

نمونه گیری

دکتر محمد اصغری جعفر آبادی
دانشیار گروه آمار و اپیدمیولوژی
دانشگاه علوم پزشکی تبریز

مطالب این جلسه

- تعریف نمونه گیری
- چرا نمونه گیری می کنیم؟
- مفهوم نمونه “معرف”
- روش های نمونه گیری
 - غیر احتمالی
 - احتمالی
- محاسبه حجم نمونه

چرا نمونه گیری می کنیم؟

- کاهش هزینه ها
- افزایش سرعت
- افزایش دقت
- غیر عملی بودن سرشماری در بعضی شرایط

تعریف

نمونه گیری فرآیندی است که در طی آن بعضی
از اعضای یک جمعیت به عنوان نمایندگان
معرف کل آن جمعیت انتخاب می شوند.

Sample

نمونه :

گروه یا مواردی که اطلاعات از آنها به دست می آید .

Population

جامعه :

گروه بزرگتری که درصدد هستیم یافته های حاصل از بررسی و مطالعه را به آنها تعمیم دهیم .

جمعیت نمونه برداری شده

- * محیط پژوهش (جامعه آماری یا جامعه تحت مطالعه): جمعیتی که نمونه ها واقعا از آن انتخاب شده است.
- * جمعیت هدف: مجموعه اشخاص یا عناصری است که معیارهای نمونه گیری را دارد.

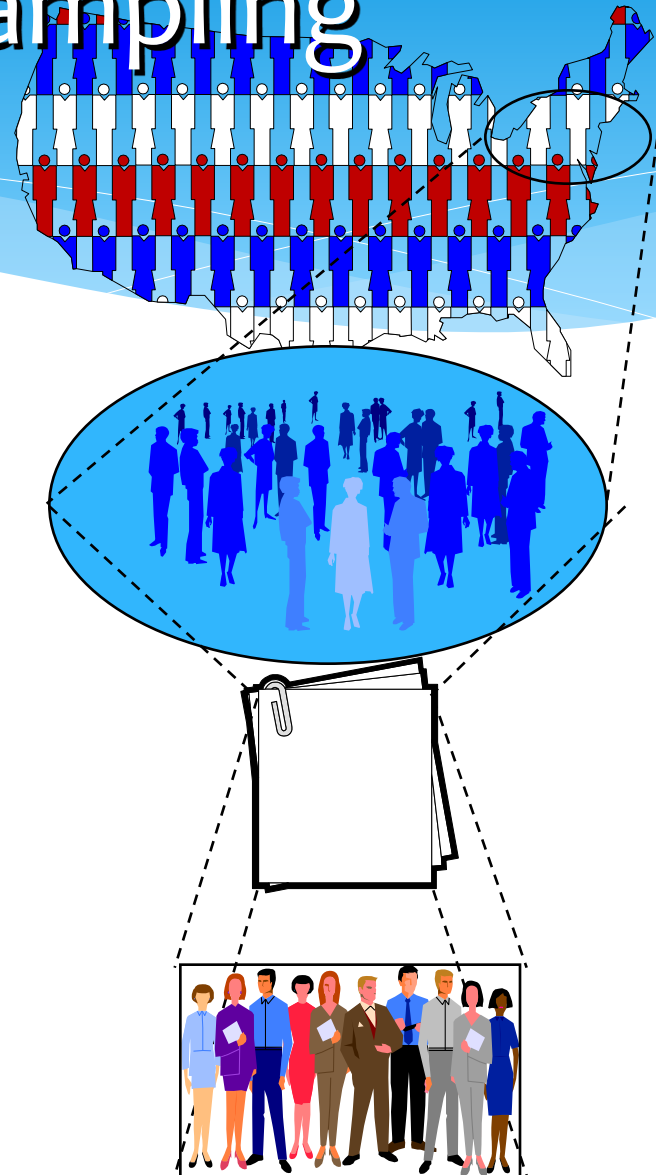
Groups in Sampling

The Target Population

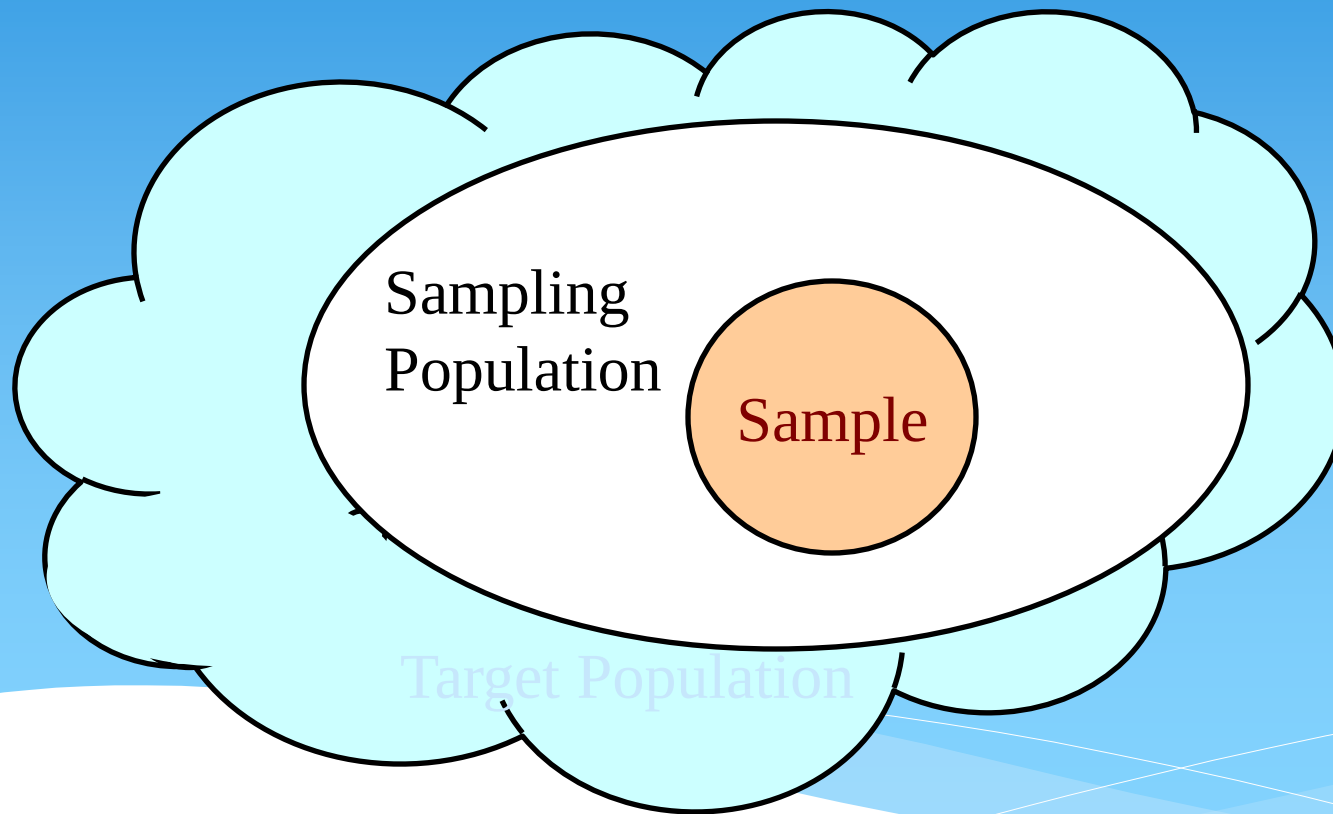
The Study Population

The Sampling Frame

Who is in your study?



نمونه گیری و معرف بودن



Target Population → Sampling Population → Sample

تعریف واژه ها

واحد نمونه گیری (sampling unit): فرد یا سوژه مورد بررسی که داده ها در مورد آن جمع آوری می شود.

نسبت نمونه گیری (sampling fraction): نسبت تقسیم تعداد نمونه به تعداد جامعه

چارچوب نمونه گیری (sampling frame): فهرست کل اعضای جامعه که شامل کل واحدهای نمونه گیری ممکن است.

رویه نمونه گیری (sampling scheme): روشی که بر مبنای آن واحدهای نمونه گیری از چارچوب نمونه گیری انتخاب می شوند

معرف بودن نمونه (representativeness)

از نظر:

■ شخص:

- ویژگی های جمعیتی (سن، جنس، ...)
- مواجهه ها

■ مکان: (شهر / روستا)

■ زمان:

• فصل

• روز هفته

• ساعت روز

روش های نمونه گیری

■ غیر احتمالی (Non-probability sampling)

■ احتمالی (Probability sampling)

نمونه گیری غیر احتمالی

- روش آسان یا در دسترس (Convenience Sampling)
- روش سهمیه ای (Quota Sampling)
- روش مبتنی بر هدف (Purposeful S.)
- روش گلوله برفی (Network S.)

احتمال انتخاب شدن مشخص نیست

روش های نمونه گیری غیر احتمالی

- *
نمونه حاصل معرف جامعه نمی باشد.
- * ممکن است بعضی از واحدها زیاد و برخی کم انتخاب شوند
- * نتایج بدست آمده را نمی توان با ادعای قوی به جامعه هدف تعمیم داد.
- * روش مناسب برای برآورد پارامترهای جامعه نمی باشد.

نمونه گیری غیر احتمالی

* نمونه گیری آسان convenience

نمونه گیری غیر احتمالی

* نمونه گیری سهمیه ای Quota sampling

* اگر پژوهشگر تلاش کند با آوردن کلیه اجزای شناخته شده جامعه در نمونه خود، که از افراد در دسترس انتخاب شده اند، به نمونه جامعیت بدهد

نمونه گیری غیر احتمالی

* نمونه گیری گلوله برفی Snowball sampling

روش های نمونه گیری احتمالی

- نمونه گیری تصادفی ساده Simple random sampling
- نمونه گیری تصادفی منظم Systematic sampling
- نمونه گیری طبقه بندی شده Stratified sampling
- نمونه گیری خوشه ای Cluster sampling
- نمونه گیری چندمرحله ای Multistage sampling

نمونه گیری احتمالی

■ شانس هر فرد برای ورود به نمونه معین، و نامساوی صفر است.

■ امکان سوگیری انتخاب را کم می کند

■ اجازه می دهد که از تئوری های آماری در تحلیل ها استفاده شود

خطاهای نمونه گیری

- خطای تصادفی: خطایی که بر حسب اتفاق و شانس رخ داده است و در هر نوبت اندازه گیری با دفعه قبل فرق می کند.
- مثال اندازه گیری قد کودکان
- خطای منظم، سوگیری: خطا جهت و اندازه معینی را دارد.
- مثال ترازو در اندازه گیری قد کودکان

روش های نمونه گیری احتمالی

- نمونه گیری تصادفی ساده Simple random sampling
- نمونه گیری تصادفی منظم Systematic sampling
- نمونه گیری طبقه بندی شده Stratified sampling
- نمونه گیری خوشه ای Cluster sampling
- نمونه گیری چندمرحله ای Multistage sampling

نمونه گیری تصادفی ساده

■ شانس انتخاب همه واحدهای نمونه گیری مساوی است

■ روش اجرا:

- تهیه فهرست کل اعضای جامعه (چارچوب نمونه گیری)

- انتخاب تصادفی هر نمونه

نمونه گیری تصادفی ساده

■ مزایا

- آسان است: برآوردها به سادگی محاسبه می شوند.
- خطای تصادفی (sampling error) به سادگی قابل اندازه گیری است

معایب

- * احتیاج به چارجوب نمونه گیری دارد
- * ممکن است نمونه بهترین معرف جامعه نباشد. (مثال میانگین هموگلوبین خون دانشجویان)
- * واحدهای نمونه انتخاب شده پراکنده هستند

برای انجام نمونه گیری تصادفی ساده باید:

- * فهرستی شماره دار از واحدهای جامعه که قرار است نمونه از آنها انتخاب شود تهیه گردد.
- * اندازه نمونه مشخص می شود.
- * برای هر واحد جامعه نمونه گیری یک عدد در نظر گرفته می شود.
- * واحدهای مورد نیاز با قرعه کشی یا جدول اعداد تصادفی انتخاب شوند.

نمونه گیری تصادفی ساده

■ مثال: می خواهیم شیوع پوسیدگی دندان را در بین ۱۲۰۰ نفر دانش آموزان یک مدرسه تعیین کنیم.

- فهرست دانش آموزان مدرسه را تهیه کنیم

- دانش آموزان را از ۱ تا ۱۲۰۰ شماره گذاری کنیم

- اگر حجم نمونه = ۶۰ نفر باشد، بایستی ۶۰ عدد تصادفی را بین ۱ تا

۱۲۰۰ انتخاب کنیم.

چگونه؟؟



آقای فردوسی پور در برنامه نود

قرعه کشی

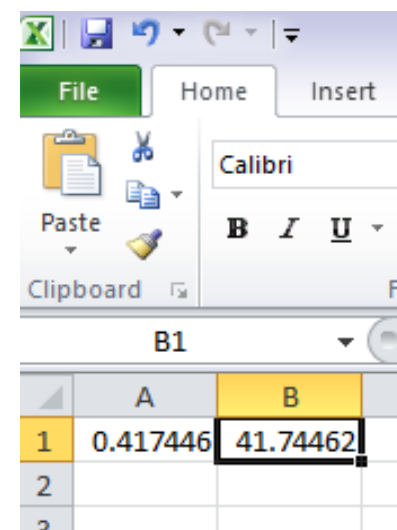
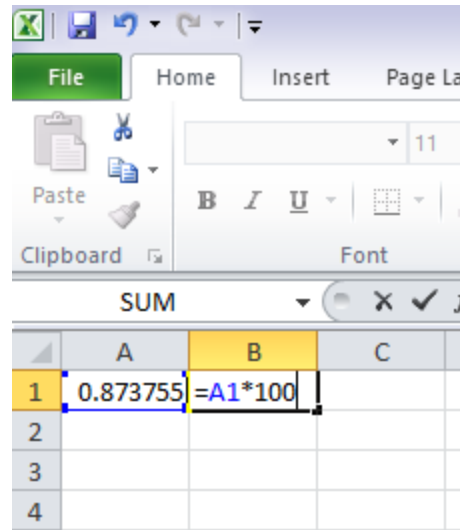
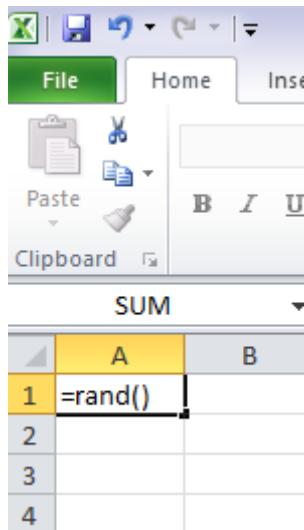
جدول اعداد تصادفی

57172	42088	70098	11333	26902	29959	43909	49607
33883	87680	28923	15659	09839	45817	89405	70743
77950	67344	10609	87119	15859	74577	42791	75889
11607	11596	01796	24498	17009	67119	00614	49529
56149	55678	38169	47228	49931	94303	67448	31286
80719	65101	77729	83949	83358	75230	56624	27549
93809	19505	82000	79068	45552	86776	48980	56684
40950	86216	48161	17646	24164	35513	94057	51834
12182	59744	65695	83710	41125	14291	74773	66391
13382	48076	73151	48724	35670	38453	63154	58116
38629	94576	48859	75654	17152	66516	78796	73099
60728	32063	12431	23898	23683	10853	04038	75246
01881	99056	46747	08846	01331	88163	74462	14551
23094	29831	95387	23917	07421	97869	88092	72201
15243	<u>21100</u>	48125	05243	16181	39641	36970	99522
53501	58431	68149	25405	23463	49168	02048	31522
07698	24181	01161	01527	17046	31460	91507	16050
22921	25930	79579	43488	13211	71120	91715	49881
68127	<u>00501</u>	37484	99278	28751	80855	02035	10910
55309	<u>10713</u>	36439	65660	72554	77021	46279	22705
92034	<u>90892</u>	69853	06175	61221	76825	18239	47687
50612	84077	41387	54107	09190	74305	68196	75634
81415	98504	32168	17822	49946	37545	47201	85224
38461	44528	30953	08633	08049	68698	08759	45611
07556	24587	88753	71626	64864	54986	38964	83534
60557	50031	75829	05622	30237	77795	41870	26300

MS Excel

چگونه می توان با استفاده از نرم افزار اکسل عدد تصادفی تولید کرد؟

- ۱) Open soft
- ۲) In a cell write: “=rand()” and enter
- ۳) In the next cell write: “cell before *N of study population and enter: the round number is your first sample’s number
- ۴) Repeat this until reach the n of your sample



RANDBETWEEN

Returns a random number between the numbers you specify. A new random number is returned every time the worksheet is calculated.

RANDBETWEEN (bottom,top)

Bottom: is the smallest integer RANDBETWEEN will return.

Top: is the largest integer RANDBETWEEN will return.

Example

=RANDBETWEEN(1,1200)

نمونه گیری تصادفی ساده

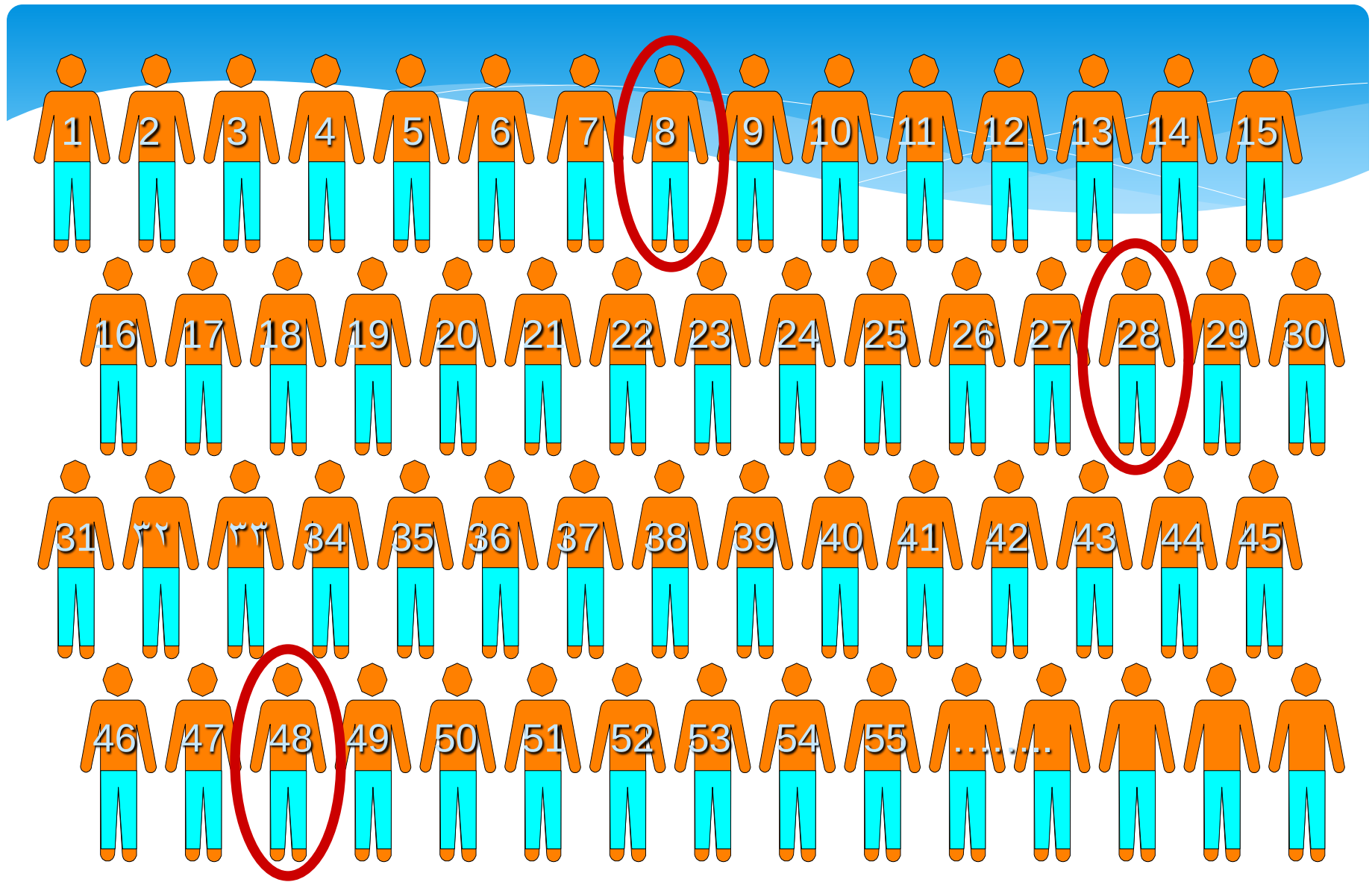
725	268	546	535	775	726
590	564	1021	468	666	1045
1161	48	933	1090	85	1141
1134	1019	740	557	55	228
290	1111	50	214	258	545
679	543	648	361	158	477
1092	771	1052	187	522	917
710	677	541	744	163	1115
473	1067	456	859	978	1136
247	718	3	110	819	242

نمونه گیری تصادفی منظم یا سیستماتیک

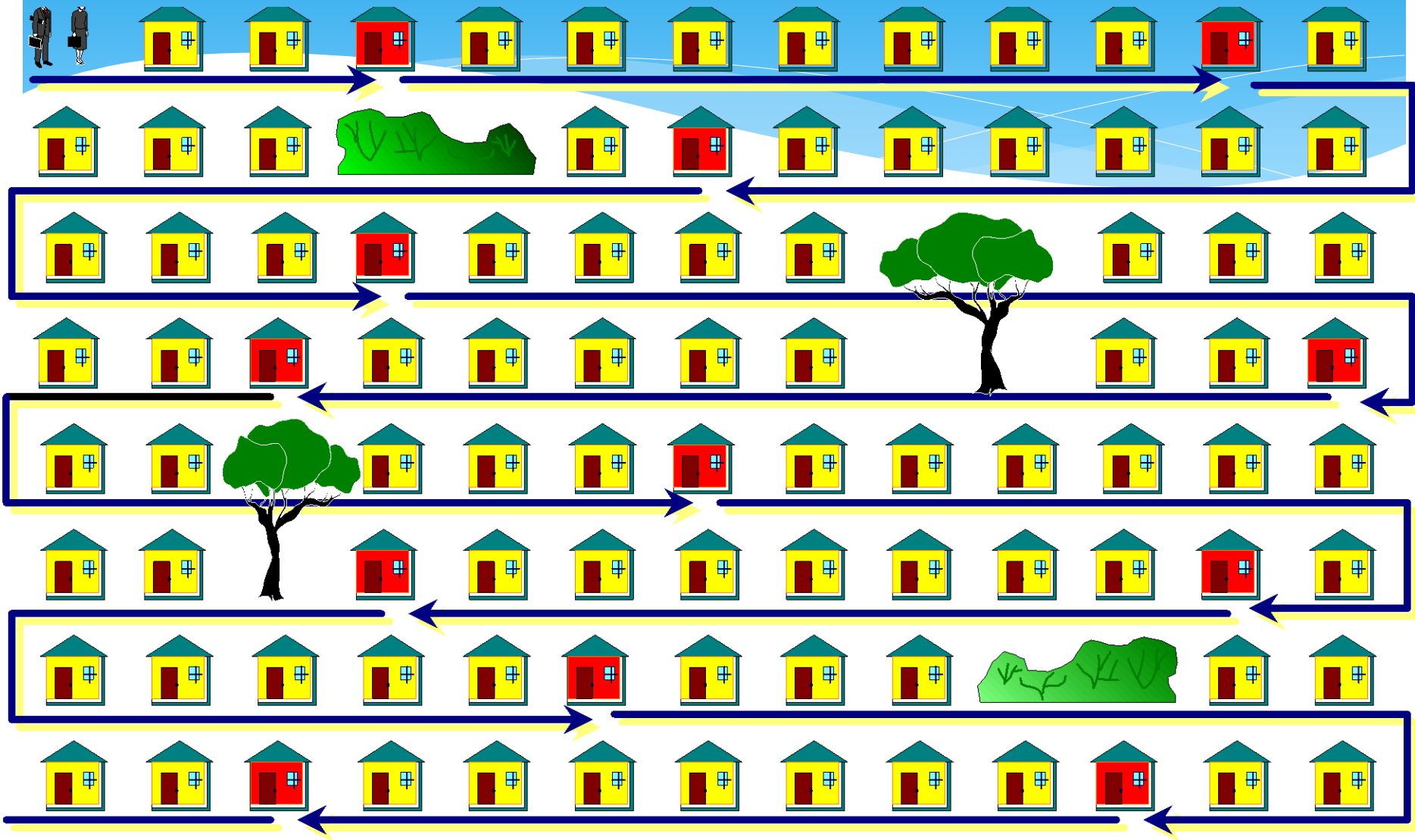
- تهیه فهرست واحد های جامعه (چارچوب نمونه گیری)
- تقسیم اندازه جامعه به تعداد نمونه مورد نیاز (X)
- انتخاب یک عدد تصادفی بین ۱ و X
- افزودن X به عدد انتخاب شده

نمونه گیری تصادفی منظم یا سیستماتیک

- مثال: می خواهیم شیوع پوسیدگی دندان (برمبنای DMF) را در بین ۱۲۰۰ نفر دانش آموزان یک مدرسه تعیین کنیم.
- فهرست دانش آموزان مدرسه را تهیه کنیم
- دانش آموزان را از ۱ تا ۱۲۰۰ شماره گذاری کنیم
- اگر حجم نمونه ۶۰ نفر باشد، بایستی با توجه به $1200 / 60 = 20$ ، از هر ۲۰ نفر یکی انتخاب کنیم.
- یک عدد تصادفی بین ۱ تا ۲۰ انتخاب می کنیم (مثلاً ۸)
- سپس ۲۰ تا ۲۰ تا به عدد فوق (۸) اضافه می کنیم به این ترتیب نمونه اول فرد شماره ۸، بعدی ۲۸، بعدی ۴۸ و خواهند بود.



نمونه گیری منظم یا سیستماتیک



نمونه گیری طبقه بندی شده (stratified)

■ جامعه را به تعدادی طبقه افراز می کنیم

• بین طبقه ها نباید هم پوشانی وجود داشته باشد

• هیچ فردی از جامعه نبایستی بیرون از طبقه بندی قرار بگیرد

■ از هر طبقه تعدادی نمونه انتخاب می کنیم

■ یافته های حاصل از طبقات را روی هم می ریزیم

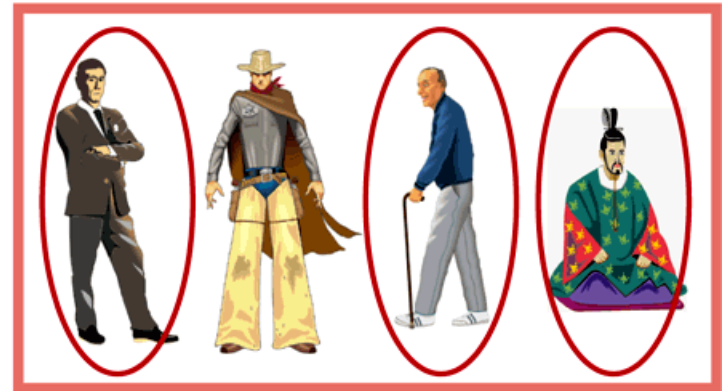
هر چقدر داخل طبقات افراد بیشتر به هم شبیه (هموژن) باشند ولی بین طبقه ها تفاوت زیاد وجود داشته باشد، نمونه گیری طبقه ای بهتر است

ادامه

Women



Men



نمونه گیری طبقه بندی شده

■ مزایا

- اگر داخل طبقه ها افراد به هم شبیه باشند، کارایی این روش از روش تصادفی ساده بیشتر است
- به علت وجود داده های مجزا از هر یک از طبقات، امکان تحلیل جداگانه برای هر یک از طبقه ها وجود دارد

■ معایب

- اگر صفت مورد نظر در داخل طبقه ها هتروژن باشد، این روش مناسب نیست
- محاسبه خطای نمونه گیری (به دست آوردن SE) مشکل است

نمونه گیری خوشه ای (cluster)

- جامعه را به تعدادی گروه (خوشه) از واحدها تقسیم می کنیم
- تعدادی از خوشه ها را به طور تصادفی انتخاب می کنیم
- در هر خوشه همه واحدها (یا تعدادی از آنها) را انتخاب می کنیم

هر چقدر داخل خوشه ها افراد کمتر به هم شبیه (هتروژنیتی بیشتر) باشند نمونه گیری خوشه ای موفقتر است

نمونه گیری خوشه ای

■ مزایا

- در این روش نیازی به داشتن فهرست کامل واحدهای نمونه گیری (چارچوب نمونه گیری) نیست.
- هزینه مسافرت و منابع مورد نیاز برای جمع آوری داده ها کمتر است

■ معایب

- اگر هموژنیتی صفت مورد نظر در داخل خوشه ها خیلی زیاد باشد، این روش مناسب نیست.
- نسبت به نمونه گیری تصادفی ساده، برای دستیابی به دقت (precision) مشابه به حجم نمونه بالاتری نیاز دارد.
- محاسبه خطای نمونه گیری (به دست آوردن SE) مشکل است

نمونه گیری خوشه ای

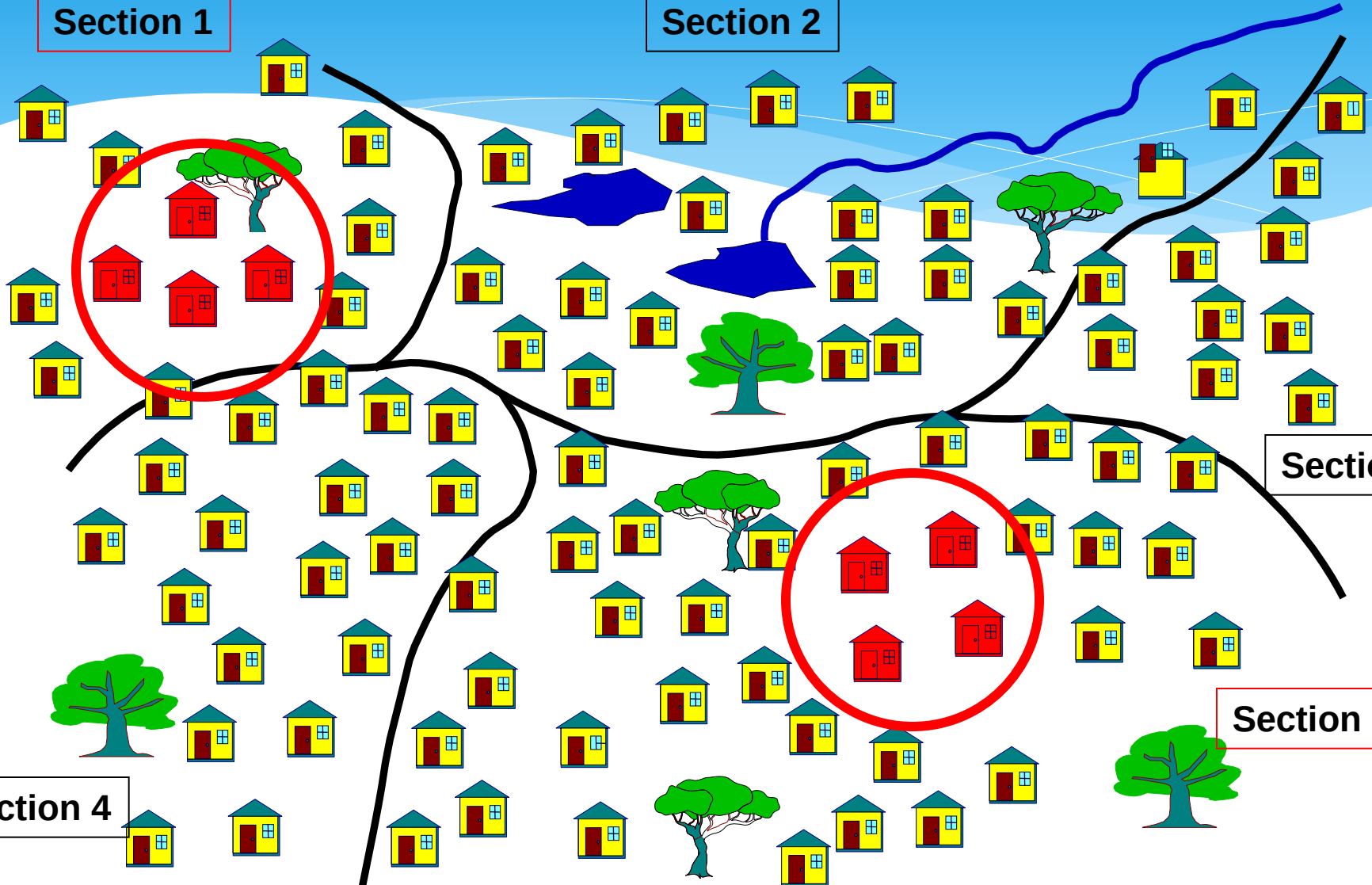
Section 1

Section 2

Section 3

Section 5

Section 4



نمونه برداری خوشه‌ای چند مرحله‌ای

Multiple cluster sampling

✓ بواسطه گستردگی بیش از حد جامعه محقق ناگزیر می‌گردد نمونه را طی دو یا چند مرحله انتخاب کند.

مراحل:

حجم نمونه

Sample Size

اهمیت حجم نمونه

- * یکی از مهمترین شاخص های تعیین کننده توان تصمیم گیری و قدرت (power) یک مطالعه است.
- مثال مطالعه دیابت در سربازان (۲۰/۱۰۰) با انتخاب نمونه ۱۵ نفر
- * حجم نمونه هم باید طوری انتخاب شود که اختلاف یا تفاوت علمی مشاهده شده از نظر علمی معنا دار باشد و هم از نظر اقتصادی با محدودیتی مواجه نشود.
- * اخلاق در پژوهش (بیشتر برای پژوهش های بالینی)

عوامل موثر بر حجم نمونه

- * میزان دقت مورد نیاز
- * روش نمونه گیری
- * نوع پژوهش
- * یکنواختی جامعه
- * هزینه و در دسترس بودن
- * محدودیت ها

عوامل موثر بر تعداد حداقل نمونه مورد نیاز

■ پراکندگی یا یکنواختی

■ دقت مورد نیاز

■ اطمینان مورد نیاز

■ اندازه جامعه؟

محاسبه حداقل حجم نمونه برای برآورد نسبت (صفت کیفی)

$$n = \frac{(z_{1-\alpha/2})^2 * p(1-p)}{d^2}$$

α : احتمال خطای نوع اول؛

اگر $\alpha = 0.05$ باشد $Z_{1-\alpha/2}$ برابر ۱.۹۶ است

p : تخمین نسبت (proportion) صفت مورد نظر

d : خطای قابل قبول در برآورد نسبت مورد نظر

محاسبه حداقل حجم نمونه برای برآورد میانگین (صفت کمی)

$$n = \left(\frac{Z_{1-\alpha/2} * \sigma}{d} \right)^2$$

α : احتمال خطای نوع اول؛

اگر $\alpha = 0.05$ باشد $Z_{1-\alpha/2}$ برابر ۹۶/۱ است

σ : انحراف معیار (SD) صفت مورد نظر

d : خطای قابل قبول در برآورد میانگین

محاسبه حداقل حجم نمونه برای مقایسه نسبت ها در دو جامعه مستقل

$$n = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2 * [p_1(1-p_1) + p_2(1-p_2)]}{(p_1 - p_2)^2}$$

α : احتمال خطای نوع اول؛ اگر $\alpha = 0.05$ باشد $Z_{1-\alpha/2}$ برابر ۱.۹۶ است

β : احتمال خطای نوع دوم؛ اگر $\beta = 0.2$ باشد $Z_{1-\beta}$ برابر ۰.۸۶ است

p_1 : نسبت در گروه اول

p_2 : نسبت در گروه دوم



A close-up photograph of a dark brown branch with several light pink cherry blossoms. The blossoms have five petals and prominent stamens. Some buds are still closed. The background is a soft-focus green field of tall grass.

**Thank
You**