UW Biostatistics Working Paper Series

8-22-1980

Sample Size Calculations and Optimal Followup Time in Health Services Research Using Utilization Rates

Paula Diehr
University of Washington, pdiehr@u.washington.edu

INTRODUCTION

One of the most frequent questions asked of a consulting statistician is "how many subjects are needed?" This question usually arises in designing an experiment where there is a cost for the collection of an additional observation, and it is desirable to collect "enough" but not "too many" subjects in order to conserve resources. Even if there is no meaningful cost associated with the collection or analysis of an additional case, the exercise of calculating a sample size is often educational. It requires the investigator to define exactly what the major questions of the analysis are, and how they will be operationalized. The calculation of a sample size helps to put things into perspective, since often the result is an impossibly large number. This might suggest that the study be abandoned, that additional funds be sought for the study, or that the level of expectation be reduced.

Often, health services research does not involve experiments or even the collection of data, but is "secondary analysis" of data which are already collected. In this case, the cost of collecting an extra case is not at issue; however, the cost of entering that extra case onto a computer might be fairly large, and the cost of analyzing a large data set may be substantially larger than the cost for a small set. Even if the investigator decides later to use a larger sample the preliminary analysis, which is where most of the mistakes are made, should routinely be performed on a small subset of the data. An additional advantage of this approach is that, if unexpected findings arise, some of the reserved data may be used to test hypotheses formed from the first

set of data.

A related problem in health services research is, how long to follow subjects. Is studying 100 people for 2 months equivalent to studying 50 people for four months, or 2 people for 100 months?

We consider two cases in the calculation of sample size: what to do when you know some parameters of the distribution of the data, and what to do when these are unknown.

HOW TO CALCULATE SAMPLE SIZES IF YOU KNOW SOMETHING ABOUT THE DATA: SOME STANDARD FORMULAS

Suppose one wishes to compare the means of two groups. Often some information or experience is available about the size of the means (m1 and m2) and standard deviations (s1 and s2) of the data. Then, the investigator chooses values for d (a difference between the two means which it is important to detect); b (the chance of failing to detect a difference as large as d when it exists); and a (the chance of declaring the two means different on the basis of observed data when in fact m1=m2). Let $z_{(1-a)}$ be the 100 (1-a) percentile of the normal distribution. Then, assuming that the statistic has a normal distribution, the optimal sample size for each group is

$$N = z^2 (s_1^2 + s_2^2) / d^2 \qquad (F1)$$

where

$$z = z_{(1-a)} + z_{(1-b)}$$

This formula can be applied to proportions, letting m1=p1 and s1=(p1 q1)**.5. If the problem can be specified as a difference

in proportions, there are useful tables in many statistics texts. The tables by Gail (5) and Feigl (4) are especially helpful.

Formula F1 assumes that an equal number of patients are required in both groups. This is not always the case. For example, if one group is more variable than the other, it would be preferable to choose a larger number from that group. (This might also suggest a transformation of the data to make the variability more equal.) If the variance of group 2 is s^2 and of group 1 is $K s^2$, then letting

$$N2 = [\, Z^2 \; s^2 \; / d^2 \,] \quad (K^{.5} + 1) \qquad \text{(F2)}$$

and

$$N1 = (K)^{.5} \; N2$$

gives the smallest total sample size which will meet the investigator's specifications. (Note that if K=1 -- the two variances are equal -- this reduces to the previous formula.)

Sometimes, as in a case-control study, one group of data has already been collected, and only the second sample size is of concern. In this case, the formula is

$$N2 = \frac{K \; N}{\frac{N \; d^2}{z^2 \; s^2}} \; -1 \qquad \text{(F3)}$$

where sample 1 is fixed, with mean = m, variance = s^2, and sample size = N; sample 2 has mean = m + d, variance = Ks^2, and sample size = N2. (Note, a negative value for N2 means that the specified difference, d, can not be detected even with an infinite number of observations in group 2, because there is too much variability in the estimate of the mean for group 1. This may suggest

obtaining more cases for group 1).

If there are more than two groups, the charts in Dixon and Massey (3) may be used. These require stating the alternative hypothesis in the form of the means of each group of interest, which may be difficult.

It might be wiser to formulate the sample size problem as an estimation problem rather than as a testing problem; i.e., how big a sample is needed to estimate this parameter to within plus or minus d with 100 (1-a)% confidence? In this case,

$$N = z^{2}_{(1-a/2)} s^{2} / d^{2} \qquad \text{(F4)}$$

where the confidence interval is (m-d, m+d). If a = .05, then z = 1.96, and we have a 95% confidence interval for the difference in the means.

WHEN YOU KNOW NOTHING ABOUT YOUR DATA

Frequently an investigator is unable to estimate the range of data expected. One possibility is to turn it into a problem involving proportions. The investigator may have little feeling for (say) the median score on a Likert scale, but might well be able to answer "about what percent will be satisfied?" or "approximately what percent will have a visit that year?" If the major question of interest can be phrased in this way, the tables mentioned above can be used. A pilot study is always a good idea to provide initial estimates of the means and the variability. It may also suggest changes in the problem formulation, methods and data instruments as well as help with choosing N.

There are often published sources for data. However, these frequently do not provide measures of variability. If it was possible to estimate the standard deviation from the mean, the investigator would be able to proceed. Here we present some insights gained from our analysis of the Seattle Prepaid Health Care Project data (7). In this study, about 8000 low income people were given free health care and their utilization was monitored for up to 48 months. We will examine the variables: visits, hospital admissions, labs and x-rays. An investigator proposing to study these measures may find some of this information helpful.

The data used here are the means and standard deviations of the various measures, calculated for 14 categories: 2 sexes by 7 age groups. This is referred to as the "uniform age/sex" data set. Only people enrolled for more than one year were included here (n = 6698). For visits and admissions, people were also subdivided by the number of months they had been followed (1 to 48 months) with the 1-month exposure group and those subgroups with fewer than 100 people excluded. This left 32 groups for analysis, which constitute the "uniform time" data. We have studied the relationship between the mean and standard deviation across these subgroups.

Other analyses have considered visits and other measures to be counts, and thus Poisson distributed. If this is so, then the mean and variance are equal. We will see that such an assumption clearly underestimates the variance of the rates and thus provides nonconservative estimates of the sample size. We examine empirically (1) the relationship of the mean to the standard deviation, (2) how the standard deviation varies as a function of follow-up time, and (3) the effect of transformations on the data. The

optimal follow-up time is also considered.

VISITS

Nationwide the mean number of physician visits per person is about 5, increasing with age and higher for females (8). As mentioned above, data on variability are not easy to find.

A. RELATION OF MEAN TO STANDARD DEVIATION

The ordinary least squares equation developed from the uniform age/sex data is

$$SD\ (VISAN) = 1.64(mean)^{.75} \qquad (r=.92) \quad (F5)$$

For the uniform time data, the relationship is approximately

$$SD(VISAN) = 1.15(mean)^{1.\emptyset4} \qquad (r=.78) \quad (F6)$$

which seems rather dissimilar. Figure 1 plots equations F5 and F6 with the uniform age/sex data. This figure shows that these two equations are very similar in the range of the data. Figure 2 shows similar results for the constant time data. The largest dots on the plot are for people enrolled less than one year ($t<12$), which are apparently the cause of the discrepancy between the two lines. Since the lines are very similar it appears that the length of the study period is not very important, at least in our range of 1 to 48 months. In the range VISAN = 3 to 1Ø, the approximation

$$SD(VISAN) = mean\ (VISAN)$$

is seen to be fairly good. Assuming that the coefficient of variation is 1.5 is also a good approximation.

The plot of mean = variance is not near the data. Thus, the annualized visit rates do not have a Poisson distribution but are closer to log-normal. Kilpatrick (6) showed that a negative binomial distribution provided a good fit to some visit data. This distribution has a larger standard deviation than mean. The coefficients of variation in the 63 practices that he examined were all in the range of 1.4 to 1.5. Our data set also included a self-reported visit rate for the previous month. The coefficient of variation was larger for these data, on the order of 2.5. This may be due to the short period of time, however. A study in Boston (1) of people over 65 years of agegave a coefficient of variation of 1.1 Thus, assuming that the standard deviation is about equal to the mean seems to be a reasonable approximation.

B. CHANGES IN STANDARD DEVIATION AS A FUNCTION OF FOLLOWUP TIME

The annualized visit rate as a function of t (number of months the person was followed) is estimated as:

$$SD(VISAN) = 10\ t^{-.204} \qquad (r=-.68) \quad (F7)$$

Thus, variability decreases as the follow-up period increases, as would be expected. Figure 3 plots equation F7 along with the data points to exhibit the fit. We would expect the sd to decrease with t**-.5, if each person month provided independent information about a person's visit rate. The observed decrease is considerably slower. If VISAN were estimated from 1000 people each observed for one month the standard error of the estimate would be

0.32; for 500 people observed for 2 months it would be 0.39; and, for 20 people observed 50 months it would be 1.01. Although all of these examples provide 1000 person-months for analysis, the estimate from the smaller number of people followed for a longer time is substantially less accurate than the two estimates which follow more people for less time. However, it may be much less expensive to follow a few people for a long time than vice-versa. The relationship in F7 can be used to optimize the design with respect to cost. Let

$$s = a\, t^{b}$$

If C is the cost of an additional person and D is the cost of following a person for one month, then the total cost of following N people for t months is

$$\text{cost} = CN + DNt,$$

which is minimized when

$$t = -2bC / [(2b + 1)D] \qquad \text{(F8)}$$

In equation F7 above, b = -.20, giving us

$$N = .67\ C/D.$$

Thus, if C = $100 per person and D= $10 per month, it would be optimal to observe each person for .67 (100) / 10 = 6.7 months. (For reasons of seasonality, it might still be best to take twelve months.) The standard error of the estimated rate is, from equation F7, 6.76/N**.5. The optimal value of N can be calculated from the conditions set on the standard error in the sample size calculation, shown in equation F1. Using the mean and time in the same equation, we have

$$SD(VISAN) = 2.38(t)^{-.13}(mean)^{.81} \quad (R=.88)$$

The power of t is still far from -.5; thus, making the subjects more uniform does not alter these findings.

C. TRANSFORMATIONS OF THE DATA

The formulas have been applied to VISAN, even though this measure will usually be highly skewed. The square root might be a better measure, since it tends to be more normally distributed for variables with long right tails, providing more efficient estimates. If we work with SVISAN (= square root of VISAN), then

$$SD(SVISAN) = 1.42(mean)^{-.258} \quad (r=-.17) \quad (F9)$$

for constant time, and

$$SD(SVISAN) = .94 \ (mean)^{.473} \quad (r=.872)$$

for constant age/sex, and finally

$$SD(SVISAN) = .81 \ (mean)^{.63} \quad (r=.87)$$

for constant age/sex but allowing only people with more than 12 months exposure. Against time,

$$SD(SVISAN) = 2.19 \ (time)^{-.205} \quad (r=-.93).$$

The mean and s.d. are fairly unrelated for the constant time data (equation F9), which is one characteristic of a normal distribution.

If the hypotheses and results could be specified in square root units, an improved analysis might be achieved. Estimating the mean m to plus or minus .1 visit is roughly equivalent to estimating root m to plus or minus 0.0250. Using equation F4 above gives N1 = 4 (s)**2 /0.1**2 using the original units com-

pared to N2 = 4(s)**2 /(.025)**2 required if the square roots are used. The sd of VISAN is 5.3, and the SD of SVISAN is 1.15. Thus, N1 = 11,236 and N2 = 8,464. The ratio of these two sample sizes is 1.33, indicating a smaller sample size required for the transformed data. This ratio holds for other reasonable values of d. However, since the mean square root is substantially lower than the square root of the mean, this may not be acceptable if answers are needed in the original units. These small savings may not be worth the problems caused by working in other units. In the example above, costs are again minimized at about 7 months observation per person if the square root is used for analysis.

ADMISSIONS

Nationwide, the rate of hospital admissions was 163 per thousand population per year in 1975 (8). Hospital admission data are usually presented as the number of admissions divided by the total number of person-years involved. The mean number of admissions per person is thus easy to estimate from published sources, but the variability cannot usually be determined.

A. RELATIONSHIP OF MEAN AND STANDARD DEVIATION

In our data, admission rates have extremely high variability between persons, which is reasonable since in a healthy group very few have any admissions, but those with one admission are likely to have more than one admission (2).

Letting NINAN be the number of admissions per thousand years for an individual = (number of admissions/time)x12000, we can estimate the relationship between the standard deviation and mean of NINAN for the individual. The best equation (calculated from 14 uniform age/sex groupings) relating an individual's standard deviation to the mean is

$$SD(NINAN) = 21.4(mean)^{.58} \qquad (r=.97) \qquad (F10)$$

Note that the exponent is near to .5 Thus, for similar people, the variance is approximately <u>proportional</u> to the mean but far from equal to it. The coefficient of variation is 2.5 at the rate of 163 admissions per thousand. The uniform time data yield the equation:

$$SD(NINAN) = .67\ (mean)^{1.30} \qquad (r=.81) \qquad (F11)$$

which looks quite different. The coefficient of variation is about 3 at 163 admissions per thousand.

Figure 4 plots equations F10 and F11 together with the means of the data which are uniform in age and sex. The two lines cross at the mean of the data, but otherwise have quite different slopes. The variability increases at a faster rate for the uniform time data than for the uniform people data. The line SD = mean is also shown. Neither line is close to the data, but the former is closer. Figure 5 shows these lines plotted against the constant time data. The two slopes are clearly different, and well above the "sd = mean" line.

Grouping our data in different ways has shown that the coefficient of variation is usually between 2 and 3 for up to 48 months. Our data included a self-reported admission rate for the previous year; the coefficient of variation for this rate was quite high, on the order of 4. This discrepancy may be due to the relatively short time period considered. For senior citizens in Boston (1) the coefficient of variation for the question "Did you have any admissions?" was 1.9. Thus, estimating that the standard deviation of the rate to be about three times the mean appears to be appropriate.

The distribution of NINAN might be handled theoretically if we could consider the number of admissions for a person followed t months to be the sum of t independent binary random variables. A person studied for "t" months who had "A" admissions, would have an estimated probability of admission (per month) of p=A/t. Here, A is assumed to be approximately Poisson distributed. This means that NINAN, which is a constant multiple of A, will not be Poisson. Using the usual binomial rules, the variance of p would be p(1-p)/t. The estimate of NINAN for the individual would be 12000p, and the variance estimate would be 12000**2 (p)(1-p)/t. We can use this theoretical variance to describe the relationship between the mean and standard deviation of NINAN, controlling for t. This equation becomes

$$SD(NINAN) = (12000/t)^{.5} * mean^{.5} \qquad (F12)$$

$$= 110\ t^{-.5} * mean^{.5}$$

which seems quite similar to F10 when t = 24 months, the average

exposure in the data. However, it provides lower estimates than F5 for all t > 12 months. It is lower by a factor of 1.4 at t=24, and a factor of 2 at 48 months. Thus, this theoretical equation though similar is a understatement of the variability. We assume that the variability in the data is higher than theoretically expected because: (1) people may not all have the same underlying "p" and (2) person-months are not independent for an individual.

B. CHANGES IN STANDARD DEVIATION AS A FUNCTION OF FOLLOWUP

The standard deviation is quite related to t, the number of months followed. The best equation is estimated as

$$SD(NINAN) = 1698\ t^{-.5} \qquad (r=-.89) \quad (F13)$$

When the mean value is taken as 163, equation F7 becomes SD = 1390 t**-.5, which is similar to F13 but provides lower estimates.

Figure 6 plots the data against time, and shows equations F12 and F13. Both provide very good fits to the constant time data. Since the power of t in F13 is -.5, it is as good to get an additional person-month on the same person as for a new person. Using equations F8 and F13 the optimal value of t is

$$t = (-2)(-.5)C/[(1-1.0)\ D\ ]$$
$$= \text{infinity}$$

Thus, a longer time is better here than for visits to the limit of our data, which is 48 months. Looking at both mean and t simultaneously, we have:

$$SD(NINAN) = 32.4\ t^{-.36}\ m^{.72} \qquad (R=.97).$$

Here, the power of t is not -.5, suggesting that there is a limit to the desirable amount of time a person is followed.

C. TRANSFORMATIONS OF DATA

If square roots are used, we have

$$SD(NINAN) = 1.10\ (mean)^{-.013} \qquad (r=-.02),\ \text{constant time}$$

$$= 5.36(mean)^{.478} \qquad (r=.97),\ \text{same age/sex}$$

$$= 4.9(mean)^{.48} \qquad (r=.99),\ \text{same age/sex, t gt 12}$$

$$= 1.58(t)^{-.142} \qquad (r=-.97)\ \text{over time.}$$

Thus, for constant time, there is no relationship between the mean and the standard deviation, suggesting that this distribution is more normal than that of the untransformed data. In the cost example, each person should be studied for four months to minimize costs, quite different from the optimal time using the untransformed data. The required sample size can be reduced to about one-third (using equation F4) if the hypotheses can be specified in square roots. N1 = 4(358)**2 /10**2, N2 = 4(9.9)**2 /(.45)**2, if the mean= 125, N1 = 5126, and N2 = 1936. The ratio of N1/N2 is 2.65. Thus, there is a large potential saving in formulating hypotheses in these units.

LABS and XRAYS

Based on our study, the standard deviation of the annualized number of laboratory tests performed is approximately proportional to the mean, and the coefficients of variation are near to 2. The best equation describing that rate was (using uniform age/sex data, any exposure amount):

$$(\text{lab sd} \times 10) = 3\,(\text{mean})^{.816} \qquad (r=.94) \qquad (F14)$$

For a mean of 3 labs per year, this gives a coefficient of variation of 1.70. Figure 7 shows these equations. The mean number of laboratory tests per person in our study was 2.9 to 3.8.

For x-rays per person per year, the equation is

$$(\text{xray sd} \times 10) = 3.6\,(\text{mean})^{.704} \qquad (r=.97) \qquad (F15)$$

Figure 8 plots these equations for x-rays.

SUMMARY

We have presented some useful formulae which can be used to estimate the necessary sample size when the approximate mean and standard deviation are known. We have also provided some information on estimating the standard deviation once the mean is known since investigators may have a better feel for the expected value than for its variability. It is interesting that for every variable considered the standard deviation was substantially larger than the mean. This is not surprising since large differences among people and long right tails are to be expected.

It suggests that families of contagious distributions should be examined when modelling such variables rather than making the typical (but clearly incorrect) assumption that variables have a Poisson distribution.

BIBLIOGRAPHY

1. Branch, L.G., Fowler, F.J.: The Health Care Needs of the Elderly and Chronically Disabled in Massachusetts. Submitted to the Massachusetts Department of Public Health by Survey Research Program, a joint facility of the University of Massachusetts /Boston and the Joint Center for Urban Studies of M.I.T. and Harvard University. March, 1975.

2. Diehr, P.: Statistical Measures for admission rates. Department of Biostatistics, University of Washington Technical Report No. 24, 1978.

3. Dixon, W and Massey, F: Introduction to Statistical Analysis. McGraw Hill, New York, 1957.

4. Feigl, P: A graphical aid for determining sample size when comparing two independent proportions. Biometrics 34: 111-122. 1978.

5. Gail, M. and Gart, J: The determination of sample sizes for use with the exact conditional test in 2 x 2 comparative trials. Biometrics 3: 441-448. 1973.

6. Kilpatrick, S.J. Jr: Consultation Frequencies in General Practice. Health Services Research Fall 1977, pp 284-298.

7. Richardson, W C et al: The Seattle Prepaid Health Care Project: Comparison of Health Services Delivery. National Technical Information Service, PB # 267488-SET, 1976.

8. USDHEW: Health - United States - 1978, DHEW Publication Number (PHS) 78-1232. 1978.

FIGURE 1

Plot of the Mean of VISAN (annualized visit rate) vs. the Standard Deviation of VISAN, for the uniform age/sex groupings. N = 6690

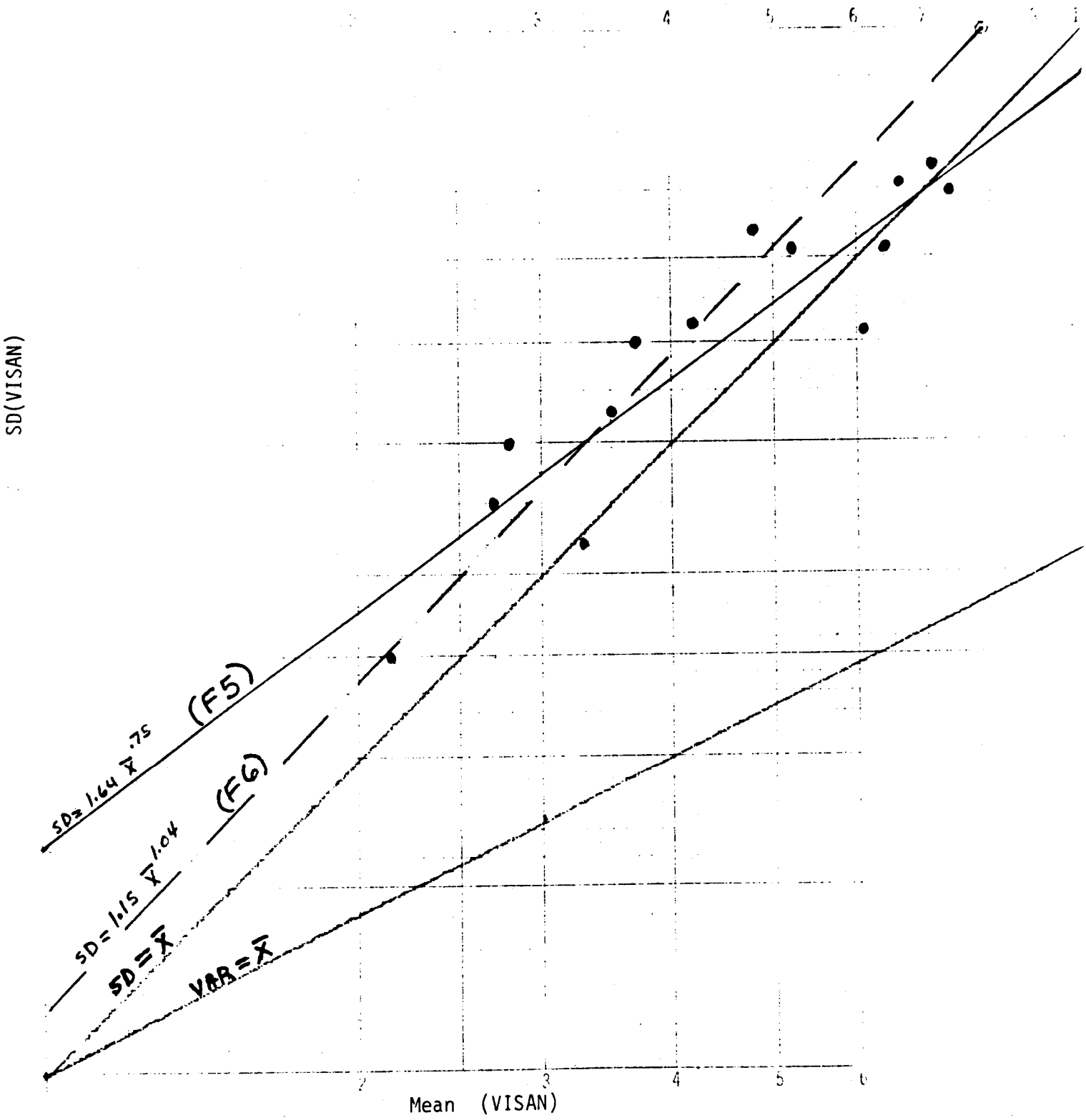

FIGURE 2
Plot of the Mean of VISAN (annualized visit rate) vs. the Standard Deviation of VISAN, for the uniform time data (n = 8288)

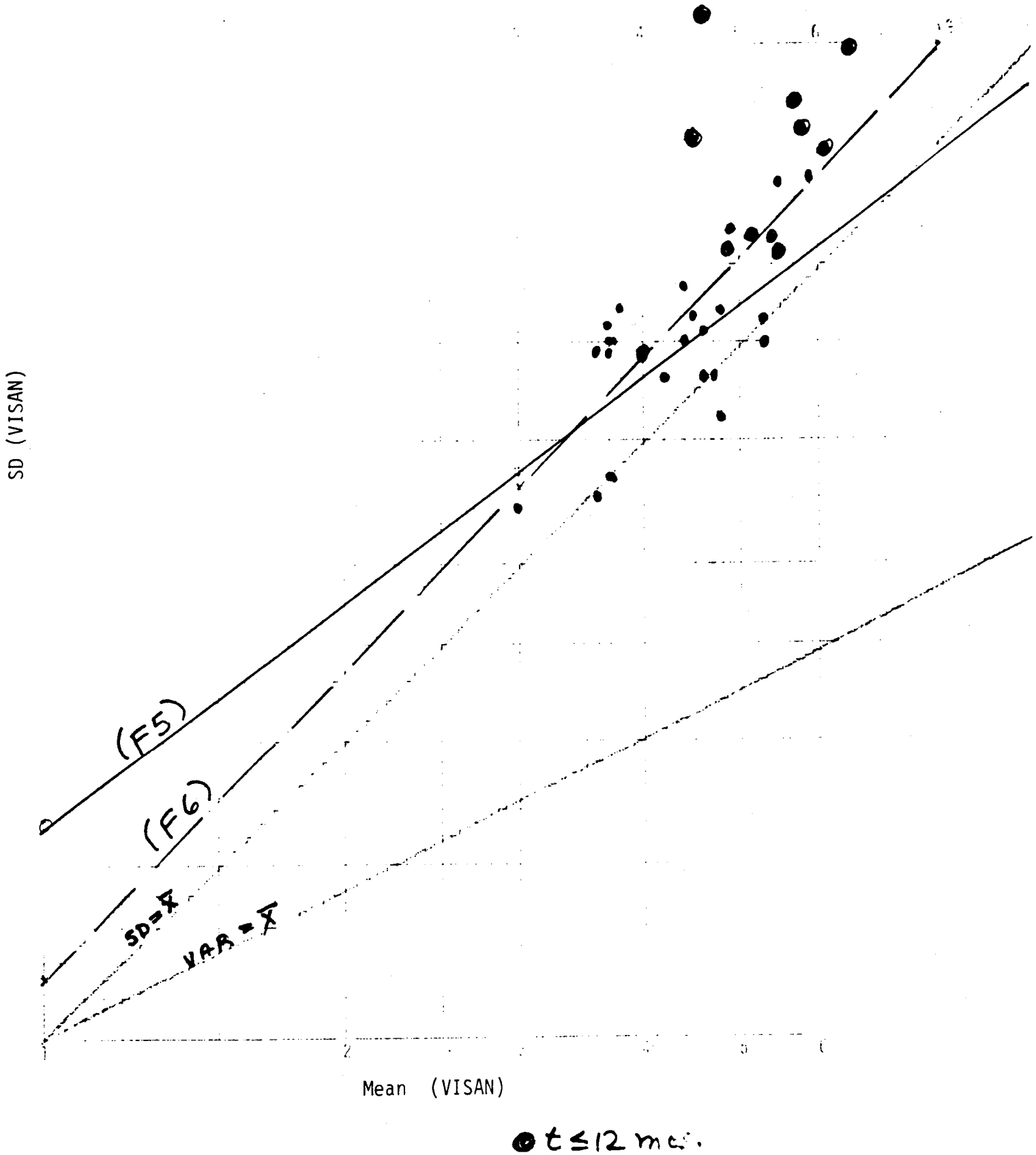

FIGURE 3

Plot of the Standard Deviation of VISAN (annualized visit rate) versus t, the number of months the person was followed. Uniform time data.

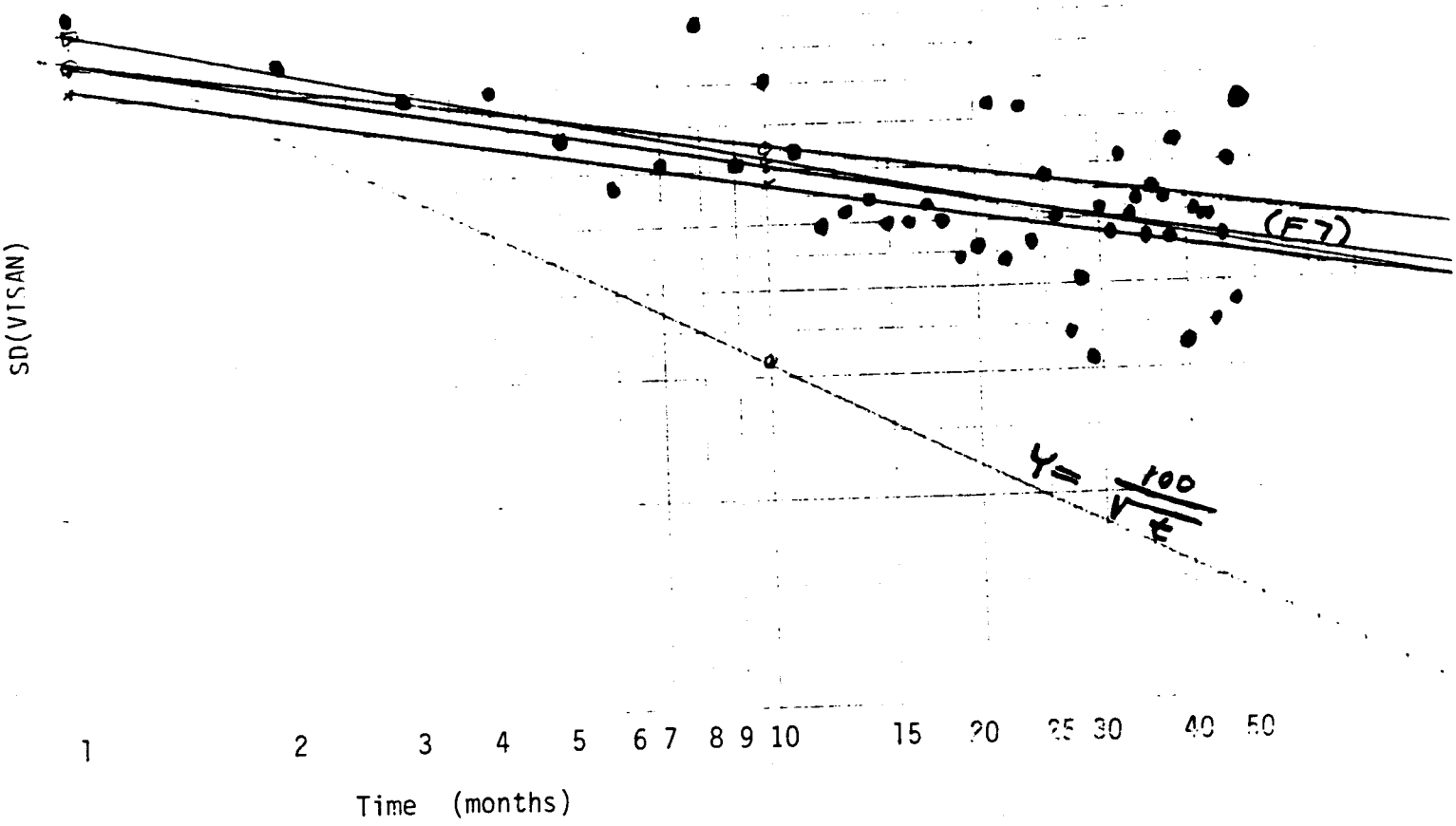

FIGURE 4

Plot of Standard Deviation of NINAN (Annualized admission rate per thousand) by the Mean of NINAN. Uniform Age/Sex Data.

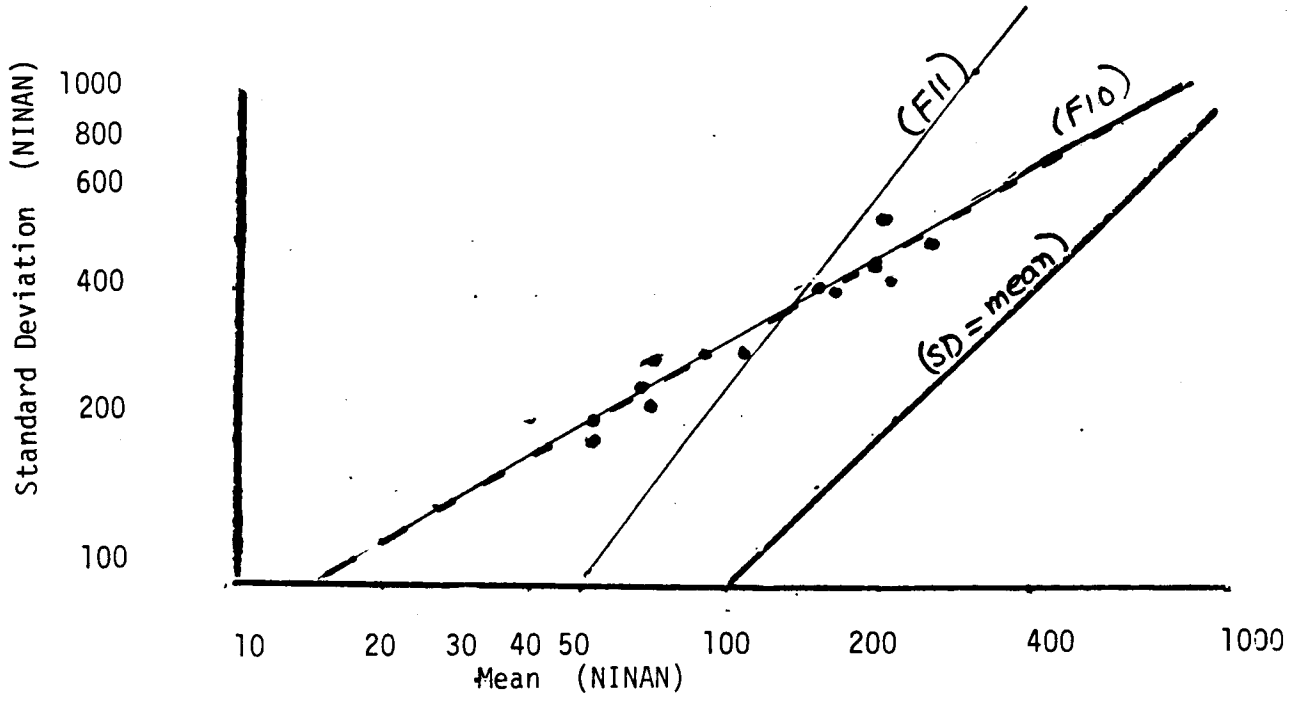

FIGURE 5

Plot of Standard Deviation of NINAN (annualized admission rate per thousand) by the Mean of NINAN. Uniform Time data.

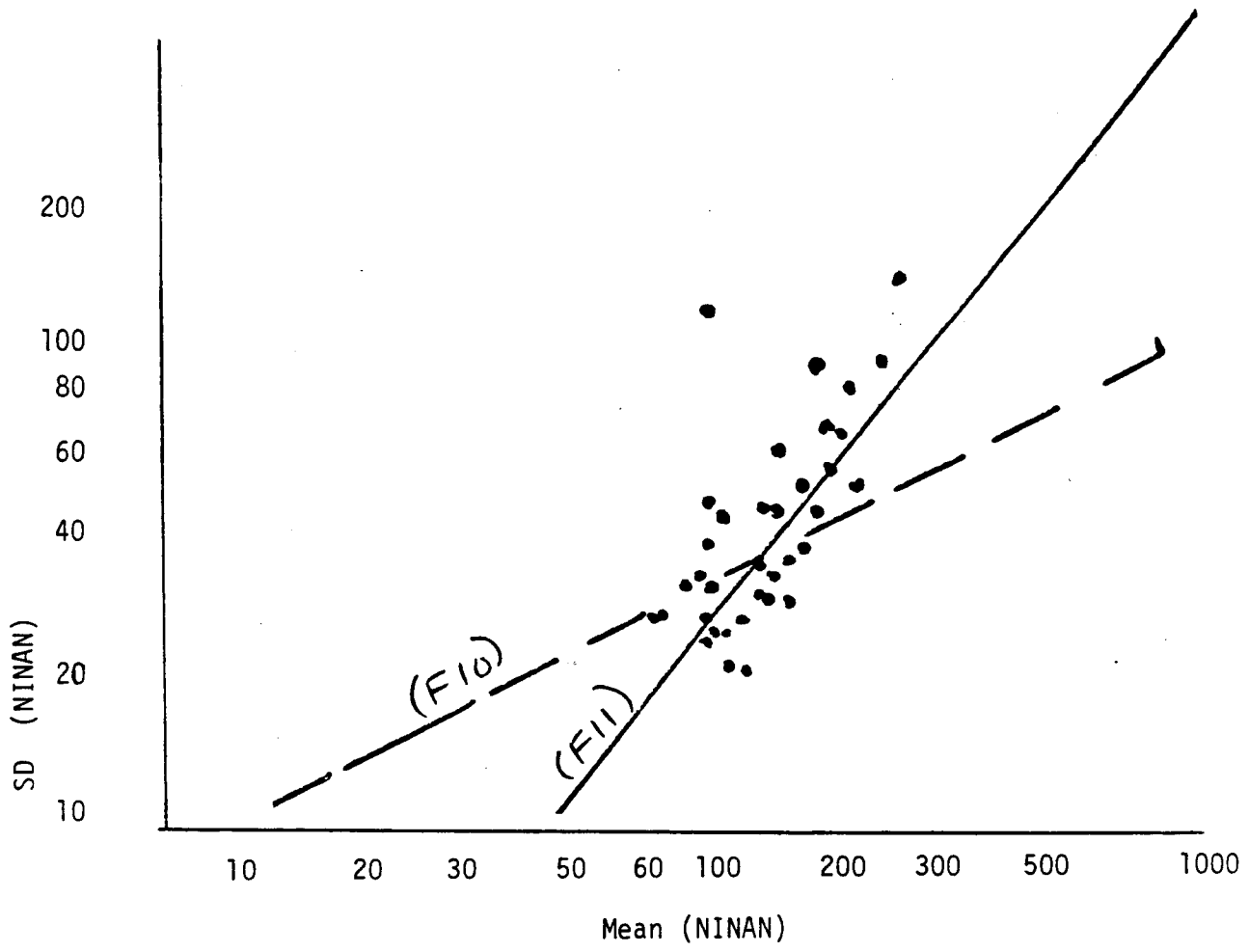

FIGURE 6

Standard Deviation of NINAN (Annualized admission rate per thousan.
a function of time (in months)

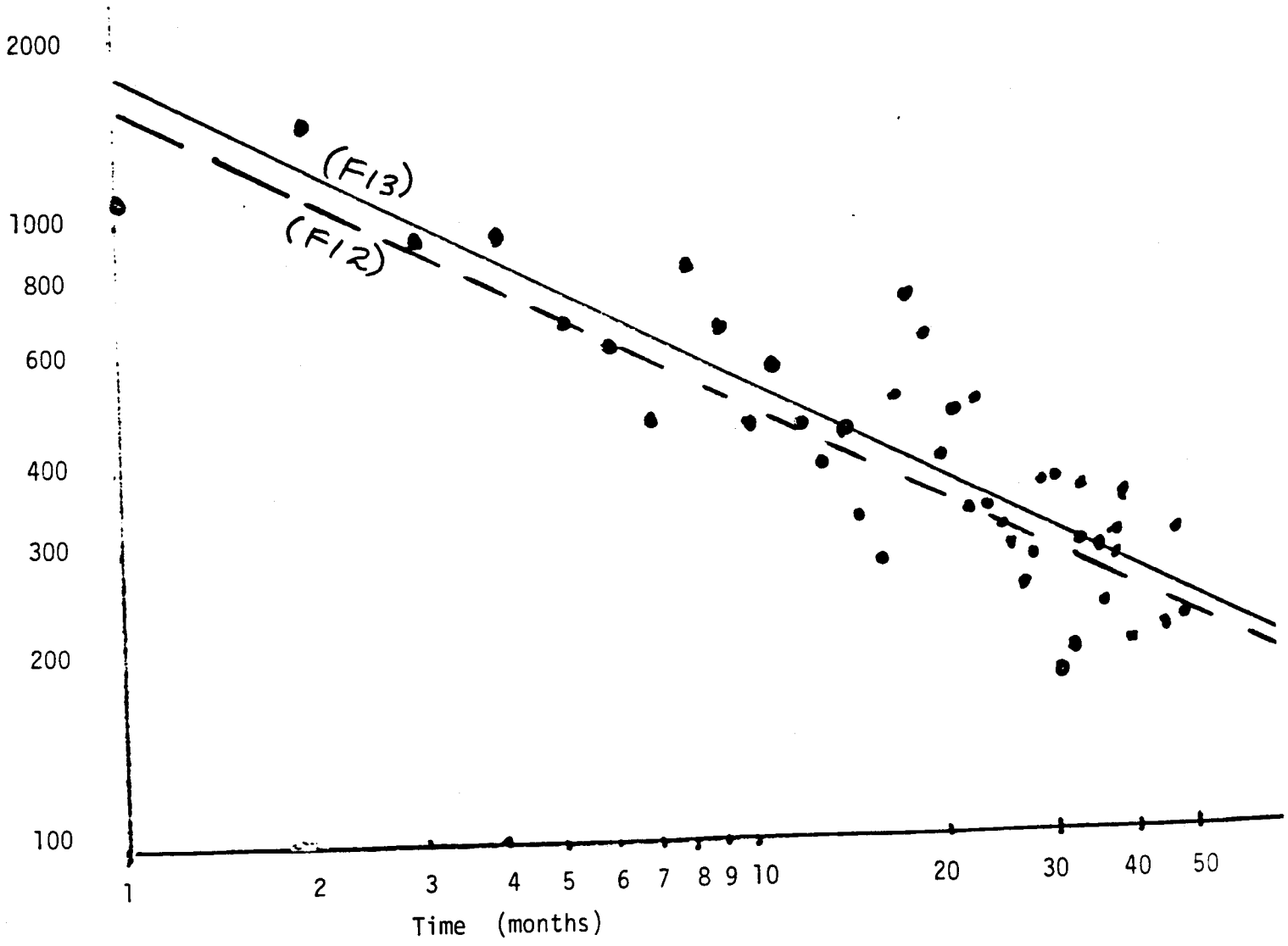

FIGURE 7

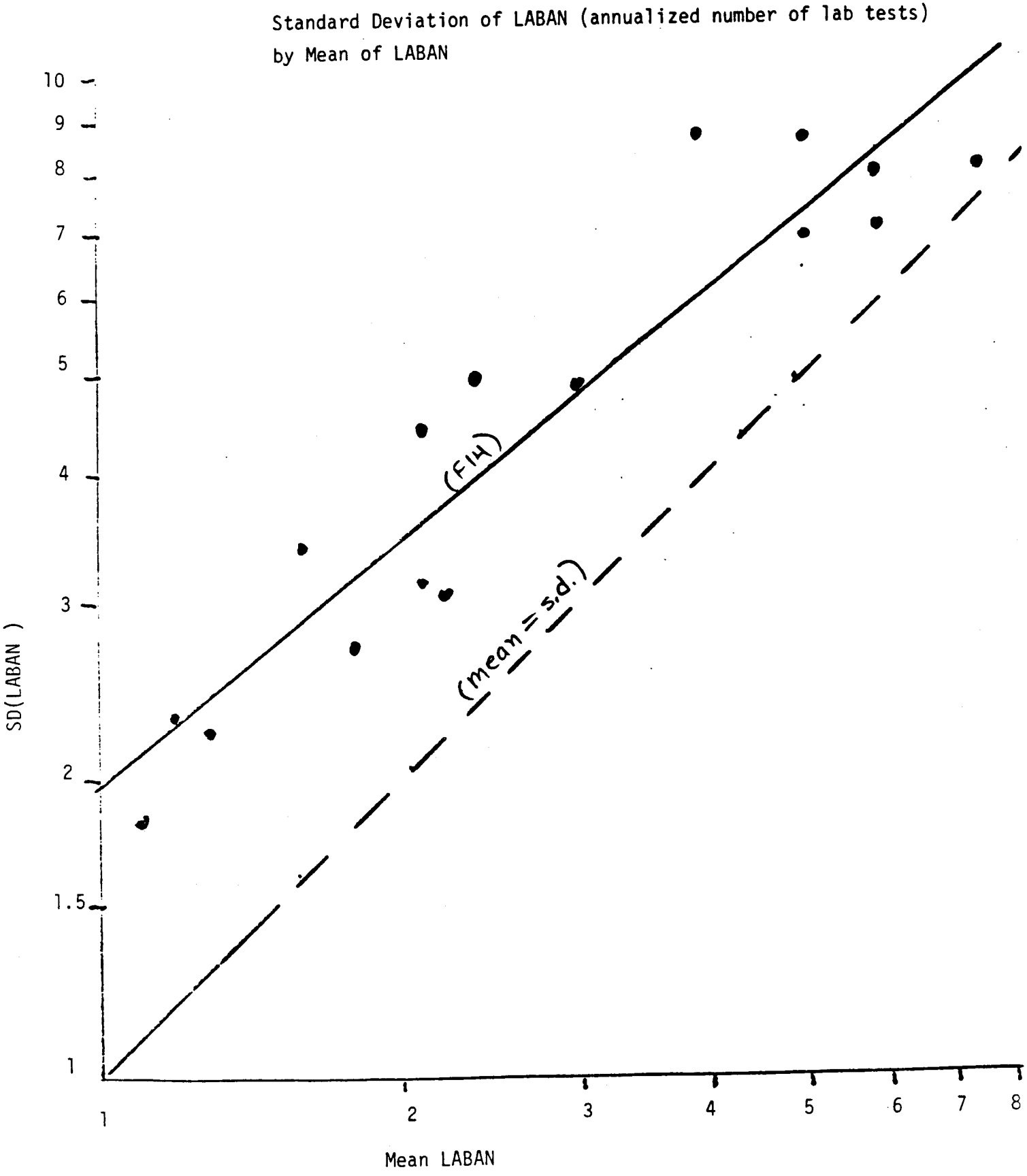

FIGURE 8

Standard Deviation of XRAYAN (Annualized number of x-rays) versus mean (XRAYAN) Uniform age/sex data