Les faits

- Font aussi référence à ce que l'on a, soit nos possessions:
  - Nombre de télés, possession d'ordinateur, d'automobile, d'appartement/maison, de vélos, d'épargnes, etc.
- Font aussi référence à ce que l'on a observé:
  - Des événements dont le répondant a été témoin ou victime,
  - Des gestes de harcèlement sexuel,
  - Des menaces ou des activités illégales,
  - Des comportements racistes ou sexistes.

16

La mesure des faits

La précision

- La précision pour les caractéristiques socio-démographiques
  - Bien s'assurer de ce que l'on veut savoir: Ex: scolarité: diplôme, nombre d'années d'étude complétées, etc.?
  - Doser précision et pertinence: année de naissance mais rarement date exacte à moins que ce soit absolument nécessaire.
- La précision dans la mesure des événements
  - Il faut que la question mentionne de qui on parle, de quel comportement/ événement exactement, l'intervalle de temps et le lieu **(Qui, quoi, quand, où).**
  - Pour les **cheminements**, la datation doit être
    - suffisamment précise pour pouvoir repérer les événements concurrents
    - mais... pas trop précise pour éviter la non-réponse ou des réponses "bidons" qui induiront en erreur

17

La mesure des faits

Neutralité

- Neutralité: ne jamais laisser voir qu'une réponse serait plus appropriée qu'une autre.
  - Exemples:
    - Il faut demander "quelle est votre langue maternelle?" et non "êtes-vous de langue maternelle française?"
    - Demander "Quel est le plus haut niveau de diplôme que vous avez obtenu?" plutôt que "Avez-vous un diplôme universitaire?".
  - Dans les préambules, éviter les phrases qui laissent entendre que... "tout le monde a ça..., a vu ça, etc"

18

La mesure des faits

Pertinence

- Pertinence:
  - Ne pas demander une information à quelqu'un qui ne peut pas la détenir (p. ex: demander à un étudiant le nombre d'étudiants exacts dans son programme).
  - Souvent respectée en utilisant **des filtres**:
    - Si une personne dit ne pas avoir de télé (question de fait), on ne lui demande pas quelle émission elle écoute à la télé (mesure de comportement).
    - Ne pas demander à un homme combien de jupes ou de robes il a achetées au cours des 3 derniers mois (SIC).

19

Les comportements

Définition

- Les actions effectuées (*et non pas l'intention d'effectuer des actions*) ...
  - Activités faites durant un laps de temps précis:
    - Consommation de breuvages ou de boissons alcooliques (par semaine, par jour)
    - Consommation de produits culturels (par mois, par année)
    - Nombre d'heures d'écoute télé par semaine
    - Nombre d'heures de navigation sur Internet par semaine
  - Ou comportements habituels:
    - Habitudes de consommation, de comportement:
      - Fréquentation habituelle des lieux de culte
      - Consommation habituelle de cigarettes.

20

La mesure des comportements

La précision

- La question doit comprendre les éléments essentiels:
  - **Qui** a fait l'action?
  - A fait **quoi**? Quelle action précisément a été effectuée?
  - **Quand**? Au cours de quelle pérdiode?
  - **Où**, Dans quel lieu, dans quel cadre, dans quelles circonstances?

21

La mesure des comportements

La neutralité: Les questions sensibles, menaçantes

- Il faut évaluer si une question est sensible, si elle peut entraîner des réactions de conformité sociale visant à cacher des comportements considérés “répréhensibles” ou, au contraire à déclarer des comportements que l’on n’a pas faits.
- **Les questions sensibles sont celles où il y a socialement une bonne et une mauvaise réponse.** Le niveau de menace, de sensibilité, de la question est socialement et historiquement ancré. Il dépend de la société et de l’époque.
- Tendance générale au sur-aveu pour les comportements “ordinaires”, désirables -- aller au Musée, voter, être informé, recycler -- et au sous-aveu pour les comportements moins acceptés socialement.
  - Donc, il faut faire apparaître tous les comportements comme également acceptables

22

La mesure des comportements

La pertinence

- S’assurer de poser les questions uniquement aux personnes auxquelles elles s’appliquent, ce qui demande d’utiliser des questions filtres.
- Les questions doivent apparaître pertinentes (validité d’apparence).
- ***L’intervalle de temps doit être approprié.***
- Attention, tendance générale au non-aveu délibéré si le répondant sait que le fait de répondre non à une question filtre lui évitera une liste de questions.
  - Donc, il est préférable d’assumer certains comportements courants.

23

La mesure des comportements

La pertinence : les problèmes de mémoire

- La source d’erreur la plus importante est reliée aux problèmes de *mémoire*...
- Principe: Plus un événement est marquant (Ex: mariage), plus on peut retourner loin dans le temps, plus il est banal (ex: prendre une bière) moins on peut retourner loin.
- La notion d’événement marquant dépend..
  - De l’importance de l’événement, s’il est inusité, inhabituel (Participation au bal des finissants comparé à prendre une bière)
  - Du coût et des bénéfices financiers et sociaux: Achat de maison, d’automobile, comparé à CD de musique
  - Des conséquences se poursuivant dans le temps (Mariage, naissance d’un enfant)

24

La mesure des comportements, en résumé...

- Le répondant a une tendance générale au sur-aveu pour les comportements “ordinaires”, désirables.
- Le répondant a une tendance au sous-aveu de comportements “sensibles”, illégaux, mal vus.
- Le répondant a une tendance générale au non-aveu délibéré “par paresse” s’il sait que le fait de répondre non à une question filtre lui évitera une liste de questions.
- **Toujours vérifier si, dans la question, on a précisé “qui fait quoi, quand et où”, lorsque nécessaire (soit la plupart du temps).**

25

Les attitudes

Définition

- Une attitude c’est ce que l’on pense, ressent, anticipe ou valorise....
  - Un état psychologique ou une prédisposition stable
  - Une réaction à une situation spécifique, stable ou non (ex : la satisfaction par rapport à son salaire, etc.)
  - Une opinion sur un sujet (ex: en sc. politiques)
  - Une valeur — soit la mesure de l’importance accordée à certains éléments dans les décisions de comportement, la projection dans l’avenir.
  - Une intention de comportement.

26

La mesure des attitudes

La précision

- Il faut prêter attention à l’**aspect**, à l’objet de la question.
  - **Affectif:** aime ou pas/être en faveur/pour ou contre
  - **Cognitif:** ce que l'on pense de, comment on conçoit les choses
  - **Conatif/ action:** les intentions de comportement
  - **Réactif versus proactif :** satisfactions et perceptions versus valeurs, importance accordée à certaines choses, aspirations, etc.
- Attention aux “réflexes” de type “échelle d’accord portant sur la satisfaction”.

27

La mesure des attitudes

La précision

- Il faut mesurer un seul concept par question :
  - Plusieurs questions plutôt que 2 questions en une, soit **un seul indicateur par question**.
- Il faut mesurer la **force** de l'attitude...
  - En utilisant des échelles de réponse (très, assez, peu, pas du tout, par exemple, **et non pas "oui ou non"**)
  - En faisant plusieurs questions indépendantes.
- Il faut donner les alternatives explicites dans la question.
  - à cause de l'étude de Rugg (1941)/ allow public speeches against democracy (21% agree)/ forbid public speeches against democracy (39% do not agree)
  - Exemple: On demandera: Lorsque vous pratiquez un sport de raquette, préférez-vous porter vos lunettes ou vos verres de contact?

28

La mesure des attitudes

La précision

- Lorsque qu'on utilise une échelle de réponse,...
  - Dans les entrevues, **les catégories de réponse doivent toujours apparaître dans la formulation de la question**, sauf dans des cas évidents (réponse oui ou non);
  - Les catégories de réponses doivent être ***exhaustives* et *exclusives***.
    - Une catégorie ne peut en "englober" une autre et donc **chaque catégorie doit être qualifiée**: tout à fait d'accord, plutôt d'accord, etc. **Et non pas:** tout à fait d'accord, d'accord, en désaccord, tout à fait en désaccord).
  - Elles doivent apparaître dans un ordre logique (logique de temps, d'importance, etc)
  - Le nombre de catégories dépend du mode de passation; on parle de 4 ou 5 catégories ou choix de réponse au maximum, surtout dans les sondages téléphoniques.

29

La mesure des attitudes

La neutralité

- La question telle que formulée ne doit jamais laisser entendre qu'une réponse serait meilleure qu'une autre, en attirant l'attention sur un point de vue, en incitant à la pitié, à la solidarité avec les gens du même groupe ethnique, etc. et ce, quelle que soit la "justesse de la cause".
- Faire attention aux réactions de prestige (désirabilité, conformité sociale).
- Faire attention à la "subtilité" des répondants et surtout faire preuve de subtilité nous-mêmes: souvent on ne peut pas poser les questions directement.

30

La mesure des attitudes

La pertinence

- On veut mesurer l'opinion, pas la créer de toutes pièces, donc l'opinion doit préexister ou pouvoir être facilement mobilisée au moment où je pose la question. C'est la "Loi de Mitofsky".
- Il est souvent pertinent de mesurer la connaissance que les gens ont d'un sujet avant de mesurer leur attitude. Par la suite, on a le choix (selon la situation):
  - Soit de poser la question d'attitude à tout le monde et ensuite vérifier s'il y a une différence d'attitude entre ceux qui disaient être au courant et les autres.
  - Soit poser la question uniquement à ceux qui disent être au courant.
  - Soit donner l'information et poser la question d'attitude ensuite.

31

La mesure des attitudes

Le choix des échelles de mesure

- Objet de débat :
- Échelles avec ou sans point milieu:
  - nombre de catégories pair ou impair?
- Échelles bi-polaires ou unipolaires:
  - satisfaction-insatisfaction ou degré de satisfaction?

32

Les connaissances

Définition

- Les questions de connaissance font référence à ce que l'on sait ou croit savoir sur une question clairement énoncée. **Il doit y avoir une réponse connue et vérifiable à la question.**
  - Elles sont souvent utilisées auprès de l'opinion publique pour vérifier l'efficacité de campagnes publicitaires d'entreprises privées ou publiques: prévention en santé, alcool au volant, taux d'augmentation de tarifs, etc.
  - Elles peuvent servir à mesurer jusqu'à quel point les gens ont compris des explications.
  - Elles sont utilisées comme mesure de perception:
    - Selon vous, quelle est la proportion d'anglophones sur l'Ile de Montréal?

33

La mesure des connaissances

La précision

- La formulation de la question doit pointer vers l'information exacte que l'on veut obtenir. La question est généralement ouverte (mais pas toujours).
- Attention:
  - Demander si la personne pense que la cigarette est dangereuse pour la santé **est une question d'attitude**;
  - Demander si la personne *connaît la position de Santé Canada sur les risques de la cigarette pour la santé et lui demander de donner cette position* **est une question de connaissance.**

34

La mesure des connaissances

La pertinence

- Des questions de connaissance peuvent précéder des mesures d'attitude pour déterminer la pertinence des réponses ou même la pertinence de poser les questions d'opinion, par exemple.
- Les sujets *trop pointus* pour les répondants ou peu adaptés à leur situation favorisent des problèmes de désirabilité sociale ainsi que des refus. Le répondant ne veut pas apparaître ignorant.
- Il faut prêter attention à maintenir l'interaction et ne pas mettre le répondant mal à l'aise. Il ne faut pas poser des questions difficiles en début de questionnaire.

35

L'ergonomie

Qu'entend-on par *ergonomie* ?

- On peut définir ***l'ergonomie*** comme la science qui examine les outils de travail pour réduire les sources de blessures, d'accidents et d'erreurs.
- Dans ce sens, on peut parler d'ergonomie du questionnaire comme on parle d'ergonomie des outils en général ou des interfaces humains-ordinateurs.
- L'ergonomie du questionnaire vise à minimiser les sources d'erreur dans la collecte des données.

36

L'ergonomie

- Voir les grands principes de Sudman et Bradburn
- Revoir les principes vus dans le cours:
  - Précodage, alignement à droite, utilisation de tableaux, etc.
  - Respect de l'ordre numérique et chronologique.
  - Regrouper ensemble ce qui va ensemble.
- Tenter de garder le questionnaire **intéressant pour le répondant** du début à la fin, entre autres en variant les styles de question.
- Respecter le cheminement normal d'une conversation.

37

L'échantillon et la gestion

- Les types d'échantillon (vu au début du cours)
- La stratégie (vu au début du cours)
- Dix-neuf fois sur 20, la marge d'erreur.
- Estimer la taille de l'échantillon théorique requise
- Estimer la taille de l'échantillon de départ requis
- Estimer la fraction de sélection
- Calculer le rendement.
- Calculter la pondération.

38

Le contexte de l'interaction, le mode d'administration du questionnaire

- Il existe deux procédés pour faire passer un questionnaire à la population sondée :
  - L'entrevue avec un interviewer:
    - Par téléphone;
    - En face à face, sur place.
  - Le questionnaire auto-administré (questionnaire que le répondant complète seul) distribué...
    - Par courriel, courrier interne dans une institution (université, entreprise, etc.) ou courrier postal (si liste d'adresses nominatives)
    - Par téléphone (Réponse Vocale Interactive) ou texto.
    - De main en main individuellement ou à un groupe de personnes pour une réponse immédiate habituellement ou parfois pour renvoi postal par le répondant.

39

Déterminer le mode d'administration le plus approprié

- Selon des critères relatifs
  - Au sujet de l'enquête (demande réflexion importante, questions intimes, désirabilité sociale, demande observation de l'environnement)
  - Au type d'informations accessibles sur la population
  - Au type de population et à sa dispersion
  - Au taux de réponse désiré ou jugé essentiel
  - Aux ressources financières disponibles

40

Relation entre mode d'administration et base de sondage

- Téléphonique:
  - Génération aléatoire de numéros de téléphone (population générale)
  - Liste de numéros de téléphone (membres d'organisations, clients, organisations, etc.)
- Par courrier électronique, postal:
  - Liste d'adresses pré-existantes ou faciles à trouver correspondant aux membres de la population
  - Note: Difficulté (impossibilité pour ce qui est de la population générale) de trouver des listes d'adresses e-mail où tous les utilisateurs seraient inclus et le seraient une seule fois... → Faire affaire avec une firme, accès à un panel web, probabiliste ou non.
- Face à face:
  - Échantillonnage par aires géographiques, cartes d'occupation du sol, liste de codes postaux, ou en fonction de la présence sur des lieux précis (commerces, entrée du métro, etc.)
- Distribué en groupe:
  - Listes de groupes-classe, d'écoles, d'équipes de travail
  - Possibilité de constituer un échantillon en grappe ou aréolaire

41

Relation entre mode d'administration et critères de base

Les critères de base (1)

- 1) Le sujet de l'enquête
  - Questions demandant réflexion poussée, intimes, demandant des rappels importants: quest. auto-administré
  - Sondage complexe, demandant observation de l'environnement: entrevue en personne (face à face, sur place)
  - Problème de désirabilité sociale, soit la tendance du répondant à donner une bonne image de lui-même : **questionnaire auto-administré de préférence, ou téléphonique.**

42

Le choix du mode d'administration du questionnaire

Les critères de base (suite)

- **2) Accessibilité de la base de sondage**
  - Aire géographique restreinte: **face à face, courrier distribué.**
  - Aire géographique étendue, population dispersée: **téléphonique. Courrier/courriel si et seulement si on a accès à une liste des adresses des personnes visées ou recours à un panel web.**
- **3) Caractéristiques de la population**
  - Niveau de scolarisation, âge, etc.
  - Intérêt pour le sujet (si peu d'intérêt, entrevue est préférable parce que l'interviewer tentera d'obtenir la collaboration).

43

Le choix du mode d'administration du questionnaire

Les critères de base (suite)

- **4) Questions financières :**
  - Moins cher: **internet, postal, courrier interne, distribution en groupe.**
  - Plus cher: **entrevue sur place, face à face (utilisé presque exclusivement par les grands organismes statistiques gouvernementaux). Mais peut être approprié et peu coûteux pour un sondage lors d'un événement ou dans une aire restreinte.**
- **5) Taux de réponse**
  - Moins bon, habituellement: **autoadministré en général**
  - Moyen (dépendant entre autres de la qualité de la firme de sondage): **entrevues téléphoniques**
  - Meilleur: **face à face, distribution avec complétion sur place**

44

Les types d'échantillon

- Probabilistes:
  - Tirage aléatoire simple ou systématique
  - Tirage à probabilité égale ou inégale
    - Inégale fonction de la taille ou de certaines caractéristiques (*échantillon stratifié*).
  - Unité d'échantillonnage peut être
    - Une unité ou un groupe d'unités (aréolaire)
    - Sélectionnée elle-même ou avec d'autres dans le même groupe (en grappe).
  - L'échantillon peut être à plusieurs degrés ou à plusieurs phases.
- Non probabilistes:
  - Par quotas: on cherche à recréer une micro-population ayant les mêmes répartitions de caractéristiques que la population à laquelle on veut généraliser.
  - Au jugé: On utilise des critères logiques et théoriques pour faire le choix des unités. On cherche à bien représenter toutes les situations.
  - De volontaires: on prend ceux qui le veulent bien.

45

Échantillon: Résumé des "notions"

- e: marge d'erreur
- e_{max}: marge d'erreur maximale quand p=0,5
- e_{diff}: marge d'erreur d'une différence de proportions
- Z_α: Valeur de Z pour un seuil de confiance 1-α
- **p: proportion de présence d'une caractéristique dans l'échantillon**
- n: taille de l'échantillon de répondants obtenue ou estimée (n théorique)
- N: taille de la base échantillonnale
- Intervalle de confiance: p ± e

46

Dix-neuf fois sur 20

Calculer la marge d'erreur

- Où : $e = Z_\alpha * \sqrt{\frac{p*(1-p)}{n}}$
  - Z_α est l'espace sous la courbe normale (1,96 quand α = 0,05; ça serait 2,58 pour α = ,01)
  - *p* est la proportion que l'on retrouve dans l'échantillon et dont on cherche la marge d'erreur
  - *1-p* est 1 moins la proportion
  - *n* est la taille de l'échantillon sur lequel porte la proportion.

47

Concrètement,

Pour une proportion de 30% (0,30) d'un échantillon comprenant 800 personnes

$$e = 1{,}96 * \sqrt{\frac{0{,}3*(1-0{,}3)}{800}} = 1{,}96 * \sqrt{\frac{0{,}21}{800}} = 1{,}96 * \sqrt{0{,}0002625} = 1{,}96 * 0{,}0162$$

$$e = 0{,}0318 \quad et \quad e\% = e*100 = 3{,}18\%$$

- ET donc 19 fois sur 20 (dans 19 échantillons sur 20), la proportion se situera entre 30% +3,18% et 30%- 3,18%, soit entre 26,82% et 33,18%. C'est **l'intervalle de confiance** de la proportion.

48

La marge d'erreur maximale

Si on calcule la marge d'erreur à 50% pour un échantillon de 1000 personnes...

- On obtient:

$$e = 1{,}96 * \sqrt{\frac{0{,}5*(1-0{,}5)}{1000}} = 1{,}96 * \sqrt{\frac{0{,}25}{1000}} = 1{,}96 * \sqrt{0{,}00025} = 1{,}96 * 0{,}0158$$

$$e = 0{,}031 \quad et \quad e\% = e*100 = 3{,}1\%$$

À 50%, la marge d'erreur est maximale. ET donc, la marge d'erreur ***maximale*** est de 3,1% pour 1000 personnes dans l'échantillon.
Et la proportion se situera 19 fois sur 20 entre 50% - 3,1% et 50% + 3,1% [46,9; 53,1].

49

La marge d'erreur pour population finie

$$e = Z_\alpha * \sqrt{\frac{p*(1-p)}{n}} * \sqrt{\frac{N-n}{N-1}}$$

- Et donc, pour une proportion de 0,50 (p), 1000 répondants (n) et une population de 5000 personnes (N):

$$e = 1{,}96 * \sqrt{\frac{0{,}5*(1-0{,}5)}{1000}} * \sqrt{\frac{5000-1000}{5000-1}} = 0{,}031 * \sqrt{\frac{4{,}000}{4{,}999}} = 0{,}031 * \sqrt{0{,}8} = 0{,}031 * 0{,}89 = 0{,}0277$$

50

Comment savoir si deux proportions sont significativement différentes l'une de l'autre?

- Une façon simple consiste à vérifier si les intervalles de confiance de deux proportions se chevauchent, par exemple...
  - Si on compare une proportion de 45% sur un échantillon de 800 (e=3,4%, comme calculé précédemment) et une proportion de 35% **sur le même échantillon** (e=3,3%), on obtient les intervalles respectifs suivants:
  - De 41,6% à 48,4% et de 31,7% à 38,3%.
  - Comme le maximum de l'intervalle de 35% (38,3%) est plus petit que le minimum de l'intervalle de 45% (41,6%), on conclut que les deux proportions sont statistiquement différentes (**les intervalles ne se recoupent pas)**. C'est la manière simple et conservatrice de calculer.

51

Marge d'erreur pour une différence de proportion dans un même échantillon

- Toutefois, pour la différence entre deux pourcentages pris sur le même échantillon, la marge d'erreur de la différence, moins conservatrice, devrait se calculer plutôt ainsi:

$$e_{diff} = Z_{\alpha} * \sqrt{\frac{(p_1 + p_2) - (p_1 - p_2)2}{n}} = 1{,}96 * \sqrt{\frac{(,45+,35)-(,45-,35)2}{800}}$$ Ce qui donne 6,1%

  - Où p_1 égale la première proportion et p_2 la deuxième et n est la taille de l'échantillon.
  - **La différence entre les deux proportions doit être plus grande que 6,1% pour être significative**. C'est le cas (diff=0,45-0,35=0,10, i.e., 10%). Il s'agit d'un test moins conservateur que celui présenté à la page précédente.

52

Marge d'erreur pour une différence de proportions entre 2 échantillons différents

- Pour la différence entre deux proportions prises sur des échantillons distincts, *comme la différence entre deux sondages d'intention de vote pour un parti, par exemple*, la marge d'erreur se calcule ainsi:

$$e_{diff} = Z_{\alpha} * \sqrt{\frac{p1*(1-p1)}{n1} + \frac{p2*(1-p2)}{n2}}$$ Ce qui s'approche de:

$$e_{diff} = Z_{\alpha} * \sqrt{\frac{2p*(1-p)}{n}}$$

  - Où "p" est égal à $(p_1+p_2)/2$, soit la moyenne des deux proportions dont on veut estimer la différence et *n* est la moyenne de n_1 et n_2
  - Cette erreur s'approche de 1,4*e.

53

L'échantillon, combien?

Estimé de la taille théorique nécessaire

La taille théorique est le **nombre de répondants que l'on souhaite avoir** à la fin pour faire nos analyses. On peut parler de taille souhaitée, attendue.

$$n = \frac{Z_{\alpha}^2 * p * (1-p)}{e^2}$$

- Où
  - n est la taille théorique souhaitée
    - Z_{α} est l'espace sous la courbe normale (1,96 quand α = 0,05; ça serait 2,58 pour α = ,01)
    - ***p* est la proportion attendue de la caractéristique qui nous intéresse** (i.e. réponses des répondants à certaines questions) et dont on veut contrôler la marge d'erreur. **On prend habituellement la proportion maximale, soit 0,5.**
    - *1-p* est 1 moins la proportion.

54

Concrètement, donc...

- Si j'estime qu'une marge d'erreur de 4% (attention, en décimales dans l'équation, soit 0,04) dans mon cas est acceptable,...
- Et si je garde le seuil de confiance habituel de 95%, avec une **proportion maximale** de 50%,...

$$n = \frac{1{,}96^2 * 0{,}5 * (1 - 0{,}5)}{0{,}04^2} = \frac{3{,}84 * 0{,}25}{0{,}0016} = 600$$

- Il faut donc 600 personnes dans l'échantillon **attendu de répondants** pour avoir une marge d'erreur de 4% lorsque la proportion est maximale (à 50%) et le seuil de confiance à 95%.

55

Si tout était parfait,...

- Si ma base échantillonnale correspondait parfaitement à ma population telle que je la définis,
- Si toutes les personnes de mon échantillon étaient en mesure de me répondre et correspondait à ma définition de la population,
- Si toutes les personnes de mon échantillon étaient rejoignables et acceptaient de répondre à mon sondage,
- Il suffirait de tirer de la base le nombre de personnes que je veux à la fin dans l'échantillon, pour obtenir ce nombre,
- MAIS tout n'est pas parfait.
- **D'où la deuxième étape qui suit...**

56

Propriétés de la base de sondage: le biais et la validité

- La base peut ne pas correspondre totalement à la population, certains membres de la population n'y sont pas représentés: c'est le biais de la base: *Des unités devraient y figurer mais n'y sont pas.*
- La base peut contenir des erreurs et des imperfections. On parle alors de *qualité de la liste* et donc de *validité des unités*.
  - Les unités non valides sont celles qui ne devraient pas figurer sur la liste si celle-ci était à jour et sans erreur et correspondait parfaitement à la population.

57

La qualité des unités/individus/ménages

L'éligibilité, l'incidence

- L'éligibilité a trait au fait que, **parmi les unités valides**, des personnes ne sont pas membres de la population telle que définie.
- On parle d'incidence pour parler de la proportion d'unités valides où l'on est susceptible de trouver au moins une personne membre de la sous-population telle que définie.
- **On parle d'incidence seulement lorsqu'il y a une sous-population** et on doit toujours estimer l'incidence dans ce cas.

58

Le taux de réponse

- Les éléments relatifs à la coopération --volontaire ou involontaire -- des personnes sélectionnées constituent le taux de réponse.
- Si l'unité collective (par exemple, un ménage) sélectionnée refuse de collaborer, on parle de *refus du ménage*.
- Si la personne sélectionnée refuse de collaborer, on parle de *refus de la personne*.
- Si, après de multiples tentatives à diverses heures sur semaine et en fin de semaine, personne ne répond aux tentatives de contact, on parle de *non-contact* ou de "*pas de réponse après plusieurs tentatives".*
- Si la personne sélectionnée est absente pour la durée de la période de collecte, on parle d'*absence prolongée.*
- Si le questionnaire a été commencé mais non terminé, on parle de *questionnaire incomplet*.

59

Comment calculer l'échantillon de départ?

- On a vu comment estimer la taille de l'échantillon théorique voulu étant donné la marge d'erreur que l'on accepte MAIS,
- ***Combien faut-il sélectionner d'unités au départ*** pour obtenir ce nombre de personnes à la fin? (Voir calculateur). La formule est la suivante:
- $$n_{Éligible} = n_{théorique} * \frac{1}{tx\ réponse}$$

$$n_{Valide} = n_{Éligible} * \frac{1}{[\ tx\,éligib. * tx\,incidence]}$$

$$n_{départ} = n_{Valide} * \frac{1}{tx\,validité}$$

$$n_{départ} = n_{théorique} * \frac{1}{tx\ réponse} * \frac{1}{tx\ éligib. * [tx\ incidence]} * \frac{1}{tx\ validité}$$

60

Calcul du n de départ

Population de très grande taille (+ de 20 fois échantillon)

1) Décider de la marge d'erreur que l'on est prêt à se permettre
2) calculer le n théorique
3) calculer le n de départ en fn des divers taux
4) Calculer le pas.

61

Calcul du n théorique

Population finie, de petite taille (moins de 20 fois l'échantillon)

1) estimer le pas
2) calculer le n de départ
3) calculer le n théorique en fn des divers taux
4) calculer la marge d'erreur (pr population finie).

62

Rendement du plan échantillonnal

- L'ensemble des facteurs mentionnés plus haut (validité, éligibilité/incidence et taux de réponse) amène à parler du **rendement du plan échantillonnal,**
  - c'est-à-dire la proportion attendue de questionnaires complétés étant donné l'échantillon de départ.
- Ce nombre peut être trouvé
  - Soit en multipliant les divers taux
  - Soit en divisant l'échantillon théorique (celui que l'on veut obtenir) par l'échantillon de départ,
- On dira, par exemple, que le rendement attendu du plan *(pour un taux de réponse de 60%, un taux d'éligibilité de 95% et un taux de validité de 80%)* est de
  - 0,6*0,95*0,8=0,456.

63

Pourquoi avoir calculé tout ça?

Pour obtenir la fraction de sélection, soit comment procéder...

- C'est la dernière étape: **J'en prends un sur combien?** (Ou la première étape pour une population finie).
- La fraction de sélection = $f = \frac{n}{N} = \frac{\text{échantillon de départ}}{\text{base échantillonnale}}$

Sauf lorsqu'on a l'information seulement sur la population. Dans ce cas:

$$f = \frac{n}{N} = \frac{\text{échantillon valide estimé}}{\text{Paramètre de la population}}$$

- **Et le pas égale l'inverse de la fraction de sélection, soit:**

$$pas = \frac{1}{f} = \frac{N}{n} = \frac{\text{base échantillonnale}}{\text{échantillon de départ}}$$

64

Synthèse des notions

- n théorique ou n attendu: Taille de l'échantillon que l'on voudrait obtenir
- Taux de validité estimé de la base: estimation de la qualité de la liste
- Taux d'éligibilité/d'incidence estimé des unités: estimation de la proportion de personnes qui correspondront à la population telle que définie.
- Taux de réponse estimé : proportion des unités éligibles --ou dont on présume l'éligibilité -- avec lesquelles il a été possible de compléter l'entrevue.
- n de départ: Taille de l'échantillon à sélectionner pour obtenir l'échantillon final voulu.
- Rendement du plan échantillonnal : proportion de questionnaires complétés sur le nombre d'unités de départ.
- Fraction de sélection/pas: proportion de l'échantillon de départ sur le nombre d'unités de la base de sondage.

65

Synthèse des étapes

Population de grande taille

- 1) Recueillir les informations sur la population et/ou sur la base de sondage et déterminer comment procéder.
- 2) Déterminer en fonction de la question de recherche et des informations obtenues la taille de l'échantillon (ou des sous-échantillons, le cas échéant) théorique souhaitée.
- 3) Estimer le taux de validité, le taux d'éligibilité (et d'incidence, le cas échéant) et le taux de réponse .
- 4) Estimer l'échantillon de départ.
- 5) Estimer la fraction de sélection.

66

Synthèse des étapes

Population finie

- Pour une population finie (base=moins de 20 fois échantillon théorique), il faut procéder à l'inverse, soit:
- 1) Combien d'unités y a-t-il dans la base?
- 2) Quelle pourrait être la fraction de sélection? Ça donnerait combien d'unités de départ? Essais-erreurs...(1 sur 2? 1 sur 3? 1 sur 10?)
- 3) Estimer le taux de validité, le taux d'éligibilité (et d'incidence, le cas échéant) et le taux de réponse.
- 4) Étant donné les taux de validité, d'éligibilité et de réponse estimés, ça donnerait combien d'unités attendues?
- 5) Quelle serait alors la marge d'erreur? Serait-elle acceptable?

67

Pondération - redressement

Étape postérieure à la collecte des données

- Pour généraliser à la population totale, après avoir collecté les données, on se demande combien de personnes "vaut" chaque répondant.
- On pondère les données en fonction de ce qui s'est passé pendant la collecte et de la (ou des) fractions de sélection appliquées.
  - Le poids d'échantillonnage est égal au pas.
  - Le poids de non-réponse est égal à l'inverse du taux de réponse.
  - Le poids total = $poids = pas * \frac{1}{tx\ de\ réponse}$

68

La gestion des opérations, pourquoi?

- La gestion des opérations est souvent le seul élément sur lequel le chercheur peut avoir un contrôle.
- Elle vise à s'assurer de la qualité des données, entre autres, en s'assurant que tous les membres de l'échantillon
  - Soient rejoints et invités à collaborer à l'enquête
  - Acceptent de collaborer à l'enquête.
- Et comme les gens ont de multiples activités, il faut s'assurer que la collaboration à notre enquête... se retrouve "sur le dessus de la pile".

69

La gestion des opérations

Sondage par entrevue

- Dans un sondage par entrevue, deux aspects apparaissent essentiels:
  - A) prendre tous les moyens pour tenter de rejoindre les personnes sélectionnées
  - B) amener ces personnes à coopérer à l'enquête
- Dans la conduite de ces opérations, deux aspects sont essentiels:
  - Il faut gérer la circulation des numéros de téléphone ou des adresses à contacter
  - Il faut gérer les personnes qui font les contacts, soit les interviewers.

70

Rejoindre les ménages

- L'idée est de contacter les répondants potentiels à divers moments de la journée sur semaine et durant le week-end.
- On fait habituellement un minimum de 5 tentatives (2 le soir, 2 le week-end, 1 de jour) mais les firmes peuvent aller jusqu'à 20 contacts et plus pour les sondages scientifiques.
- A chaque numéro de téléphone ou adresse correspond un ménage et non pas un individu (mais ça change avec les cellulaires).
- **On a donc habituellement, dans les sondages auprès de l'ensemble de la population, un échantillon à deux degrés,** ménage et individu.
  - D'où les questions: comment choisir quelqu'un au hasard dans un ménage?

71

Récupérer les refus

- Comme beaucoup de refus, particulièrement les refus du ménage, sont circonstanciels --on a appelé à un mauvais moment -- ou dû à l'inexpérience de l'interviewer, on tente de rappeler ces ménages. On parle de récupération des refus.
- Empiriquement, autour de 25% de ces refus sont récupérés sans trop de difficulté en faisant des appels à un moment différent et en les confiant à des interviewers expérimentés.
- Deux avantages à la récupération des refus:
  - Moins de biais possible de l'échantillon
  - Coût moindre puisque l'on sait déjà que ce sont des personnes qui font partie de l'échantillon et qu'elles risquent d'être présentes à la maison.

72

Quel est le rôle de l'interviewer?

- L'interviewer a deux rôles clés:
  - A) Faire le contact, sélectionner et CONVAINCRE la personne sélectionnée de collaborer.
  - B) Conduire l'entrevue de façon professionnelle et neutre et enregistrer les réponses aux questions sans erreur.
- Le premier rôle est sans doute le plus difficile et le plus ingrat, d'où beaucoup de recherches sur cette question.
- Voir sur le site web le module sur les interviewers et dans la recherche, les présentations faites sur les interviewers.

73

La gestion des opérations

Questionnaires auto-administrés

- Les modes de passation sont variés:
  - Par internet
  - Par la poste "ordinaire"
  - Par la poste interne en entreprise
  - Distribués dans un groupe et passation sur place
  - Distribués et le répondant renvoie le questionnaire par la suite.
- Les informations disponibles, outre les adresses, peuvent varier.

74

Les principes de base

- Soigner la présentation pour que le questionnaire soit intéressant à remplir.
- S'assurer d'avoir un contrôle total sur le nombre de questionnaires distribués et à qui ils ont été distribués.
- S'assurer de pouvoir suivre le retour des questionnaires et, de préférence, savoir qui a retourné ou pas son questionnaire.
- Ne pas hésiter à faire des rappels (sans être trop harcelant), à varier les messages des rappels et à envoyer de nouveaux questionnaires si nécessaire.

75

En conclusion sur la gestion des opérations

- Plus ce que l'on risque de découvrir peut être"difficile à dire", peu acceptable socialement, sensible,...
- Plus la gestion des opérations devra être sans reproche puisque...
- C'est le premier point qui sera attaqué par les personnes qui désirent mettre en doute les résultats de votre étude.

76

Rappel : les étapes

- 1. Élaboration de la problématique, question de recherche,...
- 2. Élaboration du questionnaire
  - Première version, deuxième ou troisième si nécessaire (selon commentaires du "client")
  - Version pré-test
  - Rapport de pré-test
  - Version finale corrigée suite au pré-test
  - Traduction si nécessaire
- 3. Parallèment, élaboration du plan de sondage

77

Le pré-test

- Le pré-test du questionnaire est une phase extrêmement importante qui permet
  - De repérer les erreurs, les mauvaises formulations, les questions incompréhensibles ou mal comprises.
  - D'évaluer la durée du questionnaire et de le raccourcir au besoin.
  - D'évaluer si le mode de passation est viable.

78

Rappel les étapes (suite)

- 4. Planifier les opérations de collecte
  - Selon le mode d'administration
    - Entrevue : combien d'heures d'interviews, de rappels, d'interviewers, etc.
    - Auto-administré : nombre de questionnaires, enveloppes aller, enveloppes retour (pré-timbrées et pré-adressées), rappels, etc.
  - Rapport méthodologique
- 5. Plan de codification si nécessaire:
  - Comment coder les réponses aux questions ouvertes, comment coder la non réponse, etc.
  - Plan de code
- 6. Entrée des données (si non informatisé)
  - Entrée des codes pour les questions ouvertes (même si informatisé).
- 7. Plan d'analyse
  - Rapport d'analyse

79

Rappel des objectifs du cours

- Être en mesure de bien comprendre et d'élaborer un projet de sondage du début à la fin.
- Développer une approche critique face aux sondages publiés dans les médias et dans les articles scientifiques sur les divers plans vus dans le cours, entre autres la qualité des questions et la marge d'erreur.

80

Pour l'examen:

- Comprendre l'origine, l'historique des sondages.
- Être capable
  - d'identifier le type de question (attitudes, comportements, faits et connaissances)
  - d'identifier les problèmes des questions (neutralité, précision, pertinence)
  - de corriger les questions
- Être capable de formuler adéquatement des indicateurs et des questions de divers types permettant de répondre à une question de recherche.
- Être capable d'élaborer une stratégie appropriée à la question de recherche.

81

Pour l'examen

- Être capable de calculer une marge d'erreur sur une proportion et maximale, un intervalle de confiance et estimer un échantillon théorique.
- Comprendre les notions vues dans la partie sur l'échantillonnage (validité, éligibilité, taux de réponse, seuil de confiance, ...)
- Être capable de calculer un échantillon de départ, une fraction de sélection, un rendement de plan, une pondération.
- Bien comprendre les étapes pour passer du n de départ au n théorique et vice-versa.

82

Pour l'examen

Comment se préparer

- 1. Réviser les présentations et les notes de cours.
- 2. Réviser les commentaires faits sur vos travaux et vous assurer d'avoir bien compris.
- 3. Utiliser le questionnaire de K. Marx pour vous exercer à repérer les problèmes des questions et à les corriger.
- 4. Faites les problèmes dans Studium et sur le site Web pour vous assurer d'avoir bien compris la partie sur l'échantillon.
- Vérifier avec le calculateur que vous êtes bien capable de calculer les diverses informations avec votre calculatrice.

83