Section 2. Méthodes d'échantillonnage

Pierre Duchesne

September 6, 2019

Population, base de sondage, échantillon

- On dispose d'une base de sondage de taille finie N;
- Notation: $U = \{1, 2, \ldots, k, \ldots, N\}$;
- Habituellement on va tenter d'utiliser l'indice k pour l'identificateur de l'unité k;
- La variable d'intérêt sera notée y.
  1. Exemple, variable comme le salaire; Salaire $y_k = 50500\$$ indique le salaire de l'unité k;
  2. Autre exemple, variable dichotomique: $y_k = 1$ si l'unité k est chômeur, et $y_k = 0$ sinon.

Paramètres de la population

- Ces quantités reposent sur la connaissance de l'ensemble des unités dans la population.
- Total: $t_y = \sum_{k=1}^{N} y_k$;
- Moyenne: $\bar{y}_U = \frac{1}{N} \sum_{k=1}^{N} y_k$;
- Variance: $S_{yU}^2 = \frac{1}{N-1} \sum_{k=1}^{N} (y_k - \bar{y}_U)^2$;
- D'autres exemples sont possibles; on pourrait penser à un paramètre comme la médiane, ou encore un quantile de population.

Convention pour le symbole $\sum$

- Soit A in sous-ensemble d'unités, $A \subseteq U$;
- Pour le total de la variable y sur A:

$$\sum\nolimits_A y_k = \sum_{k \in A} y_k;$$

- Avec cette notation, on écrit:

$$t_y = \sum\nolimits_U y_k;$$

et

$$\bar{y}_U = \frac{1}{N} \sum\nolimits_U y_k;$$

Échantillon

- Un échantillon s est un sous-ensemble de U, on aura donc $s \subseteq U$.
- On note que U est de cardinalité finie, N.
- Il existe un nombre fini d'échantillons, mais le nombre possible peut en fait être très grand.
- On peut dénombrer le nombre total d'échantillons:

$$\binom{N}{0} + \binom{N}{1} + \binom{N}{2} + \cdots + \binom{N}{N} = 2^N.$$

Plan d'échantillonnage

- Dans un sondage, on sélectionne un échantillon s et on observe la variable d'intérêt y pour les unités $k \in s$.
- On observe alors:

$$\{y_k; k \in s\}$$

- Échantillon probabiliste: c'est un échantillon obtenu selon un mécanisme aléatoire que l'on appelle **plan d'échantillonnage**.
- Plan d'échantillonnage: c'est un ensemble de règles strictes, qui une fois mise en oeuvre, nous donne le sous-ensemble de U, c'est-à-dire l'échantillon s.
- On requiert que la probabilité d'être inclus dans l'échantillon s est strictement positive, et ce $\forall k \in U$.

Échantillon représentatif

- Un échantillon a pour but de représenter la population, donc d'être représentatif.
- En quelque sorte, l'échantillon est un modèle de la population.
- Il n'est pas possible de déterminer si un échantillon est *représentatif* dans le sens qu'il est un *modèle réduit* de la population.
- Un bon plan d'échantillonnage nous assure que l'échantillon est obtenu selon une méthodologie que le méthodologiste peut tenir compte.
- Dans un sens, un plan d'échantillon performant contribue à éliminer des échantillons non représentatifs, donc la sens que les échantillons reflètent des caractérisques utiles de la population.

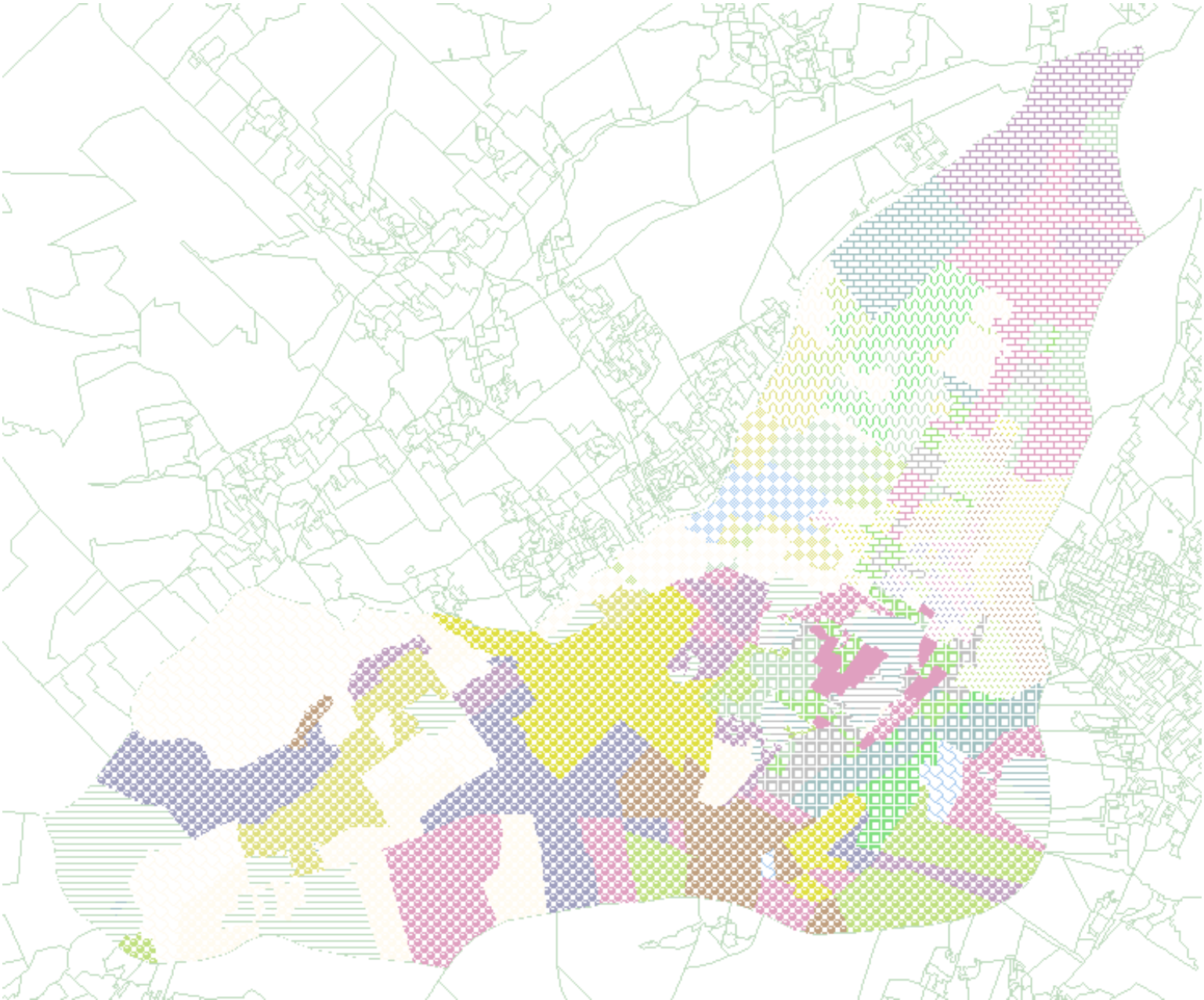
Échantillonnage probabiliste

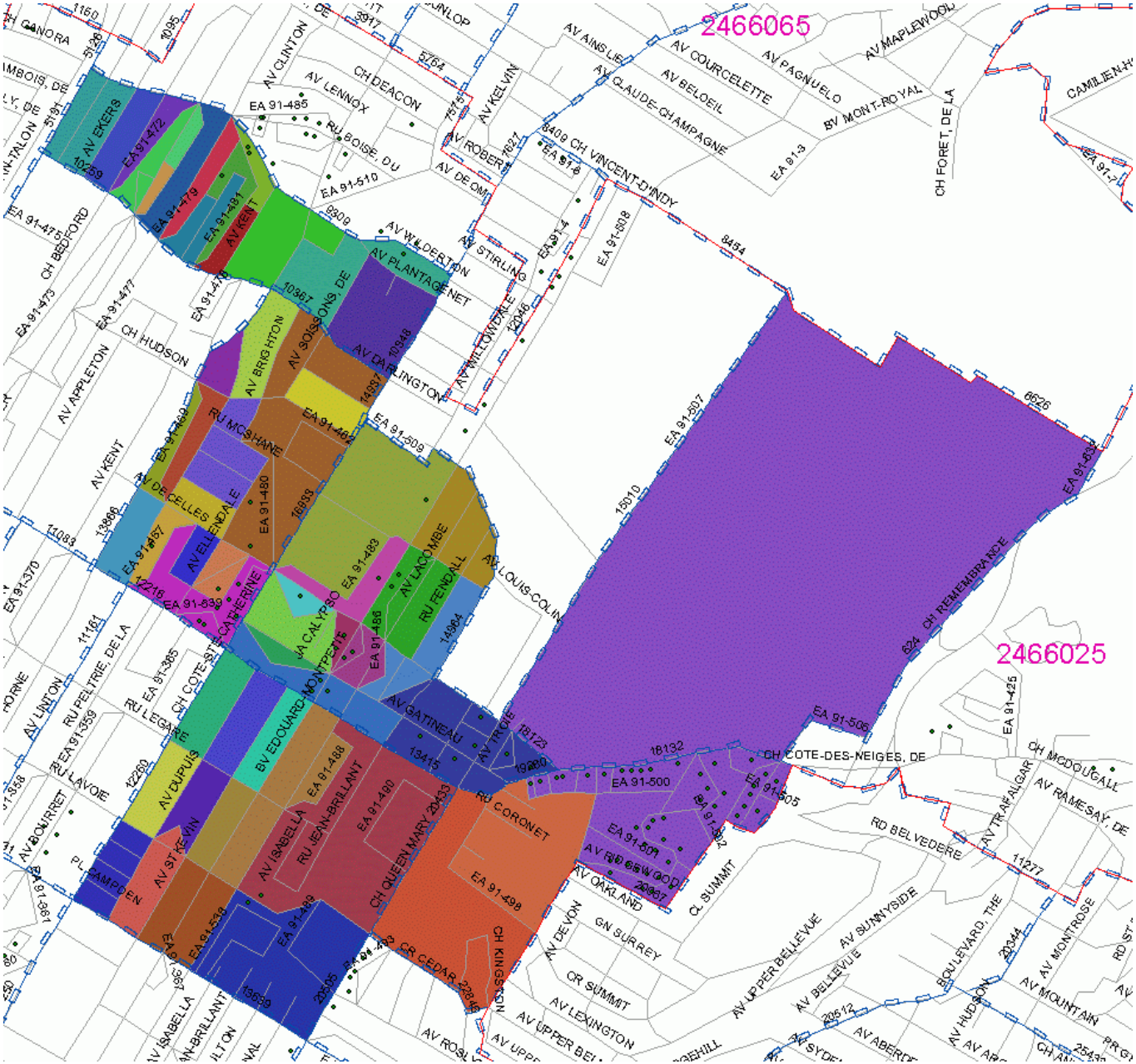
Exemples de plan de sondage:

- Tirage aléatoire simple avec remise;
- Tirage aléatoire simple sans remise (chaque échantillon de taille n, n fournie, possède les mêmes chances de survenir);
- Tirage de Bernoulli; Tirage de Poisson;
- Tirage stratifié; tirage stratifié simple;
- Tirage systématique;
- Tirage en grappes;
- Tirage à plusieurs degrés.

Exemples de sondages probabilistes

- Dans une étude de marketing, on désire sonder les ménages de la ville de Montréal.
- On suppose que l'on dispose d'une liste de 1 à M des unités géographiques sur une carte; on parle sondages aréolaires.
- On sélectionne un échantillon au hasard d'unités géographiques.
- Dans chaque unité géographique, on sélectionne et observe tous ou une partie des ménages.




2466065
2466025
AV CLINTON
CH DEACON
AV LENNOX
RU BOISE, DU
AV KELVIN
AV ROBERVAL
AV AINSLIE
AV COURCELETTE
AV BELOEIL
AV CLAUDE-CHAMPAGNE
AV PAGNUELO
AV MAPLEWOOD
BV MONT-ROYAL
CH FORET, DE LA
CH VINCENT-D'INDY
AV EKERS
AV KENT
AV WILDERTON
AV PLANTAGENET
AV STIRLING
AV WILLOWDALE
AV SOISSONS, DE
AV BRIGHTON
AV DARLINGTON
CH HUDSON
AV APPLETON
AV DECELLES
AV ELLENDALE
RU MCSHANE
AV LACOMBE
RU FENDALL
AV LOUIS-COLIN
CH CÔTE-SAINTE-CATHERINE
AV GATINEAU
AV TROIE
RU CORONET
CH QUEEN MARY
CH KINGSTON
CR CEDAR
AV ISABELLA
RU JEAN-BRILLANT
AV ST-KEVIN
PL CHAMPDEN
BV EDOUARD-MONTPETIT
AV DUPUIS
RU LAVOIE
RU PELTRIE, DE LA
AV LINTON
AV BOURRET
CH CÔTE-DES-NEIGES, DE
RD BELVEDERE
CL SUMMIT
AV OAKLAND
GN SURREY
CR SUMMIT
AV LEXINGTON
AV DEVON
AV UPPER BELLEVUE
AV SUNNYSIDE
AV BELLEVUE
BOULEVARD, THE
AV MONTROSE
AV MOUNTAIN
AV HUDSON
AV TRAFALGAR
AV RAMESAY, DE
CH MCDOUGALL
CH REMEMBRANCE
CH BEDFORD
EA 91-472
EA 91-473
EA 91-474
EA 91-475
EA 91-477
EA 91-478
EA 91-479
EA 91-480
EA 91-481
EA 91-483
EA 91-484
EA 91-485
EA 91-486
EA 91-487
EA 91-488
EA 91-489
EA 91-490
EA 91-498
EA 91-499
EA 91-500
EA 91-505
EA 91-506
EA 91-507
EA 91-508
EA 91-509
EA 91-510
EA 91-538
EA 91-639
EA 91-3
EA 91-365
EA 91-370
EA 91-425
EA 91-426
EA 91-427
EA 91-428
EA 91-429
EA 91-430
EA 91-431
EA 91-432
EA 91-433
EA 91-434
EA 91-435
EA 91-436
EA 91-437
EA 91-438
EA 91-439
EA 91-440
EA 91-441
EA 91-442
EA 91-443
EA 91-444
EA 91-445
EA 91-446
EA 91-447
EA 91-448
EA 91-449
EA 91-450
EA 91-451
EA 91-452
EA 91-453
EA 91-454
EA 91-455
EA 91-456
EA 91-457
EA 91-458
EA 91-459
EA 91-460
EA 91-461
EA 91-462
EA 91-463
EA 91-464
EA 91-465
EA 91-466
EA 91-467
EA 91-468
EA 91-469
EA 91-470
EA 91-471
EA 91-472
EA 91-473
EA 91-474
EA 91-475
EA 91-476
EA 91-477
EA 91-478
EA 91-479
EA 91-480
EA 91-481
EA 91-482
EA 91-483
EA 91-484
EA 91-485
EA 91-486
EA 91-487
EA 91-488
EA 91-489
EA 91-490
EA 91-491
EA 91-492
EA 91-493
EA 91-494
EA 91-495
EA 91-496
EA 91-497
EA 91-498
EA 91-499
EA 91-500
EA 91-501
EA 91-502
EA 91-503
EA 91-504
EA 91-505
EA 91-506
EA 91-507
EA 91-508
EA 91-509
EA 91-510
EA 91-511
EA 91-512
EA 91-513
EA 91-514
EA 91-515
EA 91-516
EA 91-517
EA 91-518
EA 91-519
EA 91-520
EA 91-521
EA 91-522
EA 91-523
EA 91-524
EA 91-525
EA 91-526
EA 91-527
EA 91-528
EA 91-529
EA 91-530
EA 91-531
EA 91-532
EA 91-533
EA 91-534
EA 91-535
EA 91-536
EA 91-537
EA 91-538
EA 91-539
EA 91-540
EA 91-541
EA 91-542
EA 91-543
EA 91-544
EA 91-545
EA 91-546
EA 91-547
EA 91-548
EA 91-549
EA 91-550
EA 91-551
EA 91-552
EA 91-553
EA 91-554
EA 91-555
EA 91-556
EA 91-557
EA 91-558
EA 91-559
EA 91-560
EA 91-561
EA 91-562
EA 91-563
EA 91-564
EA 91-565
EA 91-566
EA 91-567
EA 91-568
EA 91-569
EA 91-570
EA 91-571
EA 91-572
EA 91-573
EA 91-574
EA 91-575
EA 91-576
EA 91-577
EA 91-578
EA 91-579
EA 91-580
EA 91-581
EA 91-582
EA 91-583
EA 91-584
EA 91-585
EA 91-586
EA 91-587
EA 91-588
EA 91-589
EA 91-590
EA 91-591
EA 91-592
EA 91-593
EA 91-594
EA 91-595
EA 91-596
EA 91-597
EA 91-598
EA 91-599
EA 91-600
EA 91-601
EA 91-602
EA 91-603
EA 91-604
EA 91-605
EA 91-606
EA 91-607
EA 91-608
EA 91-609
EA 91-610
EA 91-611
EA 91-612
EA 91-613
EA 91-614
EA 91-615
EA 91-616
EA 91-617
EA 91-618
EA 91-619
EA 91-620
EA 91-621
EA 91-622
EA 91-623
EA 91-624
EA 91-625
EA 91-626
EA 91-627
EA 91-628
EA 91-629
EA 91-630
EA 91-631
EA 91-632
EA 91-633
EA 91-634
EA 91-635
EA 91-636
EA 91-637
EA 91-638
EA 91-639
EA 91-640
EA 91-641
EA 91-642
EA 91-643
EA 91-644
EA 91-645
EA 91-646
EA 91-647
EA 91-648
EA 91-649
EA 91-650
EA 91-651
EA 91-652
EA 91-653
EA 91-654
EA 91-655
EA 91-656
EA 91-657
EA 91-658
EA 91-659
EA 91-660
EA 91-661
EA 91-662
EA 91-663
EA 91-664
EA 91-665
EA 91-666
EA 91-667
EA 91-668
EA 91-669
EA 91-670
EA 91-671
EA 91-672
EA 91-673
EA 91-674
EA 91-675
EA 91-676
EA 91-677
EA 91-678
EA 91-679
EA 91-680
EA 91-681
EA 91-682
EA 91-683
EA 91-684
EA 91-685
EA 91-686
EA 91-687
EA 91-688
EA 91-689
EA 91-690
EA 91-691
EA 91-692
EA 91-693
EA 91-694
EA 91-695
EA 91-696
EA 91-697
EA 91-698
EA 91-699
EA 91-700
EA 91-701
EA 91-702
EA 91-703
EA 91-704
EA 91-705
EA 91-706
EA 91-707
EA 91-708
EA 91-709
EA 91-710
EA 91-711
EA 91-712
EA 91-713
EA 91-714
EA 91-715
EA 91-716
EA 91-717
EA 91-718
EA 91-719
EA 91-720
EA 91-721
EA 91-722
EA 91-723
EA 91-724
EA 91-725
EA 91-726
EA 91-727
EA 91-728
EA 91-729
EA 91-730
EA 91-731
EA 91-732
EA 91-733
EA 91-734
EA 91-735
EA 91-736
EA 91-737
EA 91-738
EA 91-739
EA 91-740
EA 91-741
EA 91-742
EA 91-743
EA 91-744
EA 91-745
EA 91-746
EA 91-747
EA 91-748
EA 91-749
EA 91-750
EA 91-751
EA 91-752
EA 91-753
EA 91-754
EA 91-755
EA 91-756
EA 91-757
EA 91-758
EA 91-759
EA 91-760
EA 91-761
EA 91-762
EA 91-763
EA 91-764
EA 91-765
EA 91-766
EA 91-767
EA 91-768
EA 91-769
EA 91-770
EA 91-771
EA 91-772
EA 91-773
EA 91-774
EA 91-775
EA 91-776
EA 91-777
EA 91-778
EA 91-779
EA 91-780
EA 91-781
EA 91-782
EA 91-783
EA 91-784
EA 91-785
EA 91-786
EA 91-787
EA 91-788
EA 91-789
EA 91-790
EA 91-791
EA 91-792
EA 91-793
EA 91-794
EA 91-795
EA 91-796
EA 91-797
EA 91-798
EA 91-799
EA 91-800
EA 91-801
EA 91-802
EA 91-803
EA 91-804
EA 91-805
EA 91-806
EA 91-807
EA 91-808
EA 91-809
EA 91-810
EA 91-811
EA 91-812
EA 91-813
EA 91-814
EA 91-815
EA 91-816
EA 91-817
EA 91-818
EA 91-819
EA 91-820
EA 91-821
EA 91-822
EA 91-823
EA 91-824
EA 91-825
EA 91-826
EA 91-827
EA 91-828
EA 91-829
EA 91-830
EA 91-831
EA 91-832
EA 91-833
EA 91-834
EA 91-835
EA 91-836
EA 91-837
EA 91-838
EA 91-839
EA 91-840
EA 91-841
EA 91-842
EA 91-843
EA 91-844
EA 91-845
EA 91-846
EA 91-847
EA 91-848
EA 91-849
EA 91-850
EA 91-851
EA 91-852
EA 91-853
EA 91-854
EA 91-855
EA 91-856
EA 91-857
EA 91-858
EA 91-859
EA 91-860
EA 91-861
EA 91-862
EA 91-863
EA 91-864
EA 91-865
EA 91-866
EA 91-867
EA 91-868
EA 91-869
EA 91-870
EA 91-871
EA 91-872
EA 91-873
EA 91-874
EA 91-875
EA 91-876
EA 91-877
EA 91-878
EA 91-879
EA 91-880
EA 91-881
EA 91-882
EA 91-883
EA 91-884
EA 91-885
EA 91-886
EA 91-887
EA 91-888
EA 91-889
EA 91-890
EA 91-891
EA 91-892
EA 91-893
EA 91-894
EA 91-895
EA 91-896
EA 91-897
EA 91-898
EA 91-899
EA 91-900
EA 91-901
EA 91-902
EA 91-903
EA 91-904
EA 91-905
EA 91-906
EA 91-907
EA 91-908
EA 91-909
EA 91-910
EA 91-911
EA 91-912
EA 91-913
EA 91-914
EA 91-915
EA 91-916
EA 91-917
EA 91-918
EA 91-919
EA 91-920
EA 91-921
EA 91-922
EA 91-923
EA 91-924
EA 91-925
EA 91-926
EA 91-927
EA 91-928
EA 91-929
EA 91-930
EA 91-931
EA 91-932
EA 91-933
EA 91-934
EA 91-935
EA 91-936
EA 91-937
EA 91-938
EA 91-939
EA 91-940
EA 91-941
EA 91-942
EA 91-943
EA 91-944
EA 91-945
EA 91-946
EA 91-947
EA 91-948
EA 91-949
EA 91-950
EA 91-951
EA 91-952
EA 91-953
EA 91-954
EA 91-955
EA 91-956
EA 91-957
EA 91-958
EA 91-959
EA 91-960
EA 91-961
EA 91-962
EA 91-963
EA 91-964
EA 91-965
EA 91-966
EA 91-967
EA 91-968
EA 91-969
EA 91-970
EA 91-971
EA 91-972
EA 91-973
EA 91-974
EA 91-975
EA 91-976
EA 91-977
EA 91-978
EA 91-979
EA 91-980
EA 91-981
EA 91-982
EA 91-983
EA 91-984
EA 91-985
EA 91-986
EA 91-987
EA 91-988
EA 91-989
EA 91-990
EA 91-991
EA 91-992
EA 91-993
EA 91-994
EA 91-995
EA 91-996
EA 91-997
EA 91-998
EA 91-999

Exemple (suite)

- Sondages de grappes et sondages à deux degrés.
- Sondages de grappes: On observe tous les ménages dans les unités géographiques sélectionnées.
- Sondage à deux degrés: On observe un sous-ensemble de ménages dans les unités géographiques sélectionnées. Il y aura échantillonnage (premier degré: les grappes) suivi de sous-échantillonnage (second degré: les ménages).

Base de sondage

- C'est une liste de N unités de la population;
- Dans l'échantillonnage probabiliste, chaque unité a une chance positive d'être dans l'échantillon;
- Idéalement on possède une base de sondage ou encore on en construit une;
- Sinon, on peut construire des grappes avec l'aide d'une carte géographique et observer tous les ménages dans les grappes choisies: l'échantillonnage de grappes permet de contourner (du moins en théorie) les problèmes lorsqu'il n'y a pas de base de sondage.

Exemple d'une base de sondage: population MU284 (SSW, Appendice B, p. 652)

LABEL	P85	P75	RMT85	CS82	SS82	S82	ME84	REV84	REG	CL
1	?	27	288	13	24	49	2135	2836	1	1
2	?	15	139	14	12	41	957	2035	1	1
3	?	20	196	12	14	41	1530	6030	1	1
4	?	15	159	12	19	41	1059	4704	1	1
5	?	52	536	20	27	61	3951	5183	1	1
6	?	15	134	16	12	41	918	2157	1	2
7	?	62	623	18	27	61	4367	7072	1	2
8	?	54	517	15	32	61	4345	5246	1	2
9	?	12	96	10	12	31	754	951	1	2
10	?	50	467	14	29	61	3902	6067	1	2
11	?	29	277	14	20	45	1993	3264	1	3
12	?	14	155	10	21	41	1312	1899	1	3
13	?	40	386	24	13	51	2780	5931	1	3
14	?	27	241	24	8	45	1649	3877	1	3
15	?	43	422	19	18	51	2983	4968	1	3
16	?	671	6263	34	41	101	45324	59877	1	4
17	?	78	612	14	31	61	5331	7027	1	4
18	?	54	532	23	23	61	3994	6529	1	4
19	?	28	250	9	22	41	1616	2208	1	4
20	?	55	412	20	27	61	3240	3976	1	4
...										
282	?	27	226	7	28	49	1682	2898	8	50
283	?	9	63	5	19	41	604	594	8	50
284	?	31	233	5	27	45	1788	2366	8	50

Base de sondage (suite)

- Dans l'exemple précédent, la variable P85 est inconnue et on veut faire un sondage afin d'estimer le total de la population;
- Les autres variables sont des variables auxiliaires qui sont utiles au niveau de la:
  1. Conception d'un sondage; construction d'estimateurs.
- Variables auxiliaires quantitatives: P75, RMT85, CS82, . . .,
- Variables auxiliaires qualitatives: REG, CL.

Stade d'estimation et analyse

Pour un échantillon s, on introduit

- Taille de l'échantillon: n_s;
- Moyenne échantillonnale:

$$\bar{y}_s = \frac{1}{n_s} \sum_s y_k$$

- Variance échantillonnale:

$$S_{ys}^2 = \frac{1}{n_s - 1} \sum_s (y_k - \bar{y}_s)^2$$

Utilisation de la moyenne échantillonnale

On considère $\bar{y}_s = \frac{1}{n_s} \sum_s y_k$.

- Les propriétés statistiques de cette quantité dépendent du plan d'échantillonnage;
- Pour le tirage aléatoire simple sans remise, on sait que:

$$E_{SI}(\bar{y}_s) = \bar{y}_U.$$

- Ainsi, la moyenne échantillonnale possède la propriété d'absence de biais lorsque le plan de sondage est le tirage aléatoire simple sans remise.

Biais

- Pour le tirage aléatoire simple sans remise, la moyenne de $\bar{y}_s$ sur tous les s égale la moyenne de la population $\bar{y}_U$;
- Le biais d'un estimateur est formalement défini ainsi:

$$E_p(\hat{\theta}) - \theta$$

- Biais positif: en moyenne, les estimateurs excèdent la véritable valeur du paramètre;
- Biais négatif: en moyenne, les estimateurs sont inférieurs à la véritable valeur du paramètre.