

AMOSTRAGEM POR CONGLOMERADO APLICADA À AUDITORIA E CONTABILIDADE

Edmundo Eboli Bonini

Professor Livre-Docente do Departamento de Contabilidade e Atuária da FEA-USP

INTRODUÇÃO

A amostragem estratificada é usada com o objetivo inicial de realizar um ganho na eficiência da amostragem, isto é, reduzir o erro da amostragem para um dado tamanho da amostra ou, reciprocamente, reduzir o tamanho da amostra para realizar uma confiabilidade da amostra.

Outras razões podem condicionar o uso de outras formas de amostra. Às vezes o levantamento de documentos individuais tirados da população demandam um trabalho árduo e extensivo (custoso). Um tipo de amostragem que não levante os documentos individuais, mas em

grupos (conglomerados), pode ser preferível em algumas circunstâncias.

Em certas situações, a localização e seleção das unidades individuais escolhidas através dos números aleatórios pode consumir muito tempo. É possível que em alguns casos o tempo envolvido na localização e seleção de unidades amostrais seja mais complicado do que o próprio exame dos documentos, levando evidentemente mais tempo na seleção do que no teste de auditoria. Assim, suponhamos que o nosso levantamento seja de 500 documentos e, portanto, tenhamos de selecionar 500 dígitos randômicos ou aleatórios, a fim de analisar os 500 documentos; ao invés disto, podere-

41

mos selecionar 100 números aleatórios e tomar cinco unidades a partir de cada número aleatório, ou talvez 50 números aleatórios, com 10 documentos a partir de cada número aleatório. A amostragem por Conglomerado (Grupo, Cacho) é aquela que consiste de um número de grupos de unidades contíguas a partir da seleção aleatória. Este método reduz sensivelmente o tempo envolvido na preparação e seleção das unidades amostrais.

Os conglomerados (grupos, cachos) podem ser obtidos utilizando-se números aleatórios ou amostragem sistemática como mecanismo de amostragem (levantamento dos elementos). É recomendável que tenhamos pelo menos 20 conglomerados.

Para estimar o valor do desvio padrão populacional (σ) através da amplitude média (\bar{R}) ou seja

$$\sigma = \frac{\bar{R}}{*d_2}, \text{ deveremos ter}$$

normalmente pelo menos 8 conglomerado ($K = 8$), com número de elementos em cada conglomerado (n') variando entre 2 e 12. Normalmente $n' = 5$. Outro critério será calcular através de

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{k}}$$

O erro da amostragem é obtido através da multiplicação do desvio padrão pelo fator ou fatores selecionados em função do nível de confiança, tamanho da população (partida) e tamanho da amostra desejada.

Quando utilizamos menos de 30 conglomerados, é necessário utilizar fatores relacionados com as pequenas amostras.

DIMENSIONAMENTO – TAMANHO DO CONGLOMERADO – NÚMERO DE CONGLOMERADOS (Grupos)

Variáveis – Valor Monetário

Para determinarmos o número de documentos a serem analisados para a amostragem irrestrita aleatória, devemos ter uma amostra-piloto (no mínimo 30 documentos) a fim de estimarmos o desvio-padrão. A fórmula para a amostragem aleatória irrestrita é dada por:

$$n = \frac{N}{\left(\frac{e}{\sigma}\right)^2 \frac{1}{Z^2_{\alpha/2}} (N-1) + 1},$$

onde n é o tamanho da amostra, N = tamanho do Universo, σ = desvio-padrão da população, e = precisão ou margem de confiança; α = nível de significância, $1 - \alpha$ = nível de confiança; normalmente o nível de confiança é de 90%; 95%; 99%; 99,9%; $Z_{\alpha/2}$ é o valor da variável

* d_2 = Fator que depende do número de elementos do grupo.

normal reduzida que depende do nível de significância α . Assim, para:

α :	0,10	0,05	0,01	0,001
$\frac{\alpha}{2}$:	0,05	0,025	0,005	0,0005
$Z_{\alpha/2}$:	1,645	1,96	2,58	3,30

Seja: $N = 5.000$; $G = 32$; $e = 8$, ou seja, 25% de G , isto é, $\frac{e}{G} = 0,25$; $1 - \alpha = 0,99$ ou 99%, qual o número de documentos a ser auditado?

$$n = \frac{5.000}{\left(\frac{8}{32}\right)^2 \frac{1}{2,58^2} 4.999+1} \cong 105$$

Se dividirmos o número de documentos a ser auditado pelo número de elementos em cada conglomerado (pelo menos 5), teremos o número de grupos.

Assim, por exemplo, se o número de elementos a ser levantado for 100 e cada grupo for composto de 5 documentos, deveremos ter então $\frac{100}{5} = 20$ conglomerados ($K =$ número de conglomerados), $K \geq 20$.

É importante notar que em muitas situações a amostragem por conglomerado pode ser menos eficiente do que a amostragem irrestrita aleatória (ou seja, a amostragem em que os elementos são levantados sem nenhum critério de estratificação ou manipulação, a fim de obtermos inferências sobre a população).

Depois de montar os conglomerados, devemos calcular valores característicos, tais como média, desvio-padrão, etc., com o objetivo de estimar o comportamento da partida em função do nível de confiança, número de conglomerados, número de elementos do lote, etc.

A média de cada conglomerado é \bar{X}_i ; a média total será:
$$\bar{X} = \frac{\sum_{i=1}^K \bar{X}_i}{K}$$
 onde K é o número de grupos (conglomerado).

A margem de confiança ou erro da amostragem é dada por:

$$\pm G \frac{t_{\alpha/2; K-1}}{K-1} \sqrt{\frac{1-K}{N'}}$$

onde $G = \sqrt{\frac{\sum (\bar{x}_i - \bar{x})^2}{K}}$; \bar{x} é a média

de cada conglomerado, \bar{x} é a média total, ou seja, de todos os conglomerados, K é o número de conglomerados, N' é o número de conglomerados na população, ou seja, $N' = \frac{N}{n'}$, sendo n' o número de elementos em cada conglomerado; $t_{\alpha/2; K-1}$ é o valor da variável t de Student, que depende de α e $K-1$. Assim, se tivermos $N = 5.000$; $n' = 5$; $K = 20$; $N' = 1.000$; $G = 31,61$; $1 - \alpha = 0,99 \Rightarrow \alpha = 0,01$, teremos para a margem de confiança ou

erro da amostragem o valor:

$$\begin{aligned} & \pm 31,61 \cdot \frac{t_{0,005; 19}}{\sqrt{20-1}} \sqrt{1 - \frac{20}{1000}} = \\ & = \pm 31,61 \cdot \frac{2,8609}{\sqrt{20-1}} \cdot \sqrt{1 - 0,02} = \\ & = \pm 31,61 \cdot 0,6561 \cdot 0,9899 = \\ & \pm 20,53; \text{ o valor } 31,61 \text{ é o desvio-} \\ & \text{padrão de cada documento, dando} \\ & \text{como erro da amostragem o valor} \\ & \pm 20,53 \text{ (por documento); e para o} \\ & \text{total dos documento será: } \pm 20,53 \cdot \\ & 5.000 = \pm 102.650,00//. \end{aligned}$$

Atributos – Porcentagens

Para determinar o número de documentos a ser analisados para a amostragem irrestrita aleatória, deveremos ter uma amostra-piloto (no mínimo 30 documentos) a fim de estimarmos a porcentagem de erros ou taxa de ocorrência. A fórmula para determinarmos o número de documentos a ser levantado através de amostragem aleatória irrestrita é dada por:

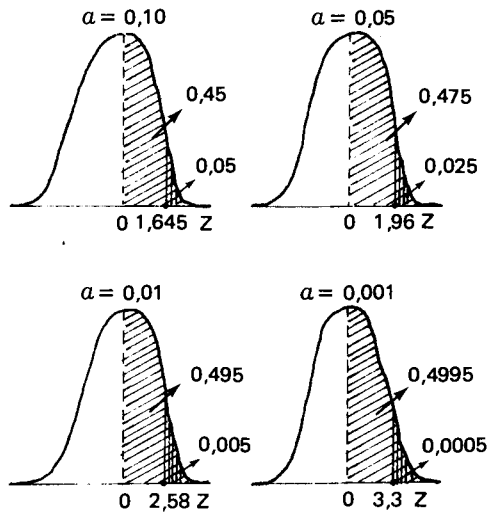
$$n = \frac{Z^2_{\alpha/2} p' q' N'}{e^2 (N-1) + Z^2_{\alpha/2} p' q'}$$

onde n é o tamanho da amostra, N = tamanho da população, p' é a proporção da amostra, dada em decimal, q' = 1 - p', q' é a proporção de insucessos da amostra, α = nível de significância, 1 - α = nível de

confiança, normalmente o nível de confiança é de 90%; 95%; 99% e 99,9% e Z_{α/2} é o valor da variável normal reduzida que depende do nível de significância α :

Os valores de Z_{α/2} são obtidos através de curva normal, sendo que Z_{α/2} são as abscissas que correspondem às áreas $\frac{1-\alpha}{2}$

Assim, teremos:



e = erro da amostragem, ou margem de confiança ou tolerância ou precisão.

Seja: N = 4.000; p' = 0,10; e = 0,04; 1 - α = 0,95 ou 95%, qual o número de documentos mínimo a ser auditado?

* t_{0,005; 19} = 2,8609, vide tabela t de Student - pg. 434 - Bonini, E.E. - Bonini, S.E. - Estatística - Teoria e Exercício.

$$n = \frac{1,96^2 \cdot 0,10 \cdot 0,90 \cdot 4.000}{0,04^2 (4.000-1) + 1,96^2 \cdot 0,1 \cdot 0,9} \cong 205$$

Se dividirmos o número de documentos a ser levantado (auditado) pelo número de elementos de cada conglomerado, teremos o número de conglomerados. Assim, por exemplo, se o número de elementos a ser levantado for 420 (n) e cada conglomerado for composto de 20 documentos, deveremos ter então $\frac{420}{20} = 21$ (K) conglomerados.

Depois de montarmos os conglomerados teremos que calcular os p_i' de cada conglomerado, em seguida \bar{p}'_i (média das taxas de ocorrência dos conglomerados), desvio-padrão de p_i' , ou seja;

$$G(p') = \sqrt{\frac{\sum (p_i' - \bar{p}'_i)^2}{K}}$$

A margem de confiança ou erro da amostragem é dada por:

$$\pm G(p') \frac{t_{\alpha/2; K-1}}{\sqrt{K-1}} \sqrt{1 - \frac{K}{N'}}$$

onde: K é o número de conglomerado, N' é o número de conglomerados na população, ou seja, $N' = \frac{N}{n'}$,

onde n' é o número de elementos em cada conglomerado; $t_{\alpha/2; K-1}$ é o valor da variável t de Student que depende de α e K.

Assim, se tivermos $N = 4.000$, $K = 21$, $n' = 20$, $N' = 200$, $1 - \alpha = 0,95 \Rightarrow \alpha = 0,05 \Rightarrow t_{0,0025; 20} = 2,0860$ (Bonini, E.E. — Bonini, S. E. — Estatística — Teoria e Exercícios), $G(p') = 6,13$, teremos para o erro da amostragem, o valor:

$$\pm 6,13 \frac{2,0860}{\sqrt{21-1}} \sqrt{1 - \frac{21}{200}} = \pm 6,13 \cdot 0,4664 \cdot 0,9460 = \pm 2,97$$

Se a média das ocorrências foi, por exemplo, de 10%, então os limites de confiança ao nível de 95% serão:

$$10\% \pm 2,97\% \text{ ou } 7,02\% \text{ até } 12,98\%$$

AVALIAÇÃO DA CONFIABILIDADE DE CONGLOMERADOS

Variáveis— Valor Monetário

Nosso objetivo é determinar a média ou o total do valor monetário de uma série de documentos através de uma parcela destes documentos, ou seja, a projeção através dos dados amostrais.

Para estabelecer o erro da amostragem quando utilizamos os conglomerados, deveremos calcular o total do conglomerado para estimar a média total do universo onde estes conglomerados foram tirados ou para estimar o valor total do universo, efetuando-se a multiplicação do valor médio por unidade amostral pelo número de itens no universo,

sendo que o primeiro passo consiste em calcular a média de cada conglomerado. O número de documentos de cada conglomerado deverá ser constante. A média total será a média de todos os conglomerados. Uma vez selecionado o elemento inicial em cada conglomerado através de uma tabela de dígitos aleatórios, os restantes documentos pertencentes ao conglomerado serão escolhidos em seqüência até preencher o número de documentos do conglomerado.

estimar o valor total de um certo grupo de 5.000 documentos. Isto será feito através da multiplicação do valor médio estimado por documento pelo número de documentos do universo (5.000). Iremos utilizar a **amostragem por conglomerado (grupos)** e decidimos tirar amostras de tamanho de 5 num total de 100 documentos, ou seja, 20 conglomerados. A partir de uma tabela de números aleatórios selecionamos 20 números, para termos os números iniciais de cada conglomerado, e admitimos a distribuição abaixo:

Vejamos uma aplicação: Desejamos

Conglomerado Nº	Nº Aleat.	Valor	Congl. Nº	Nº Aleat.	Valor	Congl. Nº	Nº Aleat.	Valor	Congl. Nº	Nº Aleat.	Valor
1	0346	65,02	2	1216	173,20	3	1436	198,30	4	0546	155,20
		93,54			184,70			226,70			162,50
		125,10			47,53			136,00			88,38
		112,20			83,80			283,70			113,30
		91,72			91,62			81,95			97,46
5	2014	178,00	6	3568	89,93	7	3236	29,38	8	2291	99,83
		314,70			261,00			107,20			190,10
		53,67			64,23			77,30			201,00
		36,18			87,62			83,04			183,10
		116,90			226,50			180,40			217,50
9	0508	191,30	10	1258	66,60	11	1420	101,80	12	3468	67,00
		27,43			178,10			48,35			67,53
		88,09			253,70			90,40			91,49
		96,80			152,70			137,30			209,80
		57,00			82,08			97,60			89,57
13	0199	93,43	14	0163	42,38	15	0137	56,47	16	4645	54,62
		116,70			14,99			92,54			69,05
		109,30			99,83			135,60			119,20
		168,40			115,70			40,83			85,35
		66,01			80,17			72,41			98,51
17	3684	90,07	18	2411	156,20	19	2550	80,10	20	4851	64,66
		212,70			73,67			93,50			163,50
		69,11			94,92			56,67			131,90
		198,10			109,80			224,90			229,00
		242,90			106,10			97,85			79,70

As médias dos conglomerados são: (\bar{X}_i)

Congl. Nº	Média	Congl. Nº	Média	Congl. Nº	Média
1	97,92	2	116,17	3	185,33
4	123,37	5	139,89	6	145,86
7	95,46	8	178,31	9	88,32
10	146,64	11	94,98	12	104,96
13	110,77	14	70,61	15	79,57
16	85,35	17	162,58	18	108,14
19	110,60	20	133,75		

A média total das médias dos conglomerados é $\bar{X} = \frac{\sum_{i=1}^{20} \bar{x}_i}{20} = 118,93$,

que representa a média por documento.

Multiplicando-se por 5.000 (número de documentos no Universo) a \bar{X} teremos uma estimativa para o total de todos os documentos, ou seja:

$$118,93 \cdot 5.000 = 594.650,00$$

O intervalo de confiança para a média, considerando-se uma pequena amostra e uma população finita, é dada por:

$$P\left\{\bar{X} - G \frac{t_{\alpha/2; K-1}}{\sqrt{K-1}} \sqrt{\frac{1-K}{N'}}\right\}$$

$$\leq \mu \leq \bar{X} + G \frac{t_{\alpha/2; K-1}}{\sqrt{K-1}} \sqrt{\frac{1-K}{N'}} = 1-\alpha$$

onde \bar{X} é a média das médias dos conglomerados, G é o desvio-padrão dos conglomerados, K é o número de conglomerados; $N' = \frac{N}{n'}$, onde

N é o tamanho do Universo, n' tamanho de cada conglomerado, $(1-\alpha)$ é o nível de confiança, $t_{\alpha/2; K-1}$ valor da variável t de Student.

Na expressão acima teremos que calcular o valor de G , que é dada por:

$$G = \sqrt{\frac{\sum (\bar{x}_i - \bar{X})^2}{K}}$$

e constituindo a seguinte tabela:

Amostragem por Conglomerado Aplicada à Auditoria e Contabilidade

Conglomerado Nº	Média (\bar{x}_i)	$\bar{x}_i - \bar{\bar{x}}$	$(\bar{x}_i - \bar{\bar{x}})^2$
1	97,92	- 21,01	441,4201
2	116,17	- 2,76	7,6176
3	185,33	66,40	4408,9600
4	123,37	4,40	19,7136
5	139,89	20,96	439,3216
6	145,86	26,93	725,2249
7	95,46	- 23,47	550,8409
8	178,31	59,38	3525,9844
9	88,32	- 30,61	936,9721
10	146,64	27,71	767,8441
11	94,98	- 23,95	573,6025
12	104,96	- 13,97	195,1609
13	110,77	- 8,16	66,5856
14	70,61	- 48,32	2334,8224
15	79,57	- 39,36	1549,2096
16	85,35	- 33,58	1127,6164
17	162,58	43,65	1905,3225
18	108,14	- 10,79	116,4241
19	110,60	- 8,33	69,3889
20	133,75	14,82	219,6324
			19981,6646

Então,
$$= \sqrt{\frac{19981,6646}{20}} = K = 20 \Rightarrow K-1 = 19; \bar{\bar{x}} = 118,93;$$

$$\sqrt{999,0832} = 31,61, \text{ sendo } N' = \frac{N}{n'} = \frac{5.000}{5} = 1.000; \text{ o valor}$$

de $t_{\alpha/2; k-1} = t_{0,005; 19} = 2,8609$ (Tabela t de Student - pg. 434 - Bonini, E.E. e Bonini, S.E., 1976) considerando o nível de 99% de confiança e 19 graus de liberdade. Substituindo os valor acima no intervalo de confiança teremos:

$$P \left\{ 118,53 - 31,61 \frac{2,8609}{\sqrt{20-1}} \sqrt{\frac{1-20}{1000}} \leq \mu \leq 118,53 + 31,61 \sqrt{\frac{2,8609}{20-1}} \sqrt{1 - \frac{20}{1000}} \right\} = 1 - 0,01$$

$$P \{ 118,53 - 31,61 \cdot 0,6561 \cdot 0,9899 \leq \mu \leq 118,53 + 31,61 \cdot 0,6561 \cdot 0,9899 \} = 0,99$$

$$P \{ 118,53 - 20,53 \leq \mu \leq 118,53 + 20,53 \} = 0,99$$

$$P \{ 98,00 \leq \mu \leq 139,06 \} = 0,99$$

Isto quer dizer que em 100 chances temos 99 de que a média de documentos esteja entre 98,00 e 139,06, e para o total dos documentos (5.000) teremos o intervalo:

$$P \{ 98,00 \cdot 5.000 \leq X \leq 139,06 \cdot 5.000 \} = 0,99$$

$$P \{ 490.000 \leq X \leq 695.300 \} = 0,99$$

Atributos -- Porcentagens

A amostragem por conglomerados pode ser utilizada para estimar atributos (porcentagens ou taxa de ocorrências).

A técnica envolve a seleção de amostras em grupos com elementos contíguos. O objetivo é estimar a taxa de ocorrência (taxa de erros) através da amostra. A fim de avaliar a confiabilidade da amostragem dos resultados é necessário determinar as porcentagens de erros de cada extrato separadamente. Uma vez estabelecido isto, o cálculo da confiabilidade da amostragem é feito da mesma forma que para variáveis.

Vejamos um exemplo: Desejamos a porcentagem de erros de 4.000 documentos. A fim de obtermos este objetivo selecionamos 21 grupos (conglomerados), com 20 ítems cada conglomerado. Normalmente devemos ter pelo menos 20 conglomerados. Em função dos documentos levantados, temos a seguinte distribuição:

Conglomerado i	Tamanho n'_i	Nº de erros no Congl. e'_i	$p'_i = \frac{e'_i}{n'_i} \cdot 100$
1	20	3	15
2	20	2	10
3	20	2	10
4	20	0	0
5	20	4	20
6	20	2	10
7	20	1	5
8	20	3	15
9	20	1	5
10	20	1	5
11	20	3	15
12	20	5	25
13	20	2	10
14	20	0	0
15	20	1	5
16	20	4	20
17	20	0	0
18	20	2	10
19	20	3	15
20	20	2	10
21	20	1	5
		\bar{p}'	10

A média da taxa de ocorrência é de 10% = $\frac{\sum_{i=1}^{21} p'_i}{21}$

O intervalo de confiança para a taxa de ocorrência p é dada por:

$$p \left\{ \bar{p}' - G(p') \frac{t_{\alpha/2; k-1}}{k-1} \sqrt{1 - \frac{K}{N'}} \leq p \leq \bar{p}' + G(p') \frac{t_{\alpha/2; k-1}}{k-1} \right.$$

$$\left. \sqrt{1 - \frac{K}{N'}} \right\} = 1 - \alpha$$

onde \bar{p}' é a média das médias das porcentagens dos conglomerados:

$$G(p') = \sqrt{\frac{(\sum p'_i - \bar{p}'_i)^2}{K}}, \text{ desvio-padrão das porcentagens dos conglomerados, } K \text{ é o número de conglomerados, } N' = \frac{N}{n'},$$

onde N é o tamanho do Universo, n' tamanho de cada conglomerado, $(1 - \alpha)$ é o nível de confiança, $t_{\alpha/2; k-1}$ valor da variável t de Student.

Inicialmente, teremos que calcular o valor de $G(p')$ através da fórmula acima, uma vez que temos menos de 30 conglomerados. A computação de $G(p')$ é mostrado na tabela seguinte:

$$G(p') = \sqrt{\frac{\sum (p'_i - \bar{p}'_i)^2}{K}} = \sqrt{\frac{950}{21}} = \sqrt{45,238} = 6,73\%$$

Se considerarmos o nível de 95% de confiança, teremos que:

$$t_{0,025; 20} = 2,0860; K = 21; n' = 20; N' = 200; N = 4.000;$$

$$1 - \alpha = 0,95 \Rightarrow \alpha = 0,05.$$

O intervalo de confiança para p (taxa de ocorrência) é dada por:

$$P \left\{ 10\% - 6,73\% \cdot \frac{2,0860}{\sqrt{21-1}} \sqrt{1 - \frac{21}{200}} \leq p \leq 10\% + 6,73\% \cdot \frac{2,0860}{\sqrt{21-1}} \right.$$

$$\left. \sqrt{1 - \frac{21}{200}} \right\} = 1 - 0,05$$

Conglom. i	p'_i	$p'_i - \bar{p}'_i$	$(p'_i - \bar{p}'_i)^2$
1	15	5	25
2	10	0	0
3	10	0	0
4	0	- 10	100
5	20	10	100
6	10	0	0
7	5	- 5	25
8	15	5	25
9	5	- 5	25
10	5	- 5	25
11	15	5	25
12	25	15	225
13	10	0	0
14	0	- 10	100
15	5	- 5	25
16	20	10	100
17	0	- 10	100
18	10	0	0
19	15	5	25
20	10	0	0
21	5	- 5	25

$$G(p') = \sqrt{\frac{\sum (p'_i - \bar{p}'_i)^2}{K}} = \sqrt{\frac{950}{21}} = \sqrt{45,238} = 6,73\%$$

$$P \{ 10\% - 2,97\% \leq p \leq 10\% + 2,97\% \} = 0,95$$

$$P \{ 7,03\% \leq p \leq m \ 12,97\% \} = 0,95$$

Isto quer dizer que há 95 chances em 100 de que a taxa de ocorrência (erros) esteja entre 7,03% e 12,97%.

COMENTÁRIOS FINAIS

Nos exemplos e discussões acima, utilizamos para a amostragem por conglomerado somente a amostragem irrestrita aleatória. De qualquer maneira, a amostragem por conglomerado pode ser usada em conjugação com a amostragem aleatória estratificada ou qualquer outro tipo.

Quando a amostragem por conglomerado é relacionada com a amostragem estratificada, o universo é dividido em extratos e os conglomerados são tirados separadamente através de cada estrato.

BIBLIOGRAFIA

- ARKIN, H.** Handbook of Sampling in Auditing and Accounting, N.Y. McGraw Hill Book Co., 1962.
- BONINI, E.E. - BONINI, S.E.** Estatística - Teoria e Exercícios. Edições Loyola, 1976, SP.
- COHRAN, G. William.** Técnicas de Amostragem. Editora Fundo de Cultura, 1965.
- HILL, Henry P. Roth, Joseph L. e Arkin H.** Sampling in Auditing, N.Y., The Ronald Press Co., 1962.
- TRUEBLOND, Robert M. e Cyert, R.M.** Sampling Techniques in Accounting. Englewood Cliffs, N.J. Prentice, Hall Inc., 1951.
- VANCE, Lawrence L. e Neter, John.** Statistical Sampling for Auditors and Accountants. N.Y. John Willey & Sons, Inc., 1956.

ABSTRACT

Can statistical methods be applied to Accounting and Auditing? In this article the author's object was to determine characteristics about the population with information from a sample. The application is faster, less costly and more reliable. This article

suggests that statistical methods are tools of great value to the accountant. It defines some questions which must be answered, and states the nature of further work which must be done before statistical sampling methods can be used in accounting and auditing.