

Chapitre 5

Echantillonnage et enquêtes épidémiologiques

claude.saegerman@ulg.ac.be



Objectifs des enquêtes épidémiologiques

- Décrire
 - les caractéristiques d'un phénomène de santé dans une population
 - son évolution dans le temps
 - sa répartition et son évolution dans l'espace
- Objectifs purement descriptifs
- Importance de la représentativité de l'échantillon étudié
- Enquêtes transversales et longitudinales

claude.saegerman@ulg.ac.be



Echantillonnage

□ Exactitude

- échantillon représentatif
- obtention par tirage au sort

□ Précision

- varie selon la taille de l'échantillon
- écart-type d'une proportion :

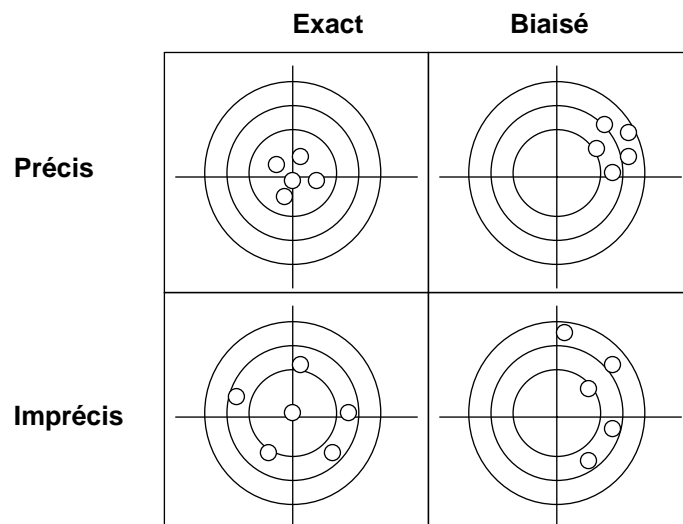
$$\sigma = \sqrt{pq/n}$$

$$\text{– IC 95\%} = p \pm 2 \sigma \quad (2 \cong 1,96)$$



claude.saegerman@ulg.ac.be

Exactitude et Précision



claude.saegerman@ulg.ac.be

Exactitude = représentativité

- Échantillon représentatif d'une population
 - tout individu de la population peut figurer dans l'échantillon, avec une probabilité non nulle, connue et égale pour tous les individus présentant les mêmes caractéristiques d'intérêt
- Représentativité globale ou pour une caractéristique
 - tirage au sort : assure la représentativité globale, càd par rapport à l'ensemble des caractéristiques de la population

claude.saegerman@ulg.ac.be



Exactitude : méthodes d'échantillonnage

- Tirage au sort : nombres aléatoires
 - table de nombres aléatoires
 - fonction de type « randomize »
 - » exemple : en Excel, ALEA()
- Sondage
 - aléatoire simple
 - échantillonnage systématique
 - sondage aléatoire stratifié
 - sondage aléatoire en grappes
 - sondage aléatoire à plusieurs degrés

claude.saegerman@ulg.ac.be



On croit souvent qu'un échantillon n'est représentatif que s'il est suffisamment grand. C'est faux. Un échantillon, si petit soit-il, est représentatif dès lors qu'il résulte d'un tirage au sort

(Schwartz, 1994)

claude.saegerman@ulg.ac.be



Précision : la taille de l'échantillon Fluctuations d'échantillonnages

- Calcul de prévalences lors de répétitions de tirage au sort d'un échantillon dans une population : dispersion des valeurs autour d'une moyenne
- Dispersion variable selon la taille de l'échantillon
- Calcul d'un intervalle de confiance (IC)

claude.saegerman@ulg.ac.be



Fraction de sondage

- Fraction de sondage (n/N) = proportion de la population introduite dans l'échantillon
- Tirage au sort « sans remise »
- Correction nécessaire si taux de sondage $n/N > 10\%$ (5 % pour Winépiscope)

claude.saegerman@ulg.ac.be



Écart-type d'une proportion

$$\sigma = \sqrt{pq/n} \quad (n/N < 10\%)$$

$$\sigma = \sqrt{(1-n/N)pq/n} \quad (n/N > 10\%)$$

- » IC 95% = $p \pm 2\sigma$ ($2 \cong 1,96$)
- » conditions d'application : $np > 5$ et $nq > 5$
- » largeur de l'IC relative à la précision de l'estimation
- » la taille de l'IC diminue lorsque la taille de l'échantillon augmente

claude.saegerman@ulg.ac.be



Précisions absolue et relative

- Imprécision de la mesure = marge d'erreur liée aux fluctuations d'échantillonnage dans une population
- Précision absolue (P_a)
 - mesure de la dispersion des valeurs de l'estimation autour de la moyenne
 - = moitié de l'intervalle de confiance
- Précision relative (P_r)
 - rapport de la précision absolue sur la valeur estimée
 - = P_a/p (p = proportion estimée)

claude.saegerman@ulg.ac.be



Evolution de la précision

- Précision importante : chiffres faibles de précision absolue ou relative
- La taille de l'IC diminue quand la taille de l'échantillon augmente
- La taille de l'IC est :
 - plus grande pour une probabilité de 99 %
 - que pour une probabilité de 95 %
- La taille de l'IC varie avec la proportion trouvée (prévalence)
 - $p = 10\%$: IC = 13 (3,5-16,5) ; $P_a = 6,5$; $P_r = 65\%$
 - $p = 50\%$: IC = 20 (40-60) ; $P_a = 10$; $P_r = 20\%$

claude.saegerman@ulg.ac.be



DÉMARCHE QUANTITATIVE

estimation de la proportion d'animaux atteints

- Déterminer le nombre d'animaux nécessaires (n)
 - fonction du nombre d'animaux dans le troupeau (N)
(*population size*)
 - fonction de la proportion d'animaux atteints (p)
(*expected prevalence*)
 - fonction du degré de précision absolue (Pa) désiré
(*accepted error, absolute error*)
 - fraction de sondage
 - » sample size (n) : si $n/N < 10\%$ ou 5% (population infinie)
 - » adjusted sample size n(a) : si $n/N > 10\%$ ou 5%

winepiscopes : samples : estimate percentage

claude.saegerman@ulg.ac.be



Utilisation des tables

- La taille de l'échantillon augmente avec une augmentation de la précision absolue (Pa)
- Pour une même Pa, la taille de l'échantillon augmente lorsque la proportion des animaux atteints (p) augmente (jusqu'à 50 %)
- Pour une Pa limitée (50 %) et une p élevée (50 %), la taille de l'échantillon augmente plus lentement que la taille du troupeau
- La précision de la mesure peut être supérieure à celle attendue, si p mesurée > p estimée pour fixer la taille de l'échantillon

claude.saegerman@ulg.ac.be



Calcul de la taille de l'échantillon

$$\square Pa = 1,96 \sqrt{pq/n} \quad (95 \%)$$

$$\square n = \frac{3,84 pq}{Pa^2} \quad \text{avec } Pa = Pr \times p$$

$$\square n = \frac{3,84 q}{p Pr^2} \quad \text{si } n/N < 10\% \quad (5\%)$$

$$\square n = \frac{1}{(1/n + 1/N)} \quad \text{si } n/N > 10\% \quad (5\%)$$

claude.saegerman@ulg.ac.be



Pour $n/N < 10 \%$ (5 %) : un paradoxe ?

- \square Pour un degré donné de précision relative
 - plus la proportion (prévalence) est faible ($< 50\%$)
 - ou plus elle est élevée ($> 50\%$)
 - plus la taille de l'échantillon doit être élevée
- \square Pour un degré donné de précision absolue
 - plus la proportion (prévalence) est proche de 50%
 - plus la taille de l'échantillon doit être élevée

claude.saegerman@ulg.ac.be



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	1 825	865	545	385	289	225	179	145	118	97
30 p. cent	4 226	2 092	1 381	1 025	812	385	242	171	129	100	80	65	53	43
40 p. cent	2 377	1 177	777	577	457	217	137	97	73	57	45	37	30	25
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

DÉMARCHE QUALITATIVE

- Détection de l'infection dans la population étudiée, en examinant un échantillon
 - qualification d'un effectif comme indemne d'une maladie
- Détection d'une maladie ou détection de troupeaux à une prévalence supérieure à un taux de prévalence limite
- Notions de
 - » taux de prévalence limite
 - » taux de prévalence acceptable

Détection d'une maladie

- On pose $Se = 1$
- Erreur par défaut :
 - manque de sensibilité du dépistage
 - conclusion : indemnité, alors que des cas existent
- Probabilités :
 - p = prévalence de l'infection
 - $1 - p$ = probabilité qu'un individu est indemne
 - $(1 - p)^n$ = probabilité que les n individus de l'échantillon soient indemnes
- Liens entre :
 - risque d'erreur par défaut
 - prévalence dans la population
 - taille de l'échantillon

claude.saegerman@ulg.ac.be



Détection d'une maladie (2)

- Poser la taille de la population
 - Poser le risque d'erreur (95 %)
 - Poser le nombre d'animaux malades attendus
 - quel seuil n ? animaux positifs = troupeau positif
 - (pour les maladies graves, utiliser $n = 1$)
 - Résultats :
 - p attendue faible : taille de l'échantillon élevée
 - risque d'erreur faible : taille élevée
- winepiscope : detection of disease, sample size

claude.saegerman@ulg.ac.be



Déclaration comme indemne de maladie

- On considère comme « indemne » tout troupeau dont le pourcentage d'animaux à réponse positive est inférieur à un niveau fixé à l'avance.
- Si tous les animaux de l'échantillon sont négatifs, est-ce que l'effectif est négatif ?
- Dépend de :
 - effectif de la population
 - taille de l'échantillon où tous les individus sont négatifs
 - risque d'erreur

winepiscope : detection of disease – maximum number of positives



claude.saegerman@ulg.ac.be

Et si tous les résultats sont négatifs : l'effectif est-il indemne ?

- NON
- Il existe un IC à droite de 0 %
- Prévalence maximale probable

winepiscope : sample, detection of disease, maximum n° of positives



claude.saegerman@ulg.ac.be

Taux de prévalence limite (TPL)

- TPL = seuil de prévalence pour la détection de la maladie
- La maladie est détectée dans la population pour des taux de prévalence $>$ TPL avec une $P > 95\%$ (ou 99%)
- Taille des échantillons
 - plus élevée si le risque d'erreur = 1% par rapport à 5%
 - diminue très vite lorsque le TPL augmente

claude.saegerman@ulg.ac.be



Taux de prévalence acceptable

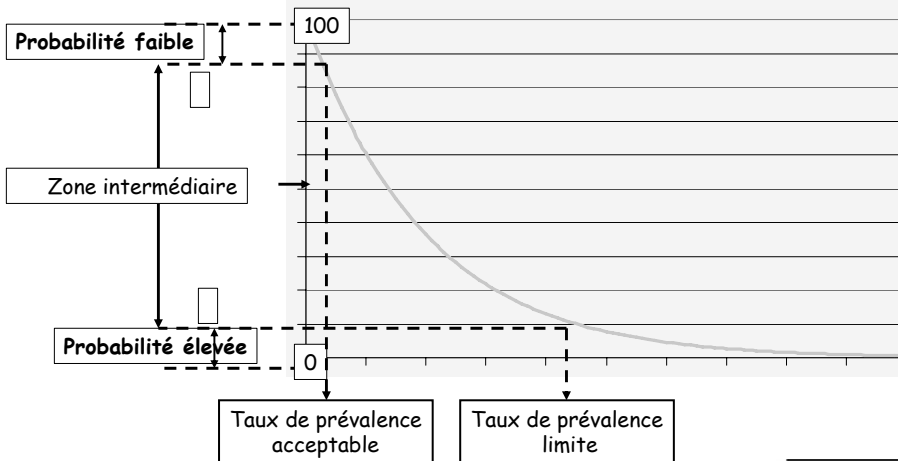
- TPA = seuil de prévalence au-dessous duquel la taille de l'échantillon ne permet plus de détecter la maladie (P : moins de 5%)
- Pourquoi « acceptable » ?
 - Avec la taille de l'échantillon retenue
 - on accepte de ne pas détecter des $p <$ TPA
 - d'accord : c'est faible

claude.saegerman@ulg.ac.be



TPL et TPA ?

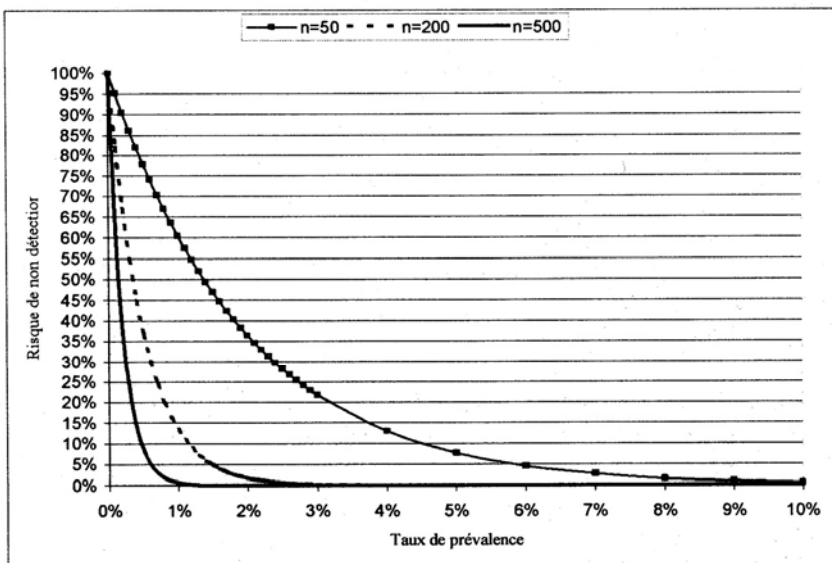
Risque de non détection



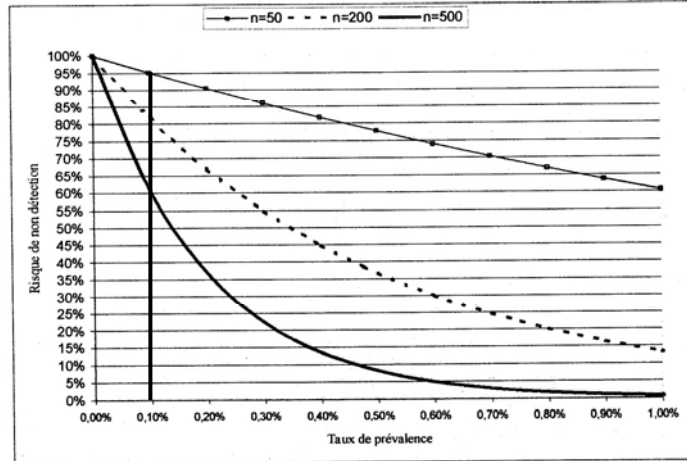
claude.saegerman@ulg.ac.be



Détermination du taux de prévalence limite (avec risque d'erreur de 5 p. cent) pour trois tailles d'échantillons



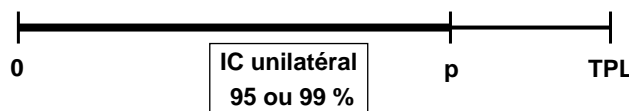
Agrandissement de la figure 3.17 pour les faibles valeurs de prévalence afin de montrer la détermination des taux de prévalence acceptable (95 p. cent de risque de non détection) pour 3 tailles d'échantillons



On constate qu'avec un échantillon de 50, on a 95% de risque de ne pas détecter la présence de la maladie lorsque sa prévalence est < 0,1%. Ce seuil peut-être appelé taux de prévalence acceptable (TPA)

Prévalence inférieure à un taux de prévalence limite

- Principe : taille d'échantillon nécessaire pour déterminer si $p < TPL$



- Winepiscopie :
 - prévalence attendue (TPL)
 - prévalence seuil
 - niveau de confiance
 - puissance du test ($1-\beta$)

winepiscopie : sample, threshold value



Table fournissant l'intervalle de confiance unilatéral à 95 p. cent en fonction de la taille de l'échantillon et du nombre d'unités à réponse positive au sein de l'échantillon

Le pourcentage indiqué dans chaque case correspond à la borne supérieure de l'intervalle de confiance unilatéral (la borne inférieure est égale à zéro).

Taille de l'échantillon	Nombre d'unités à réponse positive (p. cent)										
	0	1	2	3	4	5	6	7	8	9	10
50	5,82	9,14	12,06	14,78	17,38	19,88	22,32	24,69	27,02	29,31	31,56
100	2,95	4,66	6,16	7,57	8,92	10,23	11,50	12,75	13,97	15,18	16,37
150	1,98	3,12	4,14	5,09	6,00	6,88	7,74	8,59	9,42	10,24	11,05
200	1,49	2,35	3,11	3,83	4,52	5,18	5,83	6,47	7,10	7,72	8,33
250	1,19	1,88	2,50	3,07	3,62	4,16	4,68	5,19	5,70	6,20	6,69
300	0,99	1,57	2,08	2,56	3,03	3,47	3,91	4,34	4,76	5,18	5,59
400	0,75	1,18	1,57	1,93	2,27	2,61	2,94	3,26	3,58	3,89	4,20
500	0,60	0,95	1,25	1,54	1,82	2,09	2,35	2,61	2,87	3,12	3,37
1000	0,30	0,47	0,63	0,77	0,91	1,05	1,18	1,31	1,44	1,57	1,69
1500	0,20	0,32	0,42	0,52	0,61	0,70	0,79	0,87	0,96	1,04	1,13
2000	0,15	0,24	0,31	0,39	0,46	0,52	0,59	0,66	0,72	0,78	0,85
2500	0,12	0,19	0,25	0,31	0,37	0,42	0,47	0,53	0,58	0,63	0,68
5000	0,06	0,09	0,13	0,15	0,18	0,21	0,24	0,26	0,29	0,31	0,34

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 %)
 - corriger la p en p apparente en utilisant les valeurs de se et de sp et prendre la p apparente pour le choix de la taille d'échantillon