

Random Sampling For the Non-Statistician

Diane E. Brown

AdminaStar Solutions, Associated Insurance Companies Inc.

Random samples can be drawn based on:

- Size: an approximate number, an exact number, a percent
- Select: with or without replacement
- Universe: stratified or non-stratified
- Data Distribution: SAS random number functions

The following random number functions are available:

FUNCTION	DISTRIBUTION
NORMAL(seed)	Normal
RANBIN(seed)	Binomial
RANCAU(seed)	Cauchy
RANEXP(seed)	Exponential
RANGAM(seed)	Gamma
RANNOR(seed)	Normal
RANPOI(seed)	Poisson
RANTBL(seed)	Tabled Probability
RANTRI(seed)	Triangular
RANUNI(seed)	Uniform
UNIFORM(seed)	Uniform

- when the distribution is unknown, use RANUNI/UNIFORM
- result is number between 0 and 1; a probability
- seed is value from 0 to 2,147,483,647
 - 0 random number is generated from computer system clock; sample can NOT be replicated
 - other value is source of generation of random number; sample will be replicated

EXAMPLE 1: Select an approximate random sample of 5 without replacement; repeatable seed and universe size known:

```
DATA TRYIT;
  DO I=1 TO 20;
    X + 1;
    OUTPUT;
  END;

DATA TRYIT;
  SET TRYIT;
  IF RANUNI(78956) <= .25;

PROC PRINT DATA=TRYIT;
  VAR X;
  TITLE 'APPROXIMATE RANDOM SAMPLE OF 5 WITHOUT REPLACEMENT';
  TITLE2 'WITH REPEATABLE SEED';
  TITLE3 'UNIVERSE SIZE KNOWN';
```

PROC PRINT produces the following results:

APPROXIMATE RANDOM SAMPLE OF 5 WITHOUT REPLACEMENT
WITH REPEATABLE SEED
UNIVERSE SIZE KNOWN

OBS	X
1	4
2	5
3	11
4	16
5	18
6	20

EXAMPLE 2: Select an approximate random sample of 5 without replacement; non-repeatable seed and universe size unknown:

```
DATA TRYIT;
  DO I=1 TO 20;
    X + 1;
    OUTPUT;
  END;

DATA TRYIT;
  SET TRYIT NOBS=UNIV;
  IF N = 1 THEN DO;
    SAMPSIZ = 5;
    PROB = SAMPSIZ/UNIV;
  END;
  IF RANUNI(0) <= PROB;
  RETAIN PROB;

PROC PRINT DATA=TRYIT;
  VAR X;
  TITLE 'APPROXIMATE RANDOM SAMPLE OF 5 WITHOUT REPLACEMENT';
  TITLE2 'NON-REPEATABLE, SEED FROM SYSTEM CLOCK';
  TITLE3 'UNIVERSE SIZE UNKNOWN';
```

PROC PRINT produces the following results:

APPROXIMATE RANDOM SAMPLE OF 5 WITHOUT REPLACEMENT
NON-REPEATABLE, SEED FROM SYSTEM CLOCK
UNIVERSE SIZE UNKNOWN

OBS	X
1	4
2	7
3	11
4	17
5	20

EXAMPLE 3: Select an exact random sample of 5 without replacement;
non-repeatable, universe size unknown:

```
DATA TRYIT;
  DO I=1 TO 20;
    X + 1;
    OUTPUT;
  END;

DATA TRYIT;
  SET TRYIT NOBS=UNIV;
  IF RANUNI(0) <= SAMPSIZ/UNIV THEN DO;
    OUTPUT;
    SAMPSIZ = SAMPSIZ - 1;
  END;
  UNIV = UNIV - 1;
  IF SAMPSIZ = 0 THEN STOP;
  RETAIN SAMPSIZ 5 UNIV;

PROC PRINT DATA=TRYIT;
  VAR X;
  TITLE 'EXACT RANDOM SAMPLE OF 5 WITHOUT REPLACEMENT';
  TITLE2 'NON-REPEATABLE, SEED FROM SYSTEM CLOCK';
  TITLE3 'UNIVERSE SIZE UNKNOWN';
```

PROC PRINT produces the following results:

EXACT RANDOM SAMPLE OF 5 WITHOUT REPLACEMENT
NON-REPEATABLE, SEED FROM SYSTEM CLOCK
UNIVERSE SIZE UNKNOWN

OBS	X
1	4
2	6
3	12
4	16
5	19

EXAMPLE 4: Select a random sample of 25% without replacement;
non-repeatable and universe unknown:

```
DATA TRYIT;
  DO I=1 TO 20;
    X + 1;
    OUTPUT;
  END;

DATA TRYIT;
  SET TRYIT NOBS=UNIV;
  IF N = 1 THEN SAMPSIZ = INT(UNIV * .25);
  IF RANUNI(0) <= SAMPSIZ/UNIV THEN DO;
    OUTPUT;
    SAMPSIZ = SAMPSIZ - 1;
  END;
  UNIV = UNIV - 1;
  IF SAMPSIZ = 0 THEN STOP;
  RETAIN SAMPSIZ UNIV;

PROC PRINT DATA=TRYIT;
  VAR X;
  TITLE 'RANDOM SAMPLE OF 25% WITHOUT REPLACEMENT';
  TITLE2 'NON-REPEATABLE, SEED FROM SYSTEM CLOCK';
  TITLE3 'UNIVERSE SIZE UNKNOWN';
```

PROC PRINT produces the following result:

RANDOM SAMPLE OF 25% WITHOUT REPLACEMENT
NON-REPEATABLE, SEED FROM SYSTEM CLOCK
UNIVERSE SIZE UNKNOWN

OBS	X
1	7
2	11
3	13
4	17
5	20

EXAMPLE 5: Select an exact random sample of 5 with replacement;
repeatable seed and universe size unknown:

```
DATA TRYIT;
  DO I=1 TO 20;
    X + 1;
    OUTPUT;
  END;

DATA TRYIT;
  SELOBS = INT(RANUNI(78765) * UNIV) + 1;
  SET TRYIT NOBS=UNIV POINT = SELOBS;
  SAMPSIZ + 1;
  IF SAMPSIZ > 5 THEN STOP;

PROC PRINT DATA=TRYIT;
  VAR X;
  TITLE 'RANDOM SAMPLE OF 5 WITH REPLACEMENT';
  TITLE2 'REPEATABLE SEED';
  TITLE3 'UNIVERSE SIZE UNKNOWN';
```

PROC PRINT produces the following result:

RANDOM SAMPLE OF 5 WITH REPLACEMENT
REPEATABLE SEED
UNIVERSE SIZE UNKNOWN

OBS	X
1	12
2	1
3	4
4	11
5	1

EXAMPLE 6: Select a stratified random sample of equal size (5) without replacement; non-repeatable and universe size unknown:

```
DATA TRYIT;
  DO I=1 TO 10;
    CAT = A; X + 1;
    OUTPUT;
  END;
  DO I=1 TO 10;
    CAT = B; X + 1;
    OUTPUT;
  END;

PROC FREQ DATA = TRYIT;
  TABLES CAT / OUT = TRYCT (RENAME=(COUNT=UNIV)) NOPRINT;

PROC SORT DATA = TRYIT;
  BY CAT;

DATA TRYIT;
  MERGE TRYIT TRYCT;
  BY CAT;
  IF FIRST.CAT THEN SAMPSIZ = 5;
  IF RANUNI(0) <= SAMPSIZ/UNIV THEN DO;
    OUTPUT;
    SAMPSIZ = SAMPSIZ - 1;
  END;
  UNIV = UNIV - 1;
  RETAIN SAMPSIZ UNIV;

PROC PRINT DATA=TRYIT;
  VAR CAT X;
  TITLE 'STRATIFIED RANDOM SAMPLE OF 5';
  TITLE2 'EQUAL SIZE SAMPLE WITHOUT REPLACEMENT';
  TITLE3 'NON-REPEATABLE, SEED FROM SYSTEM CLOCK';
  TITLE4 'UNIVERSE SIZE UNKNOWN';
```

PROC PRINT produces the following result:

STRATIFIED RANDOM SAMPLE OF 5
EQUAL SIZE GROUPS WITH REPLACEMENT
NON-REPEATABLE, SEED FROM SYSTEM CLOCK
UNIVERSE SIZE UNKNOWN

OBS	CAT	X
1	A	1
2	A	4
3	A	5
4	A	8
5	A	10
6	B	14
7	B	17
8	B	18
9	B	19
10	B	20

EXAMPLE 7: Select a stratified random sample of unequal size (2/4) without replacement; non-repeatable and universe size unknown;

```
DATA TRYIT;
  DO I=1 TO 10;
    CAT = A; X + 1;
    OUTPUT;
  END;
  DO I=1 TO 10;
    CAT = B; X + 1;
    OUTPUT;
  END;

PROC FREQ DATA = TRYIT;
  TABLES CAT / OUT = TRYCT (RENAME=(COUNT=UNIV)) NOPRINT;

PROC SORT DATA = TRYIT;
  BY CAT;

DATA TRYIT;
  MERGE TRYIT TRYCT;
  BY CAT;
  IF FIRST.CAT THEN DO;
    IF CAT = 'A' THEN SAMPSIZ = 2;
    ELSE SAMPSIZ = 4;
  END;
  IF RANUNI(0) <= SAMPSIZ/UNIV THEN DO;
    OUTPUT;
    SAMPSIZ = SAMPSIZ - 1;
  END;
  UNIV = UNIV - 1;
  RETAIN SAMPSIZ UNIV;

PROC PRINT DATA=TRYIT;
  VAR CAT X;
  TITLE 'STRATIFIED RANDOM SAMPLE OF 2 FROM CAT A, 4 FROM CAT B';
  TITLE2 'UNEQUAL SIZE SAMPLE WITHOUT REPLACEMENT';
  TITLE3 'NON-REPEATABLE, SEED FROM SYSTEM CLOCK';
  TITLE4 'UNIVERSE SIZE UNKNOWN';
```

PROC PRINT produces the following result:

```
STRATIFIED RANDOM SAMPLE OF 2 FROM CAT A, 4 FROM CAT B
UNEQUAL SIZE GROUPS WITH REPLACEMENT
NON-REPEATABLE, SEED FROM SYSTEM CLOCK
UNIVERSE SIZE UNKNOWN
```

OBS	CAT	X
1	A	2
2	A	4
3	B	15
4	B	16
5	B	17
6	B	19

EXAMPLE 8: Select a stratified random sample of 25% from each group without replacement; repeatable seed and universe size unknown:

```
DATA TRYIT;
  DO I=1 TO 10;
    CAT = A; X + 1;
    OUTPUT;
  END;
  DO I=1 TO 10;
    CAT = B; X + 1;
    OUTPUT;
  END;

PROC FREQ DATA = TRYIT;
  TABLES CAT / OUT = TRYCT (RENAME=(COUNT=UNIV)) NOPRINT;

PROC SORT DATA = TRYIT;
  BY CAT;

DATA TRYIT;
  MERGE TRYIT TRYCT;
  BY CAT;
  IF FIRST.CAT THEN SAMPSIZ = INT(UNIV * .25);
  RETAIN SAMPSIZ UNIV;
  IF RANUNI(3423523) <= SAMPSIZ/UNIV THEN DO;
    OUTPUT;
    SAMPSIZ = SAMPSIZ - 1;
  END;
  UNIV = UNIV - 1;

PROC PRINT DATA=TRYIT;
  VAR CAT X;
  TITLE 'STRATIFIED RANDOM SAMPLE OF 25% FROM EACH BY GROUP';
  TITLE2 'UNEQUAL SIZE SAMPLE WITHOUT REPLACEMENT';
  TITLE3 'REPEATABLE SEED';
  TITLE4 'UNIVERSE SIZE UNKNOWN';
```

PROC PRINT produces the following result:

```
STRATIFIED RANDOM SAMPLE OF 25% FROM EACH BY GROUP
UNEQUAL SIZE SAMPLE WITHOUT REPLACEMENT
REPEATABLE SEED
UNIVERSE SIZE UNKNOWN
```

OBS	CAT	X
1	A	2
2	A	6
3	B	10
4	B	13
5	B	20

Diane E. Brown
AdminaStar Solutions
9525 Delegates Row
Indianapolis, IN 46240
Phone: (317) 581-4159
Fax: (317) 581-4113