Cette enquête est-elle croyable ?

Le rapport de plausibilité SMART, sur une enquête qui échoue à deux de ses contrôles. Un écart-type de 1,22, une équipe mesurant 0,6 score z trop bas, 69 % de ses tailles sur un demi-centimètre, et un quart des âges sur une année entière.

Pierre Robentz Cassion

Leçon 5 sur 8

Analyse nutritionnelle : PCIMA et SMART — Juger une enquête

Ce que couvre cette leçon

- Une enquête est acceptée ou ne l'est pas
- Les contrôles
- Les signalements
- L'écart-type
- La préférence de chiffres
- L'attraction des âges
- Sex-ratio et biais d'équipe
- Le verdict
- Rédigez sans accuser personne
- La suite

Une enquête est acceptée ou ne l'est pas

- La méthodologie SMART fait quelque chose d'inhabituel et de précieux — elle publie un ensemble de contrôles qu'une enquête doit passer avant que sa prévalence soit employée, et ces contrôles se calculent sur les données de l'enquête elle-même.

Les contrôles

Contrôle	Ce qu'il détecte	Acceptable
Valeurs signalées	Mesures impossibles ou extrêmes	sous 2,5 % environ
Écart-type du WHZ	Erreur de mesure gonflant la dispersion	0,8 à 1,2
Préférence de chiffres	Arrondi plutôt que lecture	faible, et homogène entre équipes
Attraction des âges	Âge estimé plutôt que documenté	faible
Sex-ratio	Biais de sélection dans qui a été mesuré	proche de 1,0
Biais d'équipe	Une équipe qui mesure autrement	moyennes proches entre équipes

Les signalements — En Python

```
computable = smart["whz"].notna()
flagged = computable & ~smart["whz"].between(-5, 5)

print(f"{computable.sum()} computable, {flagged.sum()} flagged "
      f"({flagged.sum() / computable.sum():.1%})")
```

Les signalements — En R

```r
smart |>
  filter(!is.na(whz)) |>
  summarise(n = n(), flagged = sum(!between(whz, -5, 5)),
            rate = flagged / n)
```

L'écart-type — En Python

```
analysable = smart.loc[smart["whz"].between(-5, 5), "whz"]
print(f"n = {len(analysable)}, mean = {analysable.mean():.2f}, "
      f"sd = {analysable.std():.2f}")
```

L'écart-type — En R

```
smart |> filter(between(whz, -5, 5)) |> summarise(n = n(), sd = sd(whz))
```

L'écart-type

- 1,22 — Hors de la bande acceptable, à sa limite haute
- **L'erreur de mesure.** Une erreur aléatoire sur le poids ou la taille élargit la distribution sans en déplacer le...
- **Une population réellement hétérogène.** Deux sous-populations très différentes échantillonnées ensemble.
- **L'erreur de saisie.** Chiffres inversés, mauvaises unités, enregistrements mélangés.
- Un écart-type gonflé fait monter la prévalence — Une distribution plus large place davantage d'enfants au-delà d'un...

La préférence de chiffres — En Python

```
smart["terminal"] = (smart["height_cm"] * 10).round() % 10
by_team = (
    smart.assign(rounded=smart["terminal"].isin([0, 5]))
    .groupby("team")
    .agg(rounded=("rounded", "mean"), n=("rounded", "size"))
)
print((by_team["rounded"] * 100).round(0))
```

La préférence de chiffres — En R

```r
smart |>
  mutate(rounded = round(height_cm * 10) %% 10 %in% c(0, 5)) |>
  summarise(share = mean(rounded), n = n(), .by = team)
```

La préférence de chiffres

Équipe	Tailles finissant par ,0 ou ,5
1	18 %
2	**69 %**
3	18 %
4	23 %

La préférence de chiffres

- Échec — et cela n'exige aucun test de signification — la discipline d'étalonnage du cours d'évaluation s'applique aux...

L’attraction des âges — En Python

```
ages = smart["age_months"].dropna()
whole_years = (ages % 12 == 0).mean()

print(f"{whole_years:.0%} of ages on an exact whole year")
print(ages.value_counts().reindex([22, 23, 24, 25, 26]).to_dict())
```

L'attraction des âges — En R

```r
smart |>
  filter(!is.na(age_months)) |>
  summarise(whole_years = mean(age_months %% 12 == 0))
```

L'attraction des âges

- 24 % sur une année entière, contre environ 9 % attendus — Soixante-six enfants à exactement 24 mois, contre 15 et 19 de...
- Échec — et la cause n'est presque jamais la négligence — c'est que les actes de naissance ne sont pas universels et...

Sex-ratio et biais d'équipe — En Python

```
ratio = (smart["sex"] == "m").sum() / (smart["sex"] == "f").sum()
print(f"sex ratio m:f = {ratio:.2f}")

by_team = (
    smart[smart["whz"].between(-5, 5)]
    .groupby("team")
    .agg(mean_whz=("whz", "mean"), n=("whz", "size"))
)
print(by_team.round(2))
```

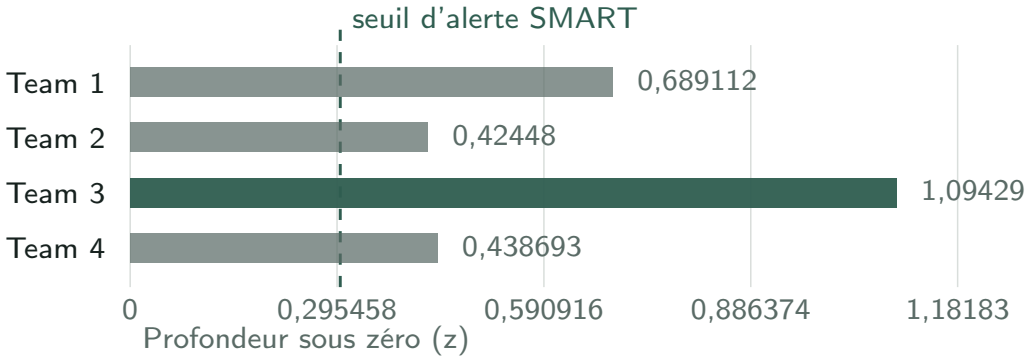
Sex-ratio et biais d'équipe — En R

```
smart |>
  filter(between(whz, -5, 5)) |>
  summarise(mean_whz = mean(whz), n = n(), .by = team)
```

Sex-ratio et biais d'équipe

Équipe	WHZ moyen	n
1	-0,69	204
2	-0,42	228
3	**-1,09**	224
4	-0,44	196

Sex-ratio et biais d'équipe



Score z poids-pour-taille moyen par équipe de mesure, tracé en profondeur sous zéro. Les grappes ayant été attribuées aux équipes indépendamment de l'état nutritionnel, un écart de cette ampleur entre équipes est un défaut de mesure et non une différence réelle entre populations.

Sex-ratio et biais d'équipe — En Python

```
gam_by_team = (
    smart[smart["whz"].between(-5, 5)]
    .assign(gam=lambda d: (d["whz"] < -2) | (d["oedema"] == True))
    .groupby("team")["gam"].mean()
)
print((gam_by_team * 100).round(1))
```

Sex-ratio et biais d'équipe — En R

```
smart |>
  filter(between(whz, -5, 5)) |>
  summarise(gam = mean(whz < -2 | oedema), .by = team)
```

Le verdict — En Python

```
plausibility = pd.DataFrame({
    "check": ["Flagged records", "SD of WHZ", "Digit preference",
              "Age heaping", "Sex ratio", "Team bias"],
    "value": ["1.4%", "1.22", "69% (team 2)", "24% whole years", "1.05",
              "-0.42 to -1.09"],
    "acceptable": ["<2.5%", "0.8-1.2", "even across teams", "low", "~1.0",
                   "close across teams"],
    "verdict": ["pass", "fail", "fail", "fail", "pass", "fail"],
})
print(plausibility)
```

Le verdict — En R

```
tibble::tribble(
  ~check,               ~value,            ~verdict,
  "Flagged records",    "1.4%",            "pass",
  "SD of WHZ",          "1.22",            "fail",
  "Digit preference",   "69% (team 2)",    "fail",
  "Age heaping",        "24%",             "fail",
  "Sex ratio",          "1.05",            "pass",
  "Team bias",          "-0.42 to -1.09",  "fail"
)
```

Le verdict

- **Rejeter l'enquête.** Défendable avec quatre échecs, et coûteux — six semaines et un budget, perdus.
- **Accepter avec des limites énoncées.** Rapporter la prévalence avec le tableau de plausibilité à côté et une mention…
- **Réanalyser en excluant l'équipe 3.** Défendable seulement si vous le dites bien en vue, et cela coûte un quart de…

Rédigez sans accuser personne

N'écrivez pas	Écrivez
« Les mesures de l'équipe 2 ne sont pas fiables »	« Les tailles de l'équipe 2 se terminent par ,0 ou ,5 dans 69 % des cas contre 18 à 23 % pour les autres équipes, ce qui est compatible avec une lecture de la toise au demi-centimètre »
« L'équipe 3 a falsifié des mesures »	« Le poids-pour-taille moyen de l'équipe 3 est de 0,4 à 0,7 score z sous celui des autres équipes. Les grappes ayant été attribuées indépendamment de l'état nutritionnel, cela est surtout compatible avec une différence de mesure ou d'étalonnage »
« Les âges sont inexacts »	« 24 % des âges tombent sur une année entière exacte contre 9 % attendus, ce qui indique que l'âge a été estimé plutôt que documenté pour une part substantielle des enfants »

La suite

- Vous savez si l'enquête est croyable et sous quelles réserves.

Lire la leçon complète, avec le code exécutable

https://data-analysis.cassion.dev/fr/cours/nutrition-analysis-cmam-smart/nutrition-05-plausibility