

Une proportion de zéro avec un intervalle jusqu'à 24 %

Douze enfants, aucun en retard d'âge. L'intervalle du manuel dit 0 % à 0 %. Le bon dit 0 % à 24,3 %, et l'écart entre ces deux réponses décide si vous agiriez sur cette cellule.

Pierre Robentz Cassion

Leçon 2 sur 8

Statistiques appliquées aux programmes — Regardez-la d'abord

Ce que couvre cette leçon

- Pourquoi une estimation ponctuelle n'est pas un résultat
- Wilson, non Wald
- Lisez la largeur, pas seulement les bornes
- L'écrire dans une phrase
- Quand l'intervalle change la décision
- Trois intervalles que ce cours ne calculera pas ainsi
- Rapportez-le comme une fourchette
- La suite

Pourquoi une estimation ponctuelle n'est pas un résultat — En Python

```
import pandas as pd
import numpy as np

wash = pd.read_csv("wash-household-survey-2024.v1.csv")
open_defecation = wash["sanitation_facility"].eq("open-defecation")
k, n = open_defecation.sum(), len(wash)
print(f"{k}/{n} = {k / n:.1%}")
```

Pourquoi une estimation ponctuelle n'est pas un résultat — En R

```
library(dplyr)
wash |> summarise(k = sum(sanitation_facility == "open-defecation"), n = n())
```

Pourquoi une estimation ponctuelle n'est pas un résultat

- 319 sur 2 403 — 13,3 % — L'intervalle là-dessus est étroit, parce que l'échantillon est grand

Wilson, non Wald — En Python

```
def wald(k, n, z=1.96):  
    p = k / n  
    se = np.sqrt(p * (1 - p) / n)  
    return p - z * se, p + z * se  
  
def wilson(k, n, z=1.96):  
    p = k / n  
    denom = 1 + z**2 / n  
    centre = (p + z**2 / (2 * n)) / denom  
    half = z * np.sqrt(p * (1 - p) / n + z**2 / (4 * n**2)) / denom  
    return centre - half, centre + half  
  
for k, n in [(0, 12), (1, 20), (3, 11)]:  
    print(f"{k}/{n} = {k/n:.1%}")  
    print(f"    Wald    [{wald(k, n)[0]:.1%}, {wald(k, n)[1]:.1%}]")  
    print(f"    Wilson [{wilson(k, n)[0]:.1%}, {wilson(k, n)[1]:.1%}]")
```

Wilson, non Wald — En R

```
# R has this built in and it is the Wilson interval by default.  
prop.test(0, 12)$conf.int  
binom.test(1, 20)$conf.int
```

Wilson, non Wald

Cellule	Wald	Wilson
0 sur 12	[0,0 %, 0,0 %]	[0,0 %, 24,3 %]
1 sur 20	[−4,6 %, 14,6 %]	[0,9 %, 23,6 %]
3 sur 11	[1,0 %, 53,6 %]	[9,7 %, 56,6 %]

- Wald affirme qu'une cellule sans événement n'a aucune incertitude, et il produit des probabilités négatives — Les deux...
- Employez Wilson — ou `prop.test` en R et `statsmodels.stats.proportion.proportion_confint(method="wilson")` en Python

Lisez la largeur, pas seulement les bornes — En Python

```
cases = [  
    ("Open defecation", 319, 2403),  
    ("Cholera case fatality", 38, 974),  
    ("Cholera CFR, Nord district", 24, 381),  
    ("Referral completion, disability reported", 54, 202),  
    ("Over-age, grade 6", 67, 132),  
]  
  
for name, k, n in cases:  
    lo, hi = wilson(k, n)  
    print(f"{name:42} {k/n:6.1%} [{lo:.1%}, {hi:.1%}] width {hi-lo:.1%}")
```

Lisez la largeur, pas seulement les bornes — En R

Same five, and the width is the column to read.

Lisez la largeur, pas seulement les bornes

Estimation	Valeur	Intervalle à 95 %	Largeur
Défécation à l'air libre	13,3 %	12,0–14,7 %	2,7 pts
Létalité du choléra	3,9 %	2,9–5,3 %	2,5 pts
Létalité, district Nord	6,3 %	4,3–9,2 %	4,9 pts
Aboutissement, handicap signalé	26,7 %	21,1–33,2 %	12,1 pts
Retard d'âge, sixième année	50,8 %	42,3–59,1 %	16,8 pts

Lisez la largeur, pas seulement les bornes

- Le retard d'âge en sixième vaut 50,8 % et pourrait se situer entre 42 % et 59 % — Le cours d'éducation a rapporté ce. . .
- Deux choses gouvernent la largeur et une seule dépend de vous — La taille d'échantillon, décidée par le plan ; et la. . .

L'écrire dans une phrase

- Pas ceci — « 50,8 % des élèves de sixième sont en retard d'âge
- Ni ceci — « 50,8 % (IC à 95 % 42,3–59,1) des élèves de sixième sont en retard d'âge
- Ceci — « Entre quatre et six élèves de sixième sur dix sont en retard d'âge (50,8 %, IC à 95 % 42,3–59,1, n=132)
- Mettez l'intervalle en mots d'abord, puis donnez les nombres — Les mots sont ce que lit un non-analyste et ils portent. . .

L'écrire dans une phrase — En Python

```
def sentence(label, k, n):  
    lo, hi = wilson(k, n)  
    return (f"{label}: {k/n:.1%} (95% CI {lo:.1%} to {hi:.1%}, n={n})")  
  
print(sentence("Grade 6 over-age", 67, 132))
```

L'écrire dans une phrase — En R

Generate the string in code so the number and its interval cannot drift apart.

Quand l'intervalle change la décision — En Python

```
lo, hi = wilson(38, 974)
threshold = 0.01                                # Sphere: cholera CFR below 1%
print(f"CFR {38/974:.1%}, interval [{lo:.1%}, {hi:.1%}]")
print(f"entire interval above the {threshold:.0%} threshold: {lo > threshold}")
```

Quand l'intervalle change la décision — En R

```
prop.test(38, 974)$conf.int
```

Quand l'intervalle change la décision

- La létalité du choléra vaut 3,9 % avec un intervalle de 2,9 % à 5,3 %, et le seuil Sphère vaut 1 % — Toute valeur de...
- C'est là qu'on peut agir sur une estimation ponctuelle : quand tout l'intervalle dit la même chose — Là où il enjambe...

Trois intervalles que ce cours ne calculera pas ainsi

- Une proportion issue d'un échantillon en grappes — Le cours sur les enquêtes a établi qu'un plan en grappes élargit. . .
- Une médiane ou une moyenne asymétrique — Les intervalles présentés ici sont pour des proportions
- Un effectif sans dénominateur — « 142 cas ont été signalés » n'a pas d'intervalle, car ce n'est l'estimation de rien

Rapportez-le comme une fourchette — Exemple

Over-age enrolment, grade 6

50.8% 95% CI 42.3 to 59.1, n = 132

Between four and six students in ten. The interval is wide because grade 6 holds 132 students; a difference of five points against another grade would not be distinguishable at this sample size.

Rapportez-le comme une fourchette

- La dernière phrase est la plus utile — Elle dit d'avance au lecteur quelles comparaisons ce nombre peut porter, ce qui. . .

- Un intervalle dit à quel point un nombre est incertain.

Lire la leçon complète, avec le code exécutable

`https://data-analysis.cassion.dev/fr/cours/
applied-statistics-for-programmes/stat-02-the-interval`