

Chapitre 4

Les techniques d'échantillonnage

17/04/2006

Claude Saegerman



Le problème à résoudre ?

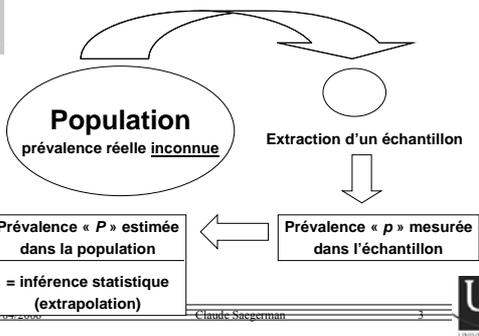
- La plupart des enquêtes épidémiologiques correspondent à des enquêtes par sondage : - concernent un échantillon de la population - afin d'en réduire les coûts
- L'étude d'une caractéristique de l'échantillon permet d'estimer celle de la population. Il s'agit donc d'une « extrapolation » dont la validité sera tributaire de la qualité du choix de l'échantillon

17/04/2006

Claude Saegerman



Estimation de la prévalence réelle d'une maladie dans une population à partir de la prévalence mesurée sur un échantillon de la population



17/04/2006

Claude Saegerman



Qualités d'un échantillon : notions d'exactitude et de précision



centré

↑

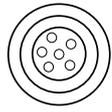
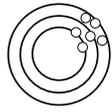
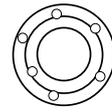
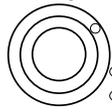
OUI

décentré

↑

NON

Tir exact

groupé → OUI		
Tir précis		
dispersé → NON		

17/04/2006

Facteurs conditionnant l'exactitude

- L'échantillon doit être représentatif de la population : tout individu de la population à représenter peut figurer dans l'échantillon, avec une probabilité non nulle, connue et égale pour tous les individus présentant les mêmes caractéristiques d'intérêt.
- La représentativité peut être globale ou pour une caractéristique. Le tirage au sort permet d'assurer la représentativité globale.
- Un échantillon, si petit soit-il, est représentatif dès qu'il résulte d'un tirage au sort.



17/04/2006 Claude Saegerman

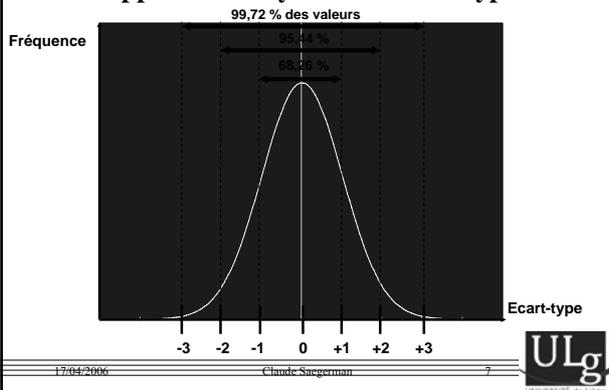
Exactitude et précision

<p>Exactitude</p> <ul style="list-style-type: none"> <input type="checkbox"/> échantillon représentatif <input type="checkbox"/> par tirage au sort : <ul style="list-style-type: none"> - table des nombres aléatoires - fonction [Alea] dans Excel 	<p>Précision</p> <ul style="list-style-type: none"> <input type="checkbox"/> varie selon la taille de l'échantillon <input type="checkbox"/> IC 95% = $p \pm 2 \sigma$ <ul style="list-style-type: none"> - avec p: la proportion et - avec σ: l'écart type <input type="checkbox"/> en augmentant l'effectif testé (taille)
--	--



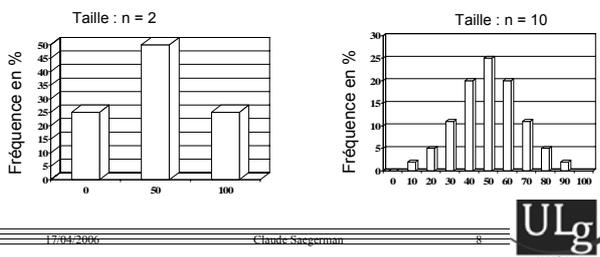
17/04/2006 Claude Saegerman

Répartition du pourcentage des valeurs par rapport à la moyenne et l'écart-type



Facteurs conditionnant la précision

- Lorsque la taille d'échantillon est suffisamment grande, la dispersion des résultats obtenus par tirage au sort peut être représentée par une courbe en cloche



Calcul de l'écart-type

- L'écart type est donné par la formule suivante :

$$\sigma = \sqrt{[(p \cdot q) / n]}$$

avec :

- σ : écart-type
- p : proportion estimée comprise entre 0 et 1 (par exemple, la prévalence)
- q : complément à 1 de la proportion = (1-p)
- n : nombre d'unités dans l'échantillon

- Cette formule ne s'applique que lorsque la taille de l'échantillon est faible par rapport à celle de la population ($n/N < 5$ à 10 % ; population dite infinie)

Types d'échantillonnage

- sondage aléatoire simple
- échantillonnage systématique
- sondage aléatoire stratifié
- sondage aléatoire en grappes
- sondage aléatoire à plusieurs degrés

17/04/2006

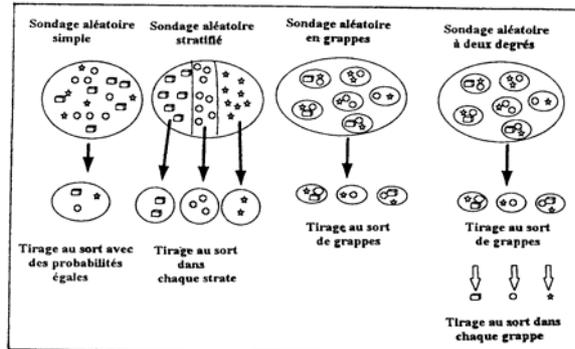
Claude Saegerman

10



FIGURE 3.8

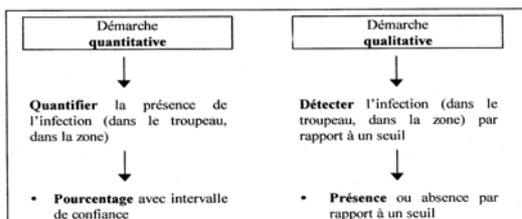
Les principaux échantillons aléatoires



Objectif d'une enquête descriptive

- soit quantitatif (infecté à un certain degré)
- soit qualitatif (infecté ou indemne par rapport à un seuil)

Différence d'objectif et de démarche (quantitative, qualitative) au sein d'un troupeau ou d'une zone



Démarche qualitative et quantitative

Démarche quantitative

- quel nombre d'animaux étudier pour estimer la proportion d'atteints dans une population donnée, avec une précision donnée ?
- fonction du degré de précision (relative) désiré
- fonction du nombre d'animaux dans le troupeau
- fonction de la proportion d'animaux atteints

Démarche qualitative

- détecter les troupeaux dont le pourcentage de réponse positive est supérieur au niveau retenu et
- identifier les troupeaux indemnes ou ceux dont le taux de prévalence est inférieur à un taux de prévalence choisi

17/04/2006

Claude Saegerman

13



Illustration de la différence entre la démarche quantitative et la démarche qualitative

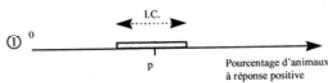
Dans la première démarche (1), on veut évaluer le pourcentage d'animaux à réponse positive (p), avec un certain niveau de précision représenté par l'intervalle de confiance (I.C.).

Dans la seconde (2), on désire simplement détecter l'infection, avec un certain risque d'erreur, par rapport à un niveau défini (p_0). Le nombre d'unités nécessaire peut être déterminé pour une réponse de type présence/absence de la maladie ou pour situer la prévalence par rapport à une fourchette allant de 0 à p_0 .

Dans le premier cas^a, toute réponse positive fait conclure à la présence de la maladie.

Dans le second cas^b, on compare le nombre de réponses positives obtenues k à un nombre limite de réponses positives c :

lorsque $k \leq c$, la prévalence est probablement inférieure à p_0 ;
lorsque $k > c$, la prévalence est probablement supérieure à p_0 .



- a. Uniquement des réponses négatives sur l'échantillon : le pourcentage d'animaux ou de cheptels à réponse positive dans la population est inférieur au pourcentage choisi. Une ou plusieurs réponses positives sur l'échantillon : l'infection est détectée.
- b. • nombre de réponses positives $k \leq c$: taux de prévalence inférieur à p_0 .
• nombre de réponses positives $k > c$: taux de prévalence supérieur ou égal à p_0 .

17/04/2006



Détermination la taille des échantillons: démarche quantitative (1)

- Fluctuation d'échantillonnage => intervalle de confiance
- Précision absolue
- Précision relative
- Fraction ou taux de sondage

17/04/2006

Claude Saegerman

15



Détermination la taille des échantillons: démarche quantitative (2)

- **Précision absolue (Pa)** : « mesure de la dispersion des valeurs de l'estimation autour de la moyenne » = moitié de l'intervalle de confiance
- **Précision relative** : « rapport de la précision absolue sur la valeur estimée » = Pa/p (p = proportion estimée)
- Voir les tables suivantes

17/04/2006

Claude Saegerman

16



Estimation d'une proportion :

Ecart-type, intervalle de confiance, précision absolue et précision relative
Attention : un intervalle de confiance ne doit être calculé que lorsqu'il s'agit d'une estimation sur un échantillon et non pas lorsque les analyses ont porté sur l'ensemble de la population.

- Soit **p** la **proportion** déterminée sur un échantillon **n** de la population **N**
 (Exemple : $p = 0,20$ car sur un échantillon de 100 animaux, 20 ont fourni une réponse positive).

- L'**écart-type** de cette proportion est : $\sigma = \sqrt{\frac{pq}{n}}$ lorsque $\frac{n}{N} < 10$ p. cent
 p : proportion (0 à 1)
 q : complément à 1 de la proportion
 n : nombre d'unités dans l'échantillon

$$\sigma = \sqrt{\left(1 - \frac{n}{N}\right) \frac{pq}{n}} \quad \text{lorsque } \frac{n}{N} > 10 \text{ p. cent}$$

Conditions d'application* : $np > 5$ $nq > 5$

(Exemple : $\sqrt{\frac{0,2 \times 0,8}{100}} = 0,04$)

- L'**intervalle de confiance (IC)** à 95 p. cent est : $p \pm 2\sigma$
 (risque d'erreur : 5 p. cent)



$$IC = 0,20 \pm 0,08 = 0,12 \text{ à } 0,28 \quad (2\sigma = 0,08)$$

- La **précision absolue (Pa)** à 95 p. cent est : $Pa = 2\sigma$
 (moitié de l'intervalle de confiance)

(Exemple : $Pa = 2 \sqrt{\frac{0,2 \times 0,8}{100}} = 0,08$)

- La **précision relative (Pr)** est : $Pr = \frac{Pa}{p}$
 (quotient de la précision absolue par la proportion)

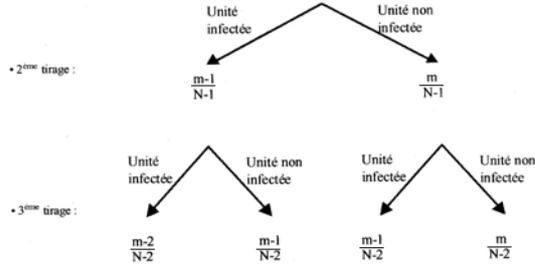
$$Pr = \frac{0,08}{0,20} = \frac{8}{20} = 40 \text{ p. cent}$$

* Ces formules permettant de calculer l'intervalle de confiance ne peuvent être utilisées que si les échantillons sont grands (c'est-à-dire $np > 5$ et $nq > 5$). Les tableaux III.2 et III.3 peuvent être utilisés aussi bien pour les grands échantillons que pour les petits car, pour ces derniers, on tient compte de la dissymétrie de la variation d'échantillonnage.

Schéma illustrant les modifications de probabilité dans un tirage sans remise

Probabilité de tirer au sort une unité infectée

• 1^{er} tirage : $p = \frac{m}{N}$ où m est le nombre d'infectés dans la population de taille N



Taille des échantillons correspondant à un faible taux de sondage (<10%)

Table de détermination du nombre de sujets nécessaire en fonction de la prévalence attendue et de la précision relative souhaitée, dans une population « infinie » (taux de sondage < 10 p. cent)
 Si le nombre obtenu conduit à un taux de sondage supérieur à 10 p. cent, consulter les tableaux III.5 et III.6.
 Pour les prévalences attendues supérieures à 50 p. cent, utiliser le complément à 100 p. cent.

Précision relative	Prévalence attendue (p. cent)													
	1	2	3	4	5	10	15	20	25	30	35	40	45	50
10 p. cent	3 8032	18 824	12 422	9 220	7 300	3 458	2 177	1 537	1 153	897	714	577	470	385
20 p. cent	9 508	4 706	3 106	2 305	1 875	865	545	385	289	225	179	145	118	97
30 p. cent	4 226	2 992	1 981	1 025	812	385	242	171	129	100	80	65	53	43
40 p. cent	2 777	1 777	1 177	777	577	437	317	237	177	137	107	87	70	57
50 p. cent	1 522	753	497	369	292	139	88	62	47	36	29	24	19	16
60 p. cent	1 057	523	346	257	203	97	61	43	33	25	20	17	14	11
70 p. cent	777	385	254	189	149	71	45	32	24	19	15	13	11	10
80 p. cent	595	295	195	145	115	55	35	25	20	17	14	13	11	10
90 p. cent	500	250	167	125	100	50	33	25	20	17	14	13	11	10
100 p. cent	500	250	167	125	100	50	33	25	20	17	14	13	11	10

Taille des échantillons correspondant à un taux de sondage >10%

Table donnant le nombre de sujets nécessaire pour un échantillon correspondant à un taux de sondage supérieur à 10 p. cent, à partir du nombre n donné par le tableau III.4 pour un échantillon dans une population infinie (taux de sondage <10 p. cent), pour différentes valeurs de n de 10 à 1000 et pour différentes tailles de population de 50 à 20 000
 Le nombre n' d'unités nécessaires, dans un échantillon lors de taux de sondage supérieur à 10 p. cent, est situé à l'intersection entre la colonne indiquant la taille de la population et la rangée correspondant au nombre n d'unités nécessaires pour un taux de sondage inférieur à 10 p. cent dans cette population (n est fourni dans le tableau III.4).
 Le coin inférieur droit de ce tableau est en partie commun avec le coin supérieur gauche du tableau III.6.

n	Taille de la population																						
	50	100	150	200	250	300	350	400	450	500	600	700	800	900	1 000	1 500	2 000	2 500	3 000	4 000	5 000	10 000	20 000
10	9	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
20	15	17	18	19	19	19	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
30	19	24	25	27	27	28	28	28	29	29	29	29	30	30	30	30	30	30	30	30	30	30	30
40	23	29	32	34	35	36	36	37	37	38	38	38	39	39	39	39	40	40	40	40	40	40	40
50	25	34	38	40	42	43	44	45	45	46	47	47	48	48	49	49	50	50	50	50	50	50	50
60	28	38	43	47	49	50	52	53	53	54	55	56	56	57	57	58	59	59	60	60	60	60	60
70	30	42	48	52	55	57	59	60	61	62	63	64	65	65	66	67	68	69	69	70	70	70	70
80	33	48	55	63	67	70	72	74	75	77	79	80	81	82	83	85	87	87	88	89	89	90	90
90	34	48	57	63	67	70	72	74	75	77	79	80	81	82	83	85	87	87	88	89	89	90	90
100	34	50	60	67	72	75	78	80	82	84	86	88	89	90	91	94	96	97	97	98	99	100	100
110	35	53	64	71	77	81	84	87	89	91	93	96	97	99	100	103	105	106	107	108	108	109	110
120	36	55	67	75	82	86	90	93	95	97	100	103	105	106	108	112	114	115	116	117	118	119	120
130	37	57	70	79	86	91	95	99	101	104	107	110	112	114	116	120	123	124	125	126	127	129	130
140	38	59	73	83	90	95	100	104	108	112	116	120	123	125	128	132	134	135	136	137	139	140	140
150	38	60	75	86	94	100	105	110	115	120	124	127	131	134	136	138	145	149	151	152	154	156	158
160	39	62	78	89	98	105	110	115	120	124	127	131	134	136	138	145	149	151	152	154	156	158	159

nombre de sujets nécessaire sera moins élevé

La taille d'un échantillon peut être calculée avec la formule suivante :

$$n = \frac{z^2 \cdot P_1 \cdot (1 - P_1)}{d^2}$$

- avec :
 - n = taille de l'échantillon
 - z = probabilité α pour que l'écart-réduit égale ou dépasse, en valeur absolue, une valeur donnée (exemple : si $\alpha = 0,05 \rightarrow z = 1,96$)
 - P_1 = prévalence attendue
 - d = précision absolue désirée sur la prévalence attendue
- si le rapport entre la taille de l'échantillon et la taille de la population (taux de sondage) dépasse 5 %, il est recommandé de calculer une taille ajustée comme suit :

$$n_{adj} = \frac{N \times n}{N + n}$$

avec N = taille de la population



17/04/2006

Claude Saegerman

25

Détermination la taille des échantillons: démarche qualitative (1)

- Objectifs :
 - Détection de l'infection ou de la maladie dans la population étudiée en n'examinant qu'un échantillon
 - Définir la taille minimale de l'échantillon permettant d'obtenir au moins une réponse positive au test de dépistage, avec un risque d'erreur donné (p.e. 5 %)
- Soit
 - détection de la maladie
 - comparaison d'une prévalence à une prévalence limite



17/04/2006

Claude Saegerman

26

Objectif : détection de la maladie ou du trouble de la santé (1)

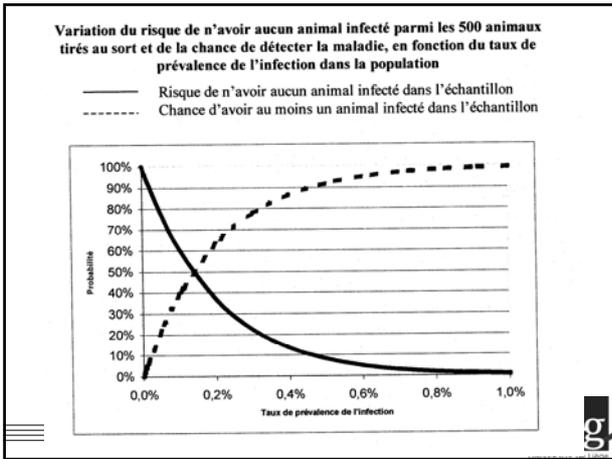
- Si on identifie, dans l'échantillon, un individu malade, on considère que la maladie y est présente.

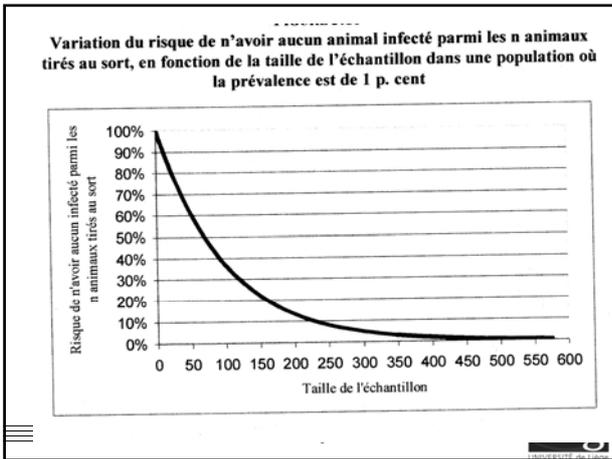


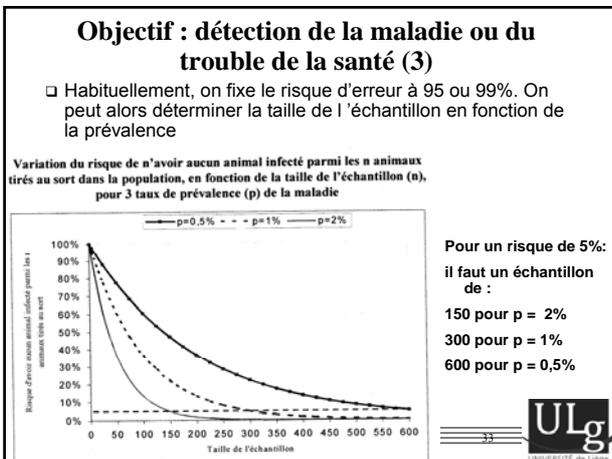
17/04/2006

Claude Saegerman

27







Objectif : détection de la maladie ou du trouble de la santé (4)

- On doit donc choisir le taux de prévalence que l'on veut détecter (taux de prévalence limite : TPL) tous les taux de prévalence qui lui sont supérieurs seront détectés avec une $p > 95\%$ et ceux qui lui sont inférieurs avec $p < 95\%$

17/04/2006

Claude Saegerman

34



Taille des échantillons permettant la détection d'une maladie dans une population infinie (taux de sondage < 10 p. cent) en fonction du taux de prévalence limite et pour un risque d'erreur de 5 p. cent et de 1 p. cent

Taux de prévalence limite	Risque	
	1 p. cent	5 p. cent
0.001 p. cent	460 515	299 572
0.01 p. cent	46 050	29 956
0.05 p. cent	9 209	5 990
0.1 p. cent	4 603	2 995
0.2 p. cent	2 301	1 497
0.4 p. cent	1 149	748
0.5 p. cent	919	598
1 p. cent	459	299
2 p. cent	228	149
3 p. cent	152	99
4 p. cent	113	74
5 p. cent	90	59
6 p. cent	75	49
7 p. cent	64	42
8 p. cent	56	36
9 p. cent	49	32
10 p. cent	44	29
20 p. cent	21	14
30 p. cent	13	9
40 p. cent	10	6
50 p. cent	7	5

17/04/20



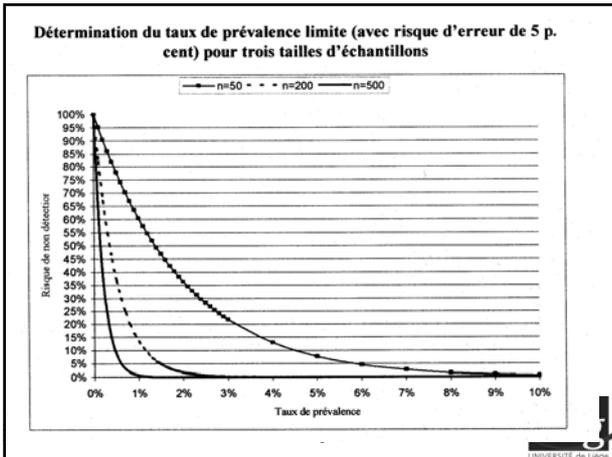
Taille des échantillons nécessaires pour la détection d'une maladie dans une population finie (taux de sondage >10%), en fonction de la taille de la population et du TPL, pour un risque d'erreur de 5%

Nombre d'unités dans la population	Taux de prévalence limite (p. cent)											
	1	2	5	10	15	20	30	40	50			
10				10	10	8	6	5	4			
20				19	16	13	10	7	6	5		
30				29	19	16	11	8	6	5		
40				31	21	15	12	8	6	5		
50				48	39	22	17	12	8	6	5	
60				57	38	23	16	13	8	6	5	
70				67	44	24	17	13	8	6	5	
80				76	42	24	17	13	9	6	5	
90				86	47	25	18	13	9	6	5	
100				95	78	45	25	17	13	9	6	5
120				114	93	47	26	18	13	9	6	5
140				133	109	48	26	18	13	9	6	5
160				152	101	49	27	18	13	9	6	5
180				171	114	50	27	18	13	9	6	5
200				155	105	51	27	18	14	9	6	5
250				194	112	55	27	19	14	9	6	5
300				189	117	54	28	18	14	9	6	5
350				221	121	56	28	19	14	9	6	5
400				211	124	55	28	19	14	9	6	5
450				237	127	56	28	19	14	9	6	5
500				225	129	56	28	19	14	9	6	5
600				235	132	56	28	19	14	9	6	5
700				243	134	57	28	19	14	9	6	5
800				249	136	57	28	19	14	9	6	5
900				254	137	57	29	19	14	9	6	5
1000				258	138	57	29	19	14	9	6	5

17/04/20

36





Objectif : détection de la maladie ou du trouble de la santé (5)

Si $Se < 1$, il convient de prendre la prévalence apparente (et non réelle) pour calculer la taille de l'échantillon

17/04/2006 Claude Saegerman 38 

Quelques éléments supplémentaires

- Ne pas prendre les nombres des tables de manière trop strictement mathématique
- le nombre d'animaux est également dépendant du budget alloué
- les tables sont données pour des tests parfaits : $se = 100\%$, $sp = 100\%$
- Il est souvent nécessaire de stratifier son échantillon (par exemple sur la taille des exploitations) et de le standardiser (par exemple sur les provinces)

17/04/2006 Claude Saegerman 39 

Objectif : comparaison du taux de prévalence à un taux de prévalence limite

Principe de la démarche qualitative de vérification du taux de prévalence d'une maladie dans une population par rapport à un taux de prévalence limite, par dépistage sur un échantillon

La valeur limite du nombre d'unités positives (k) est déterminée en fonction de la taille de l'échantillon (n) et la valeur du taux de prévalence limite (p_0).

