Control No. AP267

# PARAMETRIC AND NONPARAMETRIC TESTS FOR DEPENDENT DATA

by

A.H. El-Shaarawi[1] and Eivind Damsleth[2]

[1]National Water Research Institute
Canada Centre for Inland Waters
Burlington, Ontario, Canada L7R 4A6

[2]Norwegian Computing Centre
Oslo, Norway

January 1987

NWRI Contribution #87-35

# PARAMETRIC AND NONPARAMETRIC TESTS FOR DEPENDENT DATA

A.H. El-Shaarawi[1] and Eivind Damsleth[2]

[1]National Water Research Institute
Canada Centre for Inland Waters
Burlington, Ontario, Canada L7R 4A6

[2]Norwegian Computing Centre
Oslo, Norway

## ABSTRACT

Simulation and analytical results show that ignoring serial dependance can have serious effects on the performance of the t, sign and Wilcoxen tests. In particular, the true significance levels of these tests are altered significantly from the intended nominal levels. Modifications for these tests are given and shown to have the correct significance levels. Furthermore, an estimate of serial correlation is suggested for binary data and evaluated by simulation. An application to the Niagara River's toxic contaminants data concludes the paper.

Key Words: sign test, Wilcoxen test, serial Correlation, binary data

# TESTS PARAMÉTRIQUES ET NON PARAMÉTRIQUES POUR LES DONNÉES DÉPENDANTES

A.H. El-Shaarawi[1] et Eivind Damsleth[2]

[1]Institut national de recherche sur les eaux
Centre canadien des eaux intérieures
Burlington (Ontario), Canada L7R 4A6

[2]Centre informatique norvégien
Oslo, Norvège

## RÉSUMÉ

Les résultats obtenus par la simulation et l'analyse montrent que le fait d'ignorer la dépendance sérielle peut avoir des effets graves sur le rendement du test t, du test de signe et du test de Wilcoxen. En particulier, le seuil vrai de signification de ces tests peut être modifié de façon notable par rapport au niveau nominal prévu. Nous présentons des modifications de ces tests et montrons qu'ils ont alors un seuil correct de signification. De plus, nous proposons une estimation de la corrélation sérielle pour des données binaires, et nous l'évaluons par simulation. L'étude s'achève par une application aux données sur les polluants toxiques de la rivière Niagara.

Mots clés : test de signe, test de Wilcoxen, corrélation sérielle, données binaires

# PARAMETRIC AND NONPARAMETRIC TESTS
# FOR DEPENDENT DATA

A.H. El-Shaarawi[1] and Eivind Damsleth[2]

## SUMMARY

Most water quality data sets are generated as a result of sampling over time and or space with the possibility of serial correlations among successive values. Ignoring this serial dependence in performing statistical interpretations can be very serious (i.e. over-estimating or under-esstimating the true significance level).

The objectives of this paper are two-fold. The first is to assess the performance of the student t test, the sign test and the Wilcoxen signed test for dependent data. The second is to modify these tests in a manner which takes serial correlation into account. A special attention is given to data generated from autoregressive and moving average processes. Simulation experiments are used to show the effect of serial correlation and to evaluate the proposed modifications of the tests. The paper is concluded by application of the methods to the Niagara River's toxic contaminants data.

# TESTS PARAMÉTRIQUES ET NON PARAMÉTRIQUES POUR LES DONNÉES DÉPENDANTES

A.H. El-Shaarawi[1] et Eivind Damsleth[2]

## RÉSUMÉ

La plupart des séries de données sur la qualité de l'eau sont le produit d'un échantillonnage réalisé dans le temps et/ou dans l'espace avec possibilité de corrélation sérielle entre des valeurs successives. Le fait d'ignorer cette dépendance sérielle dans l'interprétation des statistiques peut être très grave (sous-estimation ou surestimation du seuil vrai de signification).

L'objet de cette étude est double. Premièrement, il s'agit d'évaluer le rendement du test t, du test de signe et du test de Wilcoxen pour des donnees dépendantes. Le deuxième objectif est de modifier ces tests de façon à tenir compte de la corrélation sérielle. Nous accordons une attention particulière aux données produites par le processus autorégressif et par les moyennes mobiles. La simulation sert à montrer l'effet de la corrélation sérielle et à évaluer les modifications proposées aux tests. L'étude s'achève par une application des méthodes aux données sur les polluants toxiques de la rivière Niagara.

## INTRODUCTION

Many water quality data are generated as a result of measuring the physical and chemical characteristics of water samples which are collected over time and or space. For example, the Canadian Department of the Environment (DOE) has been routinely measuring the concentrations of several contaminants in water samples which were collected at the head (Fort Erie, FE) and the mouth (Niagara on the Lake, NOTL) of the Niagara River (Figure 1) since 1983. The basic objectives are to use the data for making inferences about the differences between the quality of the water at the head and the mouth of the river and to estimate the additional load to the river along its course. Inspection of the data shows that: (1) the water samples were collected either weekly or biweekly, (2) the two stations were sampled on the same day; and (3) many of the measured concentrations were below the level of detection.

The difference between the FE and NOTL concentration levels can be evaluated using a paired comparison test such as the student t or the Wilcoxen signed rank test. However, due to the fact that some of the concentrations are below the detection level, the t-test is very difficult to compute and the Wilcoxen test can not be computed due to ambiguities in assigning ranks to the differences among the concentrations. The difficulties become more serious if we realize that the data represent a time series with the possibility of serial correlations among the successive values. If the differences are independent, then the sign test is the only exact test available for evaluating the significance of the difference in this case. It is not advisable, however, to ignore serial correlation in interpreting the results of these tests. As it was pointed out by Scheffé (1959), ignoring serial correlations when making inferences about the means of several populations can be serious.

The objectives of this paper are two-fold. The first is to assess the performance of the student t test, the sign test and Wilcoxen signed rank test for serially correlated data. The second is to modify these tests in a manner which takes serial correlation into account. A special attention is given to data generated from a first order autoregressive process (AR(1)), and from a moving average of order one (MA(1)). Simulation experiments are used to show the effect of serial correlation and to evaluate the proposed modifications of the student t and sign test.

## THEORETICAL CONSIDERATIONS

1. The effect of serial correlation on inference about the mean. Let $x_1, x_2, \ldots, x_n$ be a realization from an equally spaced stationary process with the following characteristics:

$$\begin{aligned} E(x_i) &= \mu \qquad i = 1,2,\ldots,n \\ V(x_i) &= \sigma^2 \qquad i = 1,2,\ldots,n \qquad (1) \\ \text{and } C(x_i, x_{i+K}) &= \sigma^2 \rho_K \end{aligned}$$

where E, V and C refer to expected value, variance and covariance, respectively. In the present setting the $x_i$'s denote the differences between simultaneous measurements of a contaminant at the head and mouth of the river. The object is to make inference about $\mu$. It is assumed that $|\rho_\kappa|>|\rho_{\kappa+1}|$ and for $\kappa \geq S_0$, $\rho_\kappa$ can be regarded as zero, where $S_0 << n$. Models with the above properties can be found in texts on time series such as Box & Jenkins (1970). For example, a moving average MA(q) model of order q satisfies exactly the above requirements and an autoregressive AR(P) model of order p satisfies the requirements approximately.

Let $\bar{x}$, s and $r_\kappa$ ($\kappa$=1,2,...) be the sample mean, standard deviation and the $\kappa$th serial correlation, respectively, where

$$\bar{x} = \sum x_i/n \quad , \quad s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 \text{ and}$$

$$\rho_k = \frac{\sum_{i=1} (x_i-\bar{x})(x_{i+\kappa}-\bar{x})}{s^2}$$

The null hypotheses $H_0: \mu = 0$ is tested in this paper using the t test, the sign test and the Wilcoxen signed rank test which are based on the statistics

$t \quad = \sqrt{n}\ \bar{x}/s$

$S_+ \quad =$ number of positive values among the $x_i$

and $T_+ \quad =$ sum of positive ranks.

$S_+$ and $T_+$ are obtained as follows. Let $\delta_i$ be a random variable which takes the value 1 when $x_i$ is positive and 0 otherwise and let $R_i$ be the rank of the absolute value of $x_i$ among the absolute values of all the $x_i$'s. Then

$$S_+ = \sum_{i=1}^{n} \delta_i \qquad \text{and} \qquad T_+ = \sum_{i=1}^{n} \delta_i R_i$$

The use of the above test (in their standard form) requires the independence of the different values of $x_i$, while the use of t statistic requires the added assumption of the normality.

It is easy to show, using (1), that

$$E\bar{x} = \mu \quad ,$$

$$V(\bar{x}) = \frac{\sigma^2}{n} \{1 + 2 \sum_{i=1}^{n} (1 - \frac{i}{n})\ \rho_i\}, \tag{2}$$

and $$E(s^2) = \sigma^2 \{1 - \frac{2}{(n-1)} \sum_{i=1}^{n} (1 - \frac{i}{n})\ \rho_i\} \quad .$$

When $\rho_i = 0$ for all $i = 1,2,\ldots,n$ then $V(\bar{x}) = V_0(\bar{x}) = \sigma^2/n$ and $E(s^2) = \sigma^2$. Assuming that $\sum_{i=1}^{n} (1 - \frac{i}{n}) \rho_i$ converges for $n \to \infty$ to a limit A which is independent of n, then the effect of serial correlation can be seen from (2) by noting that as $n \to \infty$

$$V(\bar{x}) = V_0 (\bar{x}) \{1 + 2A\}$$

$$\text{and} \quad E(s^2) = \sigma^2 \tag{3}$$

Hence $s^2$ provides an unbiased estimate for $\sigma^2$, while the variance of $\bar{x}$ is changed by a factor which is independent of n.

Similarly under $H_0$, the expected values of $S_+$ and $T_+$ are

$$E(S_+) = n/2, \qquad \text{and} \qquad E(T_+) = n(n+1)/4.$$

The variances of these statistics, assuming independance are

$$V(S_+) = n/4 \qquad \text{and} \qquad V(T_+) = n(n+1)(2n+1)/24.$$

To determine the effects of dependance on these statistics, suppose $x_1,\ldots,x_n$ are normally distributed, then it is easy to show that the correlation between $\delta_i$ and $\delta_{i+k}$ is

$$\frac{2}{\pi} \operatorname{Sin}^{-1} \rho_K$$

Hence

$$V(S_+) = \frac{n}{4}\{1 + \frac{4}{\pi}\sum_{i=1}^{n}(1 - \frac{i}{n})\sin^{-1}\rho_i\} \tag{4}$$

Using the projection method (Hettmansperger, 1984), it is possible to show that for large n

$$V(T_+) = \frac{n^3}{12}\{1 + \frac{12}{\pi}\sum^{n}(1 - \frac{i}{n})\sin^{-1}\rho_i/2\}. \tag{5}$$

By noting that for small $\rho_i$, expressions (4) and (5) can be written approximately as

$$V(S_+) = \frac{n}{4}\{1 + \frac{4}{\pi}\sum_{i=1}^{n}(1 - \frac{i}{n})\rho_i\}$$
$$= V_0(S_+)\{1 + \frac{4}{\pi}A\}$$

$$V(T_+) = \frac{n^3}{12}\{1 + \frac{6}{\pi}\sum_{i=1}^{n}(1 - \frac{i}{n})\rho_i\}$$
$$= V_0(T_+)\{1 + \frac{6}{\pi}A\}$$

These approximate expressions are in forms which can be easily compared with the variance of $\bar{x}$ (equation 2). This indicates that the effect of serial correlation is almost the same on $V(T_+)$ and $V(\bar{x})$, (since $6/\pi \approx 2$), while this effect is less pronounced in the case of the sign test since $4/\pi < 2$.

2. <u>Modification of the tests.</u> As indicated previously, the variances of the test statistics are affected by the presence of serial correlation. It is shown by Wold (1950) that if $S_0$ is small compared with n, then the substitution of the usual estimates for serial correlation will validate the conclusions drawn from these tests. Hence the three modified tests are as follows:

(a) the modified t-test

$$t_m = \sqrt{n}\,\bar{x}/s\,\sqrt{1 + 2\sum_{i=1}^{S_0}\left(1 - \frac{i}{n}\right) r_i}$$

(b) the modified sign test

$$S = 2\left(S_+ - \frac{n}{2}\right) / \sqrt{n}\left\{1 + \frac{4}{\pi}\sum_{i=1}^{S_0}\left(1 - \frac{i}{n}\right) b_i\right\}$$

(c) the modified Wilcoxen test

$$T = 2\sqrt{3}\left(T_+ - \frac{n(n+1)}{4}\right) / n\sqrt{n}\left\{1 + \frac{6}{\pi}\sum_{i=1}^{S_0}\left(1 - \frac{i}{n}\right) d_i\right\}$$

where $b_i$ and $d_i$ are the lag i serial correlations for the binary data $\delta_1, \ldots, \delta_n$ and for the data generated by $\delta_1 R_1, \delta_2 R_2, \ldots, \delta_n R_n$ respectively.

## APPLICATIONS TO MA(1) AND AR(1) MODELS

The above results are specialized here to the cases of first order moving average (MA(1)) and autoregressive (AR(1)) models. As demonstrated by Wold (1950), the distributions of $t_m$, S, and T, for large n, coverge to the standard normal distribution, hence the effect of ignoring the serial correlation can be studied by comparing the true significance level of $\alpha_T$ with the nominal level. Table 1 gives the exact and approximate functional form for $\alpha_T$ when the nominal level is $\alpha$ for the MA(1) and AR(1) models respectively. The value of $Z_\alpha$ used in the Table is the upper $\alpha$-percentile in the standard normal distribution.

Table 1. Expressions for the true significance level $\alpha_T$ for the tests

| Tests | $\alpha_T$ MA(1) | AR(1) |
|---|---|---|
| t | $1 - \Phi\left(\frac{Z\alpha}{\sqrt{1 + 2\rho_1}}\right)$ | $1 - \Phi\left(\frac{Z\alpha}{\sqrt{1 + \frac{2\rho_1}{1-\rho_1}}}\right)$ |
| Sign | $1 - \Phi\left(\frac{Z\alpha}{\sqrt{1 + \frac{4}{\pi}\sin^{-1}\rho_1}}\right)$ | $1 - \Phi\left(\frac{Z\alpha}{\sqrt{1 + \frac{4\rho}{\pi(1-\rho)}}}\right)$ |
| Wilcoxen | $1 - \Phi\left(\frac{Z\alpha}{\sqrt{1 + \frac{12}{\pi}\sin^{-1}\rho_{1/2}}}\right)$ | $1 - \Phi\left(\frac{Z\alpha}{\sqrt{1 + \frac{6\rho}{\pi(1-\rho)}}}\right)$ |

Numerical computations assuming $\alpha=0.05$ ($Z\alpha=1.645$) for different values of $\rho_1$ are given in Table 2 which shows that ignoring serial correlations can lead to very serious errors in performing the different tests.

## THE SIMULATION STUDIES

The CDC 171 computer system was used to generate 19 runs of 100 samples with sample size n=100 from the AR(1) model

$$x_t = \rho_1 x_{t-1} + \varepsilon_t \quad ,$$

where $\varepsilon_t$ has a standard normal distribution and the serial correlation $\rho_1$ is assumed to have the values -0.9,-0.8,...,0.9. Each run corresponds to a specific value of $\rho_1$. The ordinary sign and t-tests and their corresponding modified tests as well as the estimates $r_1$ and $\hat{\rho}_1=\sin 2/\pi\ b_1$ of $\rho_1$ were computed for each run. The proportion of the 100 samples where the null hypothesis $H_0:\mu=0$ was rejected was calculated. This gives an estimate of the true significance level. Furthermore, the mean and the standard deviation of $r_1$ and $\hat{\rho}_1$ were computed. The above simulation was repeated with n=150 and n=200.

Figure 2 gives the estimates of the true significance levels and their 90% confidence limits. As expected for the usual tests, the true significance level deviates seriously from the nominal level. As expected it is below (above) the nominal level for negative (positive) values of $\rho_1$. Also the violations appear to be more serious for the t-test. On the other hand, the modified tests appear to be well estimated by the nominal level.

Table 3 gives the summary statistics for $r_1$ and $\hat{\rho}_1$ which indicate that these values give good estimates for $\rho_1$. The use of the binary series results in a smaller precision for $\hat{\rho}_1$ as compared to $r_1$ as can be seen from their estimated standard deviations.

## NUMERICAL EXAMPLE

The data used in this example are collected on a weekly basis, from the FE and NOTL stations of the Niagara River, for 37 weeks starting on the third of October 1984. The data represent the measurements of the concentrations of aluminium in water samples. As mentioned before, two of the concentrations were below the level of detection and this did not occur at FE and NOTL on the same day, $S_0$ we were only able to determine if the difference between the two stations was positive or negative. The number of positive differences was $S_+=29$ and the estimate of the serial correlation of lag one is $b_1=-0.435$. Hence the calculated values of the sign and modified sign tests are 3.452 and 5.084 respectively which indicate a significant ($p<.01$) increase in the concentration from Fort Erie to Niagara on the Lake during the study period.

In this case, the first order autocorrelation was negative, so that the effect of taking the serial correlation into account is to strengthen the conclusions. However, we believe that in environmental series it is more common to find positive serial correaltion, which will have the opposite effect.

## REFERENCES

Box, G.E.P. and Jenkins, G.M. (1976). Time Series Analysis Forecasting and Control. Holden Day, San Fracisco.

Hehmansperger, T.P. (1984). Statistical Inference Based on Ranks. Wiley, New York.

Scheffé, H. (1959). The Analysis of Variance. Wiley, New York.

Wold, H. (1950). On Least Square Regression with Autocorrelated Variables and Residuals. Bull. Inst. Int. Statist. 32, 277-289.

Table 2. True significance level for normal 5% student t, sign and Wilcoxon tests

True significance level $\alpha_T$

| | MA(1) | | | AR(1) | | |
|---|---|---|---|---|---|---|
| $\rho_1$ | t | sign | Wilcoxon | t | sign | Wilcoxon |
| -0.9 | | | | .0000 | .0045 | .0000 |
| -0.8 | | | | .0000 | .0063 | .0000 |
| -0.7 | | | | .0000 | .0085 | .0002 |
| -0.6 | | | | .0005 | .0114 | .0010 |
| -0.5 | .0000 | .0022 | .0000 | .0022 | .0151 | .0032 |
| -0.4 | .0001 | .0086 | .0003 | .0060 | .0196 | .0073 |
| -0.3 | .0046 | .0177 | .0058 | .0125 | .0251 | .0139 |
| -0.2 | .0168 | .0282 | .0181 | .0220 | .0319 | .0232 |
| -0.1 | .0329 | .0391 | .0337 | .0345 | .0401 | .0352 |
| 0.0 | .0500 | .0500 | .0500 | .0500 | .0500 | .0500 |
| 0.1 | .0666 | .0607 | .0659 | .0684 | .0618 | .0676 |
| 0.2 | .0822 | .0711 | .0809 | .0896 | .0760 | .0880 |
| 0.3 | .0967 | .0813 | .0950 | .1137 | .0929 | .1113 |
| 0.4 | .1101 | .0913 | .1081 | .1408 | .1132 | .1376 |
| 0.5 | .1224 | .1013 | .1203 | .1711 | .1376 | .1674 |
| 0.6 | | | | .2054 | .1674 | .2014 |
| 0.7 | | | | .2448 | .2045 | .2406 |
| 0.8 | | | | .2917 | .2526 | .2879 |
| 0.9 | | | | .3529 | .3206 | .3499 |

Table 3. Mean and standard deviations of $r_1$ and $\hat{\rho}_1$ for each

| $\rho_1$ | n=100 | | n=150 | | n=200 | |
|---|---|---|---|---|---|---|
| | $r_1$ | $\hat{\rho}_1$ | $r_1$ | $\hat{\rho}_1$ | $r_1$ | $\hat{\rho}_1$ |
| -0.9 | - .89 ± .048 | - .88 ± .070 | - .89 ± .043 | - .89 ± .071 | - .89 ± .033 | - .89 ± .050 |
| -0.8 | - .78 ± .064 | - .79 ± .094 | - .75 ± .044 | - .80 ± .077 | - .80 ± .045 | - .79 ± .073 |
| -0.7 | - .69 ± .073 | - .68 ± .104 | - .70 ± .068 | - .70 ± .101 | - .70 ± .054 | - .69 ± .087 |
| -0.6 | - .60 ± .093 | - .60 ± .137 | - .59 ± .069 | - .59 ± .104 | - .59 ± .060 | - .60 ± .083 |
| -0.6 | - .50 ± .100 | - .49 ± .143 | - .49 ± .072 | - .51 ± .112 | - .49 ± .059 | - .49 ± .094 |
| -0.4 | - .40 ± .094 | - .38 ± .145 | - .40 ± .070 | - .39 ± .118 | - .40 ± .060 | - .39 ± .094 |
| -0.3 | - .30 ± .099 | - .30 ± .136 | - .30 ± .071 | - .28 ± .117 | - .30 ± .075 | - .30 ± .108 |
| -0.2 | - .20 ± .101 | - .17 ± .163 | - .21 ± .100 | - .22 ± .125 | - .21 ± .071 | - .21 ± .093 |
| -0.1 | - .11 ± .115 | - .14 ± .159 | - .11 ± .084 | - .11 ± .108 | - .10 ± .065 | - .12 ± .101 |
| 0.0 | - .01 ± .100 | - .01 ± .157 | .00 ± .082 | - .02 ± .122 | .00 ± .070 | .00 ± .117 |
| 0.1 | .08 ± .087 | .11 ± .139 | .08 ± .070 | .11 ± .113 | .09 ± .064 | .10 ± .109 |
| 0.2 | .19 ± .099 | .19 ± .153 | .19 ± .078 | .20 ± .119 | .19 ± .075 | .20 ± .113 |
| 0.3 | .29 ± .098 | .29 ± .137 | .28 ± .097 | .29 ± .138 | .29 ± .069 | .28 ± .104 |
| 0.4 | .38 ± .107 | .40 ± .158 | .38 ± .082 | .40 ± .118 | .39 ± .062 | .39 ± .108 |
| 0.5 | .48 ± .092 | .48 ± .140 | .49 ± .073 | .50 ± .117 | .49 ± .062 | .50 ± .096 |
| 0.6 | .56 ± .073 | .56 ± .128 | .57 ± .064 | .56 ± .104 | .58 ± .057 | .57 ± .094 |
| 0.7 | .67 ± .081 | .67 ± .118 | .68 ± .063 | .69 ± .096 | .69 ± .056 | .70 ± .084 |
| 0.8 | .77 ± .072 | .79 ± .109 | .77 ± .062 | .79 ± .089 | .78 ± .048 | .79 ± .066 |
| 0.9 | .86 ± .057 | .88 ± .080 | .87 ± .050 | .89 ± .070 | .88 ± .038 | .90 ± .056 |

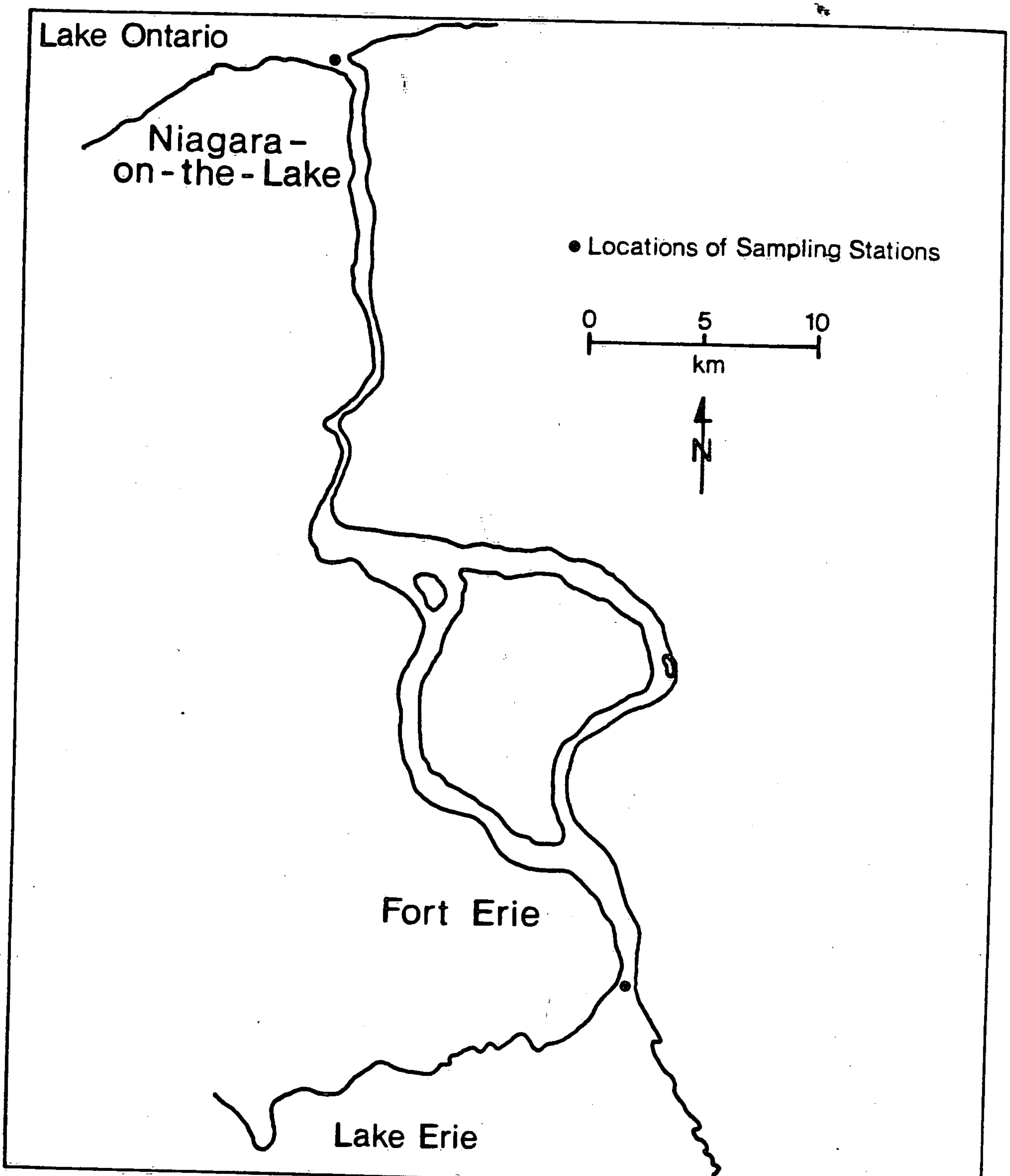
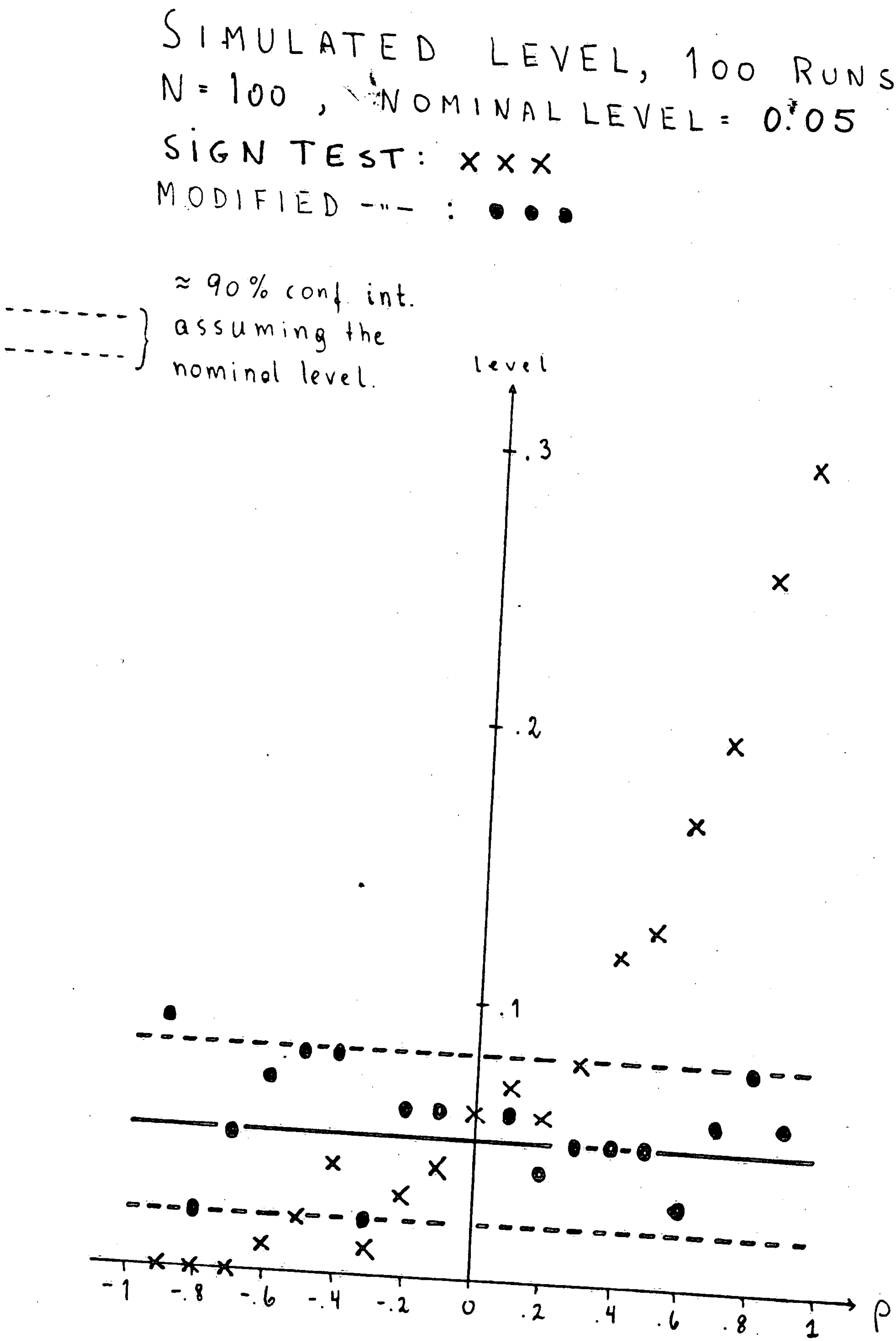
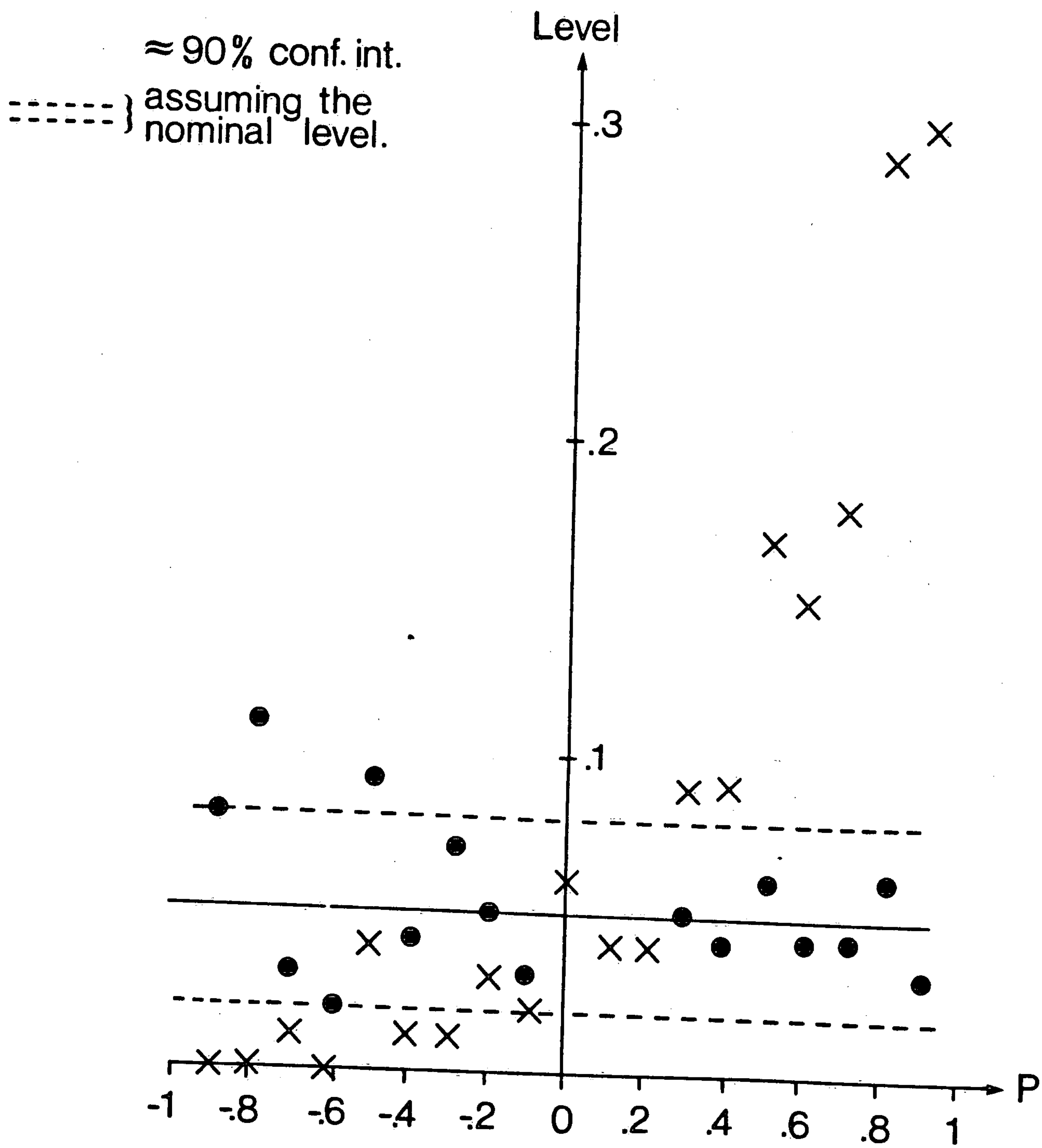



Figure 1 Location of sampling stations on the Niagara River.

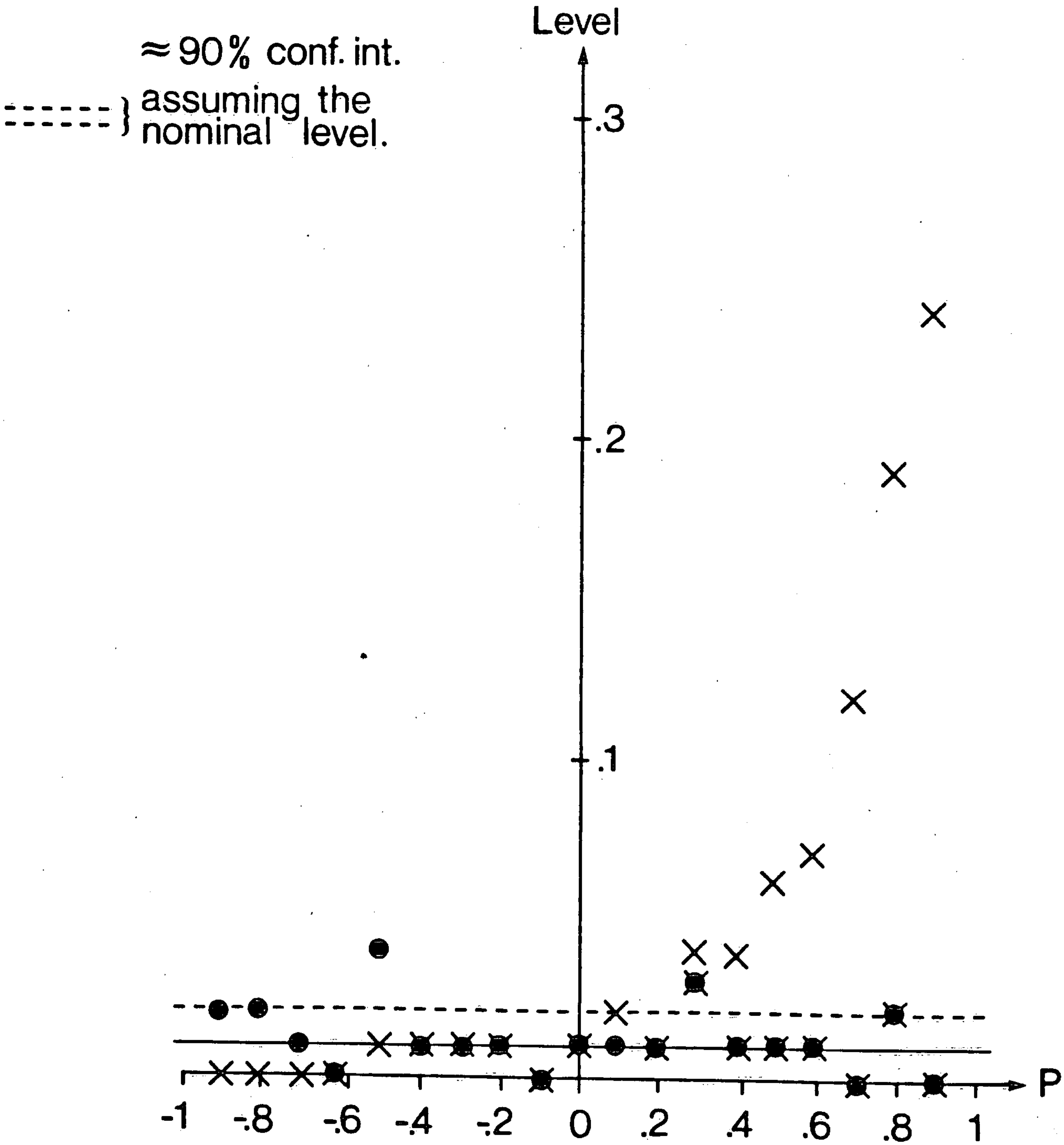
Figure 2. Simulation Results



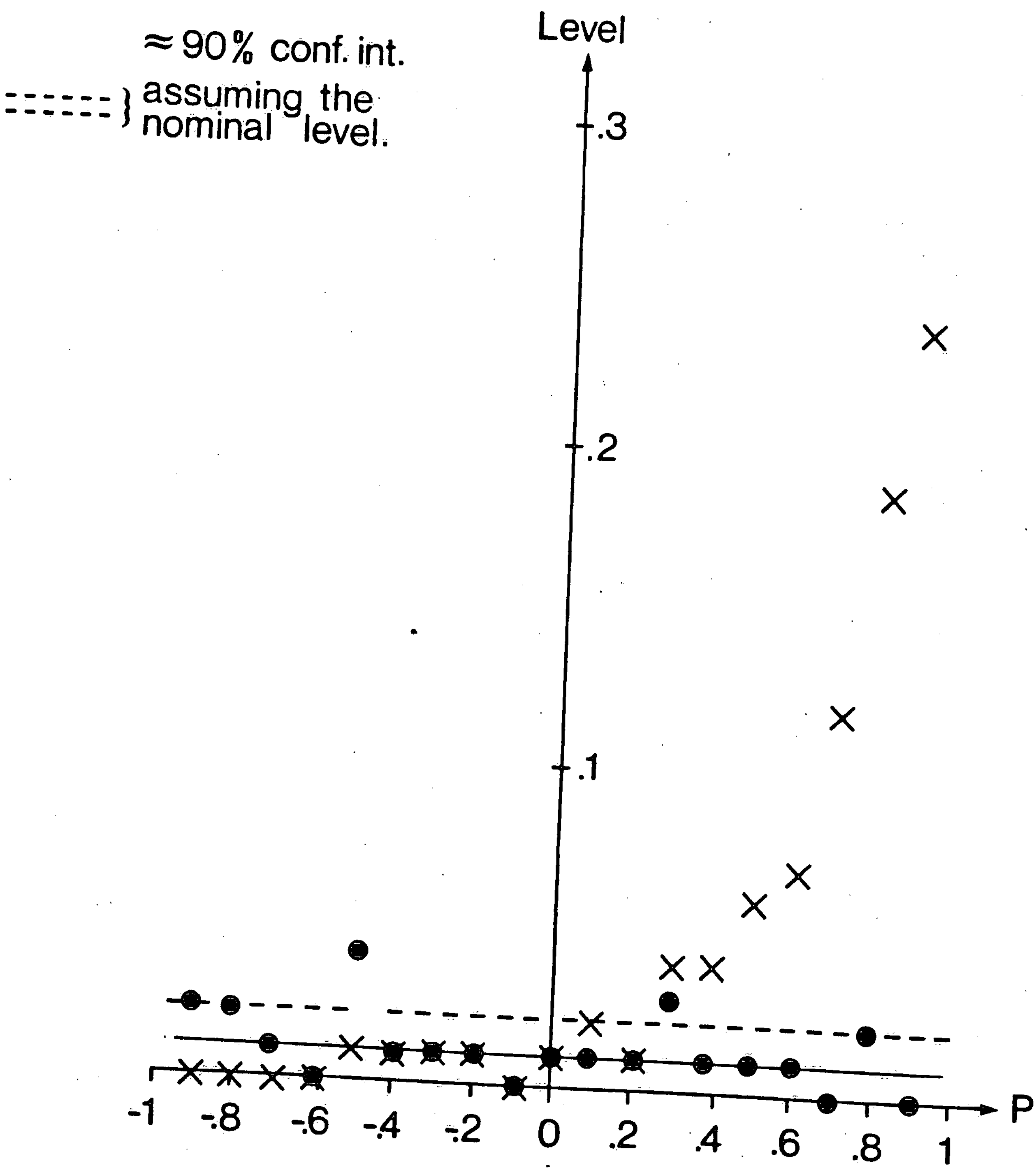
SIMULATED LEVEL, 100 RUNS
N = 200, NOMINAL LEVEL = 0.05
SIGN TEST : XXX
MODIFIED SIGN TEST : ●●●



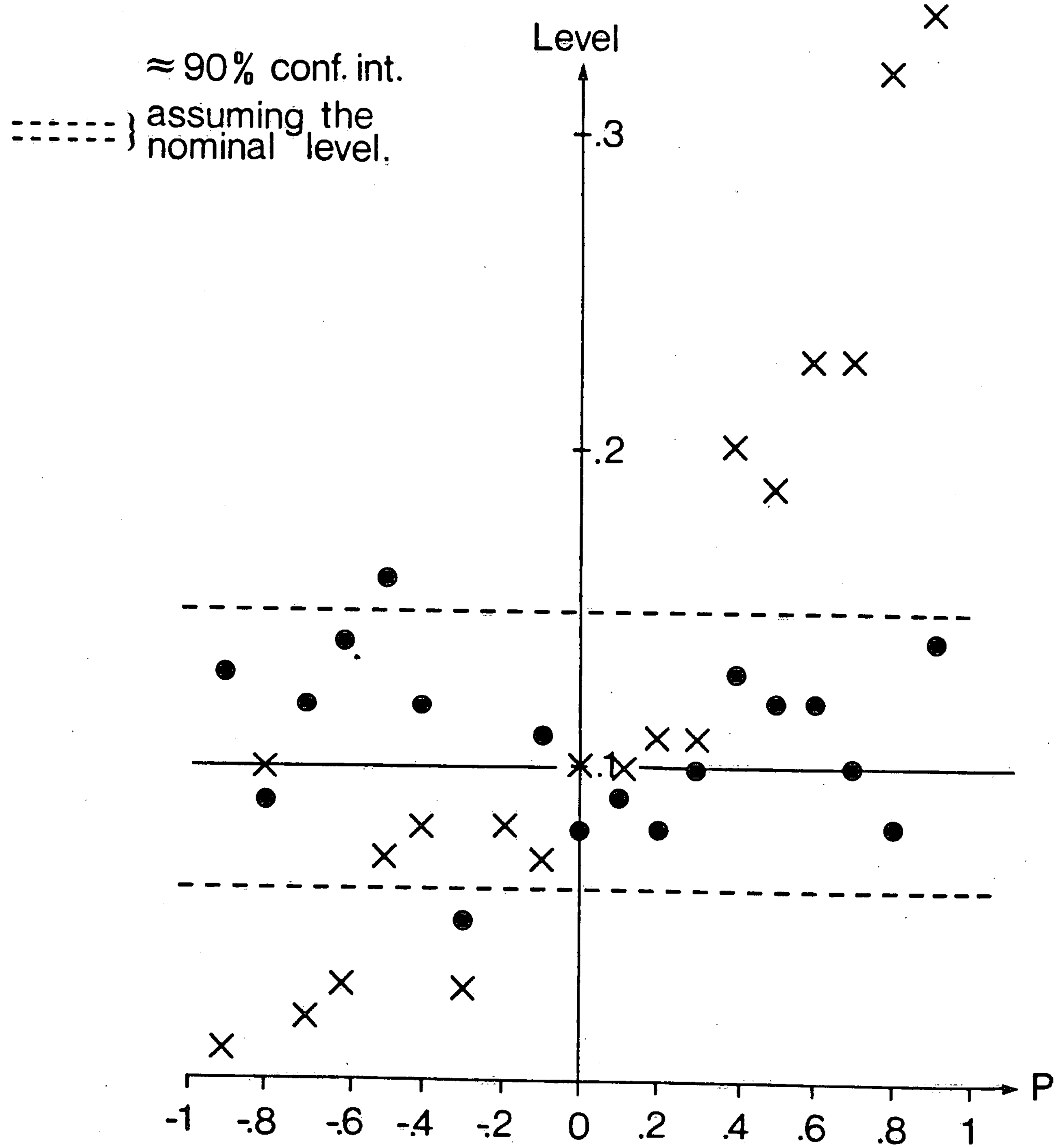
SIMULATED LEVEL 100 RUNS
N = 100, NOMINAL LEVEL = 0.01
SIGN TEST : ×××
MODIFIED SIGN TEST : ●●●

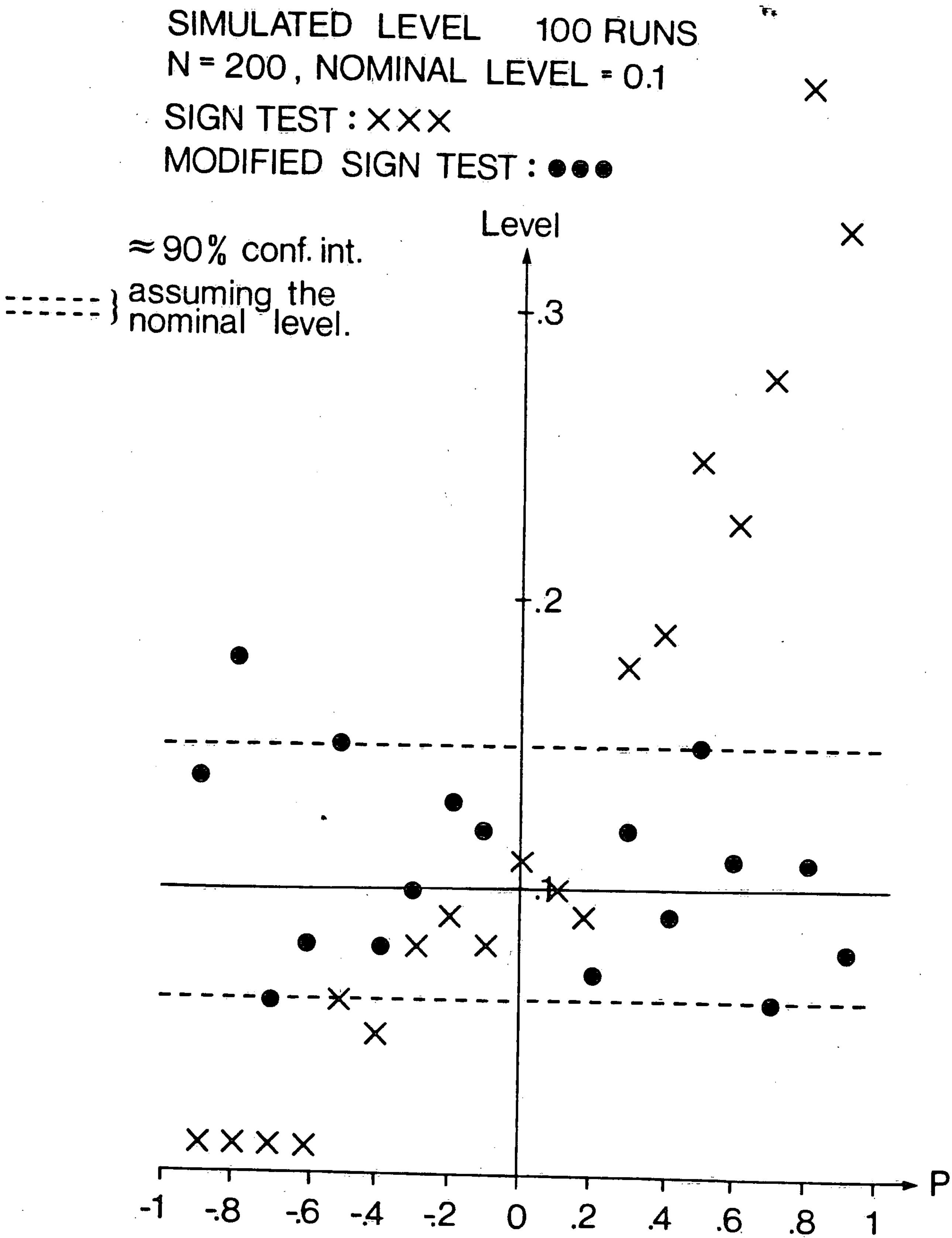


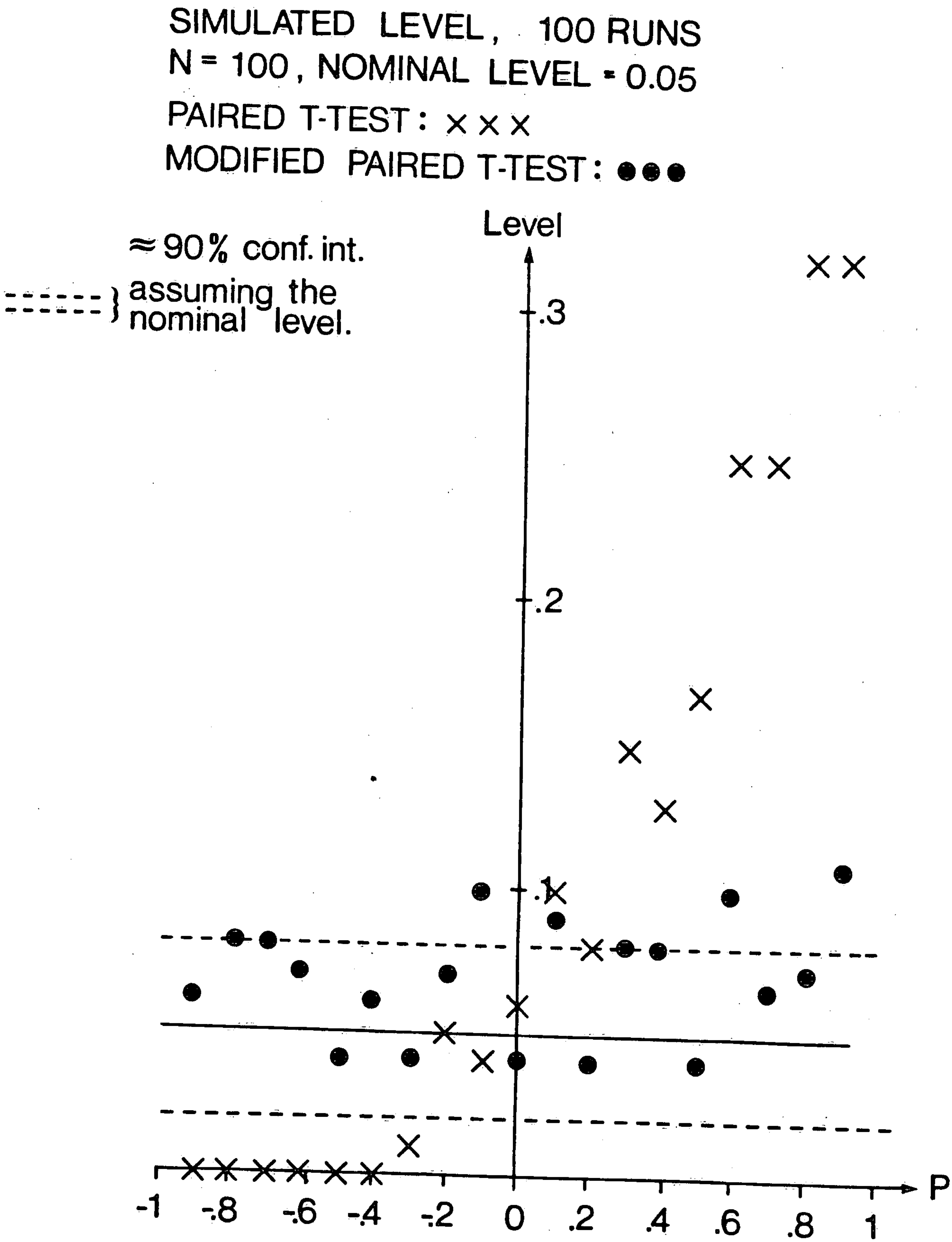
SIMULATED LEVEL, 100 RUNS
N = 200, NOMINAL LEVEL = 0.01
SIGN TEST : XXX
MODIFIED SIGN TEST : ●●●

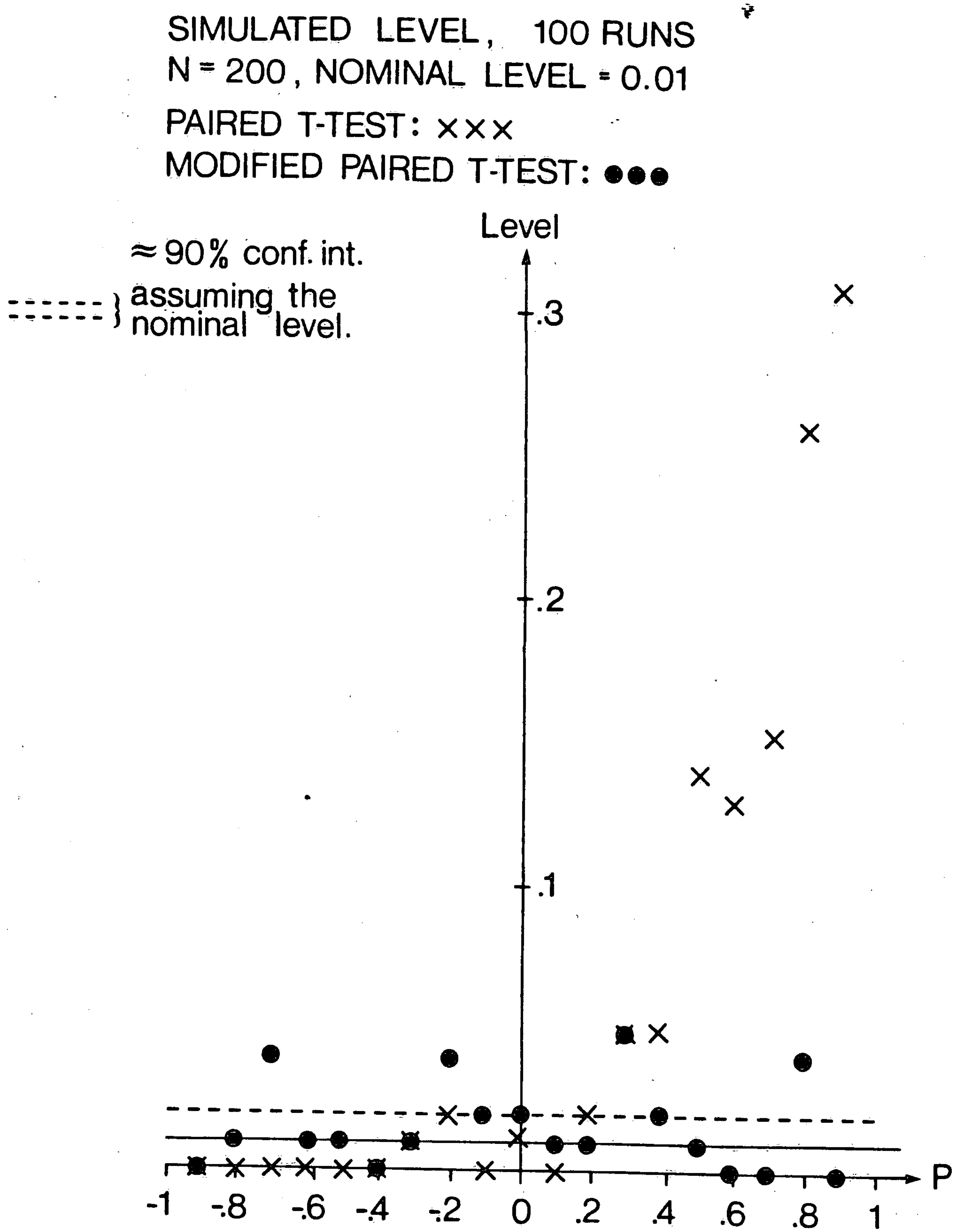


SIMULATED LEVEL, 100 RUNS
N = 100, NOMINAL LEVEL = 0.1
SIGN TEST : ×××
MODIFIED SIGN TEST : ●●●

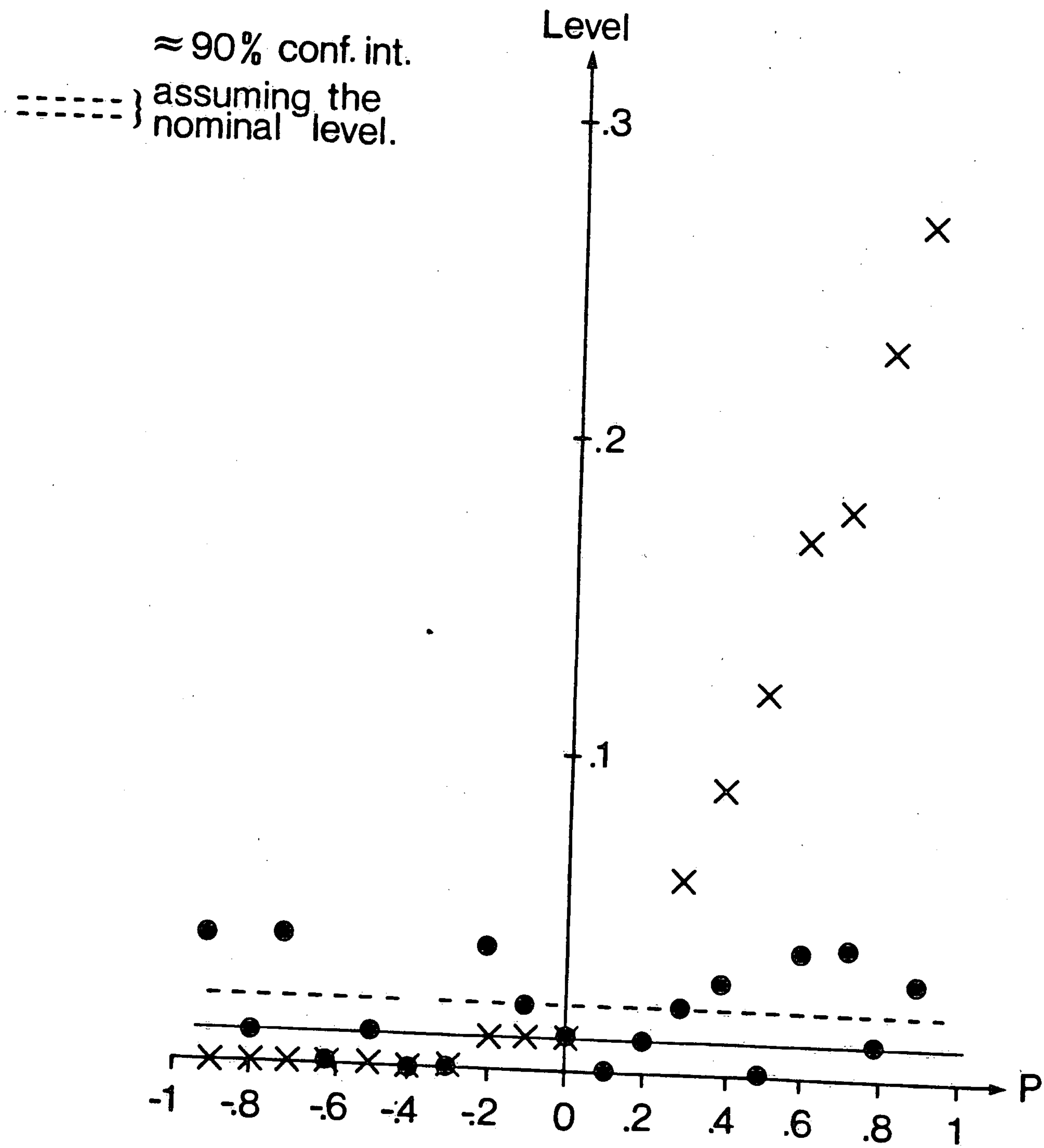


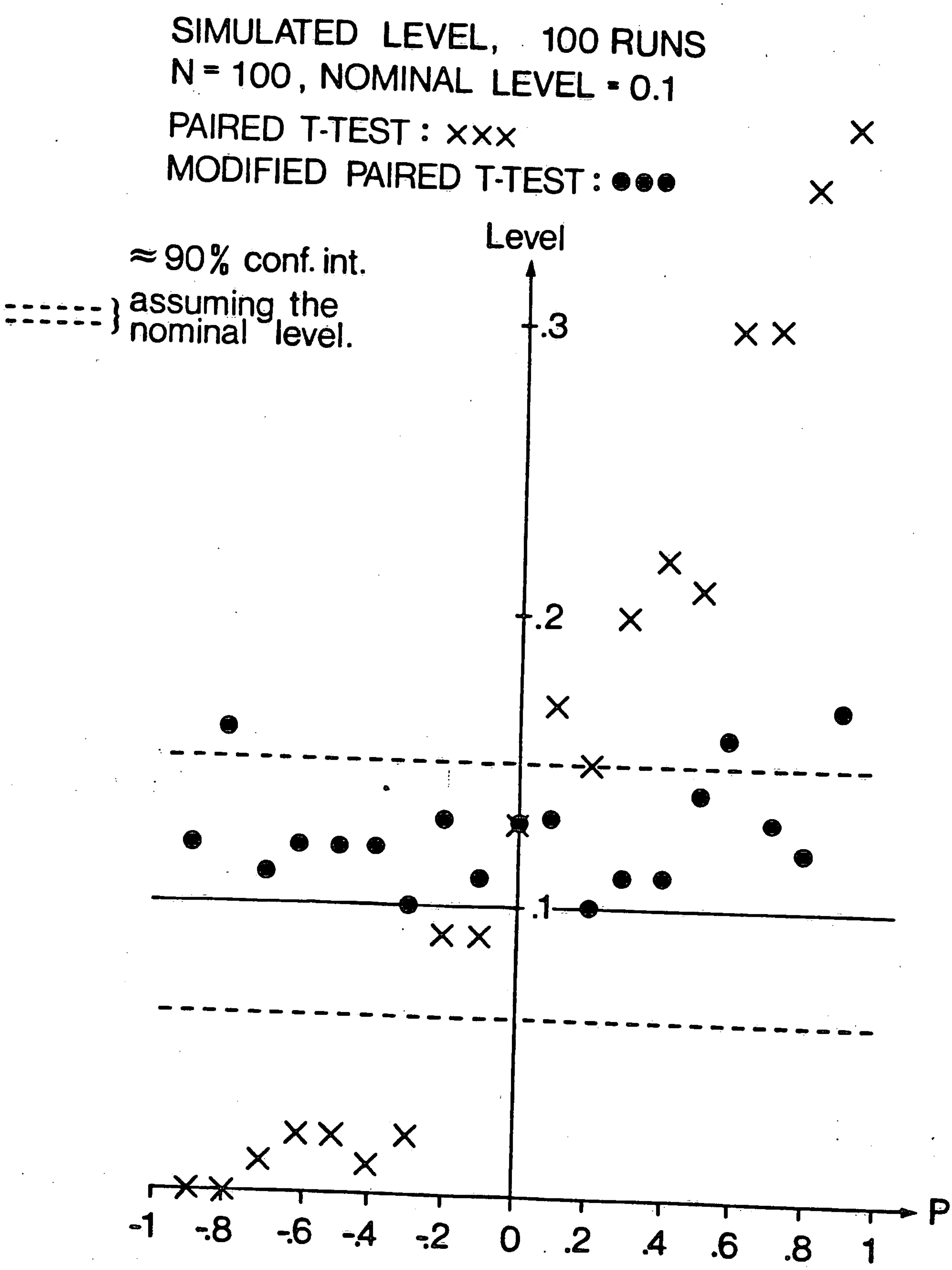

SIMULATED LEVEL 100 RUNS
N = 200, NOMINAL LEVEL = 0.1
SIGN TEST : XXX
MODIFIED SIGN TEST : ●●●
≈ 90% conf. int.
assuming the nominal level.
Level
.3
.2
.1
P
-1
-.8
-.6
-.4
-.2
0
.2
.4
.6
.8
1

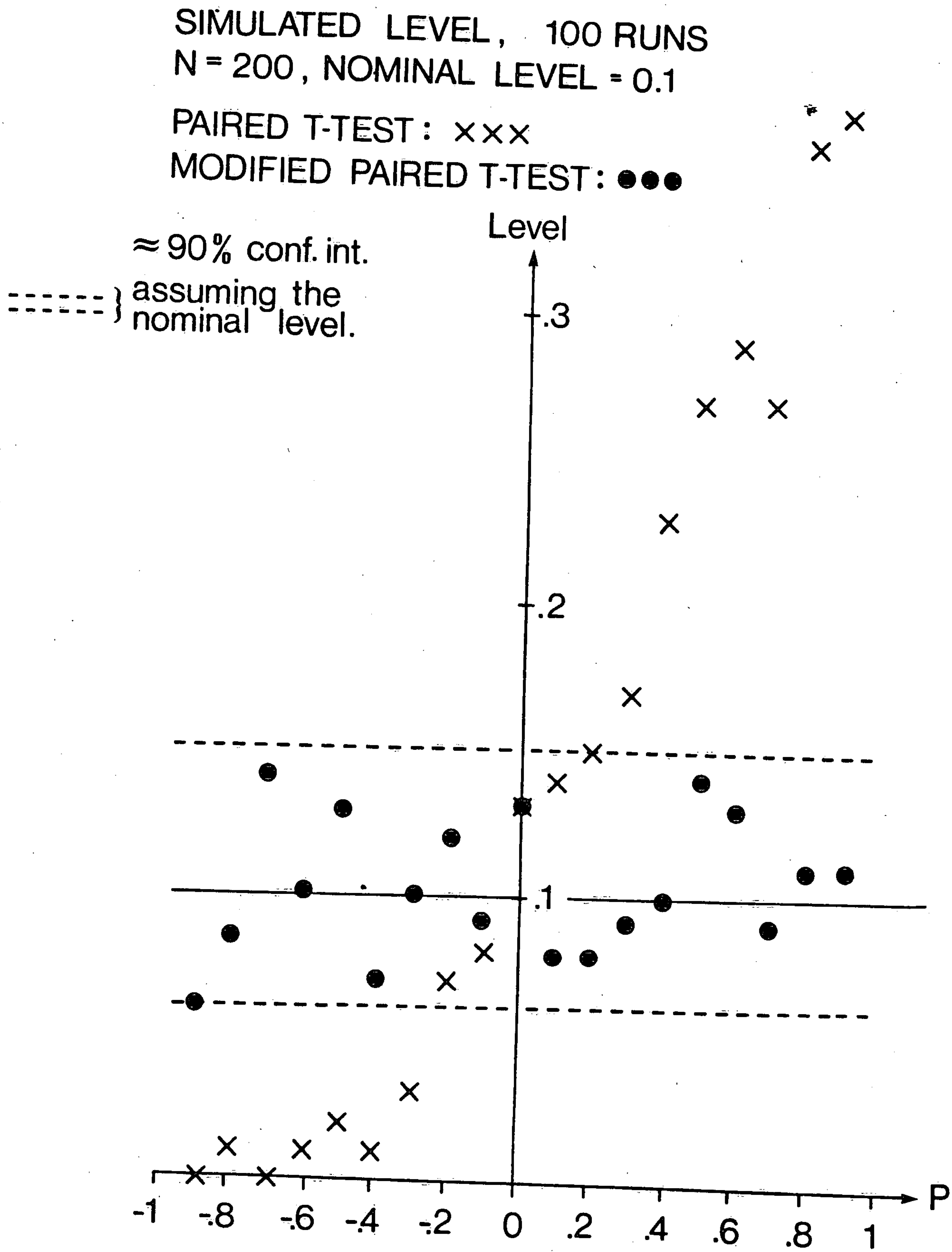

SIMULATED LEVEL, 100 RUNS
N = 100, NOMINAL LEVEL = 0.05
PAIRED T-TEST: × × ×
MODIFIED PAIRED T-TEST: ● ● ●
≈ 90% conf. int. assuming the nominal level.
Level
.3
.2
.1
P
-1
-.8
-.6
-.4
-.2
0
.2
.4
.6
.8
1


SIMULATED LEVEL, 100 RUNS
N = 200, NOMINAL LEVEL = 0.01
PAIRED T-TEST: ×××
MODIFIED PAIRED T-TEST: ●●●
≈ 90% conf. int. assuming the nominal level.
Level
.3
.2
.1
P
-1 -.8 -.6 -.4 -.2 0 .2 .4 .6 .8 1

SIMULATED LEVEL, 100 RUNS
N = 100, NOMINAL LEVEL = 0.01
PAIRED T-TEST: ×××
MODIFIED PAIRED T-TEST: ●●●




SIMULATED LEVEL, 100 RUNS
N = 100, NOMINAL LEVEL = 0.1
PAIRED T-TEST : xxx
MODIFIED PAIRED T-TEST : ●●●
≈ 90% conf. int. assuming the nominal level.
Level
.3
.2
.1
P
-1
-.8
-.6
-.4
-.2
0
.2
.4
.6
.8
1


SIMULATED LEVEL, 100 RUNS
N = 200, NOMINAL LEVEL = 0.1
PAIRED T-TEST: ×××
MODIFIED PAIRED T-TEST: ●●●
≈ 90% conf. int.
assuming the nominal level.
Level
.3
.2
.1
P
-1
-.8
-.6
-.4
-.2
0
.2
.4
.6
.8
1