Can this survey be believed?

The SMART plausibility report, on a survey that fails two of its checks. A standard deviation of 1.22, one team measuring 0.6 z-scores low, 69% of its heights on a half centimetre, and a quarter of ages on a whole year.

Pierre Robentz Cassion

Lesson 5 of 8

Nutrition Analysis: CMAM and SMART — Judging a survey

What this lesson covers

- A survey is accepted or it is not
- The checks
- Flags
- The standard deviation
- Digit preference
- Age heaping
- Sex ratio and team bias
- The verdict
- Write it up without accusing anyone
- What comes next

A survey is accepted or it is not

- The SMART methodology does something unusual and valuable: it publishes a set of checks a survey must pass before its prevalence is used, and the checks are computable from the survey's own data.

The checks

Check	What it detects	Acceptable
Flagged records	Impossible or extreme measurements	under about 2.5%
Standard deviation of WHZ	Measurement error inflating the spread	0.8 to 1.2
Digit preference	Rounding rather than reading	low, and even across teams
Age heaping	Age estimated rather than documented	low
Sex ratio	Selection bias in who was measured	near 1.0
Team bias	One team measuring differently	means close across teams

Flags — In 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%})")
```

Flags — In R

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

The standard deviation — In Python

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

The standard deviation — In R

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

The standard deviation

- 1.22 — Outside the acceptable band, at the top edge
- **Measurement error.** Random error in weight or height widens the distribution without changing its centre. The...
- **A genuinely heterogeneous population.** Two very different sub-populations sampled together.
- **Entry error.** Transposed digits, wrong units, mixed-up records.
- An inflated SD raises the prevalence — A wider distribution puts more children past a fixed cut-off, so GAM rises...

Digit preference — In 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))
```

Digit preference — In R

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

Digit preference

Team	Heights ending .0 or .5
1	18%
2	**69%**
3	18%
4	23%

Digit preference

- Fails — and it needs no significance test — the DQA course's calibration discipline applies to marginal results, and...

Age heaping — In 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())
```

Age heaping — In R

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

Age heaping

- 24% on a whole year, against about 9% expected — Sixty-six children at exactly 24 months, against 15 and 19 either side
- Fails — and the cause is almost never carelessness — it is that birth records are not universal and a carer asked a. . .

Sex ratio and team bias — In 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))
```

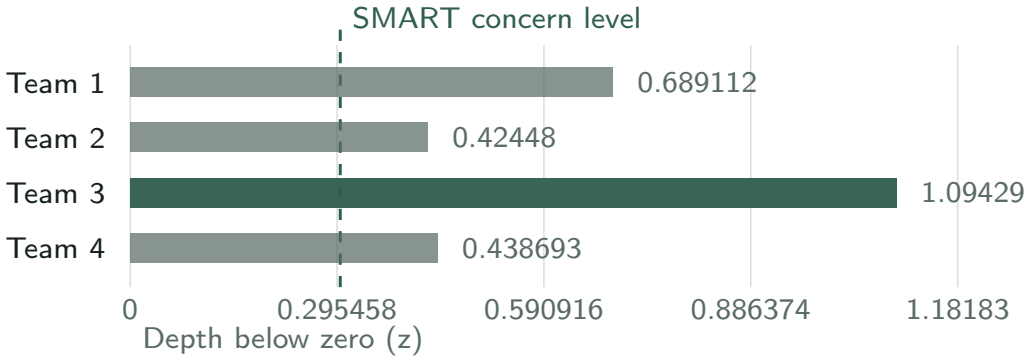
Sex ratio and team bias — In R

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

Sex ratio and team bias

Team	Mean WHZ	n
1	-0.69	204
2	-0.42	228
3	**-1.09**	224
4	-0.44	196

Sex ratio and team bias



Mean weight-for-height z-score by measurement team, plotted as depth below zero. Clusters were assigned to teams independently of nutrition status, so a spread of this size between teams is a measurement fault rather than a real difference between populations.

Sex ratio and team bias — In 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 and team bias — In R

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

The verdict — In 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)
```

The verdict — In 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"
)
```

The verdict

- **Reject the survey.** Defensible with four failures, and expensive — six weeks and a budget, gone.
- **Accept with stated limitations.** Report the prevalence with the plausibility table beside it and an explicit...
- **Reanalyse excluding team 3.** Defensible only if you say so prominently, and it costs a quarter of the sample and...

Write it up without accusing anyone

Do not write	Write
"Team 2's measurements are unreliable"	"Team 2's height readings end in .0 or .5 in 69% of cases against 18-23% for other teams, consistent with reading the board to the nearest half centimetre"
"Team 3 falsified measurements"	"Team 3's mean weight-for-height is 0.4 to 0.7 z-scores below the other teams. Clusters were assigned independently of nutrition status, so this is most consistent with a measurement or calibration difference"
"Ages are inaccurate"	"24% of ages fall on an exact whole year against 9% expected, indicating age was estimated rather than documented for a substantial share of children"

What comes next

- You know whether the survey can be believed and with what caveats.

Read the full lesson, with runnable code

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