

$p = 0,023$, et un tiers d'une journée

Les garçons sont présents à 88,77 % et les filles à 88,21 %. Sur 68 267 pointages, c'est significatif à $p = 0,023$. C'est aussi une différence d'un tiers de journée scolaire par enfant et par trimestre, sur laquelle aucun programme n'agirait et aucun ne devrait.

Pierre Robentz Cassion

Leçon 4 sur 8

Statistiques appliquées aux programmes — Comparer deux groupes

Ce que couvre cette leçon

- Un résultat significatif sur lequel personne ne devrait agir
- Convertissez-la dans l'unité de la décision
- Pourquoi p est devenu petit : c'était le n
- Rapportez une taille d'effet à côté de chaque p
- Fixez le seuil avant de tester
- Les quatre cas
- Rapportez-le en entier
- La suite

Un résultat significatif sur lequel personne ne devrait agir — En Python (suite)

```
import pandas as pd
import numpy as np
from statsmodels.stats.proportion import proportions_ztest

attendance = pd.read_csv("school-attendance-2024.v1.csv")
enrolment = pd.read_csv("school-enrolment-2024.v1.csv")
current = enrolment[enrolment["school_year"] == 2024]

MARKS = {"true": True, "Y": True, "false": False, "N": False}
marked = attendance[attendance["present"].isin(MARKS)].copy()
marked["attended"] = marked["present"].map(MARKS)
marked = marked.merge(current[["student_id", "sex"]], on="student_id")

by_sex = marked.groupby("sex")["attended"].agg(["sum", "size"])
print(by_sex)
```

Un résultat significatif sur lequel personne ne devrait agir — En Python (suite)

```
stat, p = proportions_ztest(by_sex["sum"], by_sex["size"])  
print(f"z = {stat:.2f}, p = {p:.4f}")
```

Un résultat significatif sur lequel personne ne devrait agir — En R

```
library(dplyr)
```

```
marked |> summarise(k = sum(attended), n = n(), .by = sex)
```

```
prop.test(c(29655, 30749), c(33408, 34859))
```

Un résultat significatif sur lequel personne ne devrait agir

Groupe	Présences	Pointages	Taux
Garçons	29 655	33 408	88,77 %
Filles	30 749	34 859	88,21 %

Un résultat significatif sur lequel personne ne devrait agir

- Différence 0,56 point de pourcentage. $z = 2,28$, $p = 0,023$

Convertissez-la dans l'unité de la décision — En Python

```
term_days = marked["attendance_date"].nunique()
gap = 0.0056
print(f"{term_days} school days in the term")
print(f"0.56 points = {gap * term_days:.2f} days per student per term")
```

Convertissez-la dans l'unité de la décision — En R

The statistic is in percentage points. The decision is in days.

Convertissez-la dans l'unité de la décision

- Soixante jours de classe, donc 0,56 % vaut 0,34 jour — environ un tiers de journée par enfant et par trimestre
- Convertissez toujours un effet dans l'unité où se prend la décision — Les points de pourcentage sont l'unité de...

Pourquoi p est devenu petit : c'était le n — En Python

```
for scale in (0.05, 0.2, 1.0):  
    k = (by_sex["sum"] * scale).round().astype(int)  
    n = (by_sex["size"] * scale).round().astype(int)  
    stat, p = proportions_ztest(k, n)  
    print(f"n = {n.sum():>6,}    difference {k[1]/n[1] - k[0]/n[0]:+.2%}    p = {p:.3f}")
```

Pourquoi p est devenu petit : c'était le n — En R

Same proportions, smaller n , and the p -value walks back across 0.05.

Pourquoi p est devenu petit : c'était le n

Échantillon	Différence	p
5 % des pointages (n=3 413)	+0,62 %	0,570
20 % (n=13 654)	+0,55 %	0,314
100 % (n=68 267)	+0,56 %	0,023

Pourquoi p est devenu petit : c'était le n

- L'effet n'a jamais changé. Seul l'échantillon a changé — Un p est une affirmation sur la force de la preuve, non sur. . .
- Un p seul ne peut donc jamais vous dire si quelque chose compte — Il vous dit si vous pouvez écarter zéro, et zéro est. . .

Rapportez une taille d'effet à côté de chaque p

Mesure	Pour	Se lit comme
Différence en points de pourcentage	Deux proportions	La réponse directe
Rapport de risques	Deux proportions, événements rares	Combien de fois plus probable
d de Cohen	Deux moyennes	Écarts-types de séparation

Rapportez une taille d'effet à côté de chaque p — En Python

```
p_boys, p_girls = by_sex["sum"] / by_sex["size"]  
print(f"difference: {p_boys - p_girls:+.2%} points")  
print(f"risk ratio: {p_boys / p_girls:.3f}")
```

Rapportez une taille d'effet à côté de chaque p — En R

Two lines. There is no excuse for a p-value with no effect size beside it.

Rapportez une taille d'effet à côté de chaque p

- Rapport de risques 1,006 — Les garçons sont présents 0,6 % plus souvent en termes relatifs — une autre façon de dire le. . .

Fixez le seuil avant de tester — En Python

```
MEANINGFUL = 0.03      # 3 points of attendance, about 1.8 days a term
```

```
observed = p_boys - p_girls
```

```
print(f"observed {observed:+.2%}, threshold {MEANINGFUL:.0%}")
```

```
print(f"large enough to act on: {abs(observed) >= MEANINGFUL}")
```

Fixez le seuil avant de tester — En R

Write the threshold in the analysis plan, not after seeing the result.

Fixez le seuil avant de tester

- Trois points de présence font environ 1,8 jour par trimestre — c'est-à-dire l'échelle à laquelle un programme de suivi. . .

Les quatre cas — En Python

```
cases = pd.DataFrame([
    ("significant, large",      "report it",      "disability gap, -19.4 pts"),
    ("significant, small",     "report the effect size, say it is small", "attendance by sex,
    0.56 pts"),
    ("not significant, large", "report the interval; underpowered", "over-age by sex, +5.5
    pts"),
    ("not significant, small", "report as no evidence of a difference", "-"),
], columns=["case", "what to do", "example in this course"])
print(cases)
```

Les quatre cas — En R

Four quadrants, and only one of them is "publish the p-value".

- La troisième ligne est celle qu'on traite mal — L'écart de retard d'âge selon le sexe valait 5,5 points avec $p = 0,066, \dots$

Rapportez-le en entier — Exemple

Attendance by sex, February to April 2024

Boys 88.77% 29,655 of 33,408 marks

Girls 88.21% 30,749 of 34,859 marks

Difference +0.56 percentage points ($p = 0.023$, risk ratio 1.006).

Equivalent to 0.34 school days per child per term. Below the 3-point threshold set in the analysis plan and not reported as a programme finding. The p-value is small because the analysis has 68,267 marks, not because the difference is large.

- La dernière phrase est ce qui rend le bloc honnête — Un lecteur qui ne voit que « $p = 0,023$ » supposera que la...

- Chaque test jusqu'ici portait sur une comparaison.

Lire la leçon complète, avec le code exécutable

`https://data-analysis.cassion.dev/fr/cours/
applied-statistics-for-programmes/stat-04-significant-and-irrelevant`