteristic measured by the mail questionnaire, and let Y_i be the average response corresponding to the resistance i.
Suppose that the relationship of X and Y is given by (3 sample mailings)

(2) $$Y = a + bX + cX^2$$

then an estimate of μ_y, the true mean of Y, is given by

$$\hat{\mu}_y = \hat{a} + \hat{b}\hat{\mu}_x + \hat{c}\hat{\mu}_x^2$$

when $\hat{a}$, $\hat{b}$, $\hat{c}$ are the least squares solutions to the system (2) generated by the sample and $\hat{\mu}_x$ is the least squares solution to the system

(3) $$\ln X = \ln\mu_x + \sigma\hat{D}(X)$$

APPLICATION

I. The 1943 Curtis Impact Study offers an example of a combined mail questionnaire -- personal interview survey. A probability sample of 33,000 households was enumerated by the personal interview technique, and an additional questionnaire was left at those households which had in possession a recent copy of at least one of four specified magazines. The additional questionnaire was to be completed at leisure and returned by mail. The nonrespondents to this personally delivered mail questionnaire were sent an additional schedule along with a letter requesting cooperation. A third questionnaire was later sent to the nonrespondents of the previous mailing.

A particularly nice feature of this example is that certain control information is available on all recipients of the mail questionnaire. The personal interviews yielded, for example, the following data on size of family characteristics of all households receiving the mail schedule:

Family Size	Number of Families
1	242
2	1942
3	2053
4	1758
5	866
6	432
7	165
8 or more	133
	7591

av. family size = 3.4854 (using 8 or more = 8)

The mailed returns were:

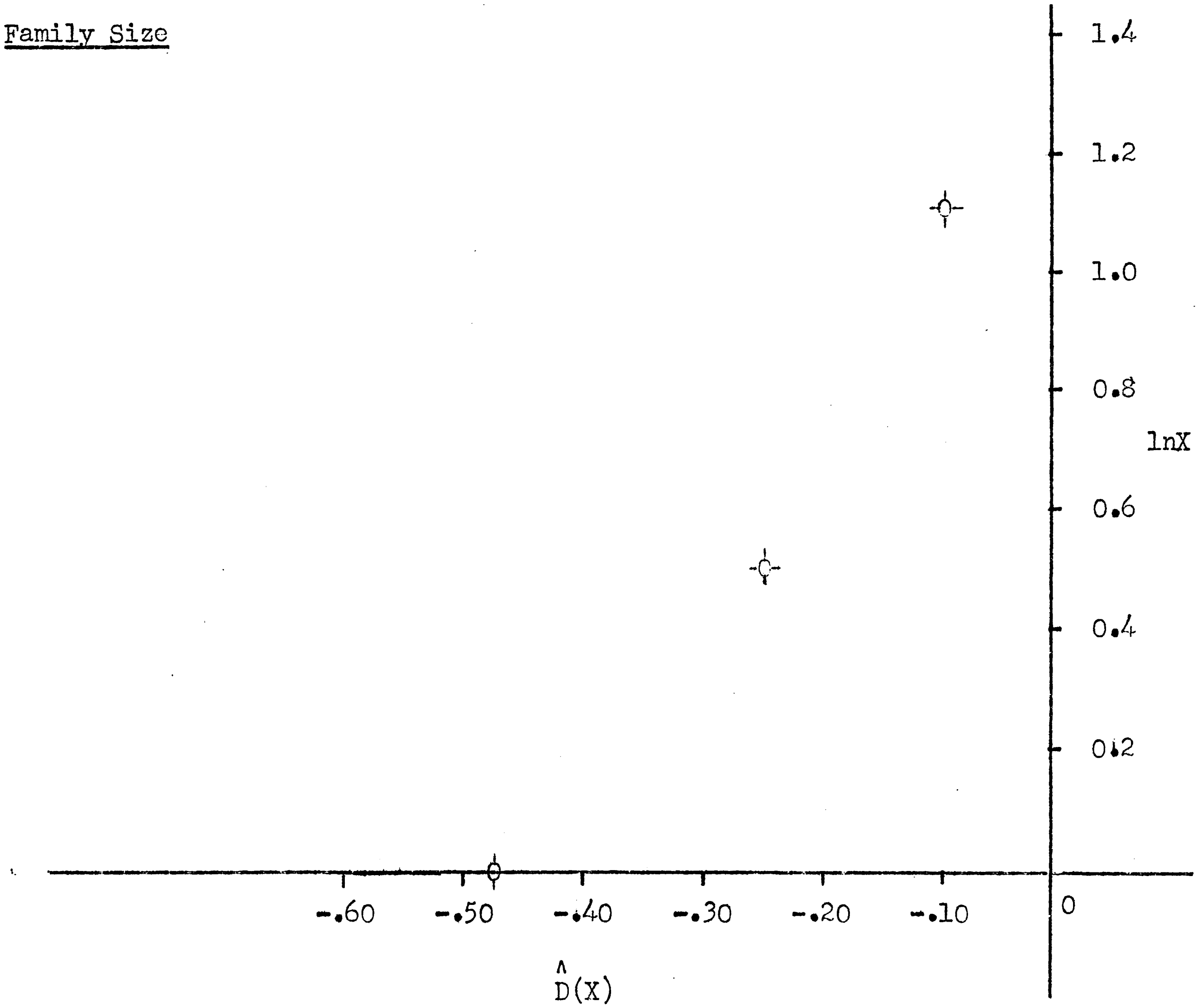
X Mailing	Number of Returns $=n_i$	Average Size of Family $=Y_i$	lnX	$\frac{n_i}{7591}$	$\hat{D}(i)$
1	2386	3.7108	0	.314	-.48
2	678	2.7271	.693	.089	-.24
3	437	3.5675	1.099	.058	-.10
	3501				

and the Hendricks solution is:

$$\hat{\sigma} = 2.8904 , \quad \ln\hat{\mu}_x = 1.387 , \quad \hat{\mu}_x = 4.00$$
$$\hat{a} = 6.5183 , \quad \hat{b} = -3.7196 , \quad \hat{c} = 0.9120$$
$$\hat{\mu}_y = 6.2319 = \text{estimated average family size}$$

Figure 1 suggests that the assumption of normality is valid ($r = \frac{.2136}{.2136}$ to 4 digits)

FIGURE 1



II. Standard of Living in the 1948 Curtis Impact Study Mail Survey.

Characteristics of total mail sample, obtained from personal interview:

Standard of Living	Number of Households
1	708
2	2619
3	3629
4	506
	7462
No Answer	129
	7591

Average standard of living = 2.527

Mailed returns:

X Mailing	n_i Number of returns	Y_i Ave. Standard of Living	lnX	$\frac{n_i}{7462}$	$\hat{D}(i)$
1	2343	2.440	0	.3140	-.485
2	664	2.596	.693	.0890	-.245
3	435	2.510	1.099	.0583	-.097

and the Hendricks solution is:

$$\hat{\sigma} = 2.837 \qquad \ln\hat{\mu}_x = 1.3803 \qquad \hat{\mu}_x = 3.98$$

$$\hat{a} = 2.6764 \qquad \hat{b} = -.0521 \qquad \hat{c} = -.0122$$

$$\hat{\mu}_y = 2.2757$$

Figure 1 serves equally well for this case ($r = \frac{.2176}{.2176}$)

III. 1950 Subscriber Familes Study.

In 1950 the National Analysts, Inc., selected a sample of 4046 subscribers of a particular magazine by choosing every k'th name from the subscribers list. A questionnaire and letter from the editor were mailed to the sample of subscribers. Later, a second wave of schedules were mailed to the nonrespondents. Finally a third wave of schedules and letters of appeal were sent to nonrespondents; in addition, a specially prepared booklet was mailed as an added incentive. The nonrespondents to the three waves of mailed questionnaires were then sampled by the personal interview technique.

The mailed returns for the character annual income were:

X	n_i	Y_i	$\frac{n_i}{4046}$	$\hat{D}$
1	987	68.6373	.2439	-.695
2	606	69.4637	.1498	-.270
3	357	72.5630	.0882	-.050
	1950			

and the Hendricks' solution is:

$\hat{\sigma} = 1.6940 \quad \ln\hat{\mu}_x = 1.1704 \quad \hat{\mu}_x = 3.22 \quad r = .9995$

$\hat{a} = 70.0833 \quad \hat{b} = -2.5826 \quad \hat{c} = 1.1364$

$$\hat{\mu}_y = 73.5499$$

The additional returns by personal interview were:

X	n_i	Y_i	$\frac{n_i}{4046}$
4 (?)	856	68.4579	.2116

A logical difficulty would be involved in combining the personal interview with the mailed questionnaire data in forming Hendricks' estimate. It is, in any case, apparent that the estimate $\hat{\mu}_y = 73.5499$ must be regarded with suspicion in light of the results obtained by personal interview.

DISCUSSION

The most remarkable feature of the preceding examples - and also of the examples Hendricks used in his paper - is the almost perfect linear relationship between log wave number and estimated normal deviate. Such a situation is not likely to arise by chance; on the other hand, it is difficult to ascribe a logical reason to its occurrence. The variable X = wave number is here equally spaced, but when X is regarded as a measure of resistance such simplicity appears questionable. The time of mailing for the 1950 example was:

Wave	Time of mailing
1	0
2	25
3	70

which is fairly discrepant from a sequence of equally spaced numbers. Also, one would expect incentive mailings to confound the effect of time of mailing and further complicate the construction of a scale.