

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

• نمونه گیری و حجم نمونه

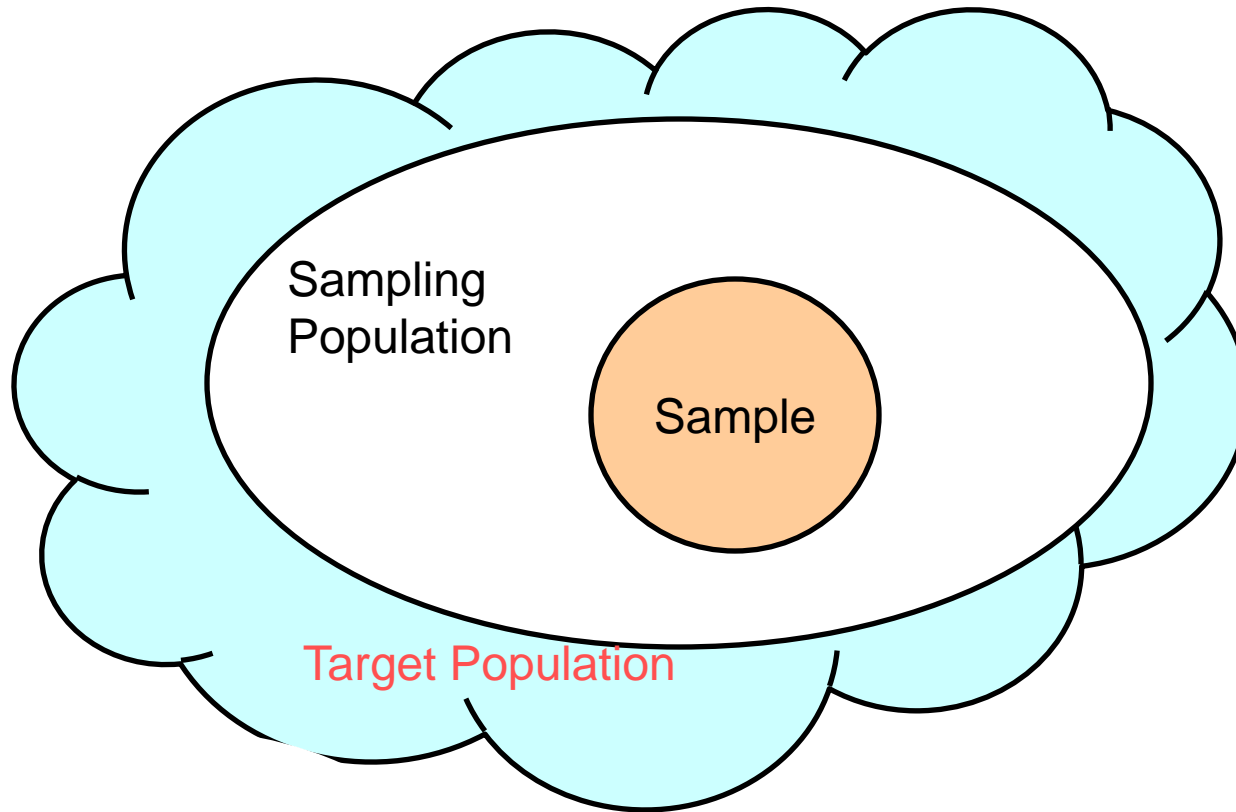
روش تحقیق

Dr. Yousef Alimohamadi
Epidemiologist

جامعه مورد مطالعه

- جامعه مورد مطالعه ، جمعیتی است که مطالعه بر روی آن انجام می شود .
- لزوماً این جامعه انسانها نیستند بلکه می تواند پدیده ها ، اشیا و موجودات زنده باشند.
- مناسب ترین جامعه برای بررسی آن است که کل جامعه تحت مطالعه قرار گیرند. به عبارتی سرشماری انجام شود. در این صورت جمعیت آماری برابر با جمعیت کل جامعه خواهد بود.
- **اما :**
- معمولاً محدودیت های زمانی و اعتباری (هزینه ها) ، دقت در گردآوری داده ها و کنترل آن، نیروی انسانی و تجهیزات و امکانات سبب استفاده از نمونه گیری به جای سرشماری است .

Sampling and representativeness



Target Population → Sampling Population → Sample

نمونه گیری sampling

- نمونه (Sample):
- قسمتی از جمعیت که با ضوابطی مقبول انتخاب می شود.
- مثال: داوری از روی چند CC خون بیمار در باره کل خون بیمار
- مشت نمونه خروار
- در نمونه گیری اصل بر این قرار داده می شود که چنانچه از مناسبات آماری صحیح استفاده شود امکان تعمیم نتایج و اطلاعات بدست آمده از مطالعه بر روی نمونه به جامعه اصلی وجود دارد .

تعریف نمونه:

نمونه عبارتست از تعدادی از افراد جامعه که صفات آنها با صفات جامعه مشابهت داشته و معرف جامعه بوده از تجانس و همگنی با افراد جامعه برخوردار باشند.

Sampling

- چرا نمونه گیری؟

در مطالعه جمعیت های انسانی تمام افراد را به دلایل زیر نمی توان در مطالعه شرکت داد:

- جمعیت بزرگ در جامعه آماری

- هزینه زیاد

- وقت گیر بودن

- مشکل نیروی کارآمد و خیره

- کاهش دقت به علت حجم کار زیاد

ضوابط نمونه گیری

۱. نمونه معرف جمعیت باشد
۲. حجم نمونه به اندازه کافی بزرگ باشد
۳. نمونه تصادفی باشد

در این حالت نتایج حاصل از نمونه به کل جمعیت قابل تعمیم است.

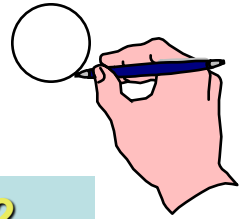
Statistical Terms in Sampling

Statistic



sample

Average = 3.72



Parameter



population

Average = 3.75

انواع نمونه

نمونه احتمالی:

هر فرد یا واحد مورد مطالعه دارای درجه احتمال معینی برای انتخاب شدن است. عامل شانس است که به جای قضاوت محقق، معین می کند که کدام واحد در نمونه وارد شود. بنا بر این، خطا عمدتاً "از مقوله اشتباهات تصادفی (شانس) است که قابل اندازه گیری است و نتایج به کل جامعه مورد نمونه برداری قابل تعمیم است.

نمونه غیر احتمالی:

به جای تکیه بر عامل شانس، نمونه به کمک قضاوت انسانی انتخاب می شود. خطا غیر تصادفی و غیر قابل اندازه گیری (تورش) است و تعمیم پذیری را مختل می کند

انواع نمونه گیری احتمالاتی

۱. نمونه گیری ساده تصادفی
۲. نمونه گیری منظم
۳. نمونه گیری طبقه ای
۴. نمونه گیری خوشه ای
۵. نمونه گیری چند مرحله ای

نمونه گیری تصادفی ساده

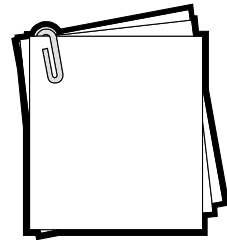
Simple Random Sampling

هر فرد شانس انتخاب برابر با بقیه افراد برای قرار گرفتن در نمونه دارد.
روش کار:

- لیستی از کلیه افراد جمعیت (N) باید در دسترس باشد
- با استفاده از جدول اعداد تصادفی یا ماشین حساب یا قرعه کشی یا کامپیوتر، تعداد نمونه مورد لزوم (حجم نمونه n) را انتخاب کنید.

Simple Random Sampling

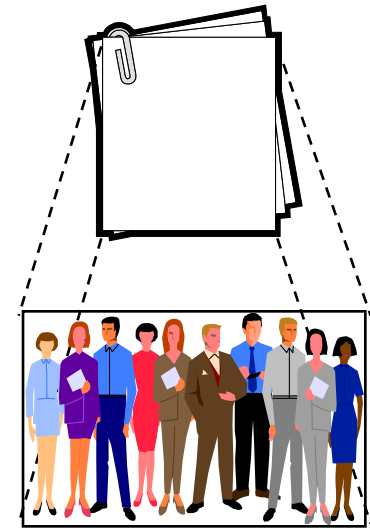
List of Residents



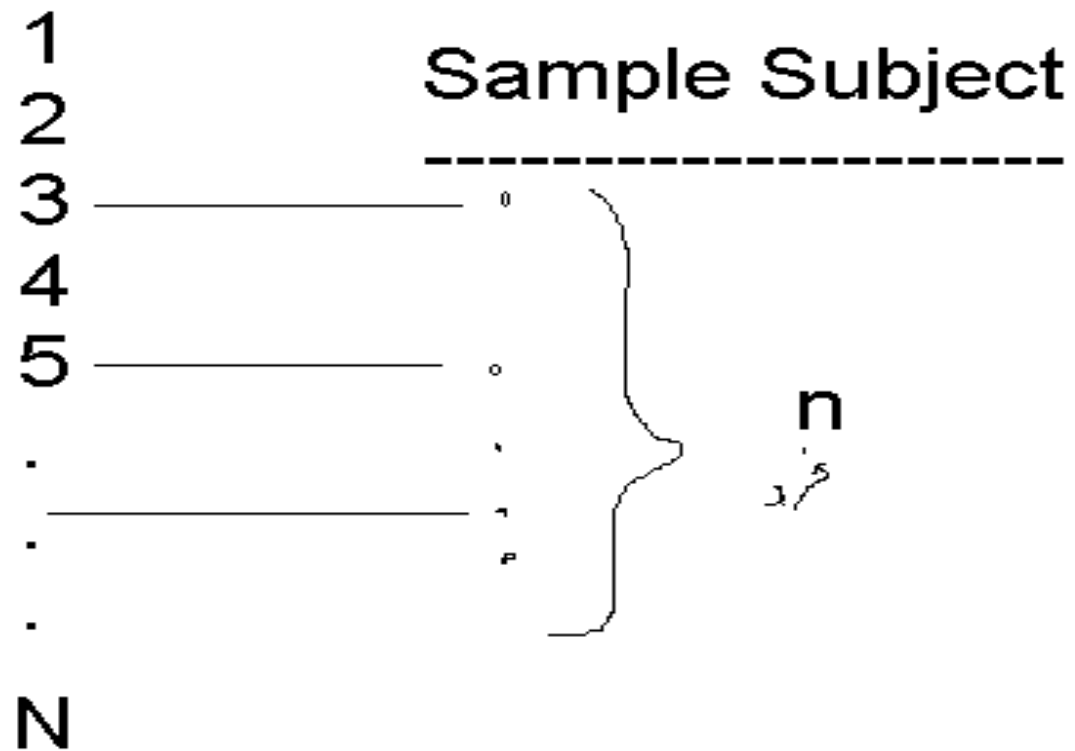
Simple Random Sampling

List of Students

Random Subsample



Pop. Subject



جدول ارقام تصادفی

جدول ارقام تصادفی

01703	49894	57579	98505	85008	98681	56862	41860
87556	95669	39885	31669	31460	96413	84398	31562
84254	60541	73290	54685	80208	77044	14771	33378
12429	43566	32578	38935	75460	98133	18386	12417
63055	26768	63609	92424	50808	95416	12795	50787
18348	79628	05778	72095	90754	90430	00791	38023

19827	95727	02372	23485	54372	89732	67768	72151
30236	52309	99971	44890	28522	92140	40703	16888
32160	42795	04959	73840	99110	07527	73725	19291
14832	30334	18047	38712	32931	85481	15378	25011
21151	02668	44154	95153	63213	70014	67531	52581
89677	82090	42211	75118	36233	25131	13314	33063

67129	12388	41678	51286	80948	91599	52652	02519
27808	23807	25424	35877	96308	45847	88287	88419
24646	88222	66395	24060	98186	81741	08675	36931
10030	79086	89464	28282	89252	14777	02033	42852
26512	51935	86185	75646	51698	89313	57145	85070

43334 27009 27879 73339 74387 14314 42078

مثال:

می خواهیم نمونه ای به حجم $n = 10$ از هفتاد نفر دانشجویان یک کلاس به صورت تصادفی انتخاب نمائید.

ابتدا به هر یک از دانشجویان شماره ای از ۰۱ تا ۷۰ اختصاص می دهیم.

۰۱ ۰۲ ۰۳ ۶۹ ۷۰

چون شماره آخرین فرد دو رقمی است با کمک جدول اعداد تصادفی ده عدد دو رقمی انتخاب می نمائیم. بدیهی است که ارقام دو رقمی بزرگتر از هفتاد در نظر گرفته نمی شود. ارقام تکراری نیز در نظر گرفته نمی شود. به عبارت دیگر:

از یک نقطه تصادفی شروع کرده و ده عدد دورقمی غیر تکراری بین ۰۱ تا ۷۰ پیدا می کنیم.

جدول ارقام تصادفی	01703	49894	57579	98505	85008	98681	56862	41860
	87556	95669	39885	31669	31460	96413	84398	31562
	84254	60541	73290	54685	80208	77044	14771	33378
	12429	43566	32578	38935	75460	98133	18386	12417
	63055	26768	63609	92424	50808	95416	12795	50787
	18348	79628	05778	72095	90754	90430	00791	38023
نقطه شروع	19827	95727	02372	23485	54372	89732	67768	72151
	30236	52309	99971	44890	28522	92140	40703	16888
	32160	42795	04959	73840	99110	07527	73725	19291
	14832	30334	18047	38712	32931	85481	15378	25011
	21151	02668	44154	95153	63213	70014	67531	52581
	89677	82090	42211	75118	36233	25131	13314	33063
	67129	12388	41678	51286	80948	91599	52652	02519
	27808	23807	25424	35877	96308	45847	88287	88419
	24646	88222	66395	24060	98186	81741	08675	36931
	10030	79086	89464	28282	89252	14777	02033	42852
	26512	51935	86185	75646	51698	89313	57145	85070
	43334	27009	27879	73339	74387	14314	42078	

شماره افراد انتخاب شده: 48, 28, 52, 29, 21, 40, 70, 31, 68, 32

بنابر این دانشجویان با شماره های زیر نمونه تصادفی را تشکیل می دهند:

48, 28, 52, 29, 21, 40, 70, 31, 68, 32

محقق می تواند به افراد مذکور مراجعه نموده و بررسی خود (تکمیل پرسشنامه، معاینه بالینی، نمونه برداری برای آزمایش های پاراکلینیکی،) را به انجام رساند.

نمونه گیری سیستماتیک

Systematic Random Sampling

از هر k نفر که همان فاصله نمونه گیری است یک نفر انتخاب می شود.

$$(N/n = k)$$

روش کار:

- از k فرد اول یک نفر به تصادف انتخاب می شود (فرد با شماره a).

- سپس فرد شماره $a+k$ ، $a+2k$ ، ... ، $a+(n-1)k$ را انتخاب کنید.

مثالی از نمونه گیری منظم یا سیستماتیک:

محقق می خواهد از بین افراد یک جامعه دانشجویی ۵۰۰ نفری نمونه ای به تعداد ۵۰ نفر را به روش منظم یا سیستماتیک انتخاب کند.

$k =$ عدد ثابت فاصله بین دو نمونه

$N =$ حجم یا تعداد جامعه

حجم یا تعداد

$n =$ نمونه

$$K = \frac{N}{n} = \frac{500}{50} = 10$$

$$p_2 = p_1 + k \quad \Rightarrow \quad p_2 = 6 + 10 = 16$$

$$p_n = p_{n-1} + k$$

مثال

- از یک جامعه ۱۰۰۰ نفری نمونه ای به حجم ۳۵ نفر نیاز داریم
 - $N=1000$
 - $n=35$
 - $K=1000/35=28.6$
۱. سپس بین ۱ تا ۲۸ یک عدد را به تصادف انتخاب کرده مثلا "عدد ۱۰"
 ۲. بعد عدد $k = 28$ را به ۱۰ اضافه کرده و دومین فرد، نفر ۳۸، سومین ۶۶ و ...
 ۳. ۱۰ و ۳۸ و ۶۶ و ۹۴ و ... و ۹۸۰
 ۴. هنگامی که به عدد ۹۸۰ برسیم ۳۵ نفر ما تکمیل شده است.

Systematic Random Sampling

N = 100

1	26	51	76
2	27	52	77
3	28	53	78
4	29	54	79
5	30	55	80
6	31	56	81
7	32	57	82
8	33	58	83
9	34	59	84
10	35	60	85
11	36	61	86
12	37	62	87
13	38	63	88
14	39	64	89
15	40	65	90
16	41	66	91
17	42	67	92
18	43	68	93
19	44	69	94
20	45	70	95
21	46	71	96
22	47	72	97
23	48	73	98
24	49	74	99
25	50	75	100

Systematic Random Sampling

$N = 100$

want $n = 20$

1	26	51	76
2	27	52	77
3	28	53	78
4	29	54	79
5	30	55	80
6	31	56	81
7	32	57	82
8	33	58	83
9	34	59	84
10	35	60	85
11	36	61	86
12	37	62	87
13	38	63	88
14	39	64	89
15	40	65	90
16	41	66	91
17	42	67	92
18	43	68	93
19	44	69	94
20	45	70	95
21	46	71	96
22	47	72	97
23	48	73	98
24	49	74	99
25	50	75	100

Systematic Random Sampling

$N = 100$

want $n = 20$

$N/n = 5$

1	26	51	76
2	27	52	77
3	28	53	78
4	29	54	79
5	30	55	80
6	31	56	81
7	32	57	82
8	33	58	83
9	34	59	84
10	35	60	85
11	36	61	86
12	37	62	87
13	38	63	88
14	39	64	89
15	40	65	90
16	41	66	91
17	42	67	92
18	43	68	93
19	44	69	94
20	45	70	95
21	46	71	96
22	47	72	97
23	48	73	98
24	49	74	99
25	50	75	100

Systematic Random Sampling

$N = 100$

want $n = 20$

$N/n = 5$

select a random number from 1-5: chose 4

1	26	51	76
2	27	52	77
3	28	53	78
4	29	54	79
5	30	55	80
6	31	56	81
7	32	57	82
8	33	58	83
9	34	59	84
10	35	60	85
11	36	61	86
12	37	62	87
13	38	63	88
14	39	64	89
15	40	65	90
16	41	66	91
17	42	67	92
18	43	68	93
19	44	69	94
20	45	70	95
21	46	71	96
22	47	72	97
23	48	73	98
24	49	74	99
25	50	75	100

Systematic Random Sampling

$N = 100$

want $n = 20$

$N/n = 5$

select a random number from 1-5: chose 4

start with #4 and take every 5th unit

1	26	51	76
2	27	52	77
3	28	53	78
4	29	54	79
5	30	55	80
6	31	56	81
7	32	57	82
8	33	58	83
9	34	59	84
10	35	60	85
11	36	61	86
12	37	62	87
13	38	63	88
14	39	64	89
15	40	65	90
16	41	66	91
17	42	67	92
18	43	68	93
19	44	69	94
20	45	70	95
21	46	71	96
22	47	72	97
23	48	73	98
24	49	74	99
25	50	75	100

نمونه گیری طبقه ای

Stratified Randomized Sampling.

نمونه گیری طبقه ای

Stratified Randomized Sampling.

جمعیت به چند طبقه تقسیم شده (به طور طبیعی) یا تقسیم می شود (به طور مصنوعی). هر طبقه دارای همگنی درونی از نظر متغیر های کنترلی و ناهمگن با طبقات دیگر می باشد. مثل دانشجویان رشته های مختلف و یا افراد جامعه در طبقات اجتماعی اقتصادی

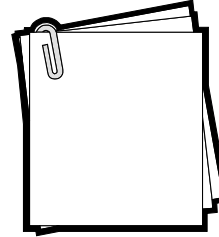
روش کار:

از هر طبقه یک نمونه جداگانه به صورت تصادفی ساده انتخاب می کنیم. حجم نمونه در طبقات می تواند متناسب با حجم جامعه در همان طبقه می باشد.

در صورت امکان بهترین روش نمونه گیری است.

Stratified Random Sampling

List of Residents



Stratified Random Sampling

List of Residents

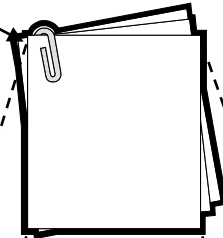
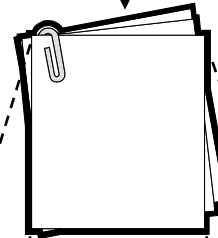
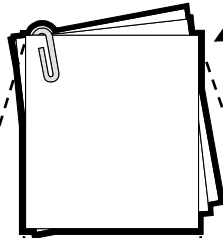


surgical

medical

Non-clinical

Strata



Random Subsamples

Stratified Random Sampling

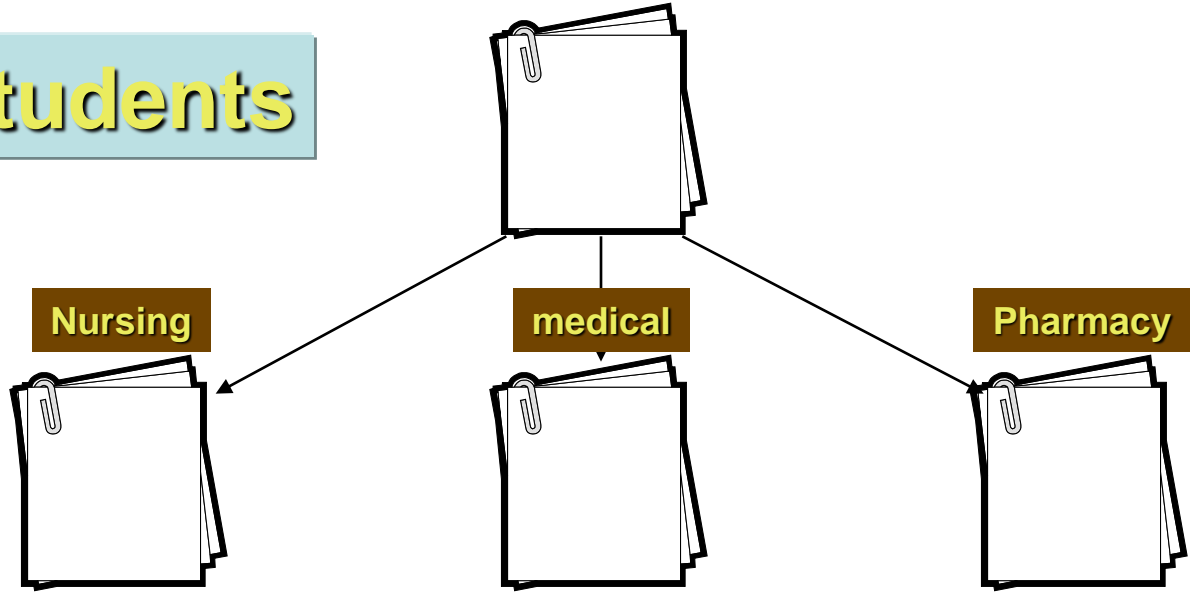
List of students

Nursing

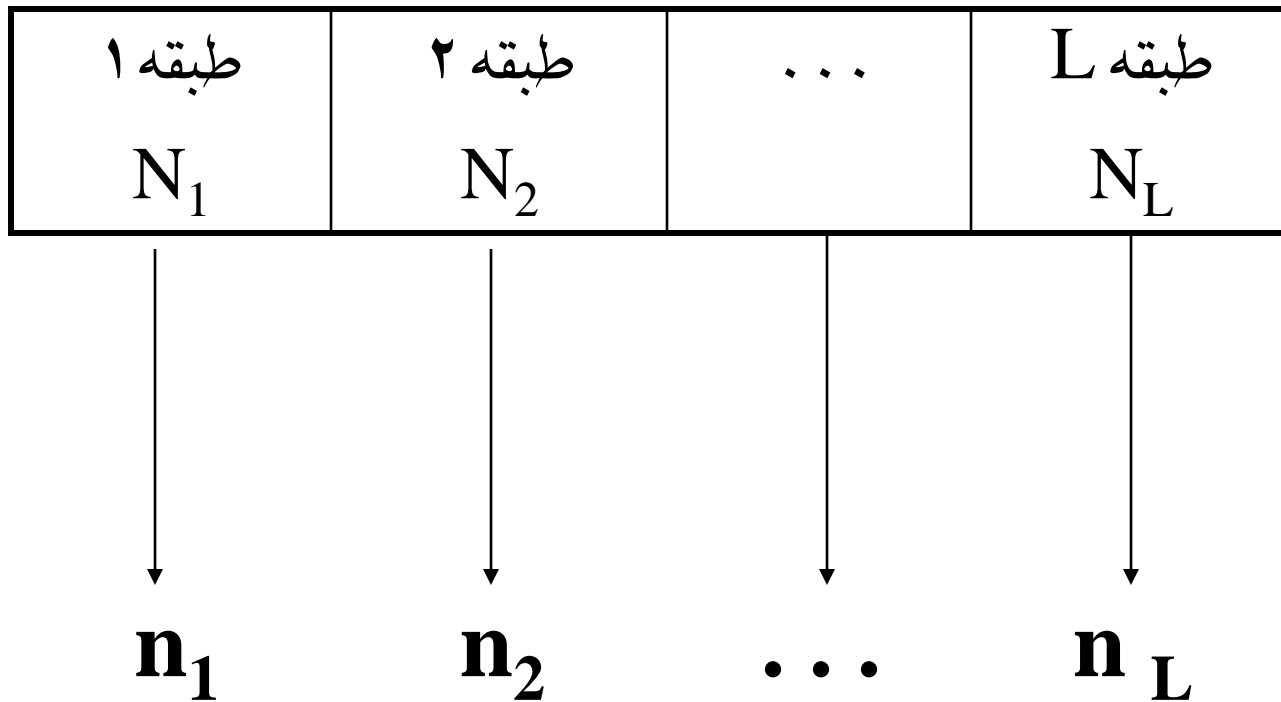
medical

Pharmacy

Strata



نمونه گیری طبقه ای



نمونه گیری خوشه ای

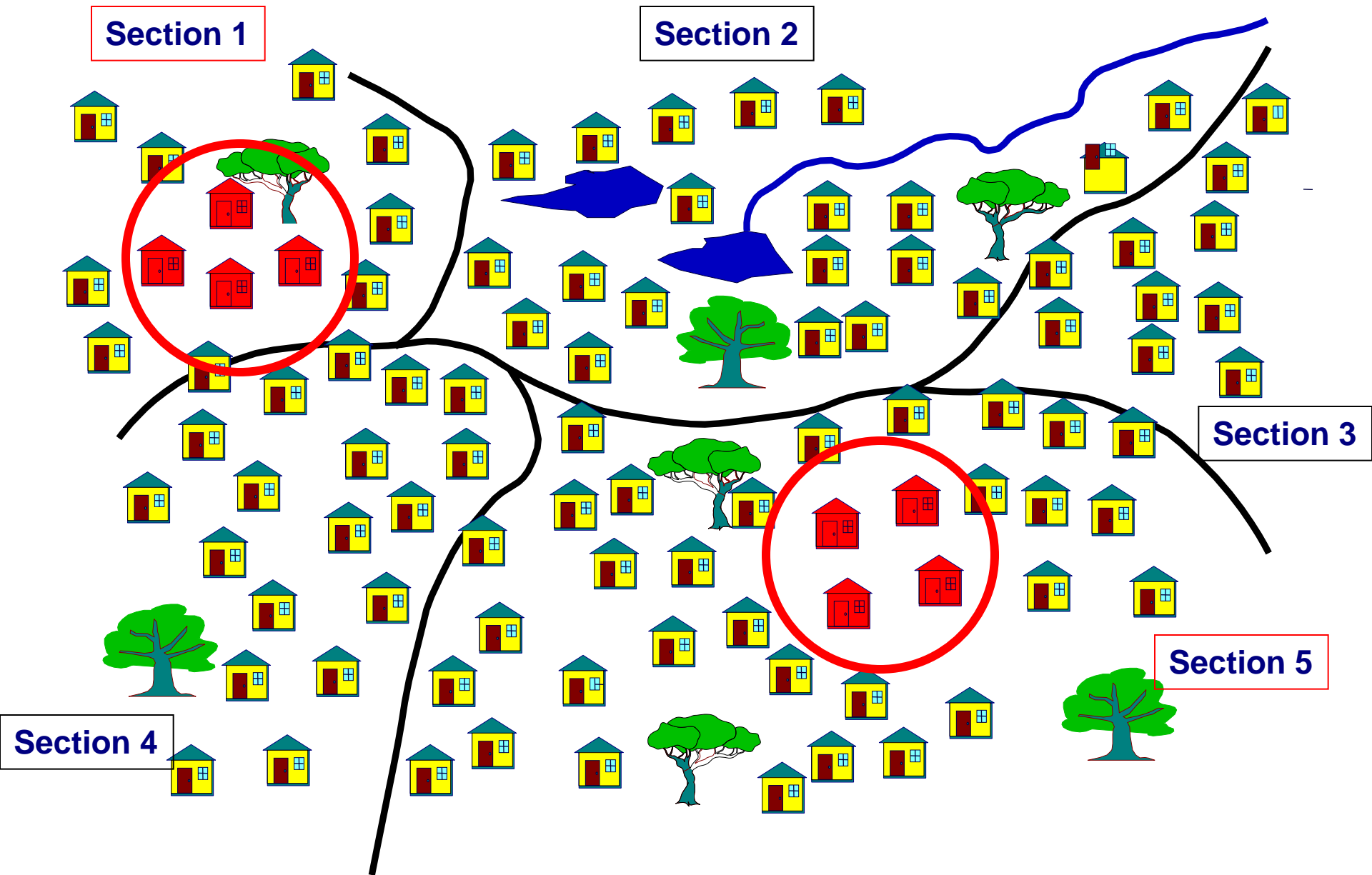
Cluster R. S.

جمعیت به تعداد زیادی خوشه تقسیم می شود. خوشه ها از لحاظ متغیر مورد مطالعه در درون خود ناهمگن و با سایر خوشه ها همگن می باشند.

روش کار:

- بر اساس ویژگیهای خوشه ها و بودجه، وقت، دقت و تعمیم پذیری خوشه ها تعدادی خوشه انتخاب شده و در مرحله بعد به نسبت تعداد نمونه لازم از درون هر خوشه با استفاده از یکی از روشهای نمونه گیری انتخاب می شود.

Example: Cluster sampling



نمونه گیری چند مرحله ای

Multi-stage Sampling

ترکیبی از نمونه گیری های مختلف است. که معمولاً "زیاد کاربرد دارد.

مثال: نمونه گیری از دانش آموزان شهر اراک

مرحله اول: انتخاب هر چهار ناحیه با انواع مدارس (دخترانه و پسرانه، ابتدائی،

راهنمائی، دبیرستان، هنرستان، پیش دانشگاهی، ...) (نمونه گیری طبقه ای یا خوشه ای)

مرحله دوم: انتخاب تعدادی مدرسه (نمونه گیری خوشه ای)

مرحله سوم: انتخاب همه کلاس های مدرسه واقع در نمونه (نمونه گیری طبقه ای)

یا تعدادی از کلاسها درون مدرسه)

مرحله چهارم: انتخاب کلیه دانش آموزان کلاس یا یک نمونه تصادفی به روش

نمونه گیری تصادفی ساده یا سیستماتیک

نمونه گیری خوشه‌ای و چند مرحله‌ای برای زمانی مناسب است:

- ۱) چهارچوب جامعه آماری در اختیار نباشد یا تهیه آن زمان و هزینه زیادی را طلب کند.
- ۲) به لحاظ گستردگی جغرافیایی واحدهای تحلیل و مطالعه، امکان گردآوری اطلاعات نباشد.

انواع نمونه گیری غیر احتمالی

۱- نمونه گیری سهمیه ای (Quota sampling) : سعی می شود ساختار نمونه شبیه به جمعیت باشد (رعایت تناسب موجود در جمعیت).

مثال : انتخاب تعدادی از افراد در گروه های سنی - جنسی

۲- نمونه گیری هدفدار یا قضاوتی (Purposive or Judgement) :
بر اساس دانش و قضاوت محقق واحدهائی انتخاب می شود که در مجموع معرف جمعیت مورد نظر باشد.

۳- نمونه گیری اتفاقی (Haphazard) : واحدهائی که بر حسب اتفاق در دسترس قرار می گیرد.

۴- نمونه گیری آسان : استفاده از نمونه های در دسترس، مثل استفاده از بیماران مراجعه کننده به مطب یک پزشک . . .

Sampling summary

- Random sampling seldom done in practice.
- Stratified sampling yields better results with smaller samples.
- Systematic sampling is easy to manage.

Sample size determination

A question?

Are Females more intelligent than Males?

- H_0 Null hypothesis: Women and Men have the same mean IQ
- H_a Alternative hypothesis: The mean IQ of Women is greater than the Men

Type 1 and 2 errors

Decision	Truth	
	H_0 true	H_0 false
Reject H_0	Type I error	Correct decision
Accept H_0	Correct decision	Type II error

Power

- The easiest ways to increase power are to:
 - increase sample size
 - increase desired difference (or effect size)

Sample size for descriptive survey

Simple random / systematic sampling

$$n = \frac{z^2 * p * q}{d^2} = \frac{1.96^2 * 0.15 * 0.85}{0.03^2} = 544$$

Cluster sampling

$$n = g * \frac{z^2 * p * q}{d^2} = \frac{2 * 1.96^2 * 0.15 * 0.85}{0.03^2} = 1088$$

z: alpha risk expressed in z-score

p: expected prevalence

q: 1 - p

d: absolute precision

g: design effect

Sample size calculation for a difference in means (equal sized groups)

$$n = \frac{(Z_{(1-\alpha/2)} + Z_{(1-\beta)})^2 (sd_1^2 + sd_2^2)}{d^2}$$

Sample size calculation for a difference in proportions (equal sized groups)

$$n = \frac{(Z_{(1-\alpha/2)} + Z_{(1-\beta)})^2 (p_1(1-p_1) + p_2(1-p_2))}{d^2}$$

اندازه نمونه در پژوهش های توصیفی و مقطعی

Sample size for Descriptive and Cross-sectional Studies

موضوع را در دو حالت مجزا مطرح می کنیم:

۱- اندازه نمونه برای برآورد میانگین جامعه

۲- اندازه نمونه برای برآورد نسبت جامعه

فرمول تعیین اندازه نمونه برای برآورد میانگین جامعه

$$n = \frac{z^2 \sigma^2}{d^2}$$

معمولا انحراف معیار جامعه (σ) معلوم نیست
بنابراین انحراف معیار نمونه (از مطالعه مقدماتی یا
مطالعات دیگران) به جای آن قرار داده می شود:

$$n = \frac{z^2 s^2}{d^2}$$

کمیت های فرمول اندازه نمونه (میانگین)

n = تعداد نمونه مورد نیاز

S = انحراف معیار صفت مورد نظر

d = حداکثر خطای قابل قبول در برآورد میانگین

Z = ضریب اطمینان

(Z با استفاده از جدول توزیع نرمال برای سطح اطمینان مشخص تعیین می شود.)

مقدار Z برای برخی سطوح اطمینان

Z^2	ضریب اطمینان (Z)	سطح خطا	سطح اطمینان
2.69	1.64	0.10	0.90
3.84	1.96	0.05	0.95
6.60	2.57	0.01	0.99

برای برآورد میانگین اندازه کالری دریافتی دانش آموزان دوره ابتدایی یک شهرستان چه تعداد نمونه لازم است تا با اطمینان ۹۵٪، فاصله بین مقدار برآورد شده با میانگین واقعی کالری دریافتی در جامعه دانش آموزان ابتدایی شهرستان بیش از ۲۰ کالری نباشد؟ در مطالعه مشابهی انحراف معیار میزان کالری دریافتی دانش آموزان ۷۵ به دست آمده است.

$$n = \frac{z^2 \sigma^2}{d^2} =$$

$$\frac{(1.96)^2 (75)^2}{(20)^2} = 54.02 \cong 55$$

فرمول تعیین اندازه نمونه

برای برآورد نسبت جامعه (با دقت ثابت)

$$n = \frac{z^2 p(1-p)}{d^2}$$

کمیت های فرمول اندازه نمونه (نسبت)

n = تعداد نمونه مورد نیاز

p = برآورد اولیه برای نسبت صفت مورد نظر

d = حداکثر خطای قابل قبول در برآورد نسبت

Z = ضریب اطمینان

(Z با استفاده از جدول توزیع نرمال برای سطح اطمینان مشخص تعیین می شود.)

مثال:

برای برآورد نسبت کودکان دبستانی مبتلا به سوء تغذیه در یک استان چه تعداد نمونه انتخاب کنیم تا با اطمینان ۹۵ درصد خطای برآورد کمتر از ۲ درصد باشد. مطالعه قبلی در استان مشابهی این نسبت را ۲۰ درصد برآورد کرده است.

حل: با توجه به موارد ارائه شده در مثال $p = 0.20$ و $Z = 1.96$ و $d = 0.02$ می باشد:

$$n = \frac{z_{1-\frac{\alpha}{2}}^2 p(1-p)}{d^2} = \frac{1.96^2 (0.2)(1-0.2)}{(0.02)^2} = 1536$$

در مثال فوق اگر خطاي قابل قبول در برآورد نسبت را
در ۰.۰۴/۰ در نظر بگيريم يعني خطا را دو برابر كنيم
تعداد نمونه مورد نياز به يكچهارم يعني ۳۸۴ نفر
تقليل پيدا خواهد كرد.

$$n = \frac{1.96^2 (0.2)(1-0.2)}{(0.04)^2} = 384$$

- مثال : حجم نمونه برای برآورد میانگین پروتئین مصرفی چقدر باشد تا با اطمینان ۹۵٪ دقتی ۵ گرمی داشته باشیم در صورتی که $S = 20$

$$n = \frac{(1/96)^2 \times 20}{5^2} = 62$$

$$n = \frac{Z_{1-\alpha/2}^2 P(1-P)}{d^2}$$

متغیر کیفی

حجم نمونه برای برآورد سوء تغذیه

$$P = 2/0$$

$$d = 0.2/0$$

$$Z = 96/1$$

$$n = \frac{(1/96)^2 \times 0/2 \times (1-0/2)}{(0/0.2)^2} = 1536$$

• اگر دو جامعه داشته باشیم :

$$Z_1 = Z_{1-\alpha/2}$$

$$Z_2 = Z_{1-\beta}$$

$$n = \frac{(Z_1 + Z_2)^2 (S_1^2 + S_2^2)}{(\mu_1 - \mu_2)^2}$$

$$n = \frac{(Z_1 + Z_2)^2 [(P_1(1 - P_1) + P_2(1 - P_2))]}{(P_1 - P_2)^2}$$

• مثال : مقایسه تعداد ضربان قلب در دو گروه ورزشکار و غیر ورزشکار

$$\bar{X}_1 = 72 \quad \text{میانگین}$$

$$\bar{X}_2 = 67 \quad \text{میانگین}$$

$$S_1^2 = 120$$

$$S_2^2 = 105$$

$$n = \frac{(1/96 + 1/84)^2 (120 + 105)}{(72 - 67)^2}$$

$$n = 70/2 \neq 71 \quad \text{برای هر گروه}$$

• مقایسه شیوع سوء تغذیه در دو منطقه

$$P_1 = 1/0$$

$$P_2 = 3/0$$

$$n = ?$$

$$n = \frac{(1/96 + 0/84)^2 (0/1 \times 0/9 + 0/3 \times 0/7)}{(0/3 - 0/1)^2}$$

$$= 58/8 \cong 59$$

