



ТЕОРИЈА УЗОРАКА 5

10. 05. '13.

СИСТЕМАТСКИ УЗОРАК

SYSTEMATIC SAMPLING

- Постоји колекција планова узорковања под заједничким именом систематски / периодични узорак. Они имају неколико предности приликом примене у пракси.
- У најкраћем, код систематског узорка, уместо избора n јединица из популације на случајан начин, одлучивање о јединицама које ће се наћи у узорку врши на основу избора само једног случајног броја.



- Linear Systematic Sampling

Идеја се састоји у следећем:

Претпостави се да се популација састоји од N јединица и да су оне означене бројевима од 1 до N , по неком реду, а да треба одабрати узорак обима n . Јединице се, затим, у узорак бирају периодично – у једнаким размацама. Наиме, ако је интервал (период, корак) избора k , од првих k јединица бира се (најчешће на случајан начин) једна јединица као почетна, нека је то нпр. јединица r , а потом свака k -та јединица.

Тако се формира узорак од јединица означених са $r, r + k, \dots, r + (n - 1)k$.

- Корак избора зависи, очигледно, од обима популације N и од обима узорка n , тако да се обично узима да је
$$k = \frac{N}{n} = \frac{1}{f}$$



- Јасно је да се систематски узорак може замислити и као стратификован узорак, код кога је популација подељена на n стратума, од којих се сваки састоји од по k јединица. При томе, из сваког стратума се у узорак бира по једна јединица.
- Стога се може очекивати да систематски узорак пружа исту прецизност као и стратификовани случајни узорак са једном јединицом у сваком стратуму.
- Ипак, систематски узорак може, у неким ситуацијама, бити и лошији од стратификованог, јер се код њега стратуми формирају без разматрања интерне хомогености јединица.
- Извесне тешкоће, могу се јавити када N није умножак од n , тј. $N \neq k \cdot n$, али се у конкретној ситуацији могу направити мања одступања од задатог обима узорка.



- Предности у односу на прост случајан узорак:
 - правило одабира јединица у узорак је сасвим једноставно, не захтева генерисање (псеудо)случајних бројева велики број пута ни потпуну нумерацију јединица у популацији
 - број могућих узорака је знатно мањи и до њих се, по правилу, једноставније долази
 - интуитивно је прихватљив – “равномерно” је распоређен на целој популацији, не допушта случајна груписања или пропуштање заступљености неких делова популације
 - у извесним ситуацијама, које се често сусрећу у пракси, има мању стандардну грешку



- Оцена тотала:

Непристрасна оцена тотала обележја популације Y код систематског узорка, са случајно одабраном почетном јединицом r је

$$\hat{Y}_{lss} = \frac{N}{n} \sum_{j=1}^n Y_{r+(j-1)k}$$

а њена дисперзија је

$$V(\hat{Y}_{lss}) = \frac{1}{k} \sum_{r=1}^k [\hat{Y}_r - Y]^2$$

где је \hat{Y}_r вредност \hat{Y}_{lss} , која одговара r -том узорку (тј. узорку који као почетну има јединицу r).

- Могућност одређивања дисперзије, код раније приказаних планова узорковања заснивала се на случајном избору јединица популације, што овде није случај.



- Ако се претпостави да је списак јединица популације на основу кога се бира узорак уређен на случајан начин, или да је обележје, које се посматра, независно од ознака јединица у популацији, систематски узорак може се поистоветити са простим случајним узорком и могу се применити истоветне формуле за оцењивање дисперзије.
- У пракси се најчешће тако и ради, понекад без довољно основа, што може довести до стицања потпуно погрешне слике о обележју од интереса (нпр. када популација има периодичан тренд).
- Systematic sampling performs badly when the list is ordered in cycles of values of the survey variables and when the sampling interval coincides with a multiple of the length of the cycle.



ГРУПНИ УЗОРАК

CLUSTER SAMPLING

- Код свих до сада описаних планова узорковања јединице су у узорак биране из целе популације или из стратума, на које је претходно подељена популација. Ово је погодно за истраживања малих размера, али не и за велика и комплексна истраживања. Главни разлог је тај што код популација великог обима обично не постоји употребљив списак свих јединица у популацији, на основу кога би се могао добити прост случајан, систематски или стратификован узорак. Чак и када постоји комплетан списак јединица, поменуте технике узорковања нису економичне за примену у популацијама великог обима.



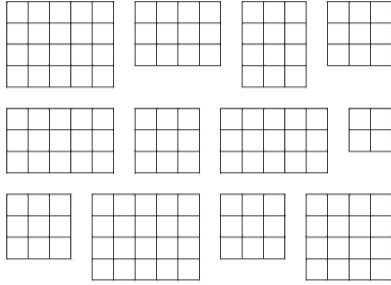
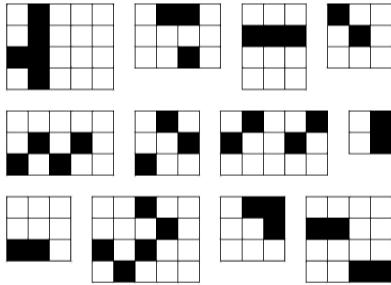
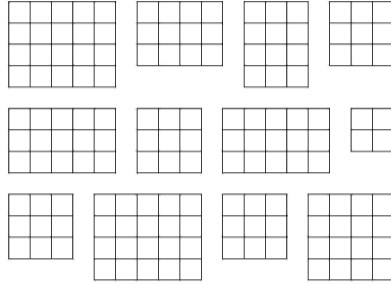
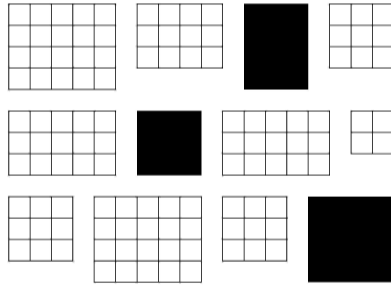
- Single-Stage Cluster Sampling

Идеја се састоји у следећем:

Популација се подели, по неком принципу, на више дисјунктних делова, који се називају **примарне јединице** (групе, скупине, серије, кластери), свака примарна јединица састоји се од **секундарних јединица** (то су, овде, јединице популације). Затим се неким вероватносним планом узорковања одабере одређени број група из којих се посматрају СВИ елементи и тако формира групни узорак.

- Кластери показују површну сличност са стратумима јер кластер, као и стратум, представља групу елемената популације. Међутим, суштинска разлика између групног и стратификованог узорка очигледна је у самом поступку селекције, тј. одабира јединица.



Stratifikovano uzorkovanje	Klaster uzorkovanje
<p>Svaki element populacije se nalazi u tačno jednom stratumu.</p>	<p>Svaki element populacije se nalazi u tačno jednom klasteru.</p>
<p>Populacija od H stratuma: stratum h ima N_h elemenata</p>  <p>Uzima se SRS iz svakog stratuma:</p> 	<p>Jednofazno klaster uzorkovanje: populacija od N klastera</p>  <p>Uzima se SRS klastera i posmatraju se svi elementi unutar izabranih klastera:</p> 
<p>Disperzija ocene srednje vrednosti populacije \bar{y}_U zavisi od promenljivosti unutar stratuma.</p>	<p>Klaster je uzoračka jedinica. Što više klastera uzorkujemo, manja je disperzija. Disperzija ocene srednje vrednosti populacije \bar{y}_U zavisi prvenstveno od promenljivosti između sredina klastera.</p>
<p>Za postizanje veće preciznosti, pojedinačni elementi unutar svakog stratuma treba da budu što sličniji, a sredine stratuma treba da se razlikuju što je više moguće.</p>	<p>Za postizanje veće preciznosti, pojedinačni elementi unutar svakog klastera treba da budu što različitiji, dok sredine klastera treba da budu što sličnije.</p>

“Кластер узорци са једнаким вероватноћама” - Мастер рад,
 Дуња Арсић, Универзитет у Новом Саду



- Резимирано, да би оцене биле прецизне потребно је да кластери по својој структури што боље одсликавају популацију. У томе се састоји и њихова главна разлика у односу на стратуме, који се формирају као интерно хомогене групе јединица.
- Стратификован узорак се користи за доношење што прецизније оцене непознатих параметара, док се групни узорак користи када је потребно смањити трошкове истраживања.
- Са друге стране, испитивањем свих јединица кластера, делимично се понавља иста информација, тј. добија се мање нових информација него што је случај код простог случајног узорка.
- Према томе, групни узорак је мање прецизан од стратификованог и простог случајног узорка.



- Претпостави се да је популација подељена на L кластера, при чему i -ти кластер садржи N_i јединица, $i = 1, 2, \dots, L$. Нека је Y_{ij} , $j = 1, 2, \dots, N_i$, $i = 1, 2, \dots, L$ вредност обележја у j -те јединице i -тог кластера а

$$Y_i = \sum_{j=1}^{N_i} Y_{ij}$$

је тотал обележја i -тог кластера, $i = 1, 2, \dots, L$.

Даље, од L кластера се врши одабир простог случајног узорка s обима n (кластера).

- Оцена тотала:

Непристрасна оцена тотала обележја популације Y код групног узорка је

$$\hat{Y}_{cls} = \frac{L}{n} \sum_{i \in s} Y_i$$



- Други могући приступ јесте да се од L кластера не врши одабир простог случајног узорка, него узорка са вероватноћама пропорционалним величини са понављањем, обима n . Наиме, идеја је да се број јединица унутар сваког кластера посматра као помоћна променљива – “величина”.

- Вероватноћа избора r -тог кластера је, тада, $p_r = \frac{N_r}{N}$
 $r = 1, 2, \dots, L$ где је

$$N = \sum_{i=1}^L N_r$$

- Систематски узорак је, заправо, специјалан случај групног узорка, код кога се од k кластера бира тачно један, као прост случајан узорак, а затим посматрају вредности испитиваног обележја на свим јединицама у том кластеру.



ВИШЕЕТАПНИ УЗОРАК

MULTISTAGE SAMPLING

- Претпоставља се да је популација подељена на одређен број, нпр. L , примарних јединица. Ако се прво бира узорак одређеног обима, нпр. n , примарних јединица, а затим бира узорак од секундарних јединица из сваке изабране примарне јединице, такав план узорковања назива се двоетапни узорак (two-stage sampling).
- Предност се даје примени двоетапног узорка у односу на групни узорак, у ситуацијама када су кластери велики (у смислу садрже велики број секундарних јединица) или када су секундарне јединице унутар кластера веома сличне, па испитивање СВИХ секундарних јединица садржаних у примарној јединици може бити скупо и непотребно.



- Понављањем описаног поступка добија се вишеетапни узорак (multistage sampling).
- Пример троетапног узорка:

Врши се анкетирање ученика средњих школа у неком граду. Прво се одабере узорак школа, затим узорак одељења из одабраних школа, и на крају узорак ученика у одабраним одељењима.
- Термин ‘примарне јединице’ резервисан је за највеће групе, тј. за делове популације из којих се избор врши у првој етапи. Наредне у хијерархији су ‘секундарне јединице’, потом ‘терцијарне јединице’ итд. У последњој етапи формира се узорак кога чине ‘праве јединице’ популације.

