



COMMUNITY MEDICINE

Notes

Done by: **Volunteer**

Sampling

By
Dr. Nanees
ghareeb

Definitions

Population:

كل الكومينتي

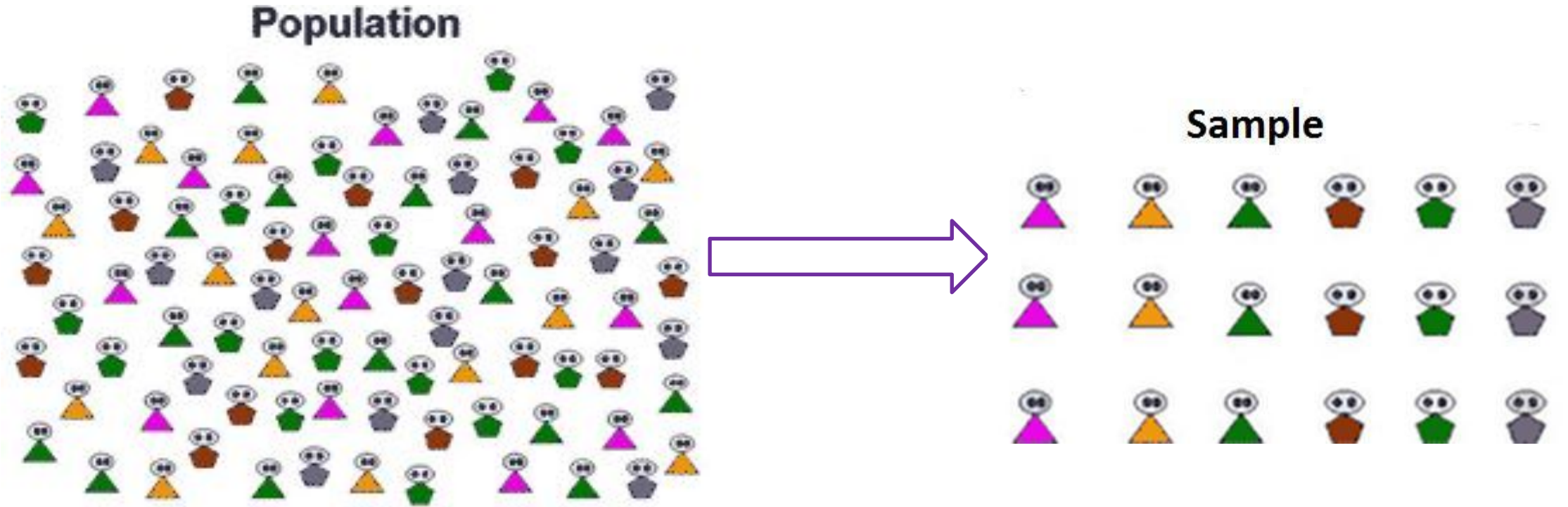
All people living in a place or any collection of individuals or things that we are interested in and their number may be finite or infinite **محدده او غير محدده**

e.g. Egyptians, students **هذا بقدر احده**, blood cells, etc.

Infinite مثلا هون بقدرش اعرف كم عدد RBC بالدم بالزبط
او عدد السكان احنا بنعرفه بطريقة تقريبية بس ما بنقدر نعرف العدد
بالتحديد .

- Ideally to carry out an epidemiologic study we should examine the whole population,
- but since this is not always possible because it is:
 1. expensive
 2. time consuming and
 3. not feasible **مش مناسب**
- So, we have to select a group from the population → sample.

عشان اعمل دراسه ع الاردنيين عن حالات TB او اي شي
ثاني حيكون صعب اعمل الدراسة على كل السكان
فلازم اخذ عينة منهم



العينة الي اخذناها هون كانت من كل الالوان والاشكال

Sample:

Group of individuals or things taken from a larger population and used to find certain information about this population.

Example: examination of 5ml of blood can diagnose liver disease. We are not in need to examine all blood.

The way that we follow in the selection of the sample will determine whether it is:

- A good representative sample → its result can be generalized on the whole population

Randomly فيه حوالي 3 او 4 طرق حنكي عنهم كمان شوي:)

يعني انا اشتغلت ع محافظة وطلعت نتائج منها بقدر احكي انه prevalence تبع المرض بهاي الدولة كان 20% لانه من الاساس اخترت العينة بطريقة صحيحة (عشوائية) ما اخترت ع حسب مزاجي او حكيت انه ما بدي اخذ هذول الناس واخذ ناس ثانيه ,, فما عندي selection bias لانه الكل الناس الموجوده كانت عندهم فرصه متساوية يدخلوا بهذا البحث

- Not good representative sample → its result can not be generalized on the whole

Sampling Units:

Each individual or thing of a population is called sampling unit.

Population =1000

Sample unite **حتكون 1000 فكل واحد منهم ممكن يدخل**
معي ع البحث

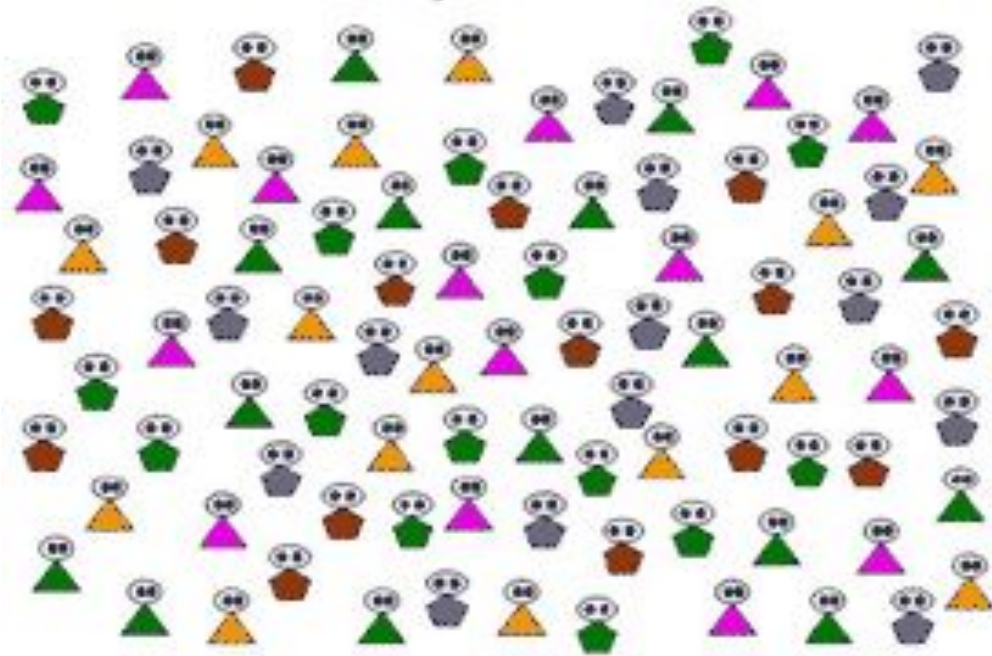
Sampling frame:

All sampling units (all individuals of the population) are known and each of them can be identified by a number or mark.

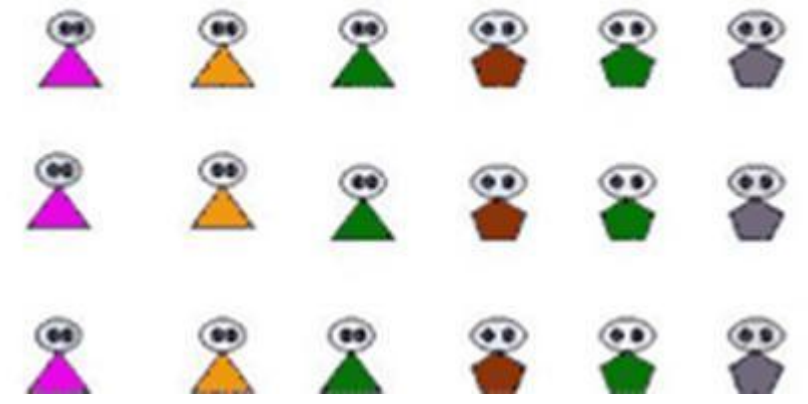
لازم اعطيهم كلهم علامة او رقم

ممكن يكون فيه sample بس ما فيه frame في حال كان عدد population كبير

Population



Sample



Why we use samples?

1. **Cheaper than examining the whole population.**

Less time consuming.

3. **Feasible and can be repeated in other areas or times.**

مناسب وبقدر اعينه بمكان ثاني وباي وقت

Types of samples

Probability samples

اهم اشئ
ولازم
الاختيار يكون
عشوائئ

1. Every individual has an equal chance (probability) of being taken in the sample before the sample is drawn.
2. It is a good representation of the population.
3. Its results can be generalized.

مش عشوائئة
بختارهم حسب مزاجئ

Non-Probability samples

1. Chance of selection not equal for all individuals → it is biased.

Types of samples

Non probability

مع انه مش كثير منيح بس في بعض حالات ممكن اضطر استخدمهم

- 1 Accessibility sample
- 2 Quota sample

Probability

- 1 Simple random sample
- 2 Systematic random sample
- 3- Stratified random sample
- 4- Cluster sample
- 5- Multi-stage random sample

Non-Probability Samples

The non-probability sample doesn't allow us to get a true representation of the population from which it is drawn.

1. Accessibility sample: الاشي المتاح عندي اخذته بغض النظر عن الطريقة

□ The investigator chooses his sample by his opinion.

□ The most convenient sample units are selected e.g. the nearest neighbors لانهم اقرب الي or relatives, volunteers, hospital cases, etc. مناسب واسهل بس مش عشوائي

مثلا يكون بدور عن مرض نادر فمش حلاقي المرض غير بالمستشفى فخلص بوخذ العينة من المستشفى لانه صعب الالقيهم بمكان ثاني

□ The sample is completed when the desired number of population is reached.

من البداية يكون مقرره بدي اخذ 100 عينة من سرطان الرئه اول ما يوصل عدد

المرضى اللى انتم 100 عااله نام بوقت

Advantages

• Cheap, quick, does not require sampling frame. **لانه حختار ع حسب مزاجي**

Disadvantages:

- Not representative of the whole population.
- It is biased due to subjective choice.

لانه اخترتهم زي ما بدك ما اخترناهم بطريقة objective

<< selection bias

.Its findings could not be generalized. So, it has to be restricted in use in scientific medical research □

مثلا لما اخترت الناس كان المكان ملوث فوقتها النتائج ما بقدر اعممها لكل الدولة

بس ممكن استخدم هاي الطريقة بالمجالات الطبية

□ Examples: sometimes we have to use this method e.g.:

1. Studying rare diseases which are available only in hospitals.
2. Studying occupational health hazards in workers exposed to that hazards.

حدرس عن سرطان صعب اللقي الحالات بين باقي الناس فوقتها بروج ع مركز السرطان

او بدى اشوف نسبة تسمم الناس من مادة lead ع سبيل المثال

2. Quota Sample

مثل الكوتا الانتخابية بكون محدد نسبة الها

- The investigator will take a sample of a certain size and structure.

نفس الي قبلها بس الفرق اني بحدد size and structure

- The choice of the actual sampling units does not follow a special scheme but left to his choice.

- The sample is completed when the desired number of population is reached.

Advantages

·
·
Cheap, quick, does not require sampling frame.

Disadvantages:

- ☐ Not a good representation of the population as it depends mainly on the investigator choice.**
- ☐ It is biased due to subjective choice.**
- ☐ Its findings could not be generalized, so seldom used in scientific medical research.**

Examples

:

1. Interview of all persons passing in a certain street at a certain time.

نزلت ع الشارع من الساعة 1 للساعة 2 واخذت المعلومات الي بدي اياها من الناس الي مروا بهذا الوقت بس وطبعا وقتها يكون محددہ العدد الي بدي اياه

1. In T.V. to know public opinion for the preferable programs.

نفس القصہ بدي اعمل مقابله لازم يكونوا عددهم 5 تين منهم شباب و 3 الي ضلوا يكونوا بنات

فاول شبين حلاقيهم حعمل معهم المقابله واول 3 بنات وخلص يكون خلصت

Probability Samples

- **Every individual (or sample unit) has an equal chance (probability) of being taken in the sample before the sample is drawn.**
- **There is minimal role for the investigator in selection of individuals or sample units. So, bias of subjective (researcher) selection is minimal. لأنه اخترت عشوائي.**
- **Results obtained from researches based on probability sampling can be generalized on population with confidence.**

Types of probability samples

- 1. Simple random sample**
- 2. Systematic random sample**
- 3. Stratified random sample**
- 4. Cluster sample**
- 5. Multi-stage random sample**

Simple Random Sample:

- **The population from which a simple random sample is drawn should be uniform or homogeneous.** بتشبه بعضها
- **A sample frame must be present, to choose the needed units from it.**
- **The units are selected by using random number tables ** (either in statistical books or generated by the computer) or by lottery or rotary depending on the size of the sample.**

احظ الارقام ع الحاسوب بعدين هو يختار لي عشوائي

□ **So, simple random sample is used when:**

- 1. Population is uniform or homogeneous and**
- 2. All sampling units are known and so sampling frame can be prepared**

عشان اعمل frame

Example

:

Selection of 5 individuals out of 15(population).

- Give number for each individual (*sampling frame*). لازم اعطي رقم لكل واحد.
- Randomly select the needed sample (5 units) by lottery from a box containing numbers from 1 to 15. بسحبهم من الصندوق.

Systematic Random

Sample:

Selection depends on an interval (K-interval) which is calculated from both the size of population and the size of the sample.

K-interval =

$$\frac{\text{total population}}{\text{Sample size}}$$

لنفرض انه عدد

population 100

و sample size 10

بقسمهم 100 على 10 بتطلع

النتيجة 10

10 هيه k-interval تمام؟

هسا حختار رقم عشوائي من

1-5 لنفرض طلع معنا 3

وقتها بنلش بالعدد 3 بعدين

بنلش نزيد 10 كل مره

فيكون الناس الـ حختار هم

Example

□ Suppose we have a population = 120 and it is required to take a sample of 12

$$\square \text{ K-interval} = \frac{\text{Total population}}{\text{Sample size}} = \frac{120}{12} = 10$$

□ So, we have to select one out of each 10.

□ Then randomly select one out of the 1st 10, say 2. **بنختاره عشوائي هذا الرقم**

□ Then repeatedly add the k-interval to the selected number.

□ So, the sample will be the individuals number: 2, 12, 22, 32, 42, 52, 62, 72, 82, 92, 102 and 112.

□ Patients can be selected from the outpatient clinic by a modified method of this sample.

□ Example: select 8 persons from an outpatient clinic:

هون بختار رقم عشوائي بعدين بضل امشي فيه بالمثال هون طلع العدد 3 خلص بنضل
كل شوي نضيف 3

المريض الثالث وبعدين المريض السادس وهكذا لتجمع عدد العينة الي بدك اياها

- We take a random number from 1-10 (or 1-5 according to the rate), suppose the 3rd.
- then we will take every 3rd person coming to the clinic i.e. 3rd, 6th, 9th, 12th, etc. till we reach the desired sample size (8 persons).

□ By this way, there is no bias in selection (no subjective selection).

Advantages:

- **Does not require sampling frame.** مش
بحاجة اعطي ارقام وترتيب الهم
- **No bias in selection.**
- **We can select sample from large scale population.**

Stratified Random Sample:

It is used when the population is not homogeneous.

مهم نفرقتها

الي قبل كانوا homogeneous

لازم اقسامهم لمجموعتين او 3 ممكن
اوزعهم الناس الكبار وناس الصغار

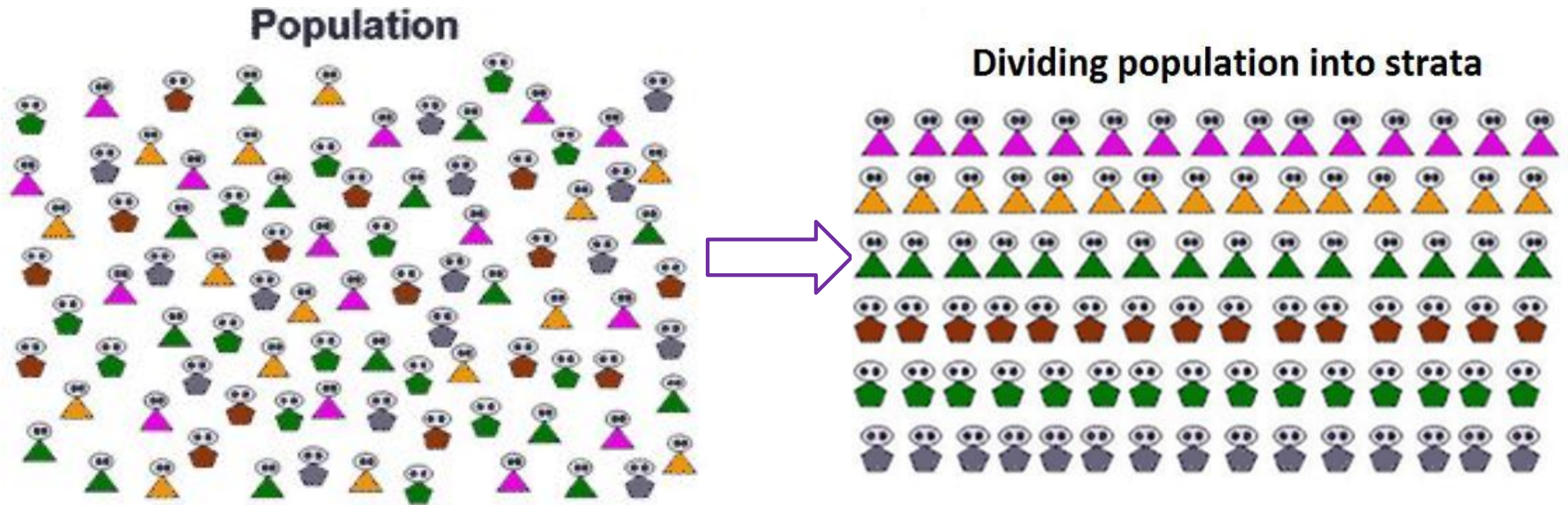
Obese or normal weight

Male ,female

Population



First: stratifying the population i.e. dividing the population into different strata each of which is as homogeneous as possible e.g. according to sex, age, residence, etc.

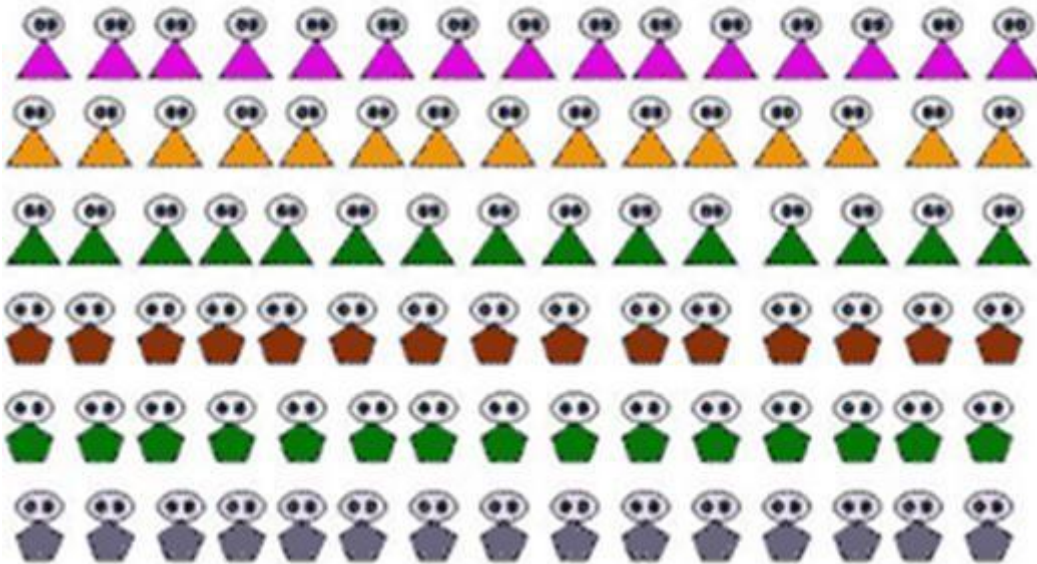


بعد ما قسمناها لstrata بختار من كل وحدة منهم random sample

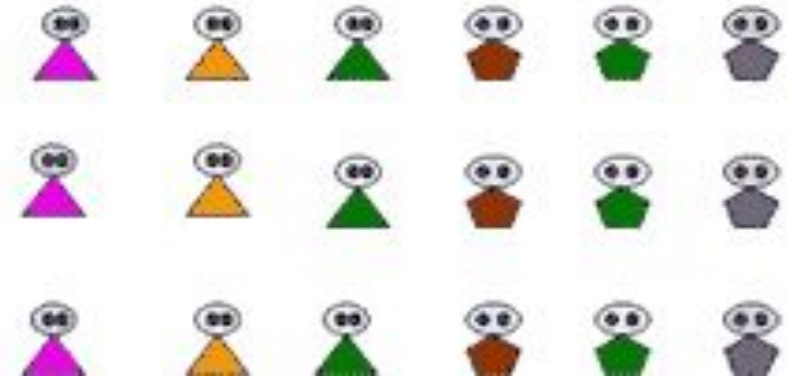
Second: selecting a simple random sample
(or systematic random sample) from each
stratum

هسا بختار من كل strata عدد محدد

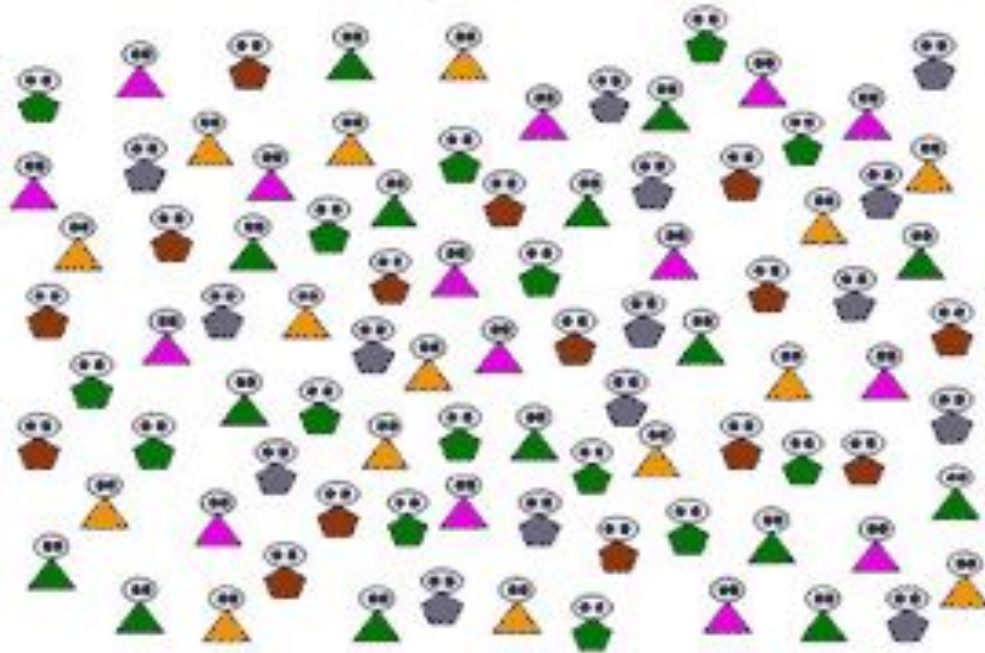
Strata



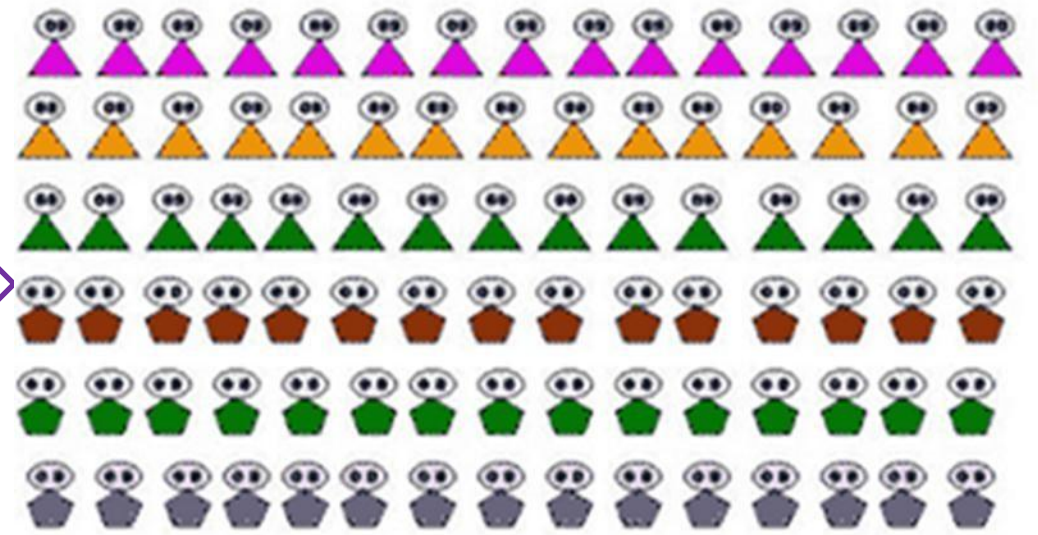
Stratified Random Sample



Population

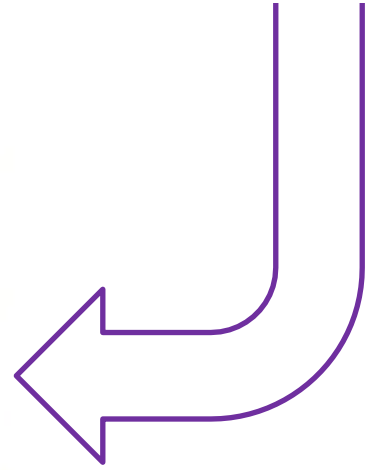
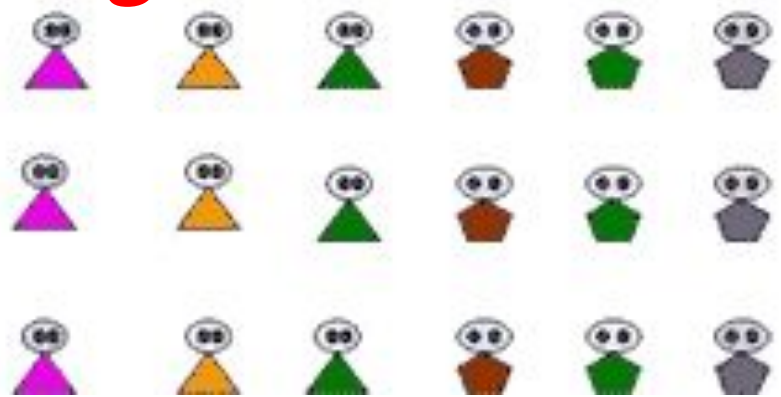


Dividing population into strata



هون مثلنا كل الناس

Stratified Random Sample



**كل strata حتوخذ
عدد ثابت من العدد الكلي
عشان sample size
مثلا عنا 2 strata ولازم
اختار عينة تكون 20
يوخذ من كل واحد 10

How many will be taken from each stratum ?

Selection can be done

using: 1- Equal allocation

Sample size \ No. of strata

method:

$$\text{No. required from each stratum} = \frac{\text{Sample size}}{\text{No. of strata}}$$

2- Proportional allocation method:

$$\text{No. required from each stratum} = \text{Sample size} \times \frac{\text{Size of stratum}}{\text{Total population}}$$

لنفرض انه قسمن الدفعه ل
male and female

وكان نسبة female اكثر

فمش حيزبط نؤخذ الطريقة الاولى وتعتمد
عليها

$$\text{العدد المطلوب من كل فئة} = \frac{\text{حجم العينة المطلوبة}}{\text{حجم الفئة}}$$

احجم المجتمع

مثال: اذا أردنا أخذ عينة ممثلة لمدرسة عدد طالبيها 300 ،منهم 120 في الصف الاول و 100 في الصف الثاني و80 في الصف الثالث والعينة المطلوبة 60 طالب؟؟

الطريقة الاولى: العدد المطلوب من كل صف $= 3 \setminus 60 = 20$ طالب
الطريقة الثانية:

العدد المطلوب من الصف الاول $= 300 \setminus 120 * 60 = 24$ طالب
من الصف الثاني $= 300 \setminus 100 * 60 = 20$
من الصف الثالث $= 300 \setminus 80 * 60 = 16$
الطريقة الثانية الي حلينا فيها اعتمدنا ع نسبة

مثال :

مصنع فيه 800 عامل , منهم 700 من الذكور , 100 من الاناث
العينة المطلوبة 80

الحل :

الطريقة الاولى : العدد المطلوب من الذكور والاناث $= 2 \setminus 80 = 40$
عامل

الطريقة الثانية

العدد المطلوب من الذكور $= 800 \setminus 700 * 80 = 70$ عامل
العدد المطلوب من الاناث : $800 \setminus 100 * 80 = 10$ عاملات

Cluster Sample

- A cluster: is a group of individuals that is present in certain locality or geographical area e.g. village, school, classroom, etc.

لازم افرقها عن Stratified

هون فرضا قسمنا الدوله لمحافظات لازم اختار محافظتين عشوائي واخذ كل الناس الي موجوده لهذول المحافظات

اما بحالة

Population



هون بنختار

2cluster

بشكل عشوائي

بعدين بوخذ

الناس الي

موجودين

بهذول

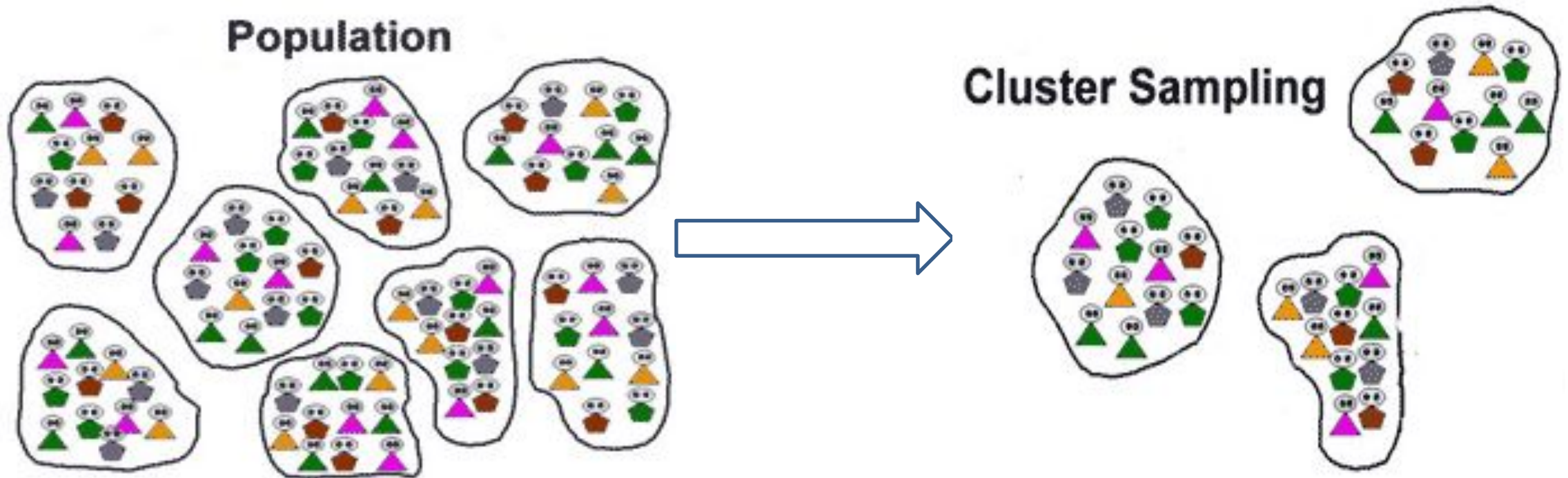
• مثال ثاني فيه مدرسه مقسمة ل 3 مراحل : ثانوي واعدادي
وابتدائي

بحاله Stratified حنقسمهم ل 3 مراحل وبعدين بوخذ من
كل مرحله عدد معين زي ما تعلمنا

اما بحاله cluster حختار مجموعه بشكل عشوائي مثلا لما
اخترنا طلع ثانوي

بوخذ العينة طلاب الثانوي كلهم

□ **First: we select a random sample of clusters.**



□ **Then: the clusters are taken as whole i.e. taking all individuals within the selected clusters.**

Examp

e:

□ If we need to select 5 districts of Al-Zarqa Governorate :

□ Prepare a list of all districts in AL-Zarqa Governorate.

يمكن يكونوا 20-25 منطقة مثلا بنحدددهم كلهم ,بعدين بنختار عشوائي 5 مناطق
وينوخذ كل السكان الموجودين بعذول الخمس مناطق

□ Then select randomly 5 districts out of the total districts

□ Then all people living in these 5 districts will be included in the study.

Exempl

e:

- **We can obtain a random sample of primary school children in an area by:**
- **Starting with a list of schools**
- **Draw a simple random sample of schools**
- **Then all children within the selected schools form the sample of children.**

Multistage Random sample

- It is usually used in case of national or widespread studies.

بحاله الدراسات الدوليه مثل الاردن كلها او مصر او اي دوله
بدكم اياها

- The field of work is arranged in levels or stages e.g. governorates, districts, villages, houses, families and individuals.

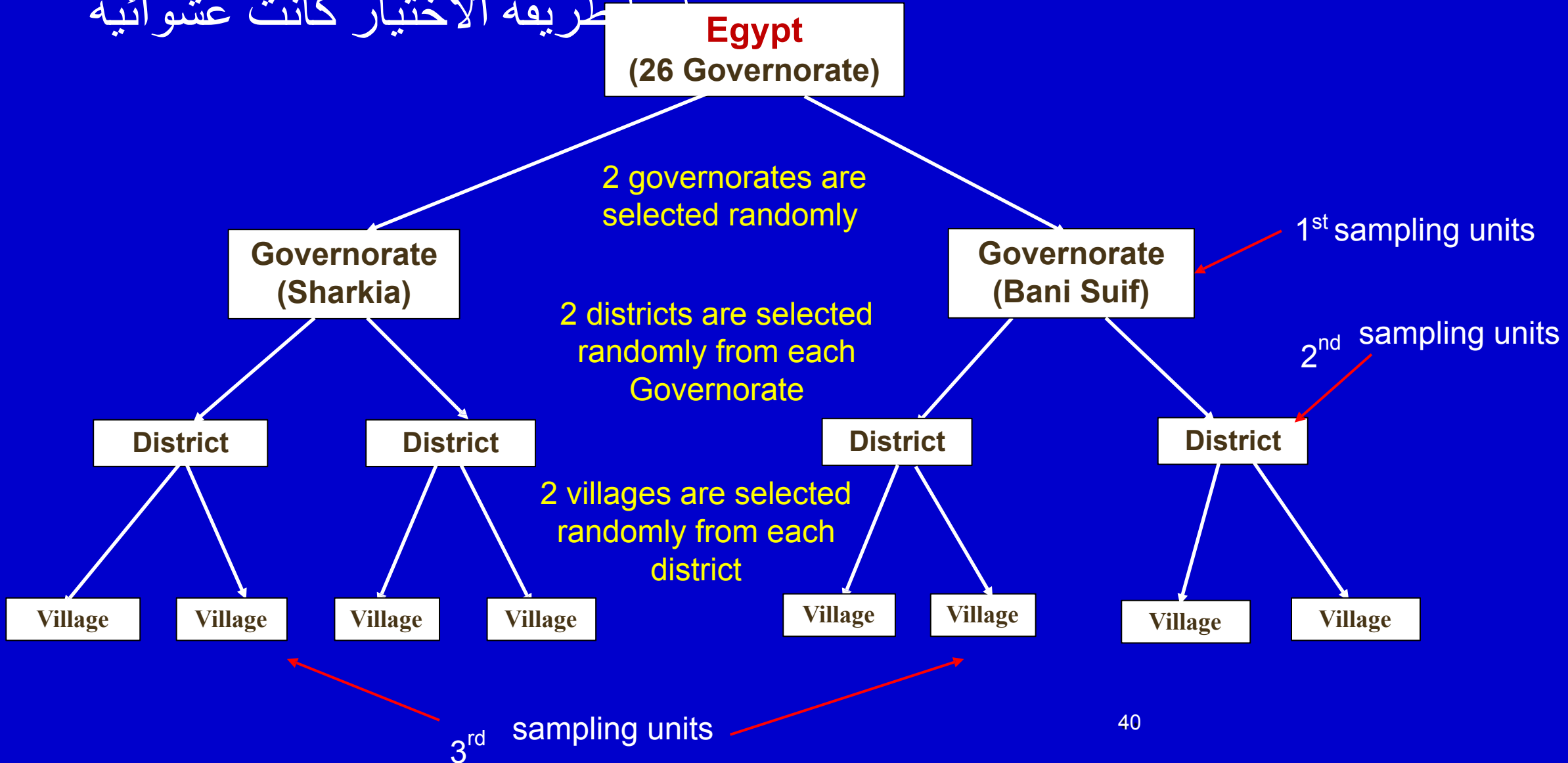
بتقسم الدوله لمحافظات بعدين مناطق بعد هيك القرى والعائلات اخر شي
بكونوا الافراد

لازم اختارهم حسب التسلسل الي فوق ويكون الاختيار عشوائي

- From each stage we select randomly the desired sample.

Example: Selection of sample of villages (8 for example) from Egypt for a morbidity Survey:

هذا المثال بمصر كيف اختاروا من 26 محافظة بس محافظتين
الطريقة الاختيار كانت عشوائية



Exempl

e:

Selection of sample of villages (8 for example) from Egypt for a morbidity Survey:

- **First, we have 26 governorates**
- **Select 2 governorates randomly (Governorates are the 1st sampling units)**
- **Then from each governorate, select 2 administrative districts (Districts are the 2nd sampling units)
→ two-stage sample.**
- **Then from each district, select 2 villages randomly (Villages are the 3rd sampling units) → three-stage sample.**

لو كان القرى فيها عدد كبير فبنرجع اعمل نفس الي عملته فوق

Sample size

How many individuals (or things) will be included in the study.

Determinants of sample size: مهم ل mcq

- 1. Available resources: man, money, materials and time.**
↑ resources → ↑ sample size and vice versa.
- 2. Number of variables affecting the study.**
↑ no. of variables → ↑ sample size and vice versa.

3. Prevalence of the problem or disease under study.

↑ prevalence → ↓ sample size and vice versa.

75% = **حتكون العينة صغيره**

5% يعني انه المرض نادر = **فلازم اخذ sample كبيرة**

4. Power of statistical test:

- It is the ability of the study to detect statistical significant relations.
- A **power of 80%** is suitable for most studies. It means that there is 20% error of missing a statistical significant difference in our selected sample.

5. Level of significance

- It is the ability of the study to detect statistical insignificant relations.
- **95 % is** usually the selected significant level.
- It means that **5% error** can occur in the study for getting significant result although it is not truly significant.

significant ع الرغم انها **Insignificant** □

- **↑ level of significance → ↑ sample size and vice versa.**

6. Effect size:

It is the difference expected between case and control groups or the strength of association.

For example: if the new treatment under study will produce percentage of cure 80% and the old treatment gives 70% (**control**) cure rate, then the effect size is 10%.

↑ effect size → decreases the sample size

- **Mean Value:** e.g. if we have mean value of 10 ± 2 for Hb of normal population and we assume that the Hb of cases of lead poisoning will be 8 ± 2.5 then the effect size will be the squared difference in the mean value divided by SD of the lead cases group

$$= \frac{(10-8)^2}{2.5}$$

2.5 الي بالمعادلة هيه standard deviation to cases

- If we have no mean value for cases, we can assume effect size of:
 - 0.2 for small suspected difference
 - 0.5 for moderate suspected difference
 - 0.8 for large suspected difference.
- We can get the mean value of the population from other previous studies or by doing a pilot study.

يا اما بنجيبها من دراسة سابقة لنفس الدراسة الي انا بشتغل عليها او بعمل دراسة استشادية من الناس ، الطيبة و الناس ، الي عندهم المرض ، طبعاً بتكون الدراسة صغيرة

7. **Type of study:** usually cross-sectional and case control study need larger samples (one reading is needed from each person) than cohort or randomized studies which need follow up and many reading for the same person.
8. **Cost of each sample:** if the cost is expensive, we have to minimize the sample.
9. **Variability in the studied population:** if great, the sample size should be larger.
10. **Reliability** كل ما اعيد التجربة يطلع عندي نفس النتيجة **and validity of the measurements:** The more valid and reliable method, the smaller is the sample.

- Sample size is calculated simply by many computer statistical packages e.g. Open Epi , Epi 6, SPSS.**
- But we have to fill some information in these statistical programs for calculation.**
- The needed information is specific for each type of study.**

In cross sectional studies (population survey):

1. **Population size**: from which the sample will be chosen.
2. **Prevalence** of the disease or factor under study in the population: from records, previous studies, websites or pilot study.
3. **Power of test** (In Epi 6: Result farthest from the prevalence rate that you would accept in your sample, higher or lower): 80% is reasonable and common level.
4. **Level of significance**: 95% is reasonable and common level.

In cohort and randomized clinical trials:

1. Two sided confidence level (level of significance, $1-\alpha$): usually 95%.
2. Power of study ($1-\beta$): usually 80%.
3. Ratio of unexposed to exposed in the sample: for equal samples use 1.

اول 3 نقاط ثابتات نسبتهم

1. Percent of disease or factor under study among unexposed (e.g. 5). **بنظريها**
من الدراسات السابقة
2. **One of the following:**
 - a. Odds ratio (e.g. 2).
 - b. Percent of disease or factor under study among exposed (e.g. 9.52)
 - c. Risk ratio or prevalence ratio = among exposed % \ among non exposed = 1.904)
 - d. Risk difference or prevalence difference = % among exposed – % among non exposed (9.52 – 5 = 4.52).

In unmatched case control study:

1. Two sided confidence level (level of significance, $1-\alpha$): usually 95%.
2. Power of study ($1-\beta$): usually 80%.
3. Ratio of control to cases in the sample: for equal samples use 1. يعني case 100, control 100

يمكن النسبة تكون 3 يكون case 300 و control 100

1. Expected frequency of exposure among controls (e.g. 40). من الدراسات السابقة
2. One of the following:
 - a. Odds ratio (e.g. 2).
 - b. Expected frequency of exposure among cases (e.g. 57.14).

Sample size for comparing two means (mean difference):

1. Two sided confidence level (level of significance, $1-\alpha$): usually 95%.
2. Power of study ($1-\beta$): usually 80%.
3. Ratio of sample size = $\frac{\text{Group 2}}{\text{Group 1}}$: for equal samples use 1.
4. Mean of group 1 and mean of group 2 (or difference between the 2 means, mean difference)
5. **One of the following:**
 - a. Standard deviations of the 2 groups
 - b. Variance of the 2 groups

<https://www.openepi.com/SampleSize/>

| Sample Size for Cross-Sectional, Cohort, & Randomized Clinical Trial Studies | | |
|--|------|----------------------------|
| Two-sided confidence level(%) | 95 | (1-alpha) usually 95% |
| Power (1-beta or % chance of detecting) | 80 | Usually 80% |
| Ratio of Unexposed to Exposed in sample | 1.0 | For equal samples, use 1.0 |
| Percent of Unexposed with Outcome | 5 | Between 0.0 and 99.9 |
| Please fill in 1 of the following. The others will be calculated. | | |
| Odds ratio | 2 | |
| Percent of Exposed with Outcome | 9.52 | Between 0.0 and 99.9 |
| Risk/Prevalence Ratio | 1.90 | |
| Risk/Prevalence difference | 4.52 | Between -99.99 and 99.99 |

هيك بنكون خلصنا اي سؤال جاهزة ان شاء الله ♥
Nour Al-zoubi ♥

