IS THE DIFFERENCE SIGNIFICANT? A NOMOGRAPHIC GUIDE TO INTERPRETING STATISTICS IN THE IRS'S TAXPAYER USAGE STUDY

Paul E. Grayson, Internal Revenue Service

INTRODUCTION

This report presents a sequence of three charts which lead to the determination of whether the difference between two percentages is statistically significant. The charts are designed for use with the Internal Revenue Service's annual Taxpayer Usage Study (TPUS). Each year the Taxpayer Usage Studies provide, by about mid-year, early indications of how taxpayers used the current-year individual income tax returns (Forms 1040, 1040A and now 1040EZ). The studies are based on samples of 6,000 to 7,000 tax returns filed between January 1 and about May 1 each year.

The sample is systematically selected from the universe of all mail that passes through automatic envelope-opening and counting machines at the ten Internal Revenue Service Centers. (A minor proportion of mail is handled by alternative means and is also subject to sampling.) Starting with a random number (C_i), assigned to each Service Center separately, machine operators withdraw a mail piece - including other than individual income tax returns - whenever the counter registers C_i+nk, where k equals the sampling interval. The designated interval was 15,000 for sampling returns filed during 1983 (for Tax Year 1982). The data base for the TPUS report consists of those selected mail pieces that contain individual income tax returns filed on the appropriate year's form (restricted to the Forms 1040, 1040A and 1040EZ). Sample data from each of the three return forms are weighted by the reciprocals of the three respective effective sampling rates. (In 1983, for example, 1,163 Form 1040EZ sample returns were selected from a population of 14,785,000 1040EZ's filed, for an effective sampling rate of 1/12,713.)

Tables in the TPUS reports present frequency information, in the form of estimated population counts and percentages, on the presence and nature of entries on various lines of the tax return, usage and number of attached forms, and other characteristics of the returns filed. The tables generally employ standard formats and cross-tabulation variables (like adjusted gross income and marital status) to permit tracking of year-to-year trends and to allow comparisons among subsets in the population of returns. An example of a typical table is provided in the following excerpt [1] from the most recent (Tax Year 1982) TPUS report. (See below)

USE OF NOMOGRAPHS

Charts A, B, and C provide approximate answers to users of the Tax Year 1982 Taxpayer Usage Study who wish to judge whether the difference between two percentage estimates is statistically significant. For example, according to the data provided in the excerpt, 25.3 percent of the Form 1040 returns in the adjusted gross income (AGI) class under $5,000 had an attached Schedule C (for sole proprietorship businesses or professions). The comparable figure in the next highest AGI class, $5,000 under $10,000, was 20.8 percent. Were the two percentages significantly different? The charts were specifically designed to test this type of situation; however, they may also be used to test year-to-year comparisons.

The conventional approach to such questions is usually based on a significance test such as the t test which involves the computation of

Table 9.--All Returns: Percentage Distribution of Selected Forms and Schedules, Classified by Size of Adjusted Gross Income, Tax Year 1982

Schedule or Form	Type of attachment	Total	Percentage of returns by size of adjusted gross income						
			Under $5,000	$5,000 under $10,000	$10,000 under $15,000	$15,000 under $20,000	$20,000 under $30,000	$30,000 under $50,000	$50,000 and over
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All Form 1040 returns filed.....	54,687	4,445	6,836	6,794	6,850	12,561	13,503	3,699
	Percent.....	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
A	Itemized Deductions.....	57.4	13.6	23.4	37.3	47.0	66.1	86.0	95.4
B	Interest and Dividend Income..	46.6	43.7	47.5	45.7	40.7	39.4	49.0	77.2
C	Profit or (Loss) From Business or Profession.....	16.4	25.3	20.8	15.1	13.8	14.9	13.8	19.8
D	[illegible]	[illegible]	9.8	[illegible]	[illegible]	[illegible]	11.2	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	Re[illegible] Indiv[illegible]etire[illegible] ment [illegible]rangement [illegible]xes.....	0.2	-	[illegible]	-	0.2	[illegible]	0.3	0.4
5695	Residential Energy Credit.....	5.2	0.6	2.6	3.9	3.7	6.4	7.6	7.2
5884	Jobs Credit.....	0.1	0.6	-	0.2	0.2	-	-	0.4
6249/ 6249A	Computation of Overpaid Windfall Profit Tax.....	0.3	-	0.2	0.2	0.4	0.3	0.1	1.1
6251	Alternative Minimum Tax Computation.....	0.6	-	-	0.2	0.2	0.1	1.1	4.2
6252	Computation of Installment Sale Income.....	1.3	-	0.8	0.6	0.8	1.0	2.1	3.4

Note: Numbers of returns in thousands.

$$t = \frac{p_1 - p_2}{\sqrt{P\,Q\,(\frac{1}{n_1} + \frac{1}{n_2})}}$$

where

n_1, n_2 = the size of the samples selected from two populations;

p_1, p_2 = percentages of n_1 and n_2, respectively, with a given characteristic;

P = average of p_1 and p_2; and

Q = 100 - P.

Since the test requires considerable arithmetic and, further, the values of n_1 are not easily available to the reader, the present nomographic approach has been developed that sidesteps the latter issue and virtually eliminates the former. Use of the nomograms is illustrated by applying them to the data in the question stated above. One starts by - -

Using Chart A:

(a) The TY 1982 TPUS shows the following:

Item	Adjusted Gross Income Class	
	Under $5,000	$5,000 under $10,000
Total Number (N_i), Form 1040 returns (thousands)	4,445	6,836
Percent (p_i) with Schedule C . .	25.3%	20.8%

Average percent (rounded), P = 23
Difference, $p_1 - p_2$ = 4.5

(b) On Chart A, lay a straight-edge connecting (0,0) to graduated scale marking at P = 23.

(c) 4.5 on the vertical scale intersects the straight-edge at

PT = 4.8 on the horizontal scale. (This is the "Preliminary t-value".)

Using Chart B:

(a) The larger of the two classes is 6.8 million; the ratio of the smaller to the larger is in the neighborhood of 0.65.

(b) 6.8 million on the vertical scale intersects an interpolated curve for .65 at

SF = 0.34 on the horizontal scale. (This is the "Size Factor".)

Using Chart C:

(a) The point where PT = 4.8 and SF = 0.34 intersect lies below (to the left of) the curve bounding the "Significance Zone."

(b) Thus the difference between the two percentages--25.3 and 20.8--is not significant at the conventional 95 percent confidence level (2 standard errors).

ADDITIONAL GUIDES TO USE OF THE CHARTS

When the average percentage, P, is greater than 50, use 100 - P. Thus, if the average percentage is 80, use 20 on the graduated scale. In Chart A, one notes that 50 is the maximum value provided for P. This is because the slope of the P lines is a function of P Q, or P (100 - P). Thus, using 100 - P yields the same answer (i.e., (80)(20) = (20)(80)).

When one of the groups being compared is one-third or less of the other group in the comparison, the procedure will have more accuracy if the average percentage (P) is weighted by the size of the groups. In the example, the AGI class with the smaller population (i.e., under $5,000) was 0.65 of the larger class. In this case, the weighted average resulted in the same (rounded) value of P.

The charts can also be used for--and between--prior years. For those years, when TPUS sampling rates were higher, the charts will give conservative statements of significance. That is, in some "borderline" cases from tax years prior to 1981, the charts could show "not significant," while computation of the "t" statistic could indicate a significant difference.

ASSUMPTIONS AND METHODOLOGY

As already stated, the test statistic, t, is employed in the charts as the basis of the test of significance. Use of the t statistic assumes independence between the values being compared, a condition that is satisfied for values taken from different years. While independence is not completely true for values for classes taken from the same sample, the actual situation is a close approximation to the theoretical. Since TPUS sample selection is systematic, with multiple random starts of a randomly ordered population, we can assume unrestricted random sampling. (This assumption is not generally applicable to the samples on which Internal Revenue's "Statistics of Income" series are based.) And, in view of the size of the sample, normality is also assumed.

Because the charts were first worked out for the Tax Year 1980 Taxpayer Usage Study, the charts presented here are, in effect, a second generation, required when the designated sampling rate was reduced in 1982. To minimize the work of updating, this meant a redrafting of Chart C only.

Chart A provides an evaluation of the "preliminary t-value,"

$$\frac{p_1 - p_2}{\sqrt{p\,q\left(\frac{1}{\bar{n}} + \frac{1}{\bar{n}}\right)}} = 45.28\,\frac{p_1 - p_2}{\sqrt{p\,q}}$$

Chart A assumes that each of the two percentages being compared was based on a sample of equal size--in the case, 4,100 returns. In 1981 when the TPUS sample was compared with population counts of returns received at the IRS service centers over the same four-month period, it was observed--as in previous years--that Form 1040 returns were sampled at a lower effective rate than the Form 1040A returns. Since a lower sampling rate leads to a lower--and hence more conservative--t value (note that the published statistics are in terms of population estimates), Chart B was based on r_i, the Form 1040 rate of 1/12,100. Chart A was, therefore, based on the sample of 4,100 Form 1040 returns (averaged for 1980 and 1981), as if it had constituted the entire sample.

Chart B takes into account unequal sample--and hence population--sizes, as in the example already described, where the smaller population was 0.65 of the larger. This process is accomplished by evaluating the "size factor."

$$SF = \frac{\sqrt{\frac{2}{n}}}{\sqrt{\frac{1}{r}}\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} = \frac{.0002008}{\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

since we already have assumed n = 4100 and r = 1/12,100.

The curve in Chart C is the locus of

$$1.96 = 1.08\ PT \times SF,$$

the two-sigma level of significance. In the Tax Year 1980 version, the product, PT(SF), was an estimate of the t statistic, and the curve represented 1.96 PT(SF). Of numerous tests, two will illustrate how well the estimate compared with the computed statistic.

(1) 35.7 percent of 52.8 million Forms 1040 (1980) contributed to the presidential campaign fund versus 32.1 percent of 36.3 million Forms 1040A (1980):

estimate of t = 3.1
computed t = 3.25;
(the difference in percentages was significant on both bases).

(2) 43.3 percent of 11.6 million returns comprising the adjusted gross income class $15,000 under $20,000 (1980) bore the signature of a commercial preparer versus 41.4 percent of 11.2 million in that class (1979) with the signature:

estimate of t = 0.85
computed t = 0.83
(the difference in percentages was not significant on both bases).

For Tax Years 1981 and 1982, as previously suggested, the designated sampling rate was lowered from the 1980 level. For 1982, the effective sampling rate for Form 1040 returns was 1/14,066, the lowest of the rates for the three types of returns filed that year. Incorporating in the charts this decrease from 1/12,100 could have been accomplished by redrawing Chart B (and leaving Chart C unchanged) or vice versa. To facilitate updating the charts, the latter course was followed, and the curve in Chart C has been replotted to reflect the relation, r_{i+1}/r_i = 14,066/12,100 = 1.08. (r_i is the effective sampling rate in a year, and r_{i+1} is the rate in the subsequent year.) The curve was moved upward and to the right. If future effective sampling rates in the Taxpayer Usage Study are further decreased, the Chart C curve will move further in the same direction. If sampling rates increase, the curve will drop and move to the left.

ACKNOWLEDGEMENTS

The author is indebted to Dr. Shien Perng of the Statistical Methods Group, Research Division (Internal Revenue Service) for providing the initial insights into the mathematics underlying the charts after they were constructed. Appreciation also goes to Esther Johnson for typing this paper.

REFERENCES

[1] Riley, Dorothea, "Individual Income Tax Returns: Selected Characteristics from the 1982 Taxpayer Usage Study," SOI Bulletin 3(1) Summer 1983, pp. 43 - 56. Internal Revenue Service, Washington, DC. Also prepared as Document 6528, Statistics of Income Division, Internal Revenue Service, 1983.

Chart A

Enter Chart With Values for (p1–p2) and P To Determine PT

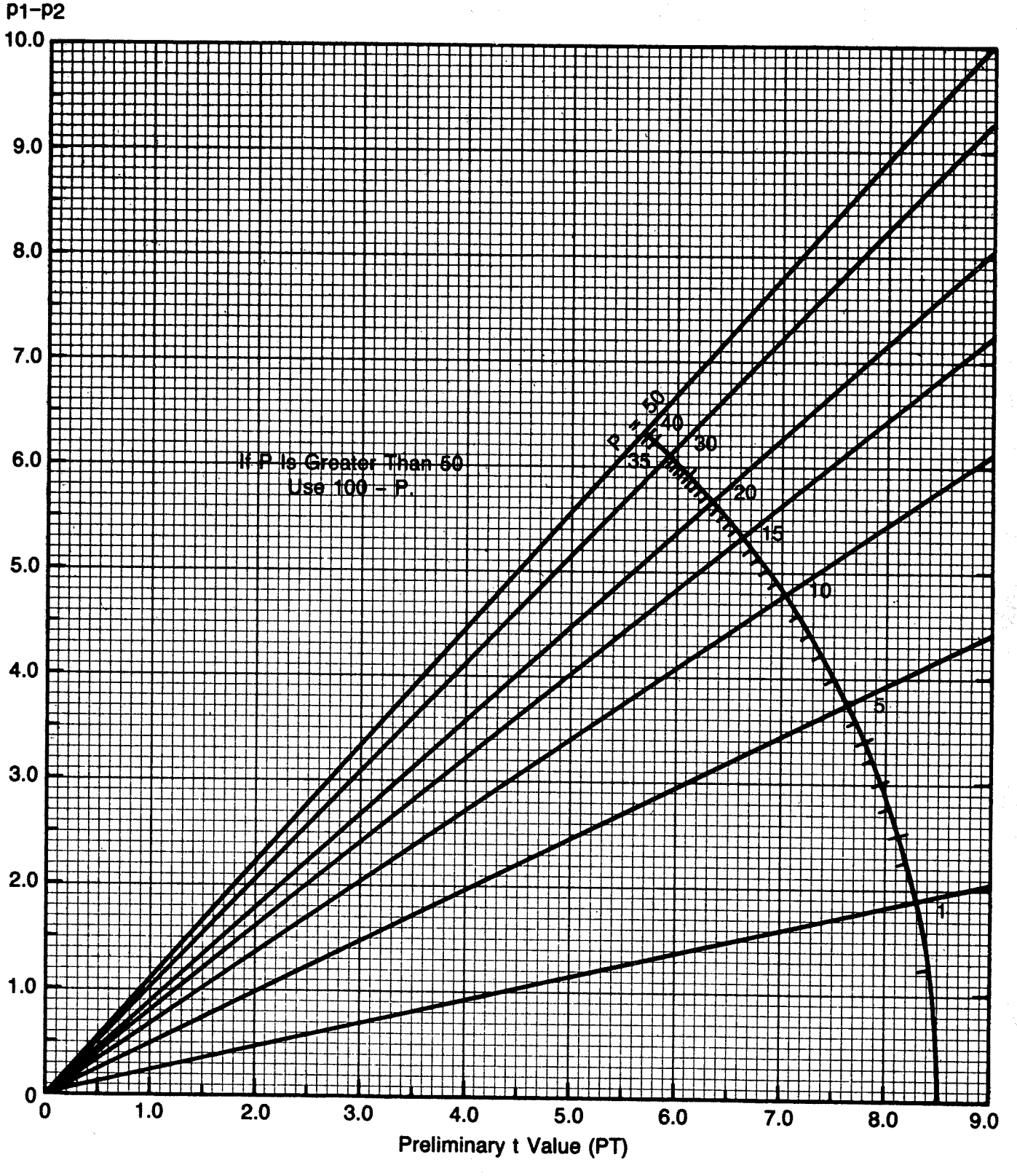

Chart B
Enter Chart With Size of Greater Population and Ratio of Smaller Population To Greater To Determine Size Factor.

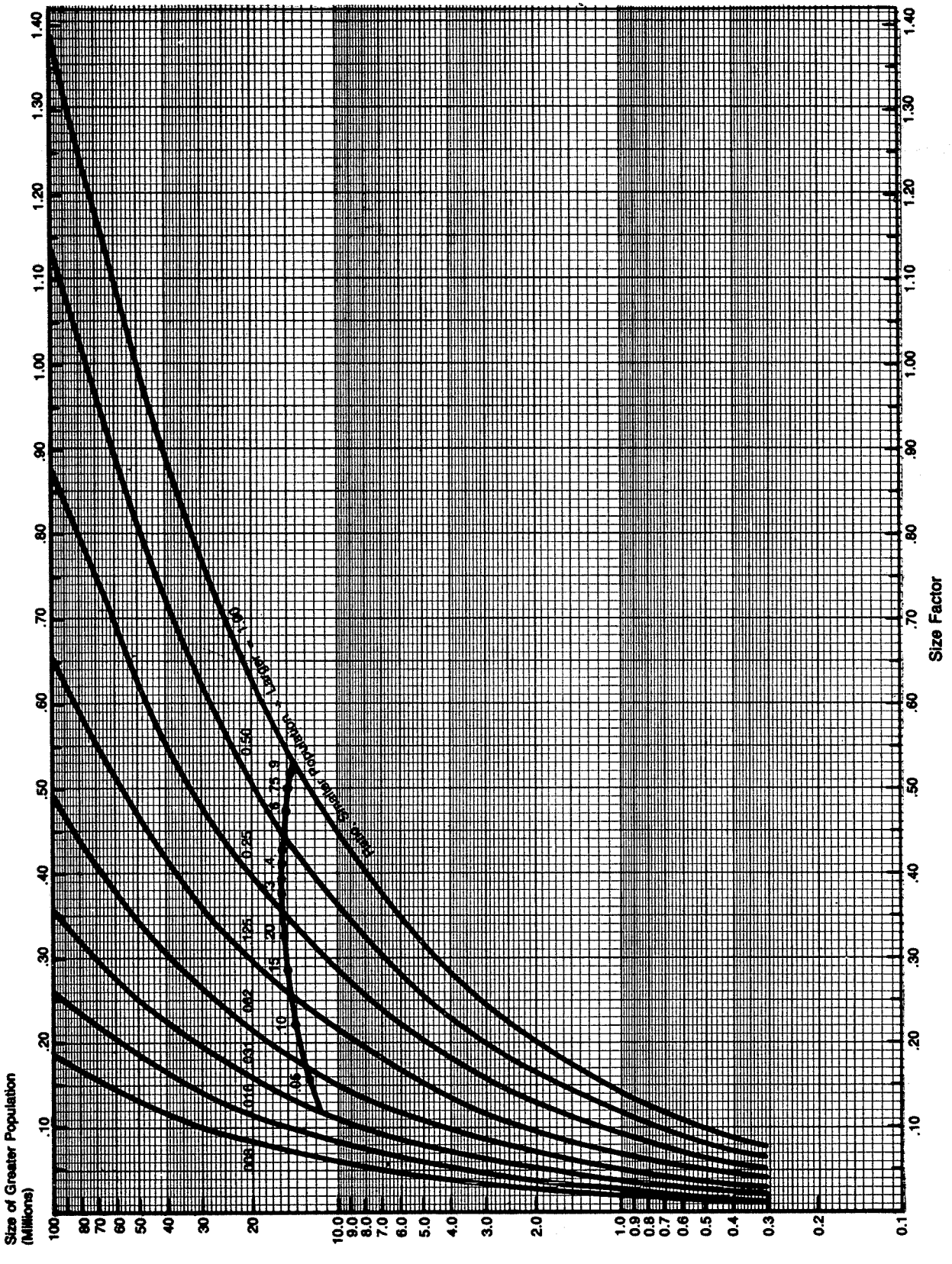

Chart C

Enter Chart With Values for PT (From Chart A) and Size Factor (From Chart B) To Determine if Answer Falls in the Significance Zone.