

## Estimation de la taille de l'échantillon

### Etapes pour calculer la taille de l'échantillon par strate

1. Choisir la stratification (par ex., régions, district)
  2. Définir la population (N) de chaque strate
  3. Décider de ou des indicateurs clefs
  4. Estimer la moyenne et la variance de l'indicateur clef
  5. Décider du niveau de précision et de confiance
  6. *Calculer la taille de l'échantillon initial (n) en fonction du budget et temps disponibles.*
  7. *Utiliser l'échantillon simple aléatoire par strate pour sélectionner votre échantillon représentatif*
- Echantillon final

### Calculer la taille de l'échantillon – échantillon aléatoire

Pour estimer la taille de l'échantillon, vous devez connaître:

- Estimation de la prévalence ou de la moyenne et de la déviation standard de l'indicateur clef (par exemple, 30% des ménages avec une consommation alimentaire pauvre)
- Niveau de précision désiré (par exemple:  $\pm 5\%$ )
- Niveau de confiance (par exemple: 95%)
- Population (seulement si inférieure à 10 000 personnes, dans le cas contraire, la taille de la population ne va pas influencer la taille requise de l'échantillon)
- Taux de réponse attendu (par exemple: 90%)
- Nombre d'individus éligibles par ménage (si applicable)

Note:

- **Précision** est la variabilité de l'estimation
- Le **niveau de confiance** est la probabilité d'obtenir le même résultat avec un autre échantillon, tous les autres paramètres restant les mêmes.
- **L'intervalle de confiance** est l'intervalle autour de l'estimation pour lequel on a le niveau de confiance désiré.
- **Choisir la taille de l'intervalle de confiance** (Précision) pour un niveau de confiance donné.
- Tant que la taille de la population ciblée est de plus de quelques milliers de ménages, elle ne va pas influencer la taille requise de l'échantillon. Si la taille de la population est moindre, la taille requise pour l'échantillon va diminuer légèrement.

### Prévalence versus moyenne

- **La prévalence** est le nombre total de cas d'une variable qui est typiquement binaire au sein d'une population divisée par sa population totale (par exemple, taux de retard de croissance, taux de chômage).
- **La moyenne** est la valeur attendue d'une variable qui est typiquement continue ou qui rentre dans une fourchette prescrite pour une population donnée (par ex., hauteur, poids, âge) .

Pour calculer la taille de l'échantillon

- Traiter les variables comme des taux de prévalence seulement quand elles sont naturellement binaires.
- **NE PAS mettre un seuil** aux indicateurs continus lors du calcul de la taille de l'échantillon même si pour l'analyse ils deviendront des indicateurs de prévalence.

*Par exemple, le Score de Consommation Alimentaire (SCA) est un indicateur continu (0-112) mais a deux seuils fixés à 21 et 35 pour la prévalence pauvre et limite.*

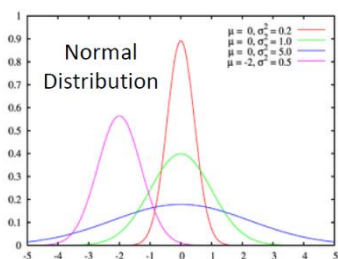
### Choisir la bonne distribution

Pour les variables continues, nous devons choisir une distribution de probabilité qui corresponde le mieux aux données.

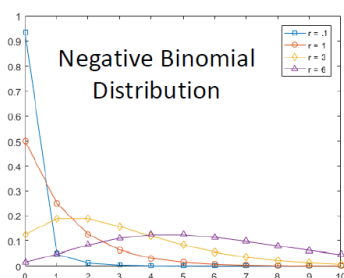
**ETAPE 1:** Préparer TOUJOURS un histogramme des données passées et choisir la meilleure forme de distribution !

Les deux distributions les plus courantes au PAM sont:

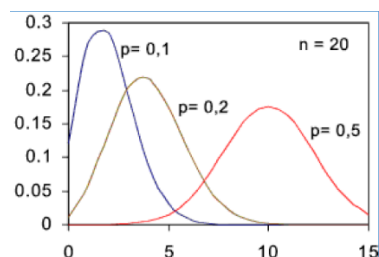
- Distribution normale :



- Distribution de type binomiale négative :



Les taux de prévalence suivent toujours une distribution binomiale ce qui explique pourquoi ils sont faciles à traiter d'un point de vue mathématique.

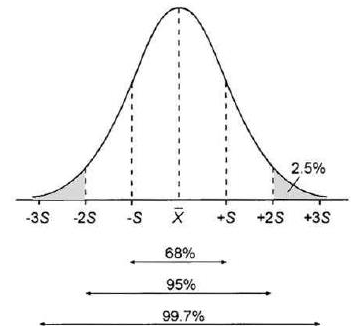


**Formule pour la taille de l'échantillon pour la prévalence**  
(enquête unique qui utilise un échantillonnage aléatoire)

Pour calculer la taille d'un échantillon pour estimer la prévalence avec une limite de confiance de 95% :

$$\hat{n} = \frac{1.96^2 \times (P)(1-P)}{d^2}$$

- 1.96 = valeur Z value pour des limites de confiance de 95%
- P = prévalence estimée (par exemple 0,3 pour 30%)
- (P)(1-P) = écart pour une variable binaire (binomiale)
- d = ½ de l'intervalle de confiance désiré (par exemple 0,025 pour ± 5%)

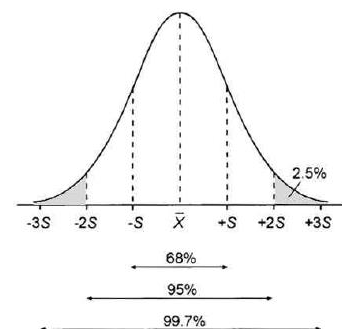


**Formule pour la taille de l'échantillon pour une variable continue**  
(enquête unique utilisant l'échantillonnage aléatoire)

Pour calculer une taille d'échantillon pour des estimations avec une limite de confiance de 95% :

$$\hat{n} = \frac{1}{\frac{1}{N} + \frac{(\mu*d)^2}{(1.96^2)\sigma^2}}$$

- 1.96 = valeur Z pour des limites de confiance de 95%
- $\mu$  = Moyenne attendue
- $\sigma$  = écart de la variable
- d = ½ de l'intervalle de confiance désiré (par exemple 0,025 pour ± 5%)
- $\mu*d$  = valeur absolue de l'intervalle de confiance
- N = population de chaque strate.



**Formule pour calculer la taille de l'échantillon pour un échantillonnage en grappe**

- Pour calculer la taille de l'échantillon pour estimer la prévalence avec un intervalle de confiance de 95% pour un échantillonnage en grappe :

$$N = \text{DEFF} \times \frac{1.96^2 \times (P)(1-P)}{d^2}$$

DEFF = Effet de plan

1.96 = valeur Z pour p = 0,05 ou 95% intervalle de confiance

P = prévalence estimée

d = précision désirée (par exemple, 0,05 pour ±5%)

### Qu'est-ce l'effet de plan ?

- Ratio de la variance actuelle selon la méthode d'échantillonnage utilisée et de la variance hypothétique avec un échantillonnage aléatoire simple.
- Pour l'échantillonnage par grappe:
  - N = # des échantillons | K = # des grappes | M = # des échantillons par grappe

$$D_{eff} = \frac{Var(x_{cluster})}{Var(x_{SRS})} = \frac{M \sum_{k=1}^K (\bar{x}_k - \bar{x})^2}{\sum_{i=1}^N (x_i - \bar{x})^2} = \frac{\text{variance between clusters}}{\text{total variance}} \quad \begin{array}{l} \text{Variance entre les grappes} \\ \text{Variance totale} \end{array}$$

- $D_{eff} > 1$  toujours pour l'échantillonnage par grappe car il ne peut jamais être plus efficace que l'échantillonnage aléatoire et inversement pour l'échantillonnage stratifié.

### Effet de plan

L'effet de plan augmente quand

- Les indicateurs clefs sont géographiquement très regroupés (par exemple, source d'eau, accès aux soins de santé)
  - Source d'eau
  - Accès aux soins de santé
- Quand le nombre de grappes diminue et la taille des grappes augmente

Pour minimiser l'effet de plan

- Inclure plus de grappes de taille plus petite
- Stratifier l'échantillon en des groupes plus homogènes
- Toutes les grappes doivent avoir la même taille

### Exemple 1 : l'indicateur clef distribué normalement

*Score de consommation alimentaire*

Exemple 1: Etude de cas sur l'Irak

1. **Choisir la stratification (Strate):** 18 Provinces de l'Irak
2. **Définir la Population (N) de chaque strate:**

Governorate(s)	Population 2015
Anbar	1723154
Babil	2008609
Najaf	1428979
Baghdad	7882807
Basrah	2822646
Diyala	1592434
Duhok	848524
Erbil	1650224
Kerbala	1183818
Kirkuk	1551670
Missan	1080392
Wassit	1340116
Muthanna	792339
Qadissiya	1254963
Ninewa	3397659
Salah al-Din	1551978
Sulaymaniyah	1858506
Thi-Qar	2035734

3. **Indicateur clé** : Score de Consommation Alimentaire (SCA)

4. **Calculer la moyenne ( $\mu$ ) et la déviation standard (STDev) du SCA (Indicateur clef)**

Governorate(s)	FCS Mean	FCS Stdev
Anbar	75.0	20.5
Babil	82.5	17.5
Najaf	82.5	17.5
Baghdad	82.5	17.5
Basrah	85.0	17.5
Diyala	80.0	17.5
Duhok	77.5	20.5
Erbil	80.0	17.5
Kerbala	82.5	17.5
Kirkuk	75.0	17.5
Missan	80.0	17.5
Wassit	80.0	17.5
Muthanna	82.5	17.5
Qadissiya	82.5	17.5
Ninewa	77.5	20.5
Salah al-Din	80.0	17.5
Sulaymaniyah	80.0	17.5
Thi-Qar	85.0	17.5

La déviation standard a été arrondie pour avoir seulement deux niveaux

5. **Décider de la précision et du niveau de confiance :**

- 90% intervalle de confiance
- $Z=1,645$

**N.B. Un intervalle de confiance de 10% à 90% est le minimum absolu !**

**Plus l'intervalle de confiance est grand, plus la capacité à détecter des tendances est mauvaise !**

6. **Taille totale de l'échantillon initial ( $n$ )** : 2 200 personnes enquêtées

La taille de l'échantillon requise **dans chaque province** calculée avec l'équation (1.2) sera de :

Governorate(s)	Population	FCS Mean	FCS Stdev	SRS 5%-90%
Anbar	1723154	75	20.5	323
Babil	3437588	82.5	17.5	195
Najaf	3437588	82.5	17.5	195
Baghdad	7882807	82.5	17.5	195
Basrah	2822646	85	17.5	184
Diyala	1592434	80	17.5	207
Duhok	848524	77.5	20.5	303
Erbil	1650224	80	17.5	207
Kerbala	1183818	82.5	17.5	195
Kirkuk	1551670	75	17.5	236
Missan	2420508	80	17.5	207
Wassit	2420508	80	17.5	207
Muthanna	2047302	82.5	17.5	195
Qadissiya	2047302	82.5	17.5	195
Ninewa	3397659	77.5	20.5	303
Salah al-Din	1551978	80	17.5	207
Sulaymaniyah	1858506	80	17.5	207
Thi-Qar	2035734	85	17.5	184
Iraq	36004552			3348

Après avoir estimé la taille de l'échantillon dans chaque province, il faut recalibrer la taille à la taille totale de l'échantillon de 2 200 (n) en utilisant la formule suivante :

$$\hat{n} = \frac{n}{n_{IRAQ}}$$

Governorate(s)	Population	FCS Mean	FCS Stdev	SRS 5%-90%	Given 5/90 -> scaling by 2200
Anbar	1723154	75	20.5	323	212
Babil	3437588	82.5	17.5	195	128
Najaf	3437588	82.5	17.5	195	128
Baghdad	7882807	82.5	17.5	195	128
Basrah	2822646	85	17.5	184	121
Diyala	1592434	80	17.5	207	136
Duhok	848524	77.5	20.5	303	199
Erbil	1650224	80	17.5	207	136
Kerbala	1183818	82.5	17.5	195	128
Kirkuk	1551670	75	17.5	236	155
Missan	2420508	80	17.5	207	136
Wassit	2420508	80	17.5	207	136
Muthanna	2047302	82.5	17.5	195	128
Qadissiya	2047302	82.5	17.5	195	199
Ninewa	3397659	77.5	20.5	303	136
Salah al-Din	1551978	80	17.5	207	207
Sulaymaniyah	1858506	80	17.5	207	136
Thi-Qar	2035734	85	17.5	184	121
Iraq	36004552			3348	2199

Exemple2: Indicateur avec une distribution binomiale négative

Indice simplifié des Stratégies de Survie

Echantillonnage aléatoire par grappe

**Exemple 2: Etude de cas du Malawi**

**1. Définir la Population totale (N):**

Malawi Total	16512568
--------------	----------

2. **Définir la population dans chaque strate (Nh):** (Strate= 10 districts agrégés du Malawi)

District(s)	Population
Blantyre-Mwanza-Neno-Balaka	1933263
Chikwawa-Nsanje	892772
Chiradzulu-Mulanje-Thyolo-Zomba-Phalombe	3170421
Dedza-Ntcheu	1406995
Dowa-Ntchisi-Kasungu-Mchinji	2322675
Lilongwe	2310728
Machinga-Mangochi	1608745
Mzimba-Karonga-Rumphi	1578519
Nkhata Bay-Chitipa-Likoma	514968
Nkhotakota-Salima	773482

3. **Indicateur clé: Indice simplifié des Stratégies de Survie (rCSI)**

4. **Calculer la moyenne ( $\mu$ ) et la déviation standard (ST Dev) de l'indice simplifié des stratégies de survie (rCSI)**

Population du District(s) rCSI Moyenne rCSI Stdev

District(s)	Population	rCSI Mean	rCSI Stdev
Blantyre-Mwanza-Neno-Balaka	1933263	12	4
Chikwawa-Nsanje	892772	16	5
Chiradzulu-Mulanje-Thyolo-Zomba-Phalombe	3170421	14	4
Dedza-Ntcheu	1406995	16	5
Dowa-Ntchisi-Kasungu-Mchinji	2322675	16	4
Lilongwe	2310728	12	4
Machinga-Mangochi	1608745	16	5
Mzimba-Karonga-Rumphi	1578519	14	4
Nkhata Bay-Chitipa-Likoma	514968	16	5
Nkhotakota-Salima	773482	14	4
Malawi Total	16512568	15	4

La déviation standard a été arrondie pour avoir juste deux niveaux

$$StDev = \sqrt{\frac{\mu + \mu^2}{n_{h,baseline}}}$$

Quand  $n_{h,baseline}$  = taille de l'échantillon de chaque strate du niveau de base avec  $h=1,/,10$

5. **Décider du niveau de précision et de confiance :**

- 90% pour l'intervalle de confiance
- $Z=1.645$

**N.B. Un intervalle de confiance de 10% à 90 % est le minimum absolu !**

**Plus l'intervalle de confiance est grand et plus la capacité à détecter des tendances est mauvaise!**

## Echantillonnage aléatoire simple

6. **Taille totale de l'échantillon (n):** 2 000 personnes enquêtées

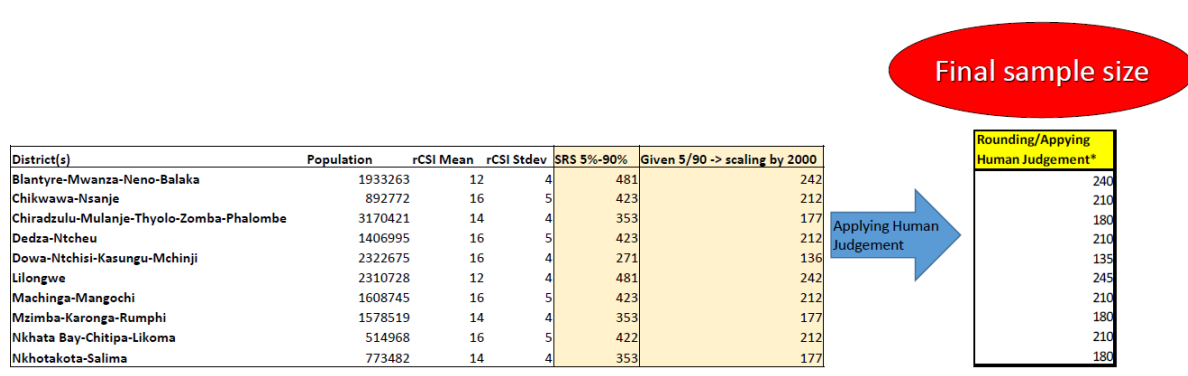
La taille de l'échantillon requise dans chaque province calculée avec l'équation (1.2) sera de :

- District(s)
- Population
- rCSI Mean = rCSI Moyen
- rCSI Stdev= rCSI déviation standard
- SRS 5%-90%

District(s)	Population	rCSI Mean	rCSI Stdev	SRS 5%-90%
Blantyre-Mwanza-Neno-Balaka	1933263	12	4	481
Chikwawa-Nsanje	892772	16	5	423
Chiradzulu-Mulanje-Thyolo-Zomba-Phalombe	3170421	14	4	353
Dedza-Ntcheu	1406995	16	5	423
Dowa-Ntchisi-Kasungu-Mchinji	2322675	16	4	271
Lilongwe	2310728	12	4	481
Machinga-Mangochi	1608745	16	5	423
Mzimba-Karonga-Rumphi	1578519	14	4	353
Nkhata Bay-Chitipa-Likoma	514968	16	5	422
Nkhotakota-Salima	773482	14	4	353

Rééchelonner pour la taille totale de l'échantillon de 2 000 personnes :

- Given 5/90 -> scaling by 2000 = Donné 5/90 -> calibré par 2000
- Applying Human Judgement = Utiliser le jugement humain
- Rounding = Arrondir/utiliser
- Final sample size = Taille finale de l'échantillon



## Nombre de tentatives – Cas du Malawi:

Donné

- Taux de réponse (25%)
- Taille estimée de l'échantillon dans chaque strate:



Aggregated Districts	Rounding/Appying Human Judgement*
Blantyre-Mwanza-Neno-Balaka	240
Chikwawa-Nsanje	210
Chiradzulu-Mulanje-Thyolo-Zomba-Phalombe	180
Dedza-Ntcheu	210
Dowa-Ntchisi-Kasungu-Mchinji	135
Lilongwe	245
Machinga-Mangochi	210
Mzimba-Karonga-Rumphi	180
Nkhata Bay-Chitipa-Likoma	210
Nkhotakota-Salima	180

**Les tentatives souhaitées = taille estimée de l'échantillon / taux de réponse**

- Aggregated Districts = Districts agrégés
- Desired Attempts = Tentatives souhaitées

Aggregated Districts	Rounding/Appying Human Judgement*	Desired Attempts
Blantyre-Mwanza-Neno-Balaka	240	960
Chikwawa-Nsanje	210	840
Chiradzulu-Mulanje-Thyolo-Zomba-Phalombe	180	720
Dedza-Ntcheu	210	840
Dowa-Ntchisi-Kasungu-Mchinji	135	540
Lilongwe	245	980
Machinga-Mangochi	210	840
Mzimba-Karonga-Rumphi	180	720
Nkhata Bay-Chitipa-Likoma	210	840
Nkhotakota-Salima	180	720