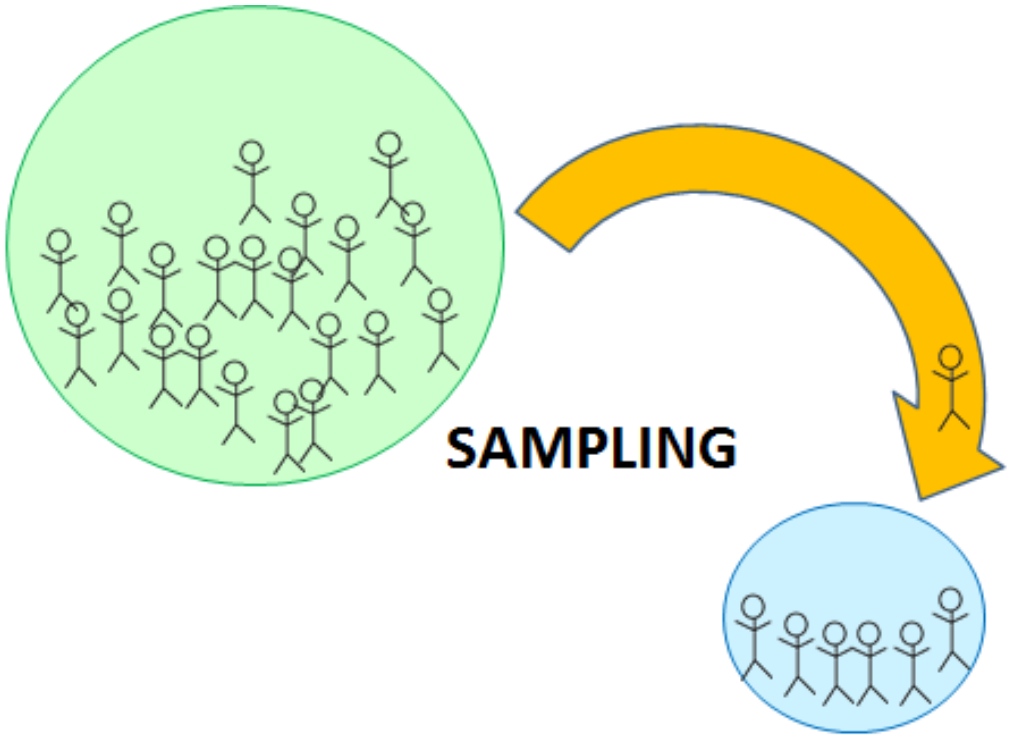
بسم الله الرحمن الرحیم

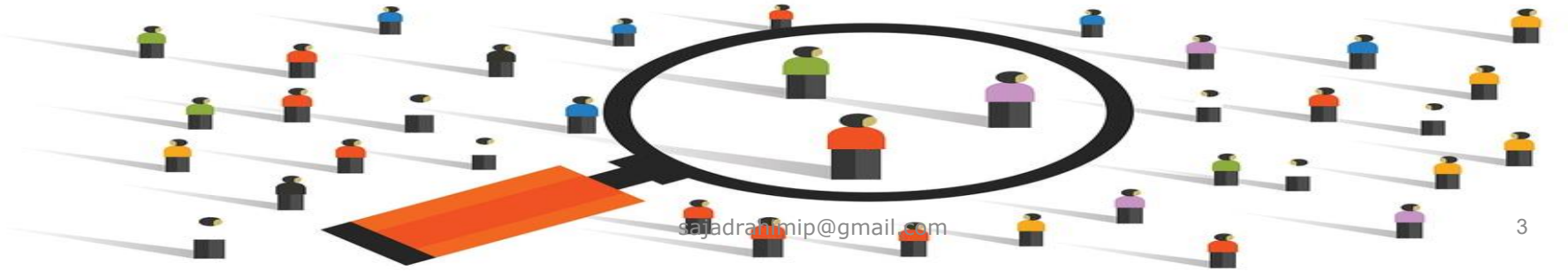
Sampling and Sample Size in Medical Research



sajadrahimip@gmail.com

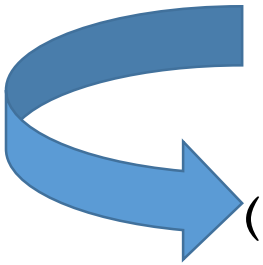
سئوال :چرا نمونه گیری می کنیم؟

- کاهش هزینه ها
- افزایش سرعت
- کاهش زمان
- غیر عملی بودن و محدودیت سرشماری در بعضی شرایط
- نتایج نمونه نسبت به سرشماری از تمام جمعیت اغلب از صحت و اعتبار بیشتری برخوردار است.
- نتایج نمونه نسبت به سرشماری از تمام جمعیت اغلب از ثبات و پایایی بیشتری برخوردار است.



اصطلاحات و مفاهیم نمونه گیری

- آماره (statistics)



- پارامتر(parameter)

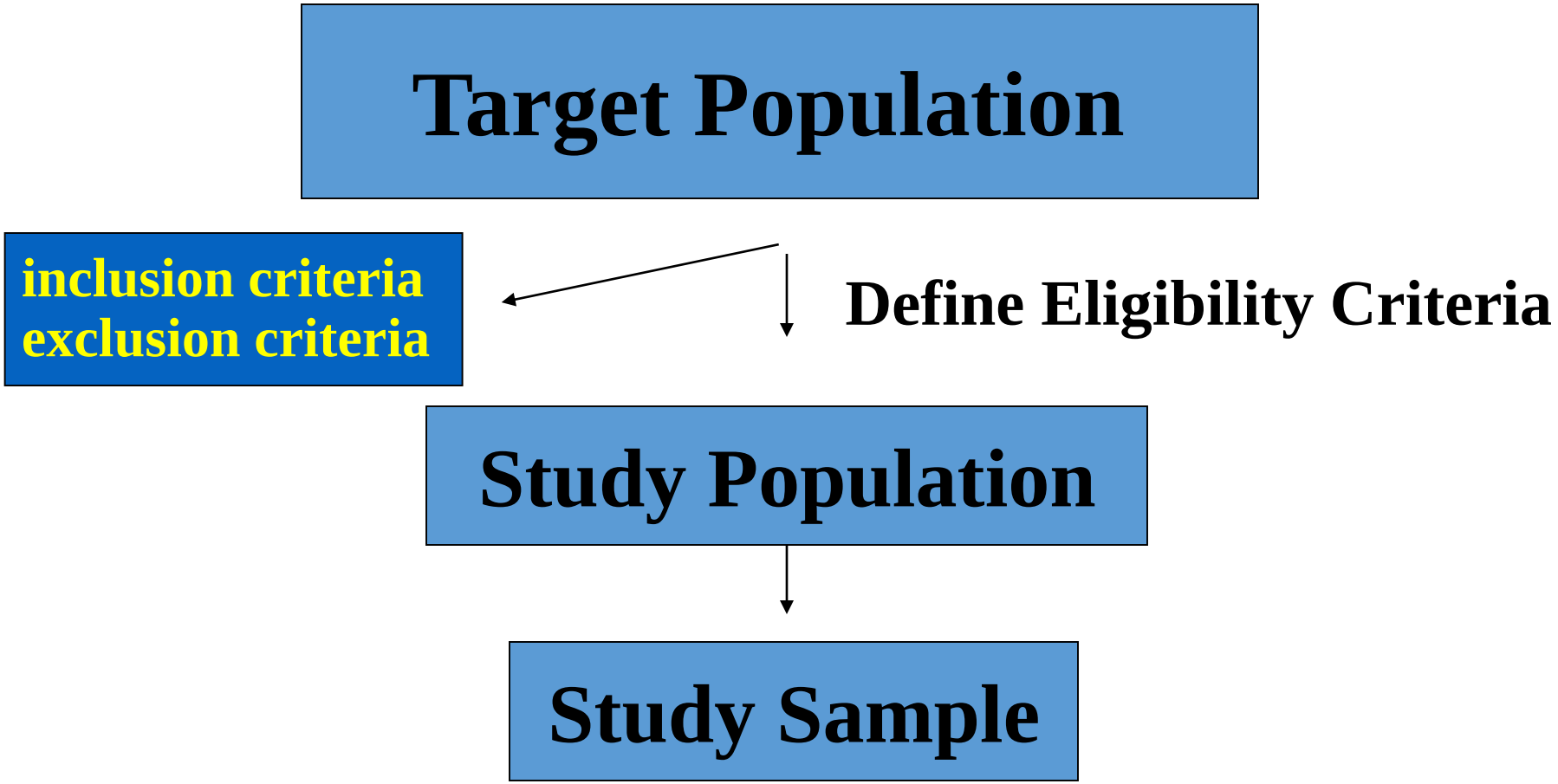
- چارچوب نمونه گیری (sampling frame): فهرستی از کل اعضای جامعه که شامل کل واحدهای نمونه گیری (sampling unit) هستند.

- طرح نمونه گیری (sampling design)

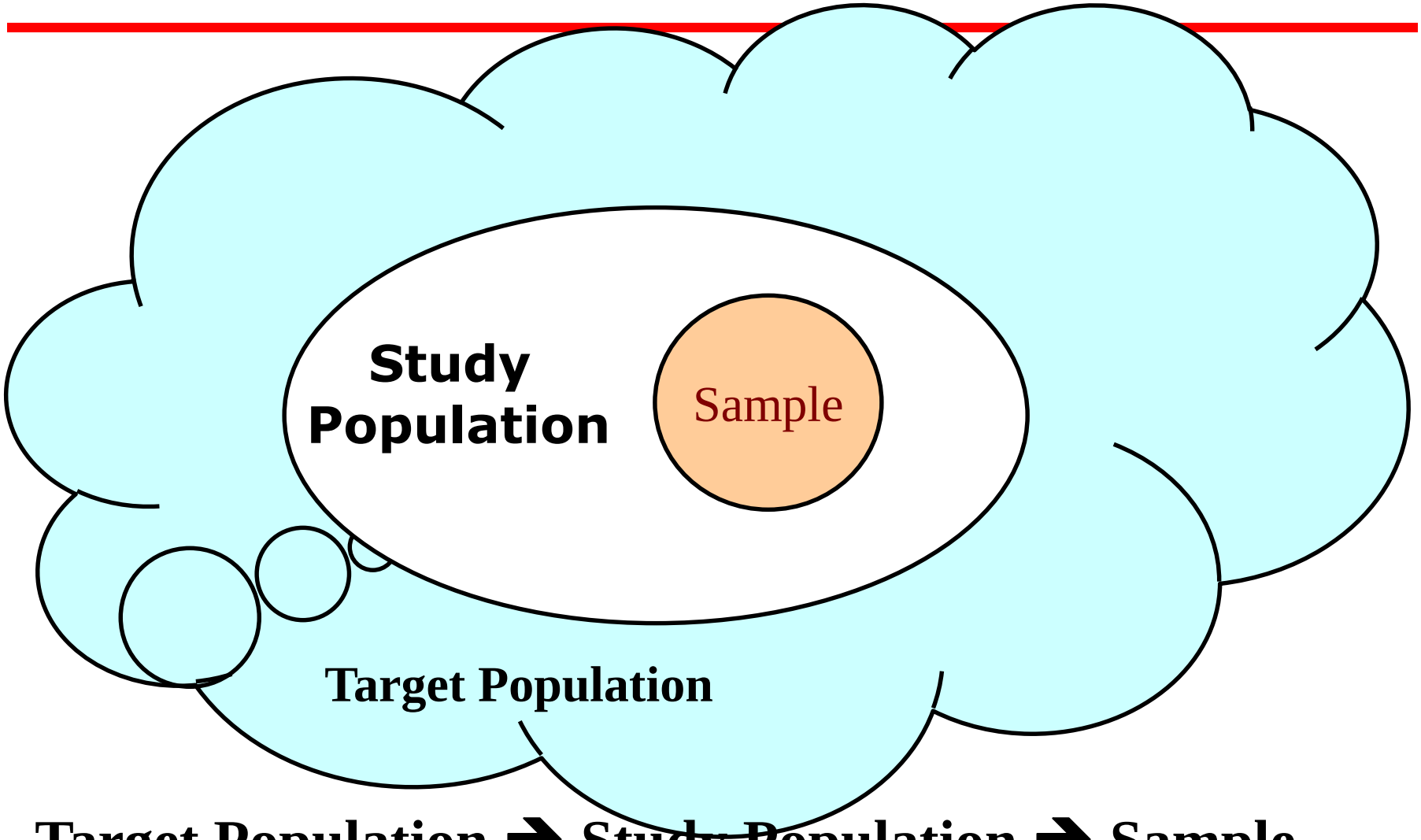
- معرف بودن نمونه (representativeness)

sajadrahimip@gmail.com

اصطلاحات و مفاهیم نمونه گیری



جمعیت هدف، جمعیت مطالعه، نمونه



Target Population ➔ Study Population ➔ Sample

روش های نمونه گیری

احتمالی (Probability sampling)

انتخاب افراد و واحدهای مطالعه به صورت تصادفی است. هر یک از عناصر جامعه، شانسی مشخص و غیر صفر برای انتخاب شدن در نمونه دارند و احتمال انتخاب همه عناصر جامعه الزاما برابر نمی باشد.

- ✓ در این روش نمونه امکان سوگیری انتخاب را کم می کند
- ✓ اجازه می دهد که از تئوری های آماری در تحلیل ها استفاده شود
- ✓ امکان تعمیم نتایج و اطلاعات بدست آمده از مطالعه بر روی نمونه به جامعه اصلی وجود دارد.

غیر احتمالی (Non-probability sampling)

انتخاب نمونه براساس قوانین احتمالات صورت نمی گیرد، نمونه ها به صورت تصادفی انتخاب نمی شوند و احتمال انتخاب شدن بستگی به نظر و قضاوت پژوهشگر دارد بنابراین

- ✓ نمونه ها معرف واقعی جامعه نیستند
- ✓ احتمال تورش انتخاب وجود دارد
- ✓ نتایج قابل تعمیم به جمعیت هدف نمی باشد

sajadrahimip@gmail.com

سوال

نمونه گیری تصادفی در چه نوع مطالعه ای اهمیت بیشتری دارد؟

استفاده از نمونه‌گیری‌های غیر احتمالی در مطالعه‌های مقطعی، که انتظار می‌رود مخرج کسر نماینده خوبی از جامعه باشد، اشتباه بزرگی است.

نمونه گیری غیر تصادفی در چه نوع مطالعاتی اهمیت پیدا می کند؟

در مطالعات مقدماتی و پایلوت مربوط به پیمایش های بزرگ کارامد هستند.در چنین مواردی که هدف تعیین واریانس و تغییرات صفت مورد اندازه گیری (برای تعیین حجم نمونه و روش نمونه گیری) و... می باشد، احتیاجی به روش های نمونه گیری احتمالی نیست

گاهی اوقات نمونه گیری غیراحتمالی بهترین روش نمونه گیری می باشد. مانند زمانی که امکان تهیه چارچوب نمونه‌گیری وجود نداشته باشد.

روشهای نمونه گیری

احتمالی

۱- نمونه گیری تصادفی ساده (Simple Random Sampling)

۲- نمونه گیری طبقه ای (Stratified sampling)

۳- نمونه گیری خوشه ای (Cluster sampling)

۴- نمونه گیری سیستماتیک (Systematic sampling)

۵- نمونه گیری چند مرحله ای (Multistage sampling)

غیراحتمالی

۱- نمونه گیری آسان یا در دسترس(Convenience sampling)

۲- نمونه گیری سهمیه ای (Quota sampling)

۳- روش هدفمند (Purposive sampling)

۴- نمونه گیری گلوله برفی(Snowball sampling)

نمونه گیری تصادفی ساده (Simple Random Sampling)

- روش نمونه گیری تصادفی ساده، ساده ترین روش نمونه گیری احتمالاتی است.
- در این نوع نمونه گیری به صورت تصادفی ساده انجام خواهد شد و احتمال انتخاب افرد برابر با فراوانی نسبی آنها می باشد. به عبارت دیگر اگر حجم افراد جامعه N و حجم نمونه را n فرض کنیم، احتمال انتخاب هر فرد جامعه در نمونه مساوی n/N است.

- **روش اجرا:**
  - **تهیه فهرست کل اعضای جامعه (چارچوب نمونه گیری)**
  - **انتخاب تصادفی هر نمونه**

انتخاب نمونه تصادفی ساده را به سه شیوه می‌توان انجام داد:

❖شیوه اول به صورت قرعه کشی

❖ شیوه دوم با استفاده از جداول اعداد تصادفی

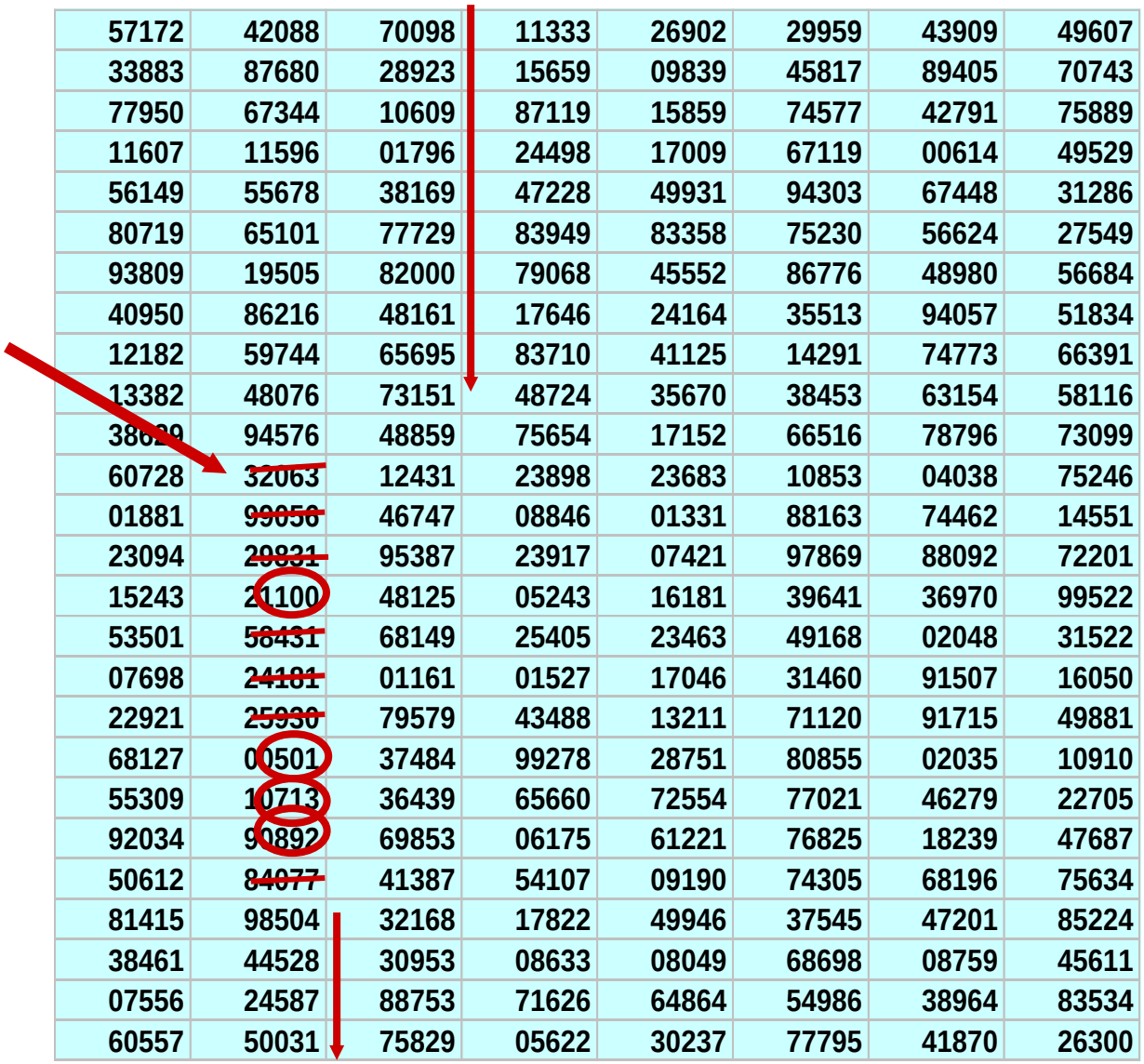
❖و شیوه سوم با نرم افزارهای رایانه ای مثلExcel

مثالی از نمونه گیری تصادفی ساده

- 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 ۶۰ عدد تصادفی را بین ۱ تا ۱۲۰۰ انتخاب کنیم.

چگونه؟؟

جدول اعداد تصادفی : جدولی از اعداد ۵ رقمی که با حرکت از سطر و ستون های آن می توان به اعداد نمونه دست یافت .



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	21100	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	00501	37484	99278	28751	80855	02035	10910
55309	10713	36439	65660	72554	77021	46279	22705
92034	90892	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

Microsoft Excel

RANDBETWEEN

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

Syntax

RANDBETWEEN (bottom,top)

Bottom is the smallest integer RANDBETWEEN will return.

Top is the largest integer RANDBETWEEN will return.

Example

=RANDBETWEEN(1,1200)

Copy and Paste

نمونه گیری تصادفی ساده (Simple Random Sampling)

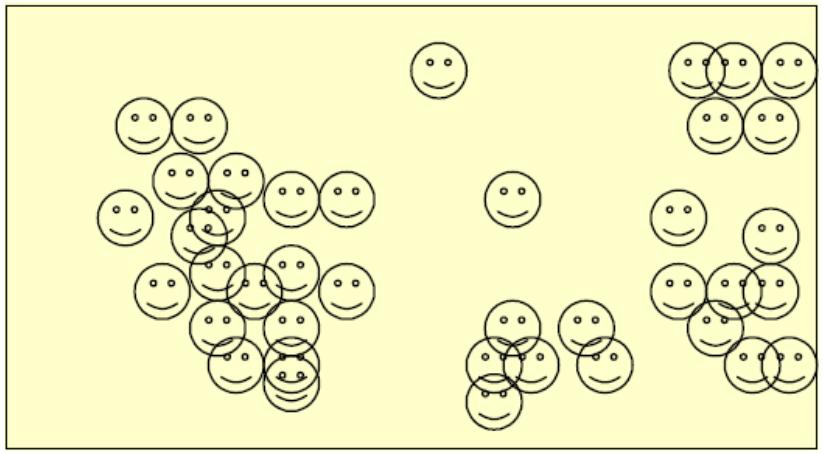
- **مزایا نمونه‌گیری تصادفی ساده:**
- سادگی روش
- خطای نمونه‌گیری به سادگی قابل اندازه‌گیری است.
- برای نمونه‌گیری از جوامع کوچک روشی مناسب و عملی است.
- سوگیری اتنخاب در این روش کمتر وجود دارد.

- **معایب نمونه‌گیری تصادفی ساده :**
- نیاز به لیست کامل و چارچوب نمونه گیری دارد.
- نمونه ممکن است بسیار پراکنده باشند و دسترسی به همه نمونه‌ها امکان‌پذیر نباشد.
- ممکن است نمونه بهترین معرف جامعه نباشد

نمونه گیری تصادفی منظم یا سیستماتیک

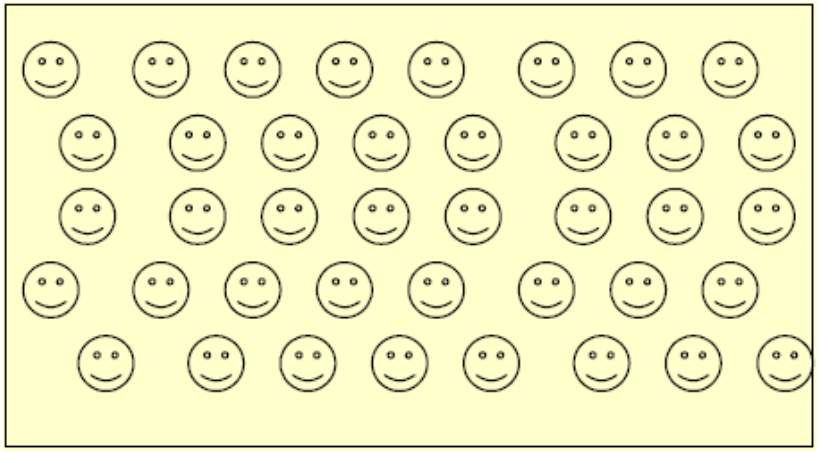
آنچه که در نمونه گیری تصادفی

ممکن است اتفاق بیافتد



انتظار ما از نمونه گیری

تصادفی ساده



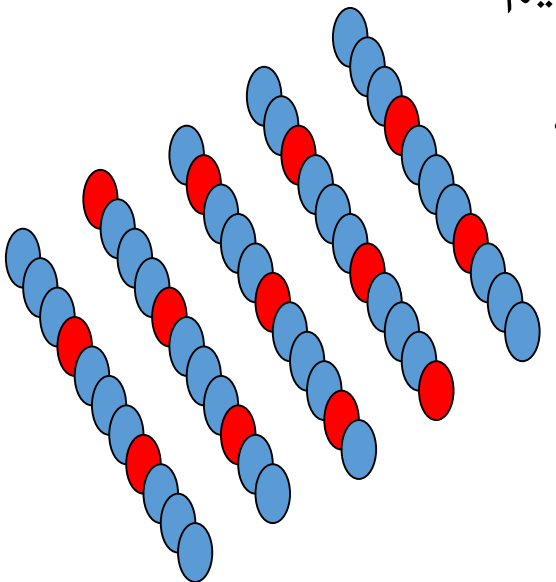
نمونه گیری منظم (سیستماتیک)
Systematic Sampling

منطبق بر یک قاعده و قانون مشخص

این روش همانند نمونه گیری تصادفی ساده است . با این تفاوت که در آن عدد کسر نمونه گیری محاسبه و نمونه به فاصله معین انتخاب می شود .

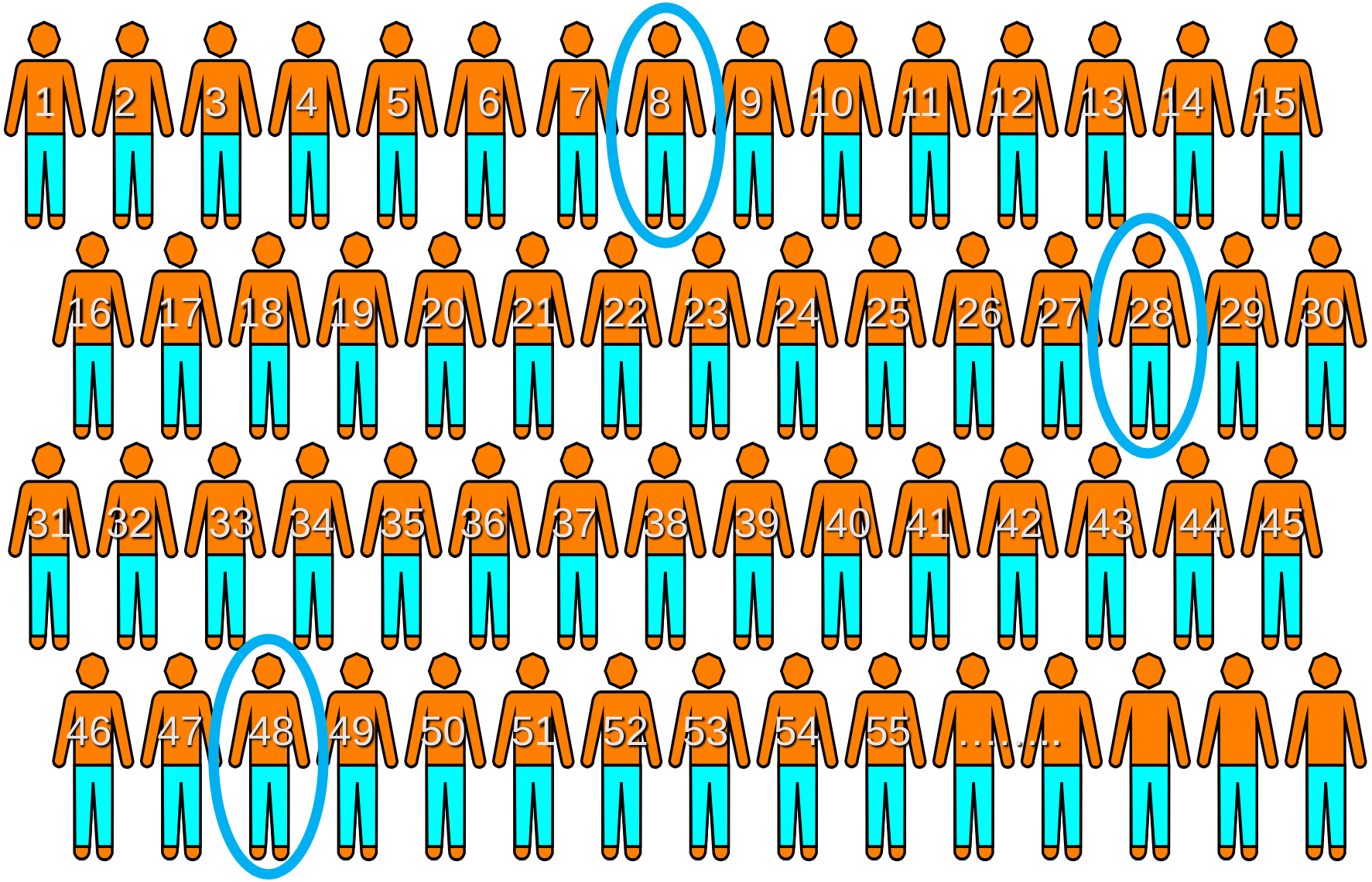
$$k = \frac{N}{n}$$

* تعداد اعضای جامعه (N) و نمونه (n) را مشخص می کنیم.
* عددی به صورت راندوم از ۱ تا k را انتخاب می کنیم .



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 ۲۰ = ۶۰ / ۱۲۰۰، از هر ۲۰ نفر یکی انتخاب کنیم.
  - یک عدد تصادفی بین ۱ تا ۲۰ انتخاب می کنیم (مثلاً ۸)
  - سپس ۲۰ تا به عدد فوق (۸) اضافه می کنیم به این ترتیب نمونه اول فرد شماره ۸، بعدی ۲۸، بعدی ۴۸ و .... خواهند بود.


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
31 32 33 34 35 36 37 38 39 40 41 42 43 44 45
46 47 48 49 50 51 52 53 54 55 .....

معایب نمونه گیری سیستماتیک

- علاوه بر مشکلات نمونه گیری ساده ، دارای مشکل تناوب در چهارچوب نمونه گیری است .
- اگر فاصله نمونه‌گیری بر تغییرات منظم در درون جامعه منطبق شود، تورش انتخاب رخ می‌دهد. (selection bias)
- مثال : اگر در شکل زیر عدد تصادفی ۲ و نسبت نمونه گیری ۴ باشد . نمونه ها عبارت اند از :

۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴
پسر	-	دختر	-	پسر	–	دختر-	پسر	–	دختر-	پسر	–	دختر-	پسر	–	دختر

- برای رفع آن بهتر است از روش نمونه گیری دیگر استفاده کنیم

مزایا نمونه گیری تصادفی منظم

- آسان و سادگی اجرا دارد.
- معمولا پراکندگی تصادفی نمونه ها از تصادفی ساده بهتر است.
- دقت نمونه گیری تصادفی سیستماتیک وقتی واحدهای جامعه به طور تصادفی مرتب شده باشند(همانند قد افراد)، مساوی با نمونه گیری تصادفی ساده و زمانی که واحدهای جامعه بر اساس صفتی مرتبط با صفت مورد برآورد مرتب شده باشند، بهتر از نمونه گیری تصادفی ساده و حتی بهتر از نمونه گیری طبقه ای است.

نمونه گیری طبقه بندی شده(Stratified Sampling)

نمونه گیری طبقه ای هنگامی مناسب است که بتوانیم جامعه آماری را نسبت به صفت مربوطه طوری تقسیم کنیم که واحدها در داخل طبقات از نظر صفت مربوطه شبیه به هم باشند.

مراحل:

ابتدا جامعه را به متغیر یا صفات مورد نظر طبقه بندی میکنیم.

چارچوب نمونه گیری و جدول توزیع فراوانی افراد جامعه را بر حسب طبقات مختلف بدست می آوریم.

نسبت درصد و سهم هر یک از طبقات را در کل جمعیت جامعه محاسبه مینماییم.

متناسب با سهم هر طبقه در جامعه ، سهم آن طبقه را در انتخاب نمونه ها نیز اعمال میکنیم.

$$P_k = N_k / N$$

با استفاده از نمونه گیری تصادفی ساده یا سیستماتیک افراد را از هر طبقه انتخاب مینماییم.

مثال

به طور مثال از یک جامعه آماری ۱۰۰۰۰ نفری که ۱۵ درصد آن دانشجو ، ۲۰ درصد کارمند اداری ، ۳۰ درصد کارگر و ۳۵ درصد کشاورز هستند می خواهیم ۴۰۰ نفر نمونه انتخاب کنیم . در مرحله اول تعداد مورد نیاز را در هر یک از این طبقات بر حسب درصدهای فوق معین می کنیم که به شرح زیر است :

تعداد افراد نمونه از بین دانشجویان ۶۰ =۰/۱۵×۴۰۰

تعداد افراد نمونه از بین کارمندان ۸۰ = ۰/۲۰ ×۴۰۰

تعداد افراد نمونه از بین کارگران ۱۲۰ = ۰/۳۰ ×۴۰۰

تعداد افراد نمونه از بین کشاورزان ۱۴۰=۰/۳۵ × ۴۰۰

مجموع افراد انتخاب شده از طبقات ۴۰۰ نفر خواهد بود

در مرحله دوم از هر یک از طبقات جمعیت و با استفاده از روش تصادفی ساده یا سیستماتیک افراد نمونه را انتخاب می کنیم .

نکات نمونه گیری طبقه بندی شده (stratified)

- این نوع نمونه گیری ، شکل اصلاح شده ای از نمونه گیری تصادفی ساده و سیستماتیک است که هدف از آن ، رسیدن به نمونه های معرف تر و دقیق تر است.
- بین طبقه ها نباید هم پوشانی وجود داشته باشد.
- هیچ فردی از جامعه نبایستی بیرون از طبقه بندی قرار بگیرد.

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

مزایای نمونه گیری تصادفی طبقه بندی شده

❖ افزایش دقت برآورد پارامترهای جامعه در صورت همگنی طبقات به علت کم بودن واریانس درون طبقات.

❖نمونه انتخاب شده بر اساس متغییر طبقه بندی شده، نماینده واقعی جامعه مورد نظر است.

❖کاهش خطای نمونه گیری

❖هزینه اجرای این نوع نمونه برداری(به سبب راحت بودن آن) به مراتب کمتر از نمونه برداری تصادفی ساده است.

❖شرایط یکسان بودن شانس انتخاب برای کل افراد درون طبقات تحقق پیدا می کند.

(Sub Group Analysis):

❖برای هر طبقه میتوان برآوردهای جداگانه از پارامترهای جامعه به دست آورد، بنابراین نتایج پژوهش را نه تنها برای کل جامعه بلکه برای بخشهای مختلف آن نیز میتوان بیان کرد.

معایب نمونه گیری تصادفی طبقه بندی شده

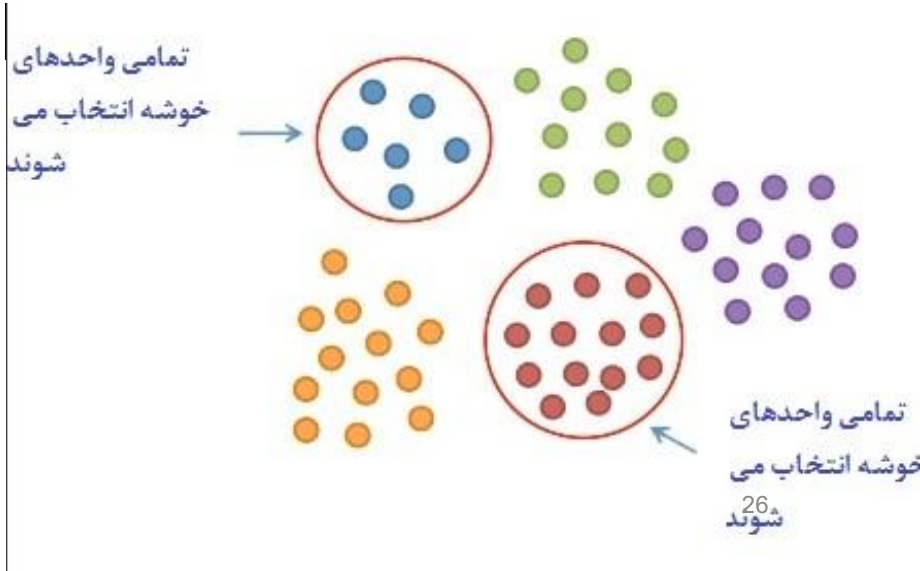
- عمده ترین مشکل این نمونه گیری ، در دست نبودن اطلاعات در مورد متغیر طبقه بندی است.
- همیشه طبقه بندی ممکن نیست.
- ممکن است تعداد نمونه در طبقات کم باشد.
- اگر صفت مورد نظر در داخل طبقه ها خیلی متفاوت باشد، این روش مناسب نیست.
- محاسبه خطای نمونه گیری (به دست آوردن SE) مشکل است.

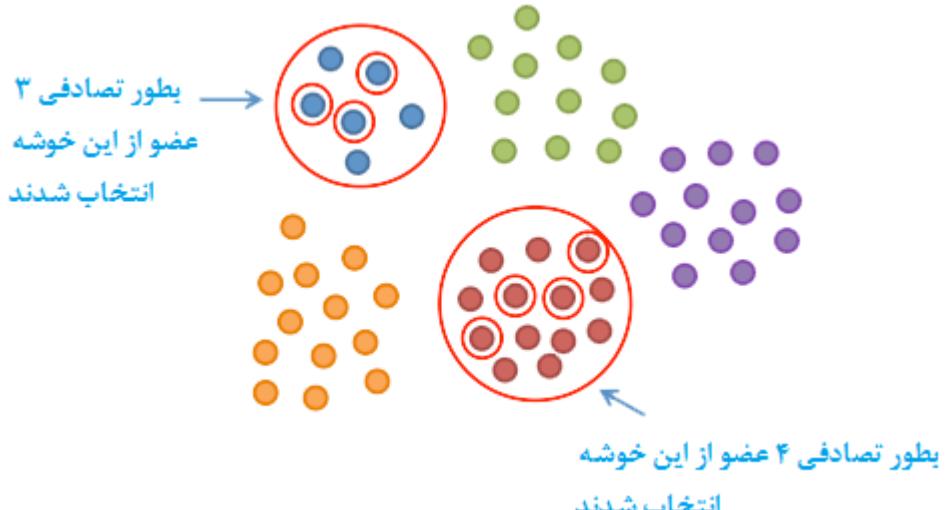
Cluster Sampling

در صورتی که فهرست کامل افراد جامعه مورد مطالعه در دسترس نباشد می‌توان از نمونه گیری خوشه بندی استفاده نمود.در این روش ابتدا جامعه مورد بررسی را به بلوک ها (خوشه های)مختلف تقسیم میکنیم سپس به صورت تصادفی چند بلوک را انتخاب می کنیم. در این مرحله دو انتخاب اساسی داریم:

❖در انتخاب اول کلیه اعضای خوشه‌های منتخب را مورد بررسی و ارزیابی قرار می‌دهیم.به این شیوه نمونه‌گیری نمونه‌گیری خوشه‌ای یک مرحله‌ای (Single stage cluster sampling) اطلاق می‌گردد.

❖در انتخاب دوم از بین کلیه اعضای خوشه یا خوشه‌های انتخاب شده، تعدادی از اعضاء به شیوه تصادفی ساده یا منظم انتخاب می‌گردند. به عبارتی، در این حالت در داخل خوشه‌ها نمونه‌گیری انجام می‌دهیم که به این شیوه، نمونه‌گیری خوشه ای دو مرحله‌ای (two stage cluster sampling) می‌گویند.





نمونه گیری خوشه ای
Cluster Sampling

- واحدهای نمونه گیری خوشه هایی شامل خانواده ها ، مدارس ، بیمارستان ها ، بلوکهای شهری ، دهکده ها و غیره هستند.
- هر چقدر داخل خوشه ها شباهت افراد کمتر و تفاوت بیشتر باشد (هتروژنیتی بیشتر) باشند نمونه گیری خوشه ای موفقتر است.
- **چگونه نگرانی از مشابه نبودن خوشه‌ها را رفع کنیم؟**

برای رفع این مشکل با استفاده از **اثر طرح** معمولاً تعداد نمونه لازم را در نمونه‌گیری خوشه ای ۱/۵ تا ۲ برابر می‌کنیم. به منظور عملی کردن میزان افزایش در اندازه نمونه در شرایط نمونه‌گیری درون خوشه‌ای، از ضریبی به نام شاخص اثر طرح یا Design Effect استفاده می‌نمایند.

A design effect(DEFF) is an adjustment made to find a survey sample size, due to a sampling method (Cluster sampling and Multistage sampling)

Design effect (DEEF) was calculated with the formula

DEEF = 1+ δ (n-1)

“δ” or the interclass correlation coefficient (ICC)
“n”average size of clusters

- معمولاً این شاخص دارای اندازه‌ای بزرگ‌تر از یک می‌باشد.
- بزرگی این شاخص رابطه مستقیم با پراکندگی یا واریانس داخل خوشه‌ای دارد هر چه این پراکندگی بیشتر باشد، مقدار عددی اثر طرح بزرگ‌تر است.
- آنجا که در زمانِ طراحی مطالعه و تعیین روش نمونه‌گیری و حجم نمونه، مقدار دقیق اثر طرح نامشخص می‌باشد، معمولاً برآورد یا پیش‌بینی تقریبی برای شاخص اثر طرح را در محاسبه اندازه نمونه نهایی لحاظ می‌نمایند. در اغلب شرایط مقدار عددی **اثر طرح را حدود ۱/۵-۲** در نظر می‌گیرند و به عبارت دیگر عدد اندازه نمونه به دست آمده از فرمول‌های محاسبه اندازه نمونه را در این مقادیر تقریبی ضرب تا اندازه نمونه نهایی در شرایطی که روش نمونه‌گیری تصادفی خوشه‌ای است، به دست آید.

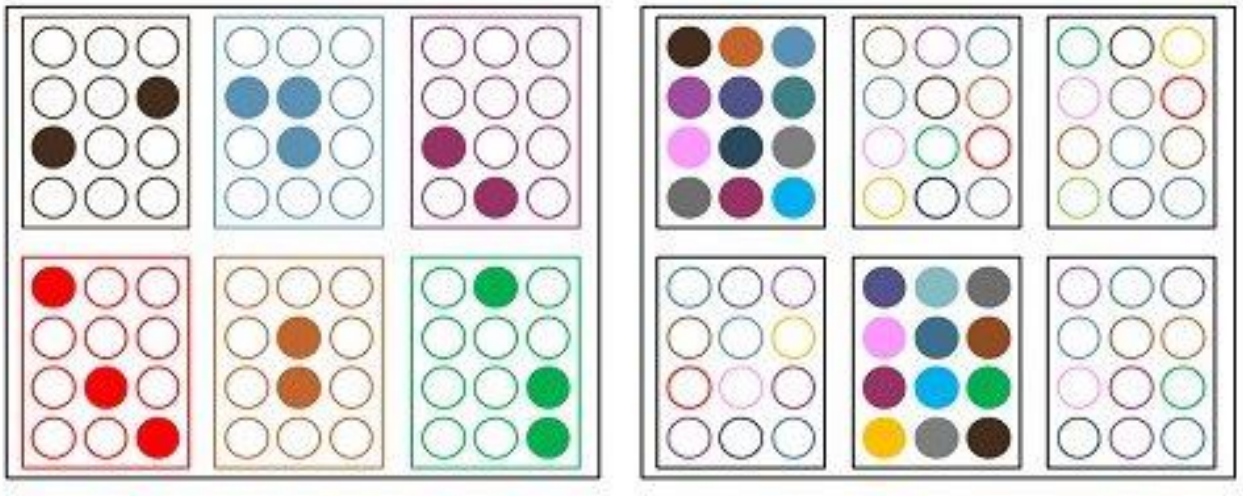
نمونه گیری خوشه ای

- مزایا
  - در این روش نیازی به داشتن فهرست کامل واحدهای نمونه گیری (چارچوب نمونه گیری) نیست.
  - هزینه و منابع مورد نیاز برای جمع آوری داده ها کمتر است.
  - صرفه جویی در وقت
- معایب
  - اگر هموژنیتی صفت مورد نظر درداخل 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

تفاوت بین نمونه گیری خوشه ای تک مرحله ای با نمونه گیری طبقه بندی شده

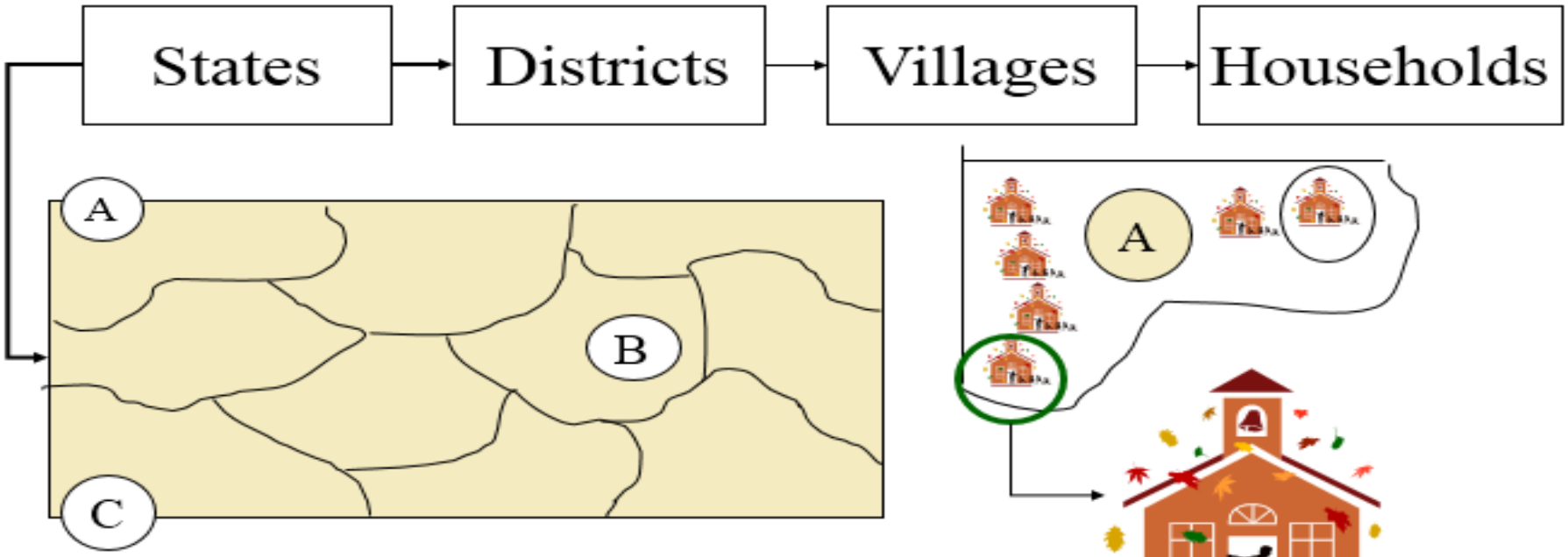
- الف-در نمونه گیری تصادفی طبقه ای از هر طبقه تعدادی از افراد را بر اساس فراوانی نسبی همان طبقه و به صورت تصادفی انتخاب می کنیم در صورتی که در نمونه گیری خوشه ای، تعدادی از خوشه ها را به تصادف انتخاب و سپس کل افراد درون خوشه را بعنوان نمونه مورد بررسی قرار می دهیم.

- ب- دقت نمونه گیری تصادفی طبقه ای در ارتباط مستقیم با همگنی درون طبقات و ناهمگنی بین طبقات است. اما دقت نمونه گیری تصادفی خوشه ای در ارتباط مستقیم با ناهمگنی درون خوشه ها و همگنی بین خوشه هاست.



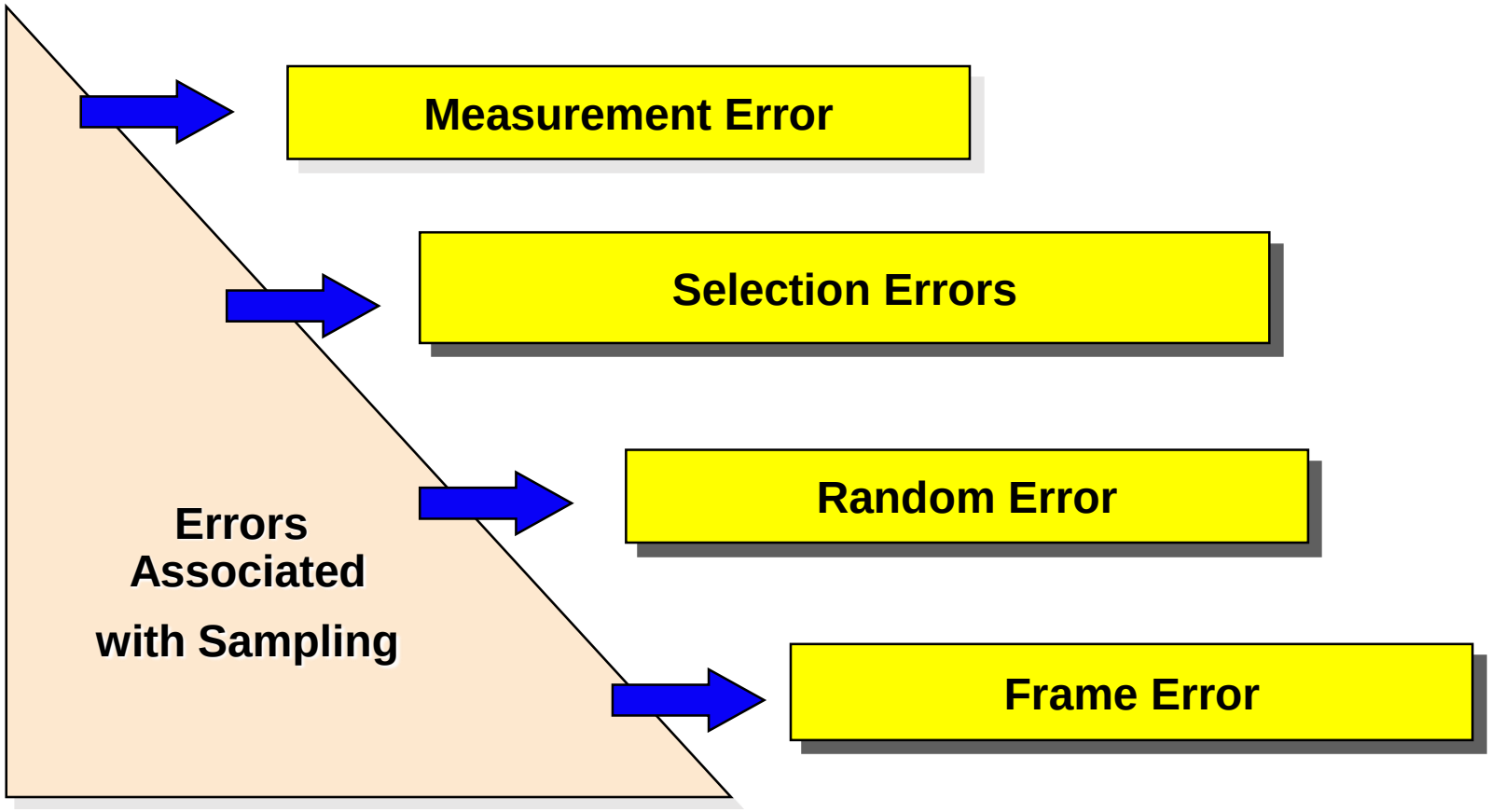
۵- نمونه گیری چندمرحله ای (Multistage sampling)

- در این روش نمونه گیری با استفاده از روشهای نمونه گیری احتمالی واحد یا سطح نمونه گیری مرحله به مرحله کوچکتر و محدودتر می شود و در آخرین مرحله نمونه گیری به واحد اصلی دسترسی پیدا می کنیم .



sajadrahimip@gmail.com

Types of Sampling Errors



Sample size

اخذ تصمیم درباره حجم نمونه, از لحاظ تامین میزان دقت نتایج نمونه گیری و صرفه جویی در مقدار وقت و هزینه, از اهمیتی خاص برخوردار است. بدیهی است که بزرگ بودن حجم نمونه موجب صرف هزینه و وقت زیاد، و کوچک بودن حجم نمونه موجب عدم دقت کافی برآوردها می شود.

برای تعیین اندازه نمونه، چند راه پیش رو داریم:

✓**تعیین اندازه نمونه بر اساس جدول های آماده مانند جدول موران**

در این جدول ها با توجه به اندازه جامعه(N)، اندازه نمونه (n)تعیین شده است.

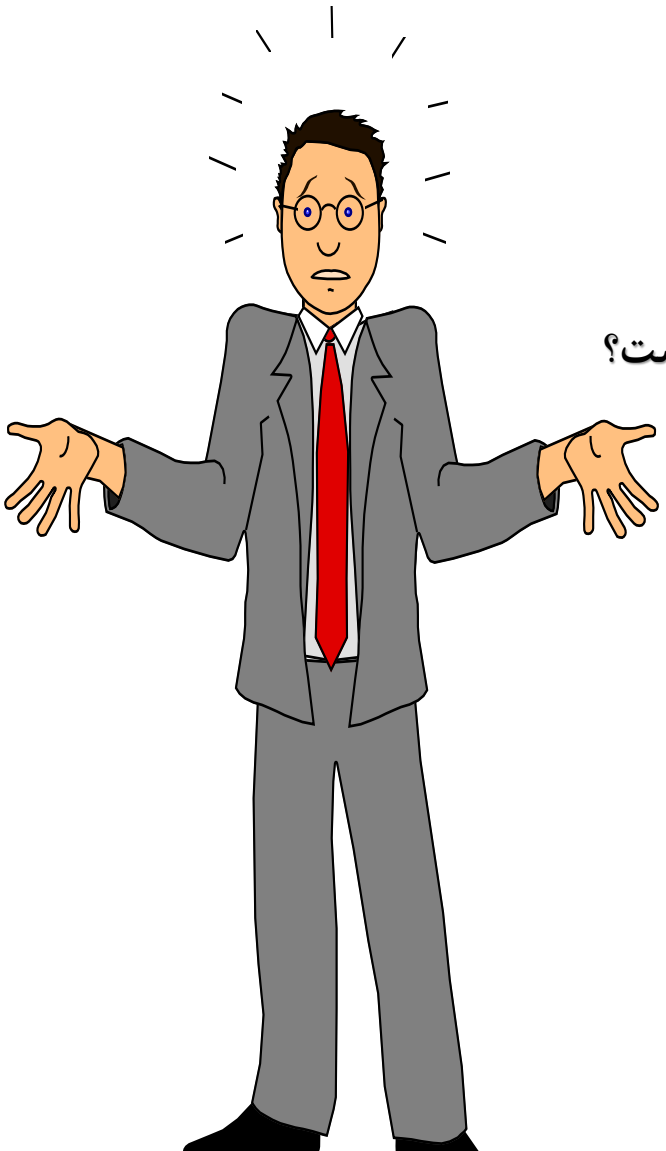
✓**تعیین حجم نمونه بر اساس نظر پژوهش گر**

در این روش پژوهش گر با در نظر گرفتن عامل هایی مثل بودجه، زمان، نیروی انسانی ماهر، امکانات، شیوع بیماری و در دسترس بودن موارد درصدی از جامعه را به عنوان نمونه انتخاب می کند. برخی از پژوهشگران حداقل اندازه نمونه را ۱۰ درصد اندازه جامعه ذکر کرده اند.

✓**تعیین اندازه نمونه بر اساس محاسبات آماری**

متداول ترین روش محاسبه اندازه نمونه، استفاده از محاسبات آماری است.

برای یافتن روش تعیین حجم نمونه نیاز است که محقق در ابتدا به چند سؤال پاسخ دهد:



- هدف اصلی پژوهشگر و نوع مطالعه (توصیفی یا تحلیلی) چیست؟
- متغیر اصلی مورد پژوهش کمی است یا کیفی؟
- چه سطح اطمینان و در صورت لزوم چه توانی برای نیل به هدف کافی است؟

$(\alpha, \beta, 1-\alpha, 1-\beta)$

- مقدار تغییرات متغییر وابسته (انحراف معیار و شیوع) چقدر است؟
- محقق چه مقدار خطا را می‌پذیرد؟(Precision)
- بودجه طرح چقدر است؟
- حجم جامعه؟

Sample Size and Population Size

- Where is N (size of the population) in the sample size determination formula?

Population Size	e=±3% Sample Size	e=±4% Sample Size
10,000	1,067	600
100,000	1,067	600
1,000,000	1,067	600
100,000,000	1,067	600

In almost all cases, the sample size is independent of the size of the population.

TABLE FOR DETERMINING SAMPLE SIZE FROM A GIVEN POPULATION or Cochran's Sample Size Formula.

	فرض صفر درست	فرض صفر غلط
رد فرض صفر	خطای نوع اول α	توان آزمون $1-\beta$
پذیرش فرض صفر	خطایی رخ نداده $1-\alpha$	خطای نوع دوم β

The quantities α, β, Δ (Effect Size) and N are all interrelated.

POWER= 1-β = P{ accept HA|HA is true)

Holding all other values constant, what happens to the power of the study if

- α decreases? Power ↓
- N increases? Power ↑
- variability increases? Power ↓
- Δ(Effect Size) increases? Power ↑

Note: Typical error rates are α = .05 and β = .1 or .2 (80 or 90% power). Why is α often smaller than β?

فلسفه حجم نمونه چیست ؟حجم نمونه از کجا می آید؟

$$\hat{p} \pm 1.96 * \sqrt{\frac{pq}{n}}$$

point estimate ± margin of error

$$\text{margin of error} = 1.96 * \sqrt{\frac{pq}{n}}$$

$$(\text{margin of error})^2 = (1.96)^2 * pq/n$$

$$n = (1.96)^2 * pq/d^2$$

sajadrahimip@gmail.com

محاسبه حجم نمونه با فرمول کوکران

- یکی از روشهای پرکاربرد در تعیین حجم نمونه فرمول کوکران است. فرمول کوکران به‌صورت زیر محاسبه می‌شود:

$$n = \frac{\frac{z^2 pq}{d^2}}{1 + \frac{1}{N}\left(\frac{z^2 pq}{d^2} - 1\right)}$$

- P=0.5 q=.5 z=1.96 d=0.05

جدول مورگان برای محاسبه حجم نمونه

مورگان چیزی جز همان کاربرد فرمول کورکران نیست. و به عبارتی دیگر هر یک از اعداد این جدول را در فرمول کوکران بگذارید همین حجم نمونه را مشاهده خواهید کرد.

جامعه آماری	حجم نمونه	جامعه آماری	حجم نمونه	جامعه آماری	حجم نمونه
10	10	220	140	1200	291
15	14	230	144	1300	297
20	19	240	148	1400	302
25	24	25	152	1500	306
30	28	260	155	1600	310
35	32	270	159	1700	313
40	36	280	162	1800	317
45	40	290	165	1900	320
50	44	300	169	2000	322
55	48	320	175	2200	327
60	52	340	181	2400	331
65	56	360	186	2600	335
70	59	380	191	2800	338
75	63	400	196	3000	341
80	66	420	201	3500	346
85	70	440	205	4000	351
90	73	460	210	4500	354
95	76	480	214	5000	357
100	80	500	217	6000	361
110	86	550	226	7000	364
120	92	600	234	8000	367
130	97	650	242	9000	368
140	103	700	248	10000	370
150	108	750	254	15000	375
160	113	800	260	20000	377
170	118	850	265	30000	379
180	123	900	269	40000	380
190	127	950	274	50000	381
200	132	1000	278	75000	382
210	136	1100	285	100000	384

اندازه نمونه در مطالعات توصیفی

- در مطالعات توصیفی، معمولاً هدف برآورد میانگین یک صفت کمی یا نسبت یک صفت کیفی است.

- حجم نمونه در مطالعات توصیفی

متغییر کیفی (برآورد نسبت)

متغییر کمی(برآورد میانگین)

محاسبه حداقل حجم نمونه برای برآورد نسبت (صفت کیفی)

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

α : احتمال خطای نوع اول؛

اگر $\alpha = 0.05$ باشد $Z_{1-\alpha/2}$ برابر 1.96 است

p : تخمین نسبت یا شیوع (prevalence) 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 1.96 است

σ : انحراف معیار (SD) صفت مورد نظر

d: خطای قابل قبول در برآورد میانگین

اندازه نمونه در مطالعات تحلیلی

در مواردی که پژوهش تحلیلی است یعنی محقق بدنبال یافتن اختلاف بین گروه‌ها با آزمون ارتباط بین متغیرها است.

نکته : علاوه بر مقوله اطمینان به نتایج، توان آزمون مورد استفاده در کشف یا نشان دادن حقیقت نیز مطرح می‌شود.

- حجم نمونه در مطالعات تحلیلی

متغییر کیفی (مقایسه نسبتها)

متغییر کمی(مقایسه میانگینها)

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 1.96 است

β : احتمال خطای نوع دوم؛ اگر $\beta = 0.2$ باشد $Z_{1-\beta}$ برابر 0.84 است

P1: نسبت یا سهم پیامد مورد نظر در گروه اول (مورد ـ مواجهه یافته یا مداخله)

P2: نسبت یا سهم پیامد مورد نظر در گروه دوم (شاهد، غیر مواجهه یافته یا مقایسه)

تفاوت دو نسبت در دو گروه در واقع همان دقت برآورد در مطالعه است.

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

$$n = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2 * (\sigma_1^2 + \sigma_2^2)}{(\bar{x}_1 - \bar{x}_2)^2}$$

α : احتمال خطای نوع اول؛ اگر $\alpha = 0.05$ باشد $Z_{1-\alpha/2}$ برابر 1.96 است

β : احتمال خطای نوع دوم؛ اگر $\beta = 0.2$ باشد $Z_{1-\beta}$ برابر 0.84 است

σ_1 : انحراف معیار صفت در جامعه اول(مورد ـ مواجهه یافته یا مداخله)

σ_2 : انحراف معیار صفت در جامعه دوم(شاهد، غیر مواجهه یافته یا مقایسه)

X_1 : میانگین صفت در جامعه اول(مورد ـ مواجهه یافته یا مداخله)

X_2 : میانگین صفت در جامعه دوم(شاهد، غیر مواجهه یافته یا مقایسه)

DETERMINING of SAMPLE SIZE PARAMETERS

it is appropriate to have a precision of 5% if the prevalence of the disease is going to be between 10% and 90%.

However, when the prevalence is going to be below 10% or more than 90%, the precision of 5% seems to be inappropriate.

Table 1 95% CI of rare diseases ($P =< 0.05$) and common diseases ($P => 0.95$) with a precision (d) set at 0.05

P	*n*	95% CI	
		Lower	Upper
0.01	16	-0.04	0.06
0.02	31	-0.03	0.07
0.03	45	-0.02	0.08
0.04	60	-0.01	0.09
0.05	73	0	0.10
0.95	73	0.90	1.00
0.96	60	0.91	1.01
0.97	45	0.92	1.02
0.98	31	0.93	1.03
0.99	16	0.94	1.04

For example, if the prevalence is 1% (in a rare disease) the precision of 5% is obviously crude and it may cause problems. The obvious problem is that 95% CIs of the estimated prevalence will end up with irrelevant negative lower-bound values or larger than 1 upper bound values as seen in the Table 1

DETERMINING of SAMPLE SIZE PARAMETERS

- Therefore, we recommend *d* as a half of *P* if *P* is below 0.1 (10%) and if *P* is above 0.9 (90%), *d* can be {0.5(1-*P*)}.
- For example, if *P* is 0.04, investigators may use *d*=0.02, and if *P* is 0.98, we recommend *d*=0.01.
- However, if there is a resource limitation, investigators may use a larger *d*. In case of a preliminary study, investigators may use a larger *d* (e.g. >10%). However, justification for the selection of *d* should be stated clearly (e.g. limitation of resources) in their research proposal.

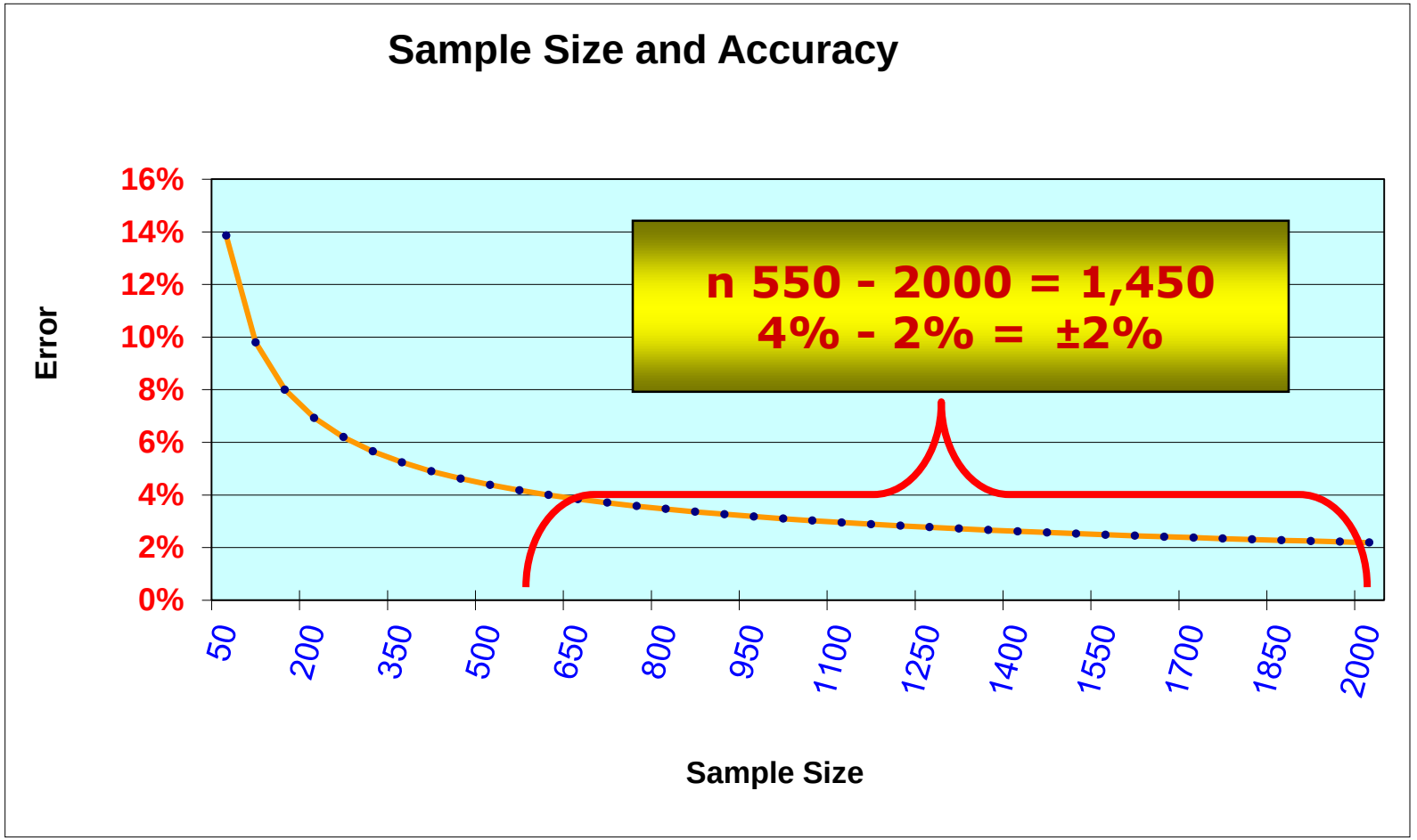
Estimating *P*

- Preferably, *P* from the studies with **pilot study, similar study design** and study population from the most recent studies would be most preferable.
- about the value of *P*, it is best to towards 50% as it would lead to a larger sample size. Setting *P*=0.5 is provide the biggest sample size.

- If we have a range of *P*, for instance, 20% to 30%, we should use 30% as it will give a larger sample size .If the range is 60% to 80%, we should use 60% as it will give a larger sample size. If the range is 40% to 60%, 50% will give a larger sample size.

Sample Size

- **A larger sample size is needed to detect the smallest meaningful difference.**
- **A larger sample size is needed when there is much variability in the population**
- **A larger sample size is required to increase the power of a study.**



Probability sample error (accuracy) can be calculated with a simple formula, and expressed as a ± % number.

متغییرها

تعریف :

مشخصه یک فرد ، شئ ، یا پدیده که قابل اندازه گیری بوده و بتواند مقادیر مختلفی را بپذیرد.(از فردی به فرد دیگر تغییر می کند)

متغیرها

اسمی (Nominal)

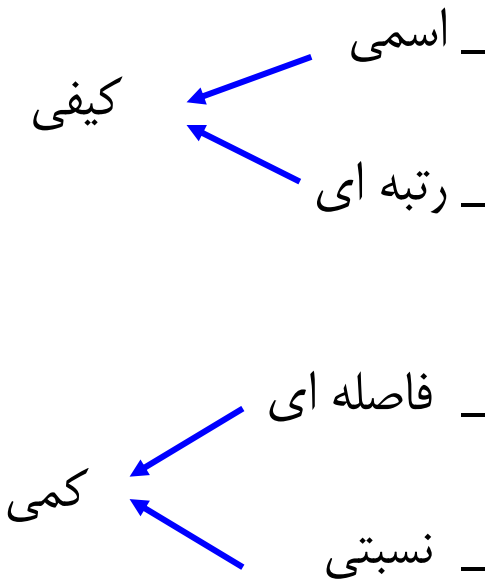
ترتیبی یا رتبه ای (Ordinal)

فاصله ای (Interval)

نسبتی (Ratio)

واژه NOIR را بخاطر بسپارید که مخفف همه انواع متغییرهاست.

نوع متغیرها
(مقیاس سنجش متغیرها)



انواع متغیر ها

- متغیر کمی: متغیری است که با عدد نمایش داده می شود این متغیر به دو دسته متغیر گسسته و پیوسته تقسیم خواهد شد که متغیر پیوسته مقادیر کسری را هم می پذیرد ولی گسسته این امکان را ندارد.

- متغیر کیفی: این متغیری است که کیفیت صفات با آن معرفی می شود.

نوع متغیر(سنجش متغیر) :

۱- اسمی (NOMINAL) :

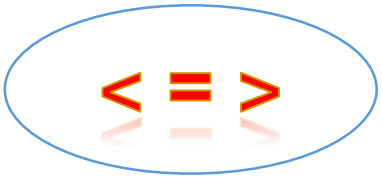
متغیری که تنها می توان نامی را بر آن نهاد ولی هیچگونه دلیلی بر برتر بودن یکی بر دیگری نداریم بنابراین تنها می توان آنها را دسته بندی و درصد گذاری کرد.

در این متغییرها هیچ یک از عملیات ریاضی را نمی توان انجام داد.

مانند: گروههای خونی، رنگ چشم ، نژاد ، جنسیت، ملیت و...

متغیر رتبه ای

دراین مقیاس اندازه گیری علاوه برداشتن نامی برای متغییر مورد بررسی می توان رتبه آنها را نسبت به یکدیگر بیان نمود مانند درجه گواتر (۰، ۱، ۲ و ۳) و شدت درد (کم،متوسط،زیاد) ، سطح تحصیلات(ابتدایی،راهنمایی،دبیرستان،دانشگاه)

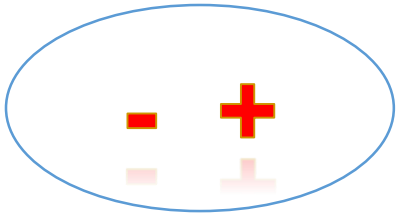
در این متغییرها عملیات ریاضی 

قابل محاسبه هستند.

نکته : دو نوع متغیر اسمی و رتبه ای را متغیر کیفی می گویند.

متغیر فاصله ای

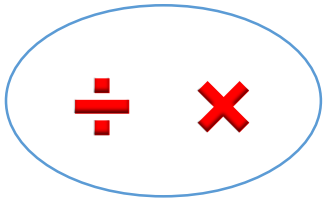
متغیری است که به صورت مقدار و عدد بیان می شود اما فاقد صفر ذاتی است و تنها بصورت صفر قراردادی مشخص می شوند. در این متغییرها علاوه بر دسته بندی و رتبه بندی متغییرها فاصله بین متغییرها را نیز می توان محاسبه کرد.

در این متغییرها عملیات ریاضی  قابل محاسبه هستند.

اعشار پذیر نیستند و اعدا صحیح را می پذیرند.
در علوم پزشکی درجه حرارت یک مقیاس فاصله ای است. مثالهای دیگر: مبداء سالهای شمسی، قمری، میلادی

متغیر نسبتی

متغیری که دارای واحد مشخص، معین، تعریف شده و نیز دارای صفر ذاتی می باشد. صفر نشان دهنده فقدان واقعی خاصیت موردنظر است.
فشارخون، جرم، میزان هموگلوبین، صدا، میزان مرگ و میر و ...

در این متغییرها عملیات ریاضی  قابل محاسبه هستند.

نکته : دو متغیر فاصله ای و نسبتی را متغیر کمی می گویند .

انواع متغیر ها (از لحاظ نقش)

- **متغیر مستقل**: آن متغیری است که محقق قصد دارد تاثیر آن را بر سایر متغیرها مورد سنجش قرار می دهد.
- **متغیر وابسته**: متغیری است که متغیر مستقل بر روی آن اثر می کنند.

بررسی تاثیر ورزش بر روی فشارخون

- **متغیر زمینه ای** (جمعیت شناسی یا دموگرافیک):این متغیرها خصوصیات جامعه مورد مطالعه را به نحوه مطلوبی توصیف می کنند و به شناخت بهتر موضوع کمک می کنند.

سن،جنس،وضعیت تاهل،محل سکونت،محل اقامت

- **متغییر مخدوشگر:** متغییری است که ارتباط بین متغییر مستقل و وابسته را مخدوش و تحریف می کند.

شروط: ۱) با متغییر وابسته رابطه علیتی داشته باشد ۲) با متغییر مستقل رابطه علی یا غیر غلی داشته باشد ۳)در مسیر بین متغییر مستقل و وابسته نباشد.

- **متغییر واسطه گر**: متغییر است که در مسیر بین متغییر مستقل و وابسته قرار می گیرد.

By Sajjad Rahimi

Ph.D in Epidemiology