

Labo 1: Statistiques et variables

Visseho Adjiwanou, PhD.

18 January 2022

Exercice 1: Du concept à la mesure

Marx a soutenu que le capitalisme est un mode de production économique basé sur le conflit entre deux classes de personnes: le capital (alias la bourgeoisie) et le travail (alias le prolétariat). L'appartenance à une classe est déterminée par sa relation avec les moyens de production. Le capital possède les moyens de production et le travail non. Au travail, les membres de la classe ouvrière vendent leur capacité de faire du travail (leur force de travail) au capital en échange d'un taux de salaire. Marx a affirmé que cela a pour effet **d'aliéner** les gens de leur travail parce que les individus ne contrôlent plus leur travail, ni le produit de leur travail.

Logiquement, cela a du sens; cependant, pouvons-nous mesurer l'aliénation d'une manière ou d'une autre pour offrir des preuves empiriques de son existence? L'aliénation de Marx existe au niveau conceptuel. Pouvez-vous penser à des moyens d'opérationnaliser l'aliénation?

Aliénation: condition dans laquelle les gens souffrent d'une «déconnexion» entre eux-mêmes et de leur travail et d'une déconnexion les uns des autres.

Réponse

Rappel: Conceptualisation et opérationnalisation

En pensant aux concepts et aux variables, il est utile de penser chronologiquement. Définir ce que nous entendons par une idée doit précéder la conception de moyens de mesurer notre idée. Décrire ou définir une idée qui, selon nous, représente quelque chose dans le monde réel est un processus appelé conceptualisation. Concevoir des moyens de mesurer les idées que nous avons conceptualisées s'appelle opérationnalisation.

1. Conceptualisation

Idée =====> Définition

Vous ne vous souciez peut-être pas beaucoup de ce que vous faites pour le travail tant que vous gagnez suffisamment d'argent. D'un autre côté, vous préférerez peut-être sacrifier un revenu pour un travail qui vous donne beaucoup de satisfaction, c'est-à-dire un travail moins aliénant. À partir de cela, *nous pourrions supposer que les personnes qui aiment le type de travail qu'elles effectuent sont moins éloignées de leur travail que les personnes qui n'aiment pas le type de travail qu'elles font.*

2. Opérationnalisation

Définition =====> Mesures (variables)

Bien que nous ne puissions pas saisir l'aliénation et la mettre sur une échelle pour voir dans quelle mesure une personne en fait l'expérience, nous pouvons demander aux gens s'ils sont susceptibles de rester à un emploi pour le revenu ou de sacrifier un revenu pour un emploi qu'ils aiment. Ceux qui indiquent qu'ils seraient disposés à prendre moins de salaire pour un emploi qu'ils aiment seraient ceux qui connaissent un niveau d'aliénation plus faible.

- **Mesure** Seriez-vous prêt à accepter une réduction de salaire pour avoir un emploi qui vous plaît plus que votre emploi actuel? - Oui - Non

Nous pourrions affirmer que ceux qui répondent «non» connaissent des niveaux d'aliénation plus importants de leur travail que ceux qui répondent «oui».

3. Quels sont les problèmes éventuels avec cette mesure?

Cependant, un problème qui se pose avec cette variable est qu'elle comporte un biais de classe. En d'autres termes, les personnes qui ne gagnent pas assez d'argent pour joindre les deux bouts sont beaucoup plus susceptibles de répondre «non». Par conséquent, nous n'avons pu comparer que les réponses des répondants de classe relativement similaire.

Plusieurs facteurs doivent être pris en compte lors de l'opérationnalisation d'une variable. Deux d'entre eux sont que les modalités (catégories) de variable doit être (1) collectivement exhaustive et (2) mutuellement exclusive.

Autres exemples avec le revenu

Lorsque nous prenons une idée et définissons ce que nous entendons exactement par elle, nous conceptualisons. Par exemple, lorsque nous prenons le concept de **revenu** et le conceptualisons, nous pourrions nous référer au montant d'argent que les gens gagnent à un emploi donné au cours d'une année donnée. Ou nous pourrions faire référence au montant d'argent que les gens gagnent à leur travail ajouté au montant d'argent qu'ils gagnent grâce aux investissements. L'une ou l'autre technique est un moyen valable de conceptualiser le revenu. Il est important de transmettre ces définitions conceptuelles à ceux qui lisent et utilisent nos résultats.

Lorsque nous prenons une idée, un concept et élaborons une mesure avec laquelle comparer un cas avec le cas suivant, nous **opérationnalisons** une variable. Par exemple, nous pourrions conceptualiser le revenu pour désigner le montant d'argent qu'une personne gagne à une année donnée dans un emploi donné, mais cela laisse plusieurs possibilités ouvertes. La variable peut faire référence au montant réel qu'une personne gagne (par exemple 32515\$), ou elle peut faire référence à une fourchette de revenus (0\$ - 19999\$, 20000\$ - 29999\$, etc.). En d'autres termes, il n'y a pas de bonne ou de mauvaise façon d'opérationnaliser les concepts en variables (cela dépend davantage de ce que nous voulons savoir). Même ainsi, parfois, une manière s'avère plus utile et valable qu'une autre."

- Autre exemple

Pensez au concept de genre: faite une petite recherche sur la conceptualisation de ce concept et sur les nombreuses façons de son opérationnalisation.

Voir texte de Jean-Guy Prévost sur le recul du Français au Québec

Exercice 2: Variable dépendantes, variable indépendantes

Parmi les exemples suivants, identifier les variables et dire si elles constituent dans le contexte une variable dépendante ou indépendante. Définissez aussi le type de variable en question.

1. Une démographe analyse l'évolution du taux de fécondité des femmes depuis 1960 selon l'origine ethnique et le degré de scolarité.
 - **VD : taux de fécondité (ratio)**
 - **VI : origine ethnique (nominale), degré de scolarité (ordinaire, ratio)**
2. Dans une expérience sur le temps de réaction visuelle, on mesure le temps nécessaire pour percevoir des mots à connotation sexuelle par rapport à des mots non sexuels. Pour cette expérience, les chercheurs prennent bien soin de s'assurer que la vision des sujets leur permet de voir adéquatement les diapositives.
 - **VD: Temps de réaction visuel (quantitative, ratio, continue)**

- **VI: Connotation des mots (qualitative, nominale)**
3. Un criminologue recherche des données sur la nature des infractions au Code criminel selon le sexe des accusés.
 - **VD : nature des infractions au Code criminel ()**
 - **VI : Sexe**
 4. Lors d'une série d'expérience en psychologie sportive, on s'est rendu compte que chez les athlètes masculins, en situation de compétition relativement à celle de non compétition, il y a une augmentation du nombre de comportements complexes émis, mais qu'il y a une augmentation du nombre de comportements complexes émis, mais qu'il y a une baisse de la qualité de ceux-ci.
 5. Des élèves du cours de psychologie expérimentale veulent étudier les sentiments de culpabilité des individus selon le type de délit qu'ils ont commis.
 6. Un journaliste fait l'examen du pourcentage des intentions de vote des différents partis politiques québécois selon la langue parlée à la maison, le groupe d'âge et le degré de scolarité des électeurs.

Exercice 3: Type de variables

Voici une série de questions issue d'un questionnaire:

Pour chacune de ces question, dites le type de mesure de la variable ?

Réponse

1. Quel est ton sexe? Homme, Femme : **qualitative, nominal, dichotomique**
2. Quel est ton âge? _____ (en années) **ratio, discret/continue**
3. Tu es en quelle année? 1e année, 2e année, junior, senior, autres **ordinaire**
4. Vis-tu sur le campus ou fais-tu la navette hors campus? Sur le campus, hors campus **nominale**
5. Quelle est ton ethnie/race? _____ **nominale**
6. Te considères-tu comme un fumeur? Oui, Non **nominale**
 - a. Si oui, combien de cigarettes fumes-tu en moyenne par semaine? _____ **ratio, discret**
 - b. Si oui, depuis combien d'années fumes-tu? _____ **ratio**
 - c. Si oui, aimerais-tu cesser de fumer? Oui, Non **nominal**
7. As-tu déjà essayé de cesser de fumer? Oui, Non **nominale**
 - a. Si oui, combien de fois as-tu essayé? _____ **ratio**

qualitative - nominale - dichotomique - - ordinaire - dichotomique -

quantitative - ratio - discret - continue

- intervalle
 - discret
 - continue

Age (continue) -> Age (discret) -> groupe d'âge (ordinaire) -> nominal (adolescent, adulte, aînée)

Question 4: Variables dépendantes, indépendantes

Jean-Marc Leger, fondateur de la firme de sondage **Leger** a dressé la liste de résultats statistiques qu'il juge "drôles" pour égayer la fin de l'été 2013 des lecteurs du Journal de Montréal (<https://www.journaldemontreal.com/2013/09/09/droles-de-statistiques>)

Voici deux résultats de sa liste:

1. Les gens qui sont beaux peuvent gagner un salaire jusqu'à 15% plus élevé. Alors que si vous êtes plus grand, vous ne pouvez qu'espérer 3% de plus.
2. Une serveuse qui porte du rouge gagnera 15% de plus en pourboires. Un homme qui porte un vêtement avec un peu de rouge est perçu comme ayant un meilleur statut, plus riche, plus ambitieux et donc plus attirant.

A votre tour de refaire ces études. Pour chaque résultat dites (le cas échéant):

- A. La variable dépendante, son type et ces catégories;
- B. La variable indépendante, son type et ces catégories;

Réponse

1. Les gens qui sont beaux peuvent gagner un salaire jusqu'à 15% plus élevé. Alors que si vous êtes plus grand, vous ne pouvez qu'espérer 3% de plus.

- A. La variable dépendante, son type et ces catégories: **salaire gagné**, quantitative continue, 1\$ - 100000\$
- B. La variable indépendante, son type et ces catégories: **beauté**, qualitative ordinaire, extrêmement beau - beau - moyen - pas beau - pas du tout beau

Remarquez ici que la beauté est un concept. Comment mesurerez-vous ce concept?

- C. Le type d'association qui donne ses résultats: **Selon les catégories de la variable indépendante, test t (si deux catégories de la variable dépendante) ou analyse de la variance (ANOVA)**
- D. Quelle est la portée publique de ces résultats? (juste en une ou deux phrases): **Je n'en vois pas. Cela permet d'accentuer les préjugés sociaux.**

2. Une serveuse qui porte du rouge gagnera 15% de plus en pourboires. Un homme qui porte un vêtement avec un peu de rouge est perçu comme ayant un meilleur statut, plus riche, plus ambitieux et donc plus attirant.

- A. La variable dépendante, son type et ces catégories: **Pouboire reçu par mois**, quantitative continue, infini (0 à 10000\$)
- B. La variable indépendante, son type et ces catégories: **Couleur de l'habit**, qualitative nominale, rouge - pas rouge
- C. Le type d'association qui donne ses résultats: **test t de la moyenne / ANOVA**
- D. Quelle est la portée publique de ces résultats? (juste en une ou deux phrases) **Aucune.**

Exercice 5: Texte de Matthews Salganik

Texte : <https://www.bitbybitbook.com/fr/1st-ed/introduction/>

1. En quoi l'ère numérique change-t-elle le métier du sociologue?
2. L'auteur parle dans son texte de "readymade" et de "customade", Que signifie ces deux notions?
3. Donnez-moi des exemple de chaque type que vous avez déjà rencontré

Réponse

- 1.

- **Natures des données** - Abondance des données - observation directe

- **Natures des questions** - Ce ne sont pas forcément/toujours de nouveaux questionnements, mais dès fois des questions réexaminées sous un autre regard/prisme

- **Natures des approches** - Expérimentation - Sur Facebook, dans les supermarchés (et particulièrement sur les sites d'achat)

Pour résumer, voici ce que Salganik affirme:

“This study is kind of like a Rorschach inkblot test: what people see depends on their background. Many social scientists see a new measurement tool that can be used to test theories about economic development. Many data scientists see a cool new machine learning problem. Many business people see a powerful approach for unlocking value in the big data that they have already collected. Many privacy advocates see a scary reminder that we live in a time of mass surveillance. And finally, many policy makers see a way that new technology can help create a better world. In fact, this study is all of those things, and because it has this mix of characteristics, I see it as a window into the future of social research.”

«Cette étude est un peu comme un test de taches d'encre de Rorschach: ce que les gens voient dépend de leurs antécédents. De nombreux **spécialistes des sciences sociales** voient un nouvel **outil de mesure** qui peut être utilisé pour tester les théories sur le développement économique. De nombreux **spécialistes des données** voient un nouveau problème d'**apprentissage automatique**. De nombreux **entrepreneurs** voient une approche puissante pour **valoriser les données volumineuses** qu'ils ont déjà collectées. De nombreux **défenseurs de la protection de la vie privée** voient un rappel effrayant que nous vivons à une **époque de surveillance de masse**. Et enfin, de nombreux **décideurs** politiques voient un moyen pour les nouvelles technologies **d'aider à créer un monde meilleur**. En fait, cette étude est tout cela, et parce qu'elle a ce mélange de caractéristiques, je la vois comme une **fenêtre sur l'avenir de la recherche sociale**." (Salganik)

Rorschach inkblot test:

- <https://fr.rorschach-inkblot-test.com/>
- <https://openpsychometrics.org/tests/HEMCR/>

Pourquoi tout ceci est différent du passé (où les progrès n'ont jamais manqué)

- “Si vous pouvez capturer une image d'un cheval, alors vous avez une photo. Et, si vous pouvez capturer 24 images d'un cheval par seconde, alors vous avez un film. Bien sûr, un film n'est qu'un tas de photos, mais seul un sceptique extrême affirmerait que les photos et les films sont identiques.” (Salganik)

2.

Readymade: reprogrammer quelque choses pour une autre utilisation (big data)

Custommade: Créer quelque chose pour l'utilisation qu'ont veut en faire (enquête)

3.

Customade: Téléphone pour appeler Readymade: Téléphone pour internet/photo

Exercice 6: Texte de Jean-Guy Prévost

Texte: <https://id-erudit-org.proxy.bibliotheques.uqam.ca/iderudit/1008237ar>

- Quelle est la difficulté de mesurer l'évolution du Français au Québec ?
- D'autres éléments seront discutés en classe
- Effets de perspectives → biais de mesures
- Définition de la variable
- Regroupement des catégories

Réponse

Exercice 7 : A votre tour

Dans votre groupe, chacun présente aux autres les sujets qui l'intéressent. Ensuite, vous vous mettez d'accord pour choisir un sujet et une question de recherche.

1. Indiquer comment vous irez chercher l'information
2. A qui adresserez-vous le questionnaire?
3. Présentez six questions importantes décrivant les différents types de variables
4. Quelles sont les limites de votre étude?
5. Vous allez présenter cela aux groupes

Tour d'horizon

1. Y a-t-il une **relation** entre deux variables ?
2. Quelle est l'**intensité** de cette relation ?
3. Quelles sont la **direction** et la **forme** de cette relation?
4. Pouvons-nous **généraliser** la relation à la **population** de laquelle est tiré l'échantillon?
5. La relation est-elle vraiment **causale**?
6. Quelles sont les **variables intermédiaires** qui relient les variables **indépendante** et **dépendante**?

Exercice à partir de vos réponses à mon sondage

Question de recherche: Est-ce que les femmes ont un niveau plus faible en statistiques que les hommes?

1. Les deux variables en questions
 - Sexe
 - Niveau en statistiques (Qu'entendons-nous par cela?)

```
library(tidyverse)

## Warning: package 'tidyverse' was built under R version 3.6.2
## -- Attaching packages ----- tidyverse 1.3.1 --
## v ggplot2 3.3.5      v purrr   0.3.4
## v tibble  3.1.6      v dplyr  1.0.7
## v tidyr   1.1.4      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1
## Warning: package 'ggplot2' was built under R version 3.6.2
## Warning: package 'tibble' was built under R version 3.6.2
## Warning: package 'tidyr' was built under R version 3.6.2
## Warning: package 'readr' was built under R version 3.6.2
## Warning: package 'purrr' was built under R version 3.6.2
## Warning: package 'dplyr' was built under R version 3.6.2
## Warning: package 'forcats' was built under R version 3.6.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(summarytools)
```

```
## Registered S3 method overwritten by 'pryr':
##   method      from
##   print.bytes Rcpp

## Warning in system2("/usr/bin/otool", c("-L", shQuote(DSO)), stdout = TRUE):
## running command ''/usr/bin/otool' -L '/Library/Frameworks/R.framework/Resources/
## library/tcltk/libs//tcltk.so'' had status 1
```

```
##
## Attaching package: 'summarytools'

## The following object is masked from 'package:tibble':
##
##   view
```

```
sondage <- read_csv("Questions_SOC2206H21.csv")
```

```
##
## -- Column specification -----
## cols(
##   Timestamp = col_character(),
##   Sexe = col_character(),
##   `Niveau d'étude` = col_character(),
##   Filière = col_character(),
##   `Avez-vous déjà suivi des cours de méthodes quantitatives` = col_character(),
##   `Sur une échelle de 1 à 10, quel est votre niveau en méthodes quantitatives (simple calcul et raiso
##   `Considérons un groupe de TD pour lequel la moyenne des notes est égale à 10 et la variance est ég
##   `Quel(s) sujet(s) vous intéresse(nt) en Sociologie et pourquoi?` = col_character(),
##   `Donnez-moi le titre d'un texte ou article qui vous a marqué dans cette thématique.` = col_charact
##   `Quels sont vos objectifs après votre formation?` = col_character()
## )
```

```
sondage <- sondage %>%
  mutate(reponse_math_subjective = `Sur une échelle de 1 à 10, quel est votre niveau en méthodes quanti
         reponse_math_objective = `Considérons un groupe de TD pour lequel la moyenne des notes est éga
```

Relation entre math et sexe

```
freq(sondage$Sexe)
```

```
## Frequencies
## sondage$Sexe
## Type: Character
##
```

	Freq	% Valid	% Valid Cum.	% Total	% Total Cum.
Autre	1	3.45	3.45	3.45	3.45
Femme	15	51.72	55.17	51.72	55.17
Homme	12	41.38	96.55	41.38	96.55
Je ne souhaite pas le préciser	1	3.45	100.00	3.45	100.00
<NA>	0			0.00	100.00
Total	29	100.00	100.00	100.00	100.00

```
freq(sondage$reponse_math_subjective)
```

```
## Frequencies
## sondage$reponse_math_subjective
## Type: Numeric
##
##          Freq  % Valid  % Valid Cum.  % Total  % Total Cum.
## -----
##          1      2      6.90         6.90      6.90         6.90
##          2      2      6.90        13.79      6.90        13.79
##          3      3     10.34        24.14     10.34        24.14
##          4      1      3.45        27.59      3.45        27.59
##          5      4     13.79        41.38     13.79        41.38
##          6      5     17.24        58.62     17.24        58.62
##          7     10     34.48        93.10     34.48        93.10
##          8      1      3.45        96.55      3.45        96.55
##          9      1      3.45       100.00      3.45       100.00
##         <NA>      0              0.00      0.00       100.00
##        Total     29     100.00       100.00    100.00       100.00
```

```
freq(sondage$reponse_math_objective)
```

```
## Frequencies
## sondage$reponse_math_objective
## Type: Character
##
##                                     Freq  % Valid  % Valid Cum.  % T
## -----
##      Je ne sais pas ce que c'est que la médiane et la moyenne      4     13.79      13.79      1
##                                     la médiane augmente      1      3.45      17.24      3
##      la médiane augmente;la moyenne augmente      5     17.24      34.48      1
##      la médiane augmente;la moyenne baisse      1      3.45      37.93      3
##                                     la médiane baisse      1      3.45      41.38      3
##      la moyenne augmente      16     55.17      96.55      5
##      la moyenne baisse      1      3.45     100.00      3
##                                     <NA>      0              0.00      0
##                                     Total     29     100.00     100.00     10
```

On ne peut pas travailler aisément avec ces variables. Il faut les transformer. C'est un travail que vous serez amené à réaliser souvent. Cela prend dès fois des jugements subjectifs. Par exemple, on peut considérer que ceux/celles qui se donnent un score de moins de 5 sont "faibles en statistiques"

```
sondage <- sondage %>%
  mutate(resultat_math_subjective_2 = if_else(reponse_math_subjective <= 5, "Faible", "Fort"),
         resultat_math_objective_2 = if_else(reponse_math_objective == "la moyenne augmente", "Vrai", "Faux"),
         sexe_2 = case_when(
           Sexe == "Femme" ~ "Femme",
           Sexe == "Homme" ~ "Homme",
           Sexe == "Autre" | Sexe == "Je ne souhaite pas le préciser" ~ NA_character_
         ))

freq(sondage$sexe_2)
```

```
## Frequencies
## sondage$sexe_2
## Type: Character
```



```
##
##           Freq  % Valid  % Valid Cum.  % Total  % Total Cum.
## -----
##      Femme    15    55.56      55.56    51.72      51.72
##      Homme    12    44.44     100.00    41.38      93.10
##      <NA>      2      0.00     100.00     6.90     100.00
##      Total    29   100.00     100.00   100.00     100.00
```

```
freq(sondage$resultat_math_subjective_2)
```

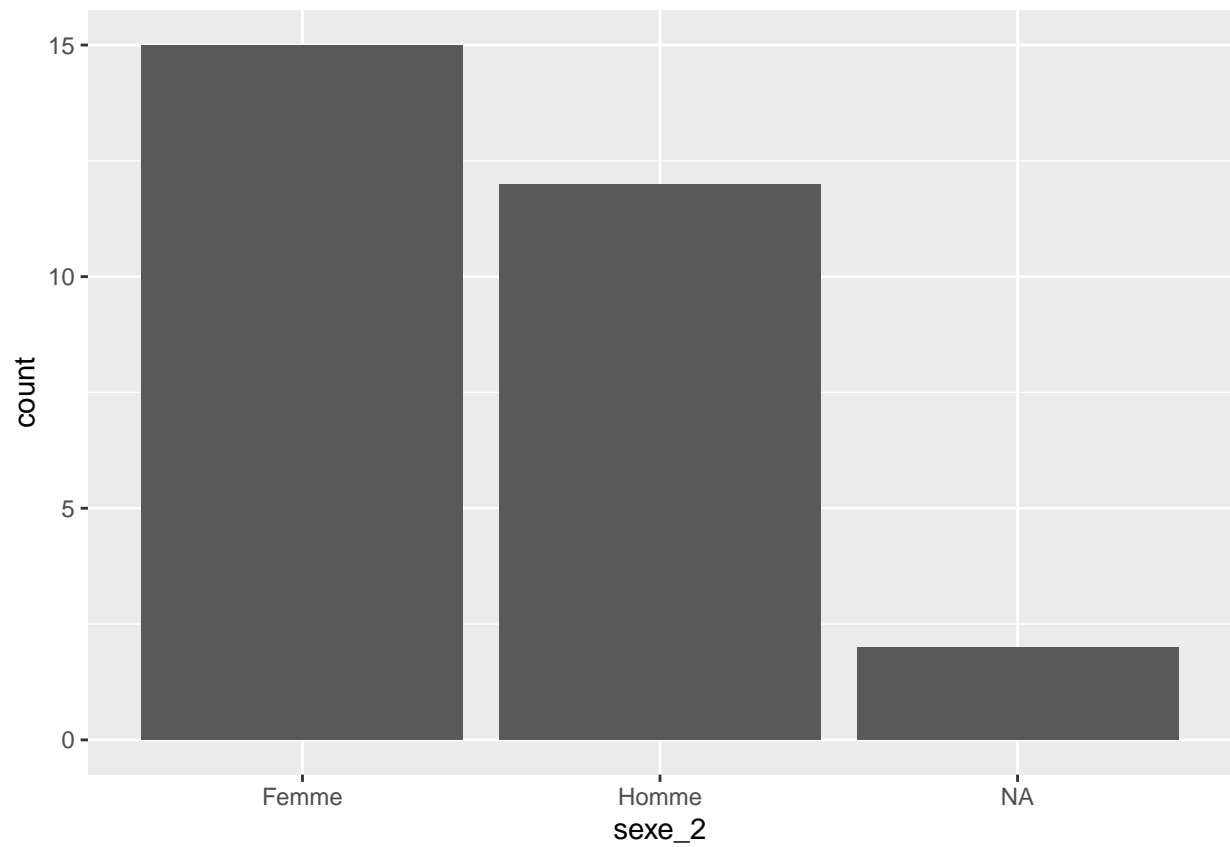
```
## Frequencies
## sondage$resultat_math_subjective_2
## Type: Character
##
##           Freq  % Valid  % Valid Cum.  % Total  % Total Cum.
## -----
##      Faible    12    41.38      41.38    41.38      41.38
##      Fort      17    58.62     100.00    58.62     100.00
##      <NA>       0      0.00     100.00     0.00     100.00
##      Total     29   100.00     100.00   100.00     100.00
```

```
freq(sondage$resultat_math_objective_2)
```

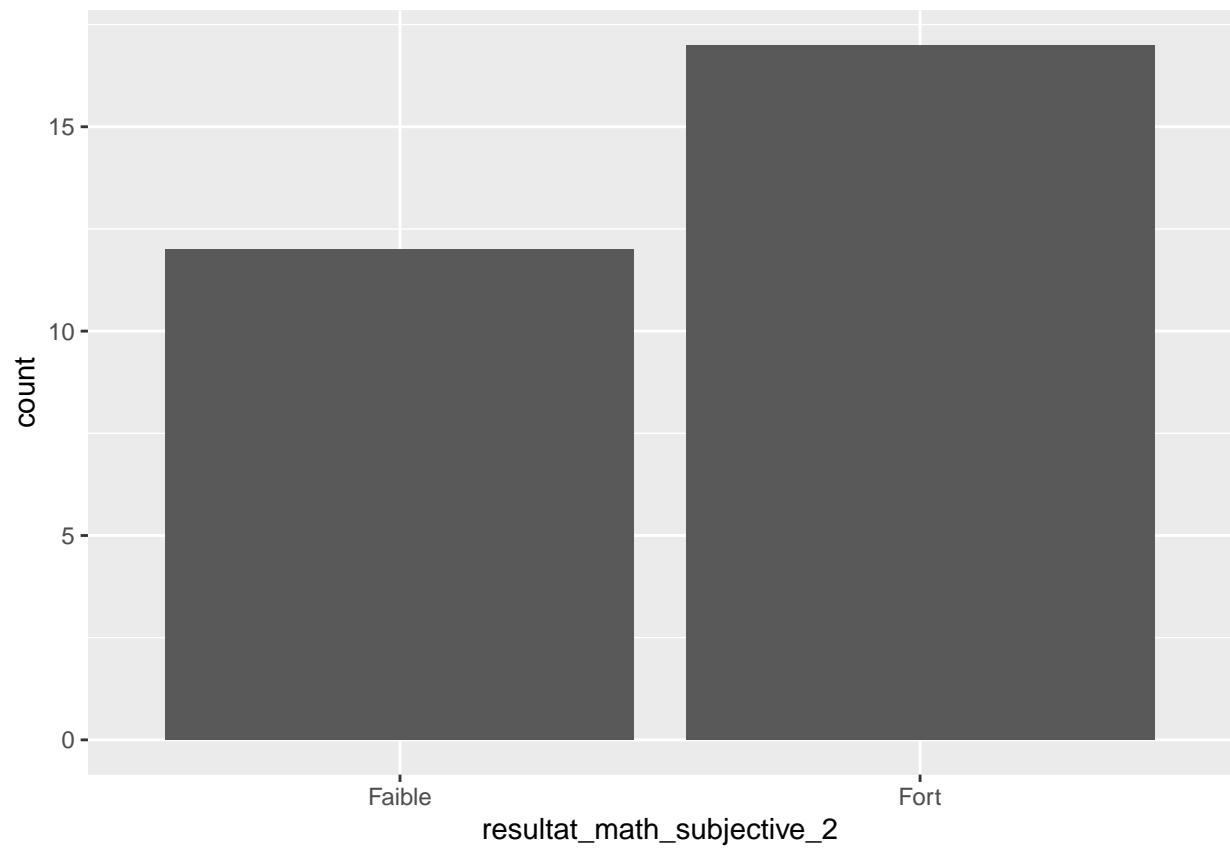
```
## Frequencies
## sondage$resultat_math_objective_2
## Type: Character
##
##           Freq  % Valid  % Valid Cum.  % Total  % Total Cum.
## -----
##      Faux      13    44.83      44.83    44.83      44.83
##      Vrai      16    55.17     100.00    55.17     100.00
##      <NA>       0      0.00     100.00     0.00     100.00
##      Total     29   100.00     100.00   100.00     100.00
```

Un graphique est toujours plus parlant qu'un tableau.

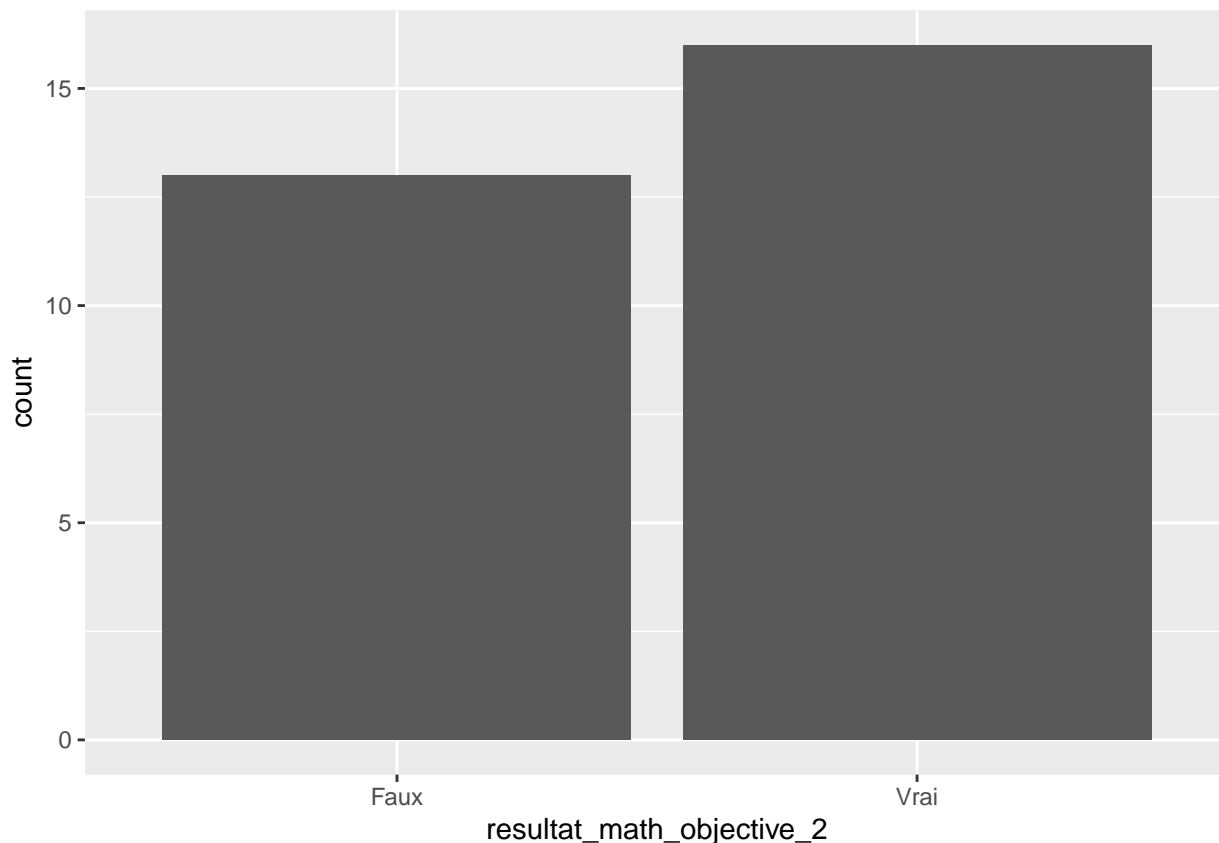
```
# Sexe
ggplot(sondage) +
  geom_bar(aes(x = sexe_2))
```



```
# Math subjectif  
ggplot(sondage) +  
  geom_bar(aes(x = resultat_math_subjective_2))
```



```
# Math objectif  
ggplot(sondage) +  
  geom_bar(aes(x = resultat_math_objective_2))
```



Maintenant, on peut vérifier la relation entre sexe et capacité en statistiques

```
ctable(sondage$sexe_2, sondage$resultat_math_subjective_2, "no")
```

```
## Cross-Tabulation
## sexe_2 * resultat_math_subjective_2
## Data Frame: sondage
##
## -----
##      resultat_math_subjective_2  Faible  Fort  Total
##  sexe_2
##    Femme                7      8    15
##    Homme                5      7    12
##    <NA>                 0      2     2
##    Total               12     17    29
## -----
```

On voit clairement qu'on ne peut pas prendre une décision ici. Que faire?

```
ctable(sondage$sexe_2, sondage$resultat_math_subjective_2, "r")
```

```
## Cross-Tabulation, Row Proportions
## sexe_2 * resultat_math_subjective_2
## Data Frame: sondage
##
## -----
##      resultat_math_subjective_2  Faible  Fort  Total
##  sexe_2
##    Femme                7 (46.7%)  8 ( 53.3%)  15 (100.0%)
```

```
##      Homme                5 (41.7%)      7 ( 58.3%)      12 (100.0%)
##      <NA>                 0 ( 0.0%)      2 (100.0%)      2 (100.0%)
##      Total                12 (41.4%)     17 ( 58.6%)     29 (100.0%)
## -----
```

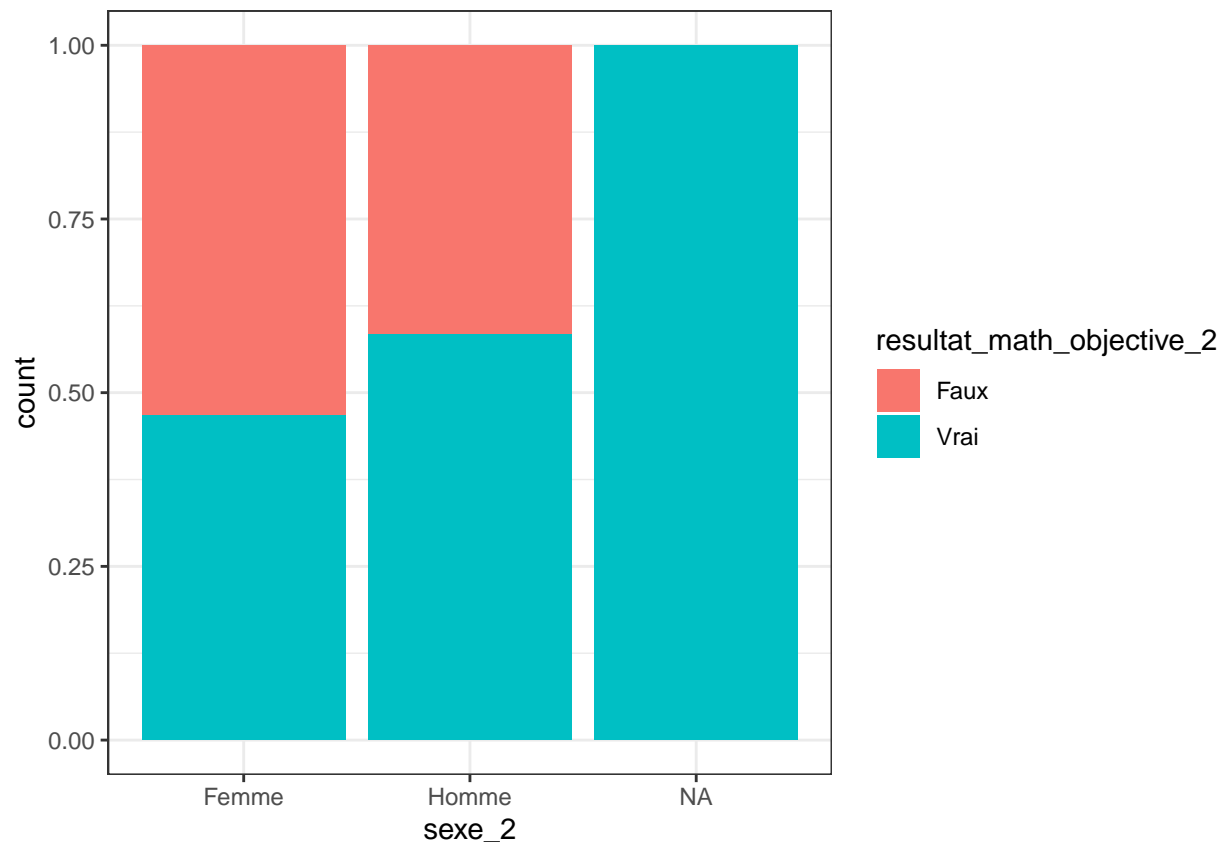
On voit ici qu'en ce qui concerne la variable subjective, les hommes semblent être plus forts que les femmes en statistiques, puisqu'il y a 58% des hommes qui estiment être fort contre 53% des femmes. Qu'en est-il de la variable objective.

```
ctable(sondage$sexe_2, sondage$resultat_math_objective_2, "r")
```

```
## Cross-Tabulation, Row Proportions
## sexe_2 * resultat_math_objective_2
## Data Frame: sondage
## -----
##      resultat_math_objective_2      Faux      Vrai      Total
##      sexe_2
##      Femme                8 (53.3%)      7 ( 46.7%)     15 (100.0%)
##      Homme                5 (41.7%)      7 ( 58.3%)     12 (100.0%)
##      <NA>                 0 ( 0.0%)      2 (100.0%)      2 (100.0%)
##      Total                13 (44.8%)     16 ( 55.2%)     29 (100.0%)
## -----
```

On peut aussi représenter ceci par un graphique

```
ggplot(sondage) +
  geom_bar(aes(x = sexe_2, fill = resultat_math_objective_2), position = "fill") +
  theme_bw()
```



D'autres questions qu'on se pose est de savoir si la relation qu'on observe ici est aussi valable dans la population générale des étudiant.es de sociologie. On doit faire dans ce cas un test statistique.

```
ctable(sondage$sexe_2, sondage$resultat_math_objective_2, chisq = TRUE)
```

```
## Warning in chisq.test(freq_table): Chi-squared approximation may be incorrect
## Cross-Tabulation, Row Proportions
## sexe_2 * resultat_math_objective_2
## Data Frame: sondage
##
##
## -----
##      resultat_math_objective_2      Faux      Vrai      Total
##  sexe_2
##  Femme      8 (53.3%)      7 ( 46.7%)     15 (100.0%)
##  Homme      5 (41.7%)      7 ( 58.3%)     12 (100.0%)
##  <NA>        0 ( 0.0%)      2 (100.0%)      2 (100.0%)
##  Total     13 (44.8%)     16 ( 55.2%)     29 (100.0%)
## -----
##
## -----
##  Chi.squared   df   p.value
## -----
##      2.112      2   0.3478
## -----
```

On voit que si on généralisait ce résultat à l'ensemble de la population, il n'est pas prouvé que les femmes sont moins fortes que les hommes en statistiques.

Tout ce que nous venons de dire, c'est qu'une partie des statistiques. Les statistiques commencent avec les questionnements que vous vous posez pour comprendre vos résultats. C'est la différence entre une sociologue qui utilise les statistiques et un analyste des données. Le/la Sociologie est outillé pour expliquer les résultats.

Voici quelques questions que vous devez vous posez alors:

- Ai-je bien mesuré ce que je veux vraiment mesuré (validité vs fiabilité)
- Quel est l'effet des gens qui n'ont pas répondu aux questionnaires? Est-ce que les gens qui n'ont pas répondu sont juste un échantillon aléatoire de tout le monde, ou s'agit-il de gens qui ont quelques choses de caractéristiques par rapport à ce que nous étudions?
- Validité versus fiabilité

Rappelez-vous ce que nous essayons de mesurer: compétence en statistique. Nous l'avons mesuré ici avec deux variables: - subjectif - objectif

Quel problème pensez-vous avoir pour ces différents mesures

- subjectif: Si les femmes sont plus gênées à répondre, elles peuvent être amenées à baisser leur compétence
- objectif:

Y'a-t-il une relation entre les deux mesures?

```
ctable(sondage$resultat_math_subjective_2, sondage$resultat_math_objective_2, "no")
```

```
## Cross-Tabulation
## resultat_math_subjective_2 * resultat_math_objective_2
## Data Frame: sondage
##
## -----
```

##	resultat_math_objective_2	Faux	Vrai	Total
##	resultat_math_subjective_2			
##	Faible	6	6	12
##	Fort	7	10	17
##	Total	13	16	29
##	-----	-----	-----	-----

Raison pour laquelle il est important d'avoir plusieurs mesures d'un même concept.

Et si l'explication est ailleurs est juste une question de confiance en soi?

Ce sont ces autres questions qui rendent les statistiques passionnantes. J'espère que insuffler un peu de cette passion en vous si vous ne l'avez pas encore.