

Статистика у библиотекама

Узорковање

Прост случајан узорак

Ако се из популације бира ограничен скуп елемената тако да се елементи бирају независно један од другог и имају подједнаке шансе да буду изабрани, тада говоримо о простом случајном узорку.

У случају када постоји списак елемената популације са редним бројевима, избор простог случајног узорак може се обавити коришћењем таблице случајних бројева (Табела 1). При томе се на случајан начин бира почетак у таблицу случајних бројева (тј. одређени ред и колона таблице), а онда се од тог места у неком смеру (рецимо слева надесно) читају цифре и групишу у бројеве који имају исти број цифара као величина цифара. Случајни бројеви који су већи од величине популације или су већ раније изабрани се одбацују. Поступак се завршава када се изабере онолико различитих случајних бројева колика је унапред одређена величина узорка.

Пример 1. Претпоставимо да бирамо узорак величине $n = 10$ из популације од $N_p = 800$ елемената. Из таблице случајних бројева (Табела 1) бирамо троцифрене бројеве мање од 800 (претпостављамо да су редни бројеви елемената популације од 000 до 799). Почетак бирамо случајно (на пример, жмурећи положимо прст на таблицу). Претпоставимо да је изабран почетак петог реда друге колоне (Слика 1).

316731	687660	240962	772199	054300	618250	297479	357381	777593	210606
263710	443996	549102	643510	011625	870996	552609	398859	466762	531123
330952	628071	476429	090470	267959	570629	273368	604753	835600	584410
303250	842365	697141	287942	345507	452166	743293	876650	995459	479288
403396	253304	644149	431214	465551	994028	121200	735312	440356	783380

Слика 1: Пети ред табеле случајних бројева (Табела 1)

Читајући тај ред бирамо са списка елементе популације са редним бројевима:

253, 304, 644, 149, 431, 214, 465, 551, 028, 121

(прескачемо број 994 јер је већи од величине популације $N_p = 800$). ◇

Систематски узорак

Систематски узорак је добио назив по томе што се његови елементи бирају по неком унапред одређеном систему. Систем користи списак објеката популације којим је сваком објекту придружен одређен редни број. Ако је популација величине N_p , за редни број rb важи $0 \leq rb < N_p$. Редни број првог члана узорка на списку (у ознаци p) бира се случајно, на пример помоћу таблице случајних бројева (Табела 1). Потом се се одређује корак систематског узорка (у ознаци k) као количник величине популације N_p и величине узорка N (ако $N_p : N$ није цео, заокружује се на мањи цео број).

Даљи поступак зависи од тога да ли је списак направљен тако да су елементи популације распоређени на случајан начин или не.

Ако елементи популације имају случајан распоред на списку, полазећи од елемента на списку са редним бројем p , бирају се редом елементи са редним бројевима $p + k$, $p + 2k$, $p + 3k$, ..., тј. сваки k -ти елемент.

Пример 2. Нека је величина популације $N_p = 250$ и из ње треба одабрати узорак величине и $N = 25$. Тада је корак $k = N_p : N = 250 : 25 = 10$. За редне бројеве важи $0 \leq rb < 250$, а такође и за почетни редни број (први елемент узорка) $0 \leq p < 250$.

Ако је распоред на списку случајан, а изабран је $p = 017$ као редни број првог члана узорка, тада се у узорак редом бирају елементи популације чији су редни бројеви на списку $17 + 10 = 27$, $27 + 10 = 37$, $37 + 10 = 47$ итд. $\diamond$

Ако елементи популације немају случајан распоред на списку, а списак је већ у форми дигиталног текста (табеле са колонама за редни број и елемент популације), онда се креира нови списак са случајним распоредом. У ту сврху се у програму за рад са табелама (на пример MS Excel) табели дода посебна колона у којој се сваком елементу популације аутоматски придружи по један случајан број (на пример, функцијом RAND у програму MS Excel¹), а онда се цела табела сортира по тој колони и поново нумерише тако да нови редни бројеви буду у складу са сортираним случајним бројевима. Потом се примењује претходно описани поступак (в. Пример 2).

¹При томе се колона са формулом за генерисање случајних бројева мора заменити колоном одговарајућих вредности иначе ће се бројеви у колони променити сваки пут кад се функција RAND поново израчунава, конкретно приликом сортирања.

Стратификован узорак

Понекад се популација састоји из мањих, хомогених група (стратума, од латинског *stratum* = слој, *strata* = слојеви) таквих да се елементи исте групе не разликују много по својству које се мери, али се елементи различитих група прилично разликују по том својству. Ако је (тачно или приближно) познат однос величина појединачних стратума у популацији, тада се стратификован узорак бира тако што се из сваког стратума бира прост случајан узорак, тако да у коначном узорку однос величина стратума буде приближно исти као у популацији и такав узорак се назива пропорционалан стратификован узорак. Ако пак однос стратума није познат, једна могућа техника је да се из сваког стратума бира подједнак број објеката и тада се добија равномеран стратификован узорак.

Пример 3. Претпоставимо да популација садржи 3.600 становника неког места, при чему својство које се мери зависи од узраста испитаника. Тада се популација може поделити према старости испитаника и претпоставимо да су уочена три стратума: испитаници млађи од 30 година, испитаници између 30 и 60 година и старији од 60 година. Нека је однос ових стратума $1 : 2 : 3$. То значи да се цела популација може поделити на $1 + 2 + 3 = 6$ једнаких делова од којих један припада првом стратуму, следећа два другом стратуму и последња три трећем стратуму, односно први стратум има $3.600 : 6 = 600$ чланова, други стратум $2 \cdot (3.600 : 6) = 2 \cdot 600 = 1.200$ чланова и трећи стратум $3 \cdot (3.600 : 6) = 3 \cdot 600 = 1.800$ чланова.

Нека је сада потребно изабрати пропорционални стратификован узорак од 60 испитаника. У том случају у узорку треба да важи исти однос величина стратума, тј. $1 : 2 : 3$, што значи да из првог стратума треба изабрати $60 : 6 = 10$ чланова, из другог стратума $2 \cdot (60 : 6) = 2 \cdot 10 = 20$ чланова и из трећег стратума $3 \cdot (60 : 6) = 3 \cdot 10 = 30$ чланова.

Ако се пак бира равномеран стратификован узорак, тада се из сваког од три стратума бира подједнак број чланова, тј. $60 : 3 = 20$ чланова. ◇

Табела 1: Табела случајних бројева

316731	687660	240962	772199	054300	618250	297479	357381	777593	210606
263710	443996	549102	643510	011625	870996	552609	398859	466762	531123
330952	628071	476429	090470	267959	570629	273368	604753	835600	584410
303250	842365	697141	287942	345507	452166	743293	876650	995459	479288
403396	253304	644149	431214	465551	994028	121200	735312	440356	783380
281743	447373	027033	149155	562678	942521	382259	723942	399303	821464
637240	053437	977798	677630	972795	171543	024961	275517	390840	979203
767678	546338	743135	432365	383001	316602	752368	553132	739997	731869
800059	567946	440546	692833	200968	197635	410333	102762	165206	204467
588259	260650	839056	210180	922441	303810	080559	095250	062426	097369
092646	127652	753748	138947	511196	345037	557439	612144	747366	608378
350966	361903	104574	356361	096867	436724	572103	133192	306916	422909
282166	086927	298801	445483	007159	232769	793236	269692	190041	600612
990750	682386	946319	736977	487980	137636	167208	018518	031513	823171
105345	225755	236640	579268	481823	365842	853386	216001	079418	654109
401755	920919	338552	411495	468937	656257	132290	455099	208351	112264
417424	317002	167757	362659	343425	896243	215305	044409	315150	823175
769986	277481	920883	077881	642033	977888	941584	509332	776849	246466
578593	954913	789727	300309	616925	583330	065849	549719	503680	649553
175222	890692	767027	449640	001926	675238	486880	234880	878285	824064
183298	368459	236601	316373	225007	675873	734301	317611	111928	206799
996425	580563	920308	483188	527181	200966	528525	173147	468832	522443
983763	891474	064571	225039	462895	336836	252771	385994	465616	525456
725205	052862	445608	664523	957700	835768	346872	362758	796084	629643
718744	063832	239448	585008	193374	915223	562086	982638	375345	003552
201226	798620	412897	707877	884106	894055	871984	919388	598549	588136
592445	411746	697280	866536	981300	760222	728070	093889	652876	484097
160161	114918	482783	880387	767062	311314	214026	151685	495462	693042
374564	157383	203599	710811	629148	098525	397655	908611	686943	644884
182630	924241	161334	802246	846367	342471	824602	745854	606161	360883
308353	720316	664931	382313	679401	848785	644178	442671	420844	930169
010926	428664	448834	762242	819273	938309	810333	963928	332113	684917
030801	837536	141488	355983	809155	621260	570054	923212	644376	723074
420853	851776	459429	499506	868386	316193	980722	370722	945539	805828
070602	408261	441909	447532	150431	085886	588139	197888	411578	711075
813403	967587	934101	177606	470034	438025	989296	202080	133433	109113
039575	446096	008458	163519	855807	448670	134298	510698	299790	364760
523600	983834	055768	820370	308425	309281	520040	406702	935576	074740
824902	903540	238658	544527	282741	027336	830937	501280	452378	043172
431220	662774	106927	179000	611553	574490	810887	577605	462910	470110
120803	337528	457468	592537	855654	253412	672616	455066	619652	214447
544127	029034	927635	305829	662844	330770	470255	528869	336024	302923
330125	612277	701006	622243	855759	264754	532121	125099	896203	476714
760170	651341	626594	766289	981261	334746	934245	304114	987612	340834
837948	439488	479982	376987	924797	268562	358008	217125	502063	939760
638560	421064	682496	537897	354957	854618	317416	243854	835161	553407
360926	847341	081757	247483	486113	906676	990753	006183	763006	904496
824115	402267	403271	210614	454844	778216	365540	571358	371861	613866
216180	625264	468535	808415	272305	517624	308697	202494	784958	274261
654198	829424	954543	815572	708967	911878	118009	462402	540546	908127