The routine of randum-sampling by computer

新聞語彙調査の

サンプリング・プログラム

田中章夫・斎藤秀紀

0. ま え が き

今回の新聞の語彙調査では，サンプリングをコンピューターによって行なった。そのさい作成したプログラムは，一般に行なわれる種々の抽出調査にも，サンプリングプログラムとして広く使えるよう配慮したつもりである。

1. 新聞の語彙調査のサンプリング方針

新聞の語彙調査のサンプリングは，つぎのような方針で行なった。

○対　　象　3紙(朝日・毎日・読売)1ヶ年分

○抽 出 比　1／60

○抽出単位　1／2段(以下1ブロック，または，1Bと呼ぶ)

○母 集 団　853,200ブロック*

朝　刊　518,400B＝30B×16P×360日×3紙
夕　刊　334,800B＝30B×12P×310日×3紙

○標　　本　14,220ブロック(853,200 B×1／60)

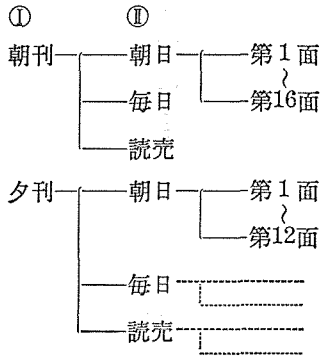
朝　刊　8,640B(518,400B×1／60)
夕　刊　5,580B(334,800B×1／60)

* 1)サービス版(日曜版・PR版ナド)をのぞく。

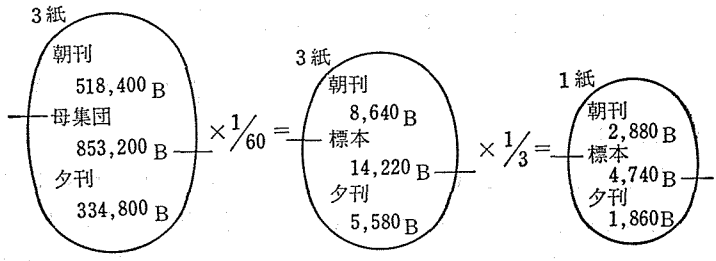
2)元日・休刊日，夕刊については日曜日をのぞく。

3)朝刊を16ページだて，夕刊を12ページだてとして計算してある。

○抽 出 法　2段抽出



以上のサンプリング方針を図示するとつぎのようになる。



2. 抽 出 手 順

2. 1 サンプリング台帳の作成(コンピューター)

3紙の朝刊，夕刊で共通に使用できる台帳を作成することが，最初の仕事となる。1ページに存在する30個のブロックに右肩から順に01•02……30の番号を与え，16ページだての新聞半年分の全ブロックにシーケンス番号をつけることにした。16ページだての新聞半年分のブロック総数は30ブロック×16ページ×31日×6個月＝89280ブロックとなる。したがって00000から89279までの番号がつくことになる。なる。

シーケンスナンバー		月	日	ページ	ブロックナンバー
00000		01	01	01	01
⌇					⌇
00029		01	01	01	30
00030		01	01	02	01
⌇		⋮	⋮	⌇	⋮
00360	*	01	01	13	01
⌇	⋮	⋮	⋮	⌇	⋮
00479	*	01	02	16	30
00480		01	02	01	01
		⌇	⌇	⋮	⋮
⌇		06	31	⋮	⋮
		⋮	⋮	⋮	⋮
89279	*	06	31	16	30

ただし，13ページ～16ページは夕刊のサンプリングの場合には，使用しないので，当該シーケンス番号に＊印を入れておく。以上のようなフォーマットの台帳をラインプリンターで OUTPUT するとともに，磁気テープに記録し，サンプリング作業に使用した。

2. 2　ランダム・サンプリング(コンピューター)

このプログラムの主要部は，コンピューターに乱数を発生させ，サンプリング台帳のシーケンス番号と照合させて標本を決定させサンプル表を作成するもので，全体の進みぐあいは次のようになる。

まず，サンプル表は3紙朝刊夕刊別すなわち6種類(半年分ずつ)作成する。その結果，1年分では，12種類となる。

次に夕刊のサンプリング過程で＊のシーケンス番号に当った時は，それをとばして照合する。

そして最後にサンプル表とともに，予備乱数表をラインプリンターで OUTPUT する。

2. 3　標本の修正(人手)

つぎの場合には，予備乱数表によって他の日の新聞に標本をふりかえる。

○元日版に当っている場合

○30日の月の31日，2月の29，30，31日に当っている場合

○夕刊のサンプルが，日曜日に当っている場合

○休刊日に当っている場合

3. プログラムの内容とプログラム作成者

RUN 1	サンプリング台帳作成	田中
	MASTER(台帳)M/T 作成	
RUN 2	乱数テーブル　SET	斎藤
	サンプリング(照合)	
	SAMPLE　M/T 作成	
	予備乱数表作成	
RUN 3	サンプル表作成	沢田

4. サンプリングプログラム使用上の注意

このプログラムを使用して一般的なサンプリングを行なうことができるが，その場合には，つぎの6点を指定すればよい。それぞれについての制限

は(　　)内の通りである。

① 乱数の桁数(最大8桁)

② 同一母集団から同時に行なうサンプリングの種類(最大9種)

③ 母集団の個数(最大99999個)

④ 標本の個数(最大9999個)*

*ただし，今のところ

(乱数の桁数)×(サンプル数)≦10000とする

⑤ 予備乱数表が必要か否か(発生個数最大9999個)

⑥ 予備乱数の初期値は任意に8桁の数字で指定する。

5. サンプリング・プログラムの概要

5．1　RUN1は，最初に1回だけ通す。

RUN2～RUN3は，1PASSごとに朝刊または，夕刊の1紙半年分のサンプリングが行なわれると同時に，200個の予備乱数をプリントする。サンプル個数は1パス(半年分)について朝刊1440個，夕刊930個である。

なお，このシステムの企画，進行には，田中，斎藤が当り，研究補助員沢田さち子が，作業を助けた。またNBC(日本ビジネス・コンサルタント)の山本武氏には，プログラム全般にわたり助言を得た。

以下の執筆は各RUNの担当者による。(以上，田中執筆)

RUN1(田中)

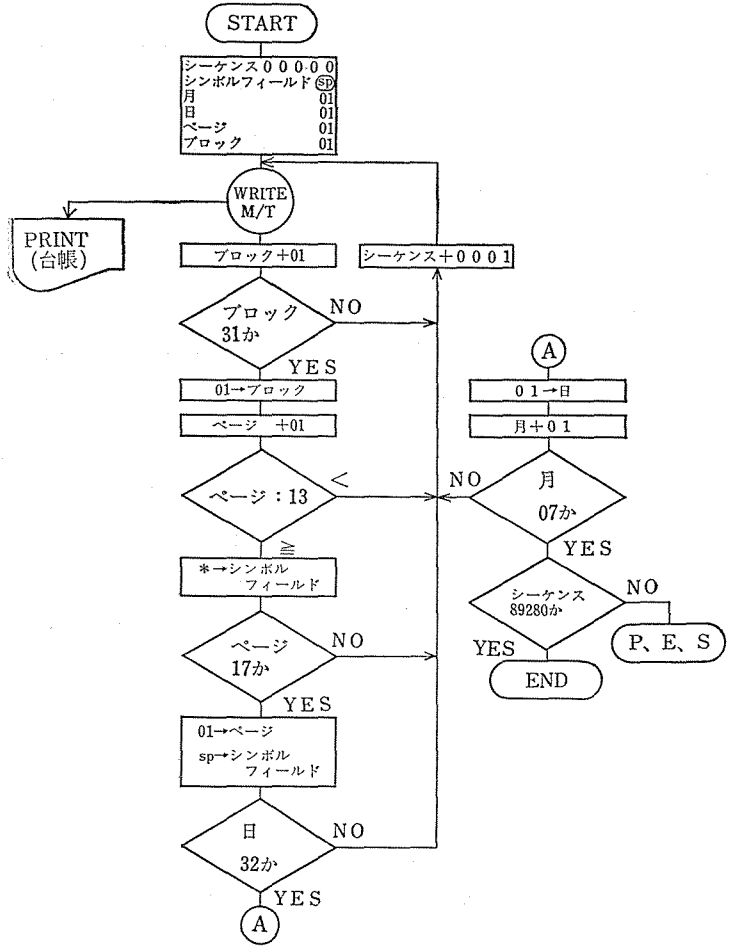
このRUNは半年分の16ページだての新聞の全ブロック(89280 B)に，日付順，ページ順の一連シーケンス番号を与えるものである。

OUTPUTはM/Tとプリントである。

○M/T　FORMAT　　20 ch/1RECORD・18 RECORDS/1 BATCH

㊩ ＃	シーケンスナンバー(5桁)	㊩ or ＊	月(2桁)	日(2桁)	ページ(2桁)	ブロック番号(2桁)	E/i	

○プリントFORMATは上記のRECORDを6個／1行で1行おきとする。



RUN 2 (斎藤)

この RUN は，大別して 4 個の独立した ROUTINE から構成されている。その内容は

1 一様乱数発生 ROUTINE

2 予備乱数表作成 ROUTINE

3 乱数・内部 SORT ROUTINE

4 乱数とサンプリング台帳コレート ROUTINE

1 乱数発生ルーチン

乱数は一様擬似乱数（Pseudo- random number)とし，合同法による乗積法によって発生させた。乱数の周期は8桁の数字で，5,882,352である。なおこのROUTINEにおける入力パラメータは

母集団総数　　5桁

標本抽出数　　4桁

予備乱数発生個数4桁

サンプリング台帳インジケータ　1桁

以上4個である。予備乱数発生個数，標本抽出数は最大2000個まで，母集団総数は朝刊用台帳89280夕刊用台帳66960（乱数の初期値は任意に8桁の数字で与えることも出来る)である。繰作卓のINTボタンがONのときP/T読込み，OFFのときHSM内部の数字を使用するようになっている。

また台帳判定用インジケータはこのサンプリングにおいては“0”を主台帳として朝刊用台帳にあて，“1”を副台帳として夕刊用台帳にあてた。しかしこのインジケータは0～9までを任意に台帳判定用インジケータとして使用出来る。

2. 予備乱数表作成ルーチン

乱数発生ルーチンで作られた一様乱数8桁の頭5桁を乱数表（最大2000個）として印字するルーチンである。ただし，ここで最小1個は予備乱数として指定しなければならない。

パラメータフォーマット

Gap	*1 桁	母集団総数5桁	抽出数4桁	予備乱数 4 桁	Gap

*台帳判定用インジケータ

乱数初期値を外部より与える場合はさらに次のパラメータを必要とする。

Gap	乱数初期値　8桁	Gap

3. 内部ソートルーチン

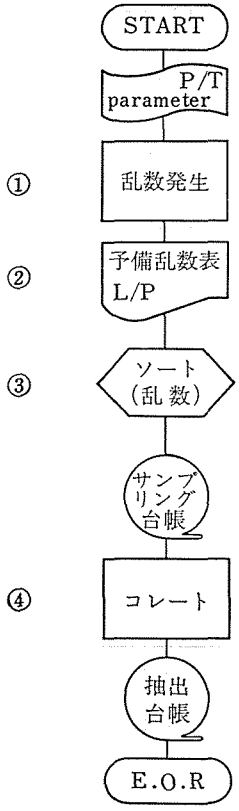
最初にN個（5桁の数字で2000個以下）のレコードの内，最小の数字を取り出し，乱数の格納されている最初の番地の内容のレコードと交換する，

次に(N－1)個のレコード中から再び最小の数字で選び出し乱数格納番地の2番目の位置におく，以下このように全部の数字がならび終るまでこの作業を続ける。

記憶容量はワークエリアを必要とせず乱数の格納されているエリアのみでよい。

1,440個をソートするに要した時間は約6分であった。時間の遅いのが欠点であるが命令ステップが17ステップ程度で出来るのが利点である。

RUN 2　BLOCK　CHART



4. コレートルーチン

内部ソートされた乱数とシーケンス順にならんでいるサンプリング台帳とをコレートさせ乱数と同一のシーケンスを持つ要素を母集団から抽出するものである。

RUN 3 （沢田）

概　　要

サンプリングプログラムにおける最後のランである。

朝刊で1,440，夕刊で930ずつサンプリングされたものを，パラメータテープの指定により，新聞名をわりあて，ラインプリンタで印刷するプログラムである。

入出力形式

データ入力　M/T

（前回までのランで作成されたもの）

	SP	# シーケンス	SP	MONTH	DAY	PAGE	BLOCK NO	E/I

パラメータテープ　P/T

（サンプル表の名称）

	N N N	SP	T_1	T_2	

NNN : ASA, MAi, YOM } 新聞名

T_1 : M　朝刊の略

E　夕刊の略

T_2 : 0　前半（1月～6月）

1　後半（7月～12月）

印字形式 （例）

サンプリング SP ヒョウ SP アサヒ SP チョウカン SP 1月

×××××× ← サンプリングされた内容

操作の概要

〔1〕 必要とする新聞名，朝夕刊の別および年の前後半の別を定めたパラメータテープを入力する。

〔2〕 パラメータで指定した内容に該当するカナ文字を，計算機内に入

れておいたテーブルから引用する。

〔3〕 月にかんしては0または1の指定により前者は1～6月，後者は7～12月のものとなるように変換する。

〔4〕 各ページのはじめは数字，カナによる，ヘッディングを行う。

〔5〕 プリントは，入力テープの1レコードをそのままの形で，1ページ25行のわりあいで印字させる。