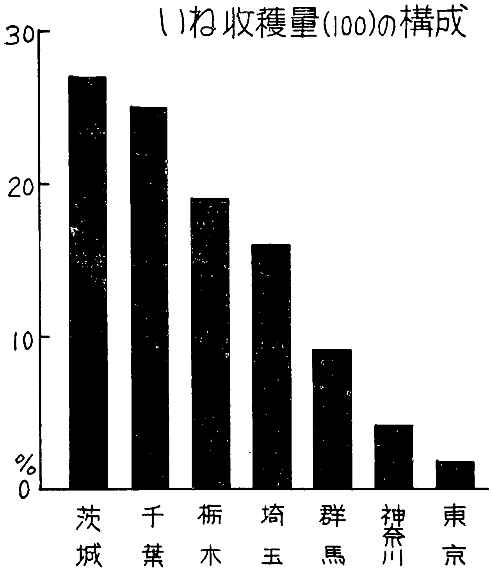
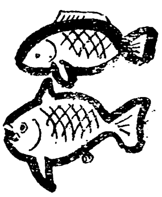
関東7都県のいね収穫量(100)の構成



いね(水稲・陸稲)の収穫

昭和	市郡別	収穫面積	10アール当り収量	収穫量
		ha	kg	kg
33	県計	110,646.92	299	331,097,850
34	〃	113,218.28	336	380,483,987
35	〃	115,392.20	368	424,157,044
	市計	29,745.0	361	107,460,523
	水戸	3,122.7	332	10,382,681
	日立	1,070.6	318	3,409,007
	土浦	1,822.2	406	7,402,457
	古河	486.0	289	1,402,145
	石岡	1,419.7	386	5,473,279
	下館	4,156.2	365	15,169,470
	結城	2,585.0	282	7,302,349
	竜ケ崎	2,448.1	435	10,638,653
	那珂湊	321.5	330	1,060,003
	下妻	2,360.5	320	7,546,824
	水海道	2,576.8	378	9,736,697
	常陸太田	2,019.4	408	8,244,069
	勝田	1,108.9	337	3,733,109
	高萩	779.9	353	2,756,713
	北茨城	1,600.2	370	5,920,886
	笠間	1,867.2	390	7,282,181
	郡計	85,647.2	370	316,696,521
	東茨城	8,369.4	467	39,050,353
	西茨城	4,191.7	324	13,596,790
	那珂	5,959.7	322	19,162,631
	久慈	3,174.1	352	11,165,500
	多賀	455.8	296	1,348,856
	鹿島	6,298.5	383	24,096,970
	行方	5,461.0	439	23,977,188
	稲敷	11,275.4	395	44,564,219
	新治	6,793.8	390	26,510,314
	筑波	8,761.4	365	32,002,074
	真壁	6,680.0	334	22,290,729
	結城	4,340.5	334	14,475,703
	猿島	9,621.1	292	28,137,385
	北相馬	4,264.8	383	16,317,809

註　昭和33.34は県表式調査，昭和35はセンサスの数字である。



標本調査への手引 (5)

総理府統計局　高　橋　史　朗

第１部　標本調査の理論（つづき）

13　無作為抽出

まず，平均営業利益をもとめる方法からかんがえてみます。

平均営業利益を，ある確率変数の実現値から推定するには，その確率変数が，第10節（９月号）で述べた，２つの条件を満たしていなければなりません。

ところで，第12節（９月号）の(1)でつくつた確率変数Xは，条件１は満たしますが，条件２は満たしません。いいかえると，そのままでは，推定子として利用できないわけです。

そこで，いま，確率変数Xから，第８節（８月号）で述べた方法にしたがつて，新たに，確率変数$\overline{X}m$を誘導してみますと，この確率変数$\overline{X}m$は，条件１を常に満たしていますし，また，条件２も，サイコロをくり返して振る回数mを，次にしめす不等式が満たされるように，大きくとることによつて，満たします。すなわち，確率変数$\overline{X}m$の標準偏差$\sigma(\overline{X}m)$および変動係数$CV(\overline{X}m)$は，26,204軒の飲食店の営業利益の標準偏差σxおよび変動係数CVによつて，それぞれ，次のように，

$$\sigma(\overline{X}m) = \frac{1}{\sqrt{m}}\sigma x \quad \text{あるいは} \quad CV(\overline{X}_m) = \frac{1}{\sqrt{m}}CV_x$$

と表わされますが，これを条件２に入れて，mについて解いてみると

$$m \geqq \left(\frac{\lambda}{\varepsilon}\right)^2\sigma_x^2 \quad \text{あるいは} \quad m \geqq \left(\frac{\lambda}{\eta}\right)^2 CV_x^2$$

となります。したがつて，この不等式を満たすように，サイコロをくり返して振る回数mをとれば，確率変数$\overline{X}m$は，平均営業利益の推定子になるわけです。いいかえると，この不等式を満たすように，サイコロをくり返してm回振り，出た目に対応するm軒の飲食店を選んで営業利益を調べ，その算術平均をもとめると，これは，あらかじめ決めておいた許容区間の絶対誤差εあるいは相対誤差ηまた危険の確率λ（ここで，λは，δに代つて，危険の確率を表わしています）という条件を満たす，平均営業利益の推定値になるわけです。

次に，１従業員当たりの平均売上高をもとめる方法をかんがえてみましょう。

第12節の(2)でつくつた確率変数（X，Y）のうち，確率変数Xの期待値E（X）は，もとめようとする１従業員当たりの平均売上高の分母に相当し，また，確率変数Yの期待値E(Y)は，分子に相当しております。

そこでさきとおなじように，確率変数（X，Y）から第８節で述べた方法にしたがつて，新たに，確率変数$\frac{\overline{Y}m}{\overline{X}m}$を誘導してみますと，この確率変数$\frac{\overline{Y}m}{\overline{X}m}$は，サイコロをくり返して振る回数$m$が十分に大きい場合に，条件１を近似的に満たしますし，また，条件２も，回数mを，さらに，次にしめす不等式が満たされるように，大きくとることによつて，満たします。すなわち，確率変数$\frac{\overline{Y}m}{\overline{X}m}$の標準偏差$\sigma\left(\frac{\overline{Y}m}{\overline{X}m}\right)$および変動係数$CV\left(\frac{\overline{Y}m}{\overline{X}m}\right)$は，回数$m$が十分に大きい場合に，26,204軒の飲食店の従業員数の算術平均M_x，変動係数CV_x，売上高の算術平均My，変動係数CVy，従業員数と売上高との相関係数ρによつて，近似的に，それぞれ，次のように

$$\sigma\left(\frac{\overline{Y}m}{\overline{X}m}\right) \fallingdotseq \frac{1}{\sqrt{m}}\frac{My}{Mx}\sqrt{CV_x^2 + CV_y^2 - 2\rho xy CVx CVy}$$

$$\text{あるいは} \quad CV\left(\frac{\overline{Y}m}{\overline{X}m}\right) \fallingdotseq \frac{1}{\sqrt{m}}\sqrt{CV_x^2 + CV_y^2 - 2\rho xy\, CVx\, CVy}$$

と表わされますが，これを条件２に入れて，mについて解いてみると

$$m \geqq \left(\frac{\lambda}{\varepsilon}\right)^2\left(\frac{My}{Mx}\right)^2(CV_x^2 + CV_x^2 - 2\rho xy\, CVx\, CVy)$$

$$\text{あるいは} \quad m \geqq \left(\frac{\lambda}{\eta}\right)^2(CV_x^2 + CV_x^2 - 2\rho xy\, CVx\, CVy)$$

となります。したがつて，サイコロをくり返して振る回数mを十分に大きく，しかも，この不等式を満たすようにとれば，確率変数$\frac{\overline{Y}m}{\overline{X}m}$は，１従業員当たりの平均

売上高の推定子になるわけです。いいかえると，サイコロを十分に大きく，しかも，この不等式を満たすようにm回くり返して振り，出た目に対応するm軒の飲食店を選んで，従業員数と売上高を調べ，従業員数の算術平均を分母とし，売上高の算術平均を分子とする比率をもとめると，これは，あらかじめ決めておいた許容区間の絶対誤差εあるいは相対誤差ηまた危険の確率λという条件を満たす，1従業員当たりの平均売上高の推定値になるわけです。

では，次に，どれだけの軒数の飲食店を選びだして，調べれば，平均営業利益および1従業員当たりの平均売上高がもとめられるかをかんがえてみましよう。

段階(1) 平均営業利益だけをもとめる場合

まず，平均営業利益だけを，たとえば，許容区間の相対誤差$\eta=5\%$，危険の確率$\lambda=2$でもとめることにしてみましよう。選びだす飲食店の軒数m（これを**標本の大きさ**といいます）を決めるには，26,204軒の飲食店の9月の営業利益の変動係数CV_xを知る必要がありますがそれは，過去の資料から，たとえば

$$CV_x \fallingdotseq 1.54$$

ぐらいになると見当つけられたとしますと，軒数mは

$$m \geqq \left(\frac{\lambda}{\eta}\right)^2 CV_x^2 = 3,794.\cdots\cdots$$

となりますから，したがつて，有効桁数をかんがえて，2桁までとり，3,800軒の飲食店を選びだして，9月の営業利益を調べ，その算術平均をもとめますと，それは平均営業利益の，許容区間の相対誤差$\eta=5\%$，危険の確率$\lambda=2$の推定値となります。なお，この飲食店の選び方を，**無作為抽出**（あるいは，**等確率抽出**）といいます。

段階(2) 1従業員当たりの平均売上高だけをもとめる場合

次に，1従業員当たりの平均売上高だけを，たとえば許容区間の相対誤差$\eta=5\%$，危険の確率$\lambda=2$でもとめることにしてみましよう。選びだす飲食店の軒数mを決めるには，26,204軒の飲食店の従業員数の変動係数CV_x，売上高の変動係数CV_yおよび従業員数と売上高との相関係数ρ_{xy}を知る必要がありますが，それは，過去の資料から，たとえば，

$$CV_x \fallingdotseq 1.12 \qquad CV_y \fallingdotseq 1.76 \qquad \rho_{xy} \fallingdotseq 0.83$$

ぐらいになると見当つけられたとしますと，軒数mは，

$$m \geqq \left(\frac{\lambda}{\eta}\right)^2 (CV_x^2 + CV_y^2 - 2\rho_{xy}\, CV_x\, CV_y$$
$$=1,727.\cdots\cdots$$

となりますから，したがつて，2桁までとり，1,800軒の飲食店を選びだして，9月の従業員数と売上高を調べ，その従業員数の算術平均を分母とし，売上高の算術平均を分子とする比率をもとめると，それは，1従業員当たりの平均売上高の，許容区間の相対誤差$\eta=5\%$，危険の確率$\lambda=2$の推定値となります。

段階(3) 平均営業利益および1従業員当たりの平均売上高を共にもとめる場合

いよいよ，平均営業利益および1従業員当たりの平均売上高をあわせて，共に，許容区間の相対誤差$\eta=5\%$，危険の確率$\lambda=2$でもめることにしてみましよう。

ここで問題になるのは，平均営業利益が，無作為抽出で選びだされた3,800軒の飲食店を必要とするのにいたして，1従業員当たりの平均売上高は，1,800軒の飲食店しか必要としないことです。

この必要とする軒数の違いを調整するには，次の3つの方法がかんがえられます。

1. いくら検討しても，平均営業利益を，許容区間の相対誤差$\eta=5\%$，危険の確率$\lambda=2$でもとめる必要があれば，3,800軒の飲食店を選びだして，営業利益，従業員数，売上高を調べ，平均営業利益および1従業員当たりの平均売上高をもとめます。このようにすると，1従業員当たりの平均売上高は，必要以上に，「正確」にもとめられますが，やむをえません。

なお，この場合，1従業員当たりの平均売上高は，危険の確率$\lambda=2$をそのままとして，許容区間の相対誤差ηがいくらでもとめられるかですが，いま，mについての不等式を，不等号を削つて，ηについて解いてみると

$$\eta = \frac{\lambda}{\sqrt{m}} \sqrt{CV_x^2 + CV_y^2 - 2\rho_{xy}\, CV_x\, CV_y}$$

となりますから，この式からもとめればよく，第7表にしめすように，$\eta \fallingdotseq 3.4\%$となります。

2. 検討した結果，平均営業利益は重要性が低く，なりゆきにまかせても，やむをえないというのであれば，1,800軒の飲食店を選びだして，平均営業利益および1従業員当たりの平均売上高をもとめます。

なお，この場合，平均営業利益は，危険の確率$\lambda=2$をそのままとして，許容区間の相対誤差ηがいくらでもとめられるかですが，いま，mについての不等式を，不等号を削つて，ηについて解いてみると

$$\eta = \frac{\lambda}{\sqrt{m}} CV_x$$

となりますから，この式からもとめればよく，第7表にしめすように，$\eta \fallingdotseq 7.3\%$となります。

3. 平均営業利益も，1従業員当たりの平均売上高と重要性に変わりないが，そのために，2,000軒も余計に必要ならば，少し譲つて，許容区間の相対誤差$\eta=6\%$でよいとすると，平均営業利益をもとめるのに必要な軒数は，2,700軒となりますから，それだけ選びだして，平均営業利益および1従業員当たりの平均売上高をもとめます。

なお，この場合，1従業員当たりの平均売上高は危険の確率$\lambda=2$をそのままとして，第7表にしめすよう

に，許容区間の相対誤差$\eta=4.0\%$でもとめられることになります。

第 7 表

η		m
平均営業利益	1従業員当たりの平均売上高	
5. %	3.4 %	**3,800**
7.3	**5.**	**1,800**
6.	4.0	**2,700**

注 1. $\lambda = 2$

2. mは切上げ，ηも切上げにしてあります。

これで，選びだす軒数の決め方は明らかになりましたが，しかし，そこには，なお，次のような問題点が残されております。

問題点 問題は，選びだす軒数が，過去の資料によつて決められるという点にあります。推定値にたいして，あらかじめ決めておく許容区間の絶対あるいは相対誤差および危険の確率のうち，危険の確率は，一般に，どのような局面になつても動かしませんから，したがつて，勢い，許容区間の絶対あるいは相対誤差が，この過去の資料から決められた軒数を調べることによつて，逆に，ただ，**目標としたもの**になつてしまい，これは，必ずしも，**達成されたもの**とは一致しなくなるわけです。

なお，達成された絶対あるいは相対誤差を，どのようにもとめるかについては，また，あとで，詳しく述べたいとおもいます。

14 信頼区間とその与え方

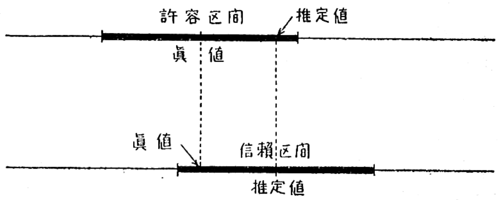
あらかじめ，決めておいた許容区間のなかにはいる推定値は，その統計を作成した目的にたいするかぎり，真値のようにみなして利用できるわけですが，ここでは，その真値それ自身が，一体どのくらいの大きさの値であるかを知る方法をかんがえてみます。

もしも，許容区間が，数値で，どこからどこまでの区間と表わせるならば，それでもつて，真値がどのくらいの大きさの値であるか分かりますが，あいにく，許容区間は，真値それ自身を中心としているので，これを，数値で表わすことは不可能です。

そこで，いま，ある推定値が，真値を中心に，左右にεの幅の，すなわち，絶対誤差εの許容区間のなかにはいるとき，逆に，真値は，その推定値を中心に，どのくらいの幅の区間のなかにはいるかをかんがえてみましょう。すぐ，お分かりになるように，真値は，その推定値を中心に，左右にεの幅の区間のなかにはいります。（第4図参照）

この，推定値を中心とした区間を，信頼区間といいます。

第 4 図



信頼区間は，許容区間とおなじように，絶対誤差あるいは相対誤差によつて与えられますが，このうち，まず絶対誤差については，いま，見たように，推定値が，絶対誤差 εの許容区間のなかにはいるとき，逆に，真値は絶対誤差 εの信頼区間のなかにはいるという関係があります。この逆も成り立ちます。

次に，相対誤差については，推定値が，相対誤差 ηの許容区間のなかにはいるとき，逆に，真値は，ηから，次の算式によつてもとめた，相対誤差η'の信頼区間のなかにはいるという関係があります。この逆もまた成り立ちます。

$$\eta' = \frac{\eta}{1-\eta} \quad \text{あるいは} \quad \eta = \frac{\eta'}{1+\eta'}$$

すこし数字をいれてみますと，たとえば，信頼区間を相対誤差$\eta'=5\%$でもとめるには，第8表にしめすように，許容区間の相対誤差$\eta \fallingdotseq 4.7\%$で推定しなければならないわけです。

第 8 表

η	η'	η'	η
%	%	%	%
1	1.1	1	0.9
2	2.1	2	1.9
3	3.1	3	2.9
5	5.3	5	4.7
10	11.2	10	9.0

注 ηは切捨て，η'は切上げにしてあります。

〔**練習問題**〕 第13節の段階(3)の３の例で，2,700軒の飲食店を選びだして，９月の営業利益，従業員数，売上高を調べ，平均営業利益を65,698円，また，１従業員当たりの平均売上高を59,284円と推定したとしますとそれぞれの信頼区間は，第９表のようになります。みなさんも確かめてください。

第 9 表

	平均営業利益	1従業員当たりの平均売上高
推定値	65,698円	59,284円
信頼区間	61,400〜70,000	56,700〜61,800
η	**6.** %	4.0 %
η'	6.4	4.2

注 $m=2,700$

むすび

これで，無作為抽出を中心とした，標本調査の理論の説明を終わります。標本調査が，確率の理論をもとにして，どのように組み立てられているかについて，おおよそのところ，分かつていただけたとおもいます。

ところで，無作為抽出は，もつとも基本的な標本選定の方法ですが，多くの制約のある標本調査を円滑におこなつてゆくためには，これだけでは不十分で，なお，いろいろの方法を工夫する必要があります。そして，実際にも，多くの方法が工夫されております。次の，第2部標本選定の技巧では，それらの方法について説明してゆきたいとおもいます。

労働力主要指標 —資料　総理府統計局編　労働力調査結果—

第1表

昭和	総人口（千人）	15才以上人口	労働力人口	就業者総数	農林業就業者	非農林業就業者	完全失業者	労働力人口比率	就業者総数/労働力人口
28	86,700	56,900	39,570	39,120	16,900	22,220	450	69.5	98.9
29	88,000	57,940	40,200	39,620	16,500	23,120	590	69.4	98.6
30	89,100	59,060	41,560	40,880	16,860	24,020	680	70.4	98.4
31	90,060	60,400	42,350	41,720	16,450	25,270	630	70.1	98.5
32	90,910	61,750	43,360	42,840	16,070	26,770	520	70.2	98.8
33	91,810	63,070	43,680	43,120	15,470	27,650	560	69.3	98.7

第2表

昭和	就業者総数	農林業	非農林業	漁業水産養殖業	鉱業	建設業	製造業	卸売・小売・金融不動産業	運輸・通信・電気ガス水道業	サービス業	公務
		%									
28	39,120	43.2	56.8	1.7	1.6	4.1	17.3	14.7	5.0	9.5	3.0
29	39,620	41.6	58.4	1.4	1.5	4,2	17.7	16.0	4.8	9.5	3.2
30	40,880	41.2	58.8	1.3	1.3	4.4	17.4	16.3	4.7	10.4	2.9
31	41,720	39.4	60.6	1.4	1.1	4.3	18.2	16.7	4.9	11.1	2.8
32	42,840	37.5	62.5	1.5	1.4	4.6	18.9	17.0	5.1	11.4	2.7
33	43,120	35.9	64.1	1.2	1.2	4.8	20.0	17.4	5.0	11.6	2.9

第3表　本県の就業者数

昭和	総人口（千人）	労働力人口比率	労働力人口	就業者総数/労働力人口	就業者数
28	2,057	69.5	1,430	98.9	1,414
29	2,066	69.4	1,434	98.6	1,414
30	2,077	70.4	1,462	98.4	1,439
31	2,081	70.1	1,459	98.5	1,437
32	2,081	70.2	1,461	98.8	1,443
33	2,082	69.3	1,443	98.7	1,424

注　第1，2表は全国の労働指標である。
第3表は本県の労働指標で本県総人口に全国の各率を乗じて，本県就業者数を求めた。また本県就業者の産業別就業者の概数は第3表就業者数に第2表の比率を乗じて求めることもできる。但しこの場合は全国の産業別就業者構成と同様であるとする前提においてである。

——編集部——