

L'intervalle, et la phrase qui l'entoure

Pourquoi l'intervalle des manuels donne 0,88 % pour une proportion qui ne peut pas descendre sous zéro, l'intervalle logit qui corrige cela, les 72 degrés de liberté dont vous disposez, et pourquoi deux intervalles qui se chevauchent ne signifient pas l'absence de différence.

Pierre Robentz Cassion

Leçon 7 sur 8

Analyse d'enquêtes, échantillonnage et pondération — Ce que vous avez le droit de publier

Ce que couvre cette leçon

- À quoi sert l'intervalle
- L'intervalle des manuels, et où il échoue
- L'intervalle logit
- Les degrés de liberté dont vous disposez réellement
- Des intervalles qui se chevauchent ne constituent pas un test
- Arrondis et fausse précision
- La phrase
- Lire un intervalle face à un seuil
- La suite

À quoi sert l'intervalle

- Une estimation d'enquête est une conjecture sur une population, faite à partir d'un échantillon.

L'intervalle des manuels, et où il échoue — Exemple

$p \pm z * se$

L'intervalle des manuels, et où il échoue — En Python

```
p, se = 0.0217, 0.0066  
print(f"Wald: {p - 1.96 * se:.2%} to {p + 1.96 * se:.2%}")
```

L'intervalle des manuels, et où il échoue — En R

```
p <- 0.0217; se <- 0.0066  
c(p - 1.96 * se, p + 1.96 * se)
```

L'intervalle logit — En Python

```
import math

logit = math.log(p / (1 - p))
se_logit = se / (p * (1 - p))
low, high = (logit - 1.96 * se_logit, logit + 1.96 * se_logit)

print(f"logit: {1 / (1 + math.exp(-low)):.2%} to {1 / (1 + math.exp(-high)):.2%}")
```

L'intervalle logit — En R

```
library(survey)  
svyciprop(~I(child_muac_mm < 115), child_design, method = "logit")
```

- 1,19 % à 3,92 % — Asymétrique — plus étendu au-dessus de l'estimation qu'en dessous — ce qui est la forme correcte pour...

Les degrés de liberté dont vous disposez réellement — Exemple

$df = \text{number of PSUs} - \text{number of strata}$

Les degrés de liberté dont vous disposez réellement — En Python

```
clusters = survey["ea_id"].nunique()
strata = survey["stratum"].nunique()
print(f"{clusters} clusters, {strata} strata, df = {clusters - strata}")
```

Les degrés de liberté dont vous disposez réellement — En R

```
c(clusters = n_distinct(survey$ea_id), strata = n_distinct(survey$stratum))  
degf(design)
```

Les degrés de liberté dont vous disposez réellement

- Employez le multiplicateur t , et rapportez les degrés de liberté — Cela ne coûte rien et cela dit au lecteur combien de...

Des intervalles qui se chevauchent ne constituent pas un test — En Python

```
for stratum in ["urban", "rural-accessible", "rural-remote"]:  
    print(stratum)    # displaced households
```

Des intervalles qui se chevauchent ne constituent pas un test

Strate	Déplacés	Intervalle 95 %
Urbain	15,6 %	12,0 – 19,2 %
Rural accessible	8,0 %	4,7 – 11,3 %
Rural reculé	15,7 %	11,0 – 20,4 %

Des intervalles qui se chevauchent ne constituent pas un test — En R

```
svyttest(I(displacement_status == "displaced") ~ I(stratum == "urban"),  
         subset(design, stratum != "rural-remote"))
```

Des intervalles qui se chevauchent ne constituent pas un test — En Python

```
# The difference of two estimates, with its own standard error
diff = p_urban - p_rural
se_diff = math.sqrt(se_urban**2 + se_rural**2)
print(f"difference {diff:.1%}, 95% CI {diff - 1.99 * se_diff:.1%} to "
      f"{diff + 1.99 * se_diff:.1%}")
```

Arrondis et fausse précision — En Python

```
print(f"{p:.1%} (95% CI {low:.1%} to {high:.1%})")
```

```
sprintf("%.1f%% (95%% CI %.1f-%.1f)", 100 * p, 100 * low, 100 * high)
```

- **Une décimale** pour un pourcentage issu d'une enquête de cette taille. Deux laissent entendre une précision de 0,01...
- **Arrondissez l'intervalle vers l'extérieur**, jamais vers l'intérieur. 25,47-32,71 devient 25,4-32,8.
- **N'arrondissez jamais l'estimation ponctuelle jusqu'à un seuil**. Une estimation de 14,96 % rapportée « 15 % » à côté...

L'insécurité alimentaire des ménages est estimée à **29,1 %** (IC 95 % 25,5-32,7), sur 996 ménages répartis dans 75 grappes, pondérés selon la base de sondage et ajustés pour la non-réponse ; effet de plan 1,60, échantillon effectif 623.

L'insécurité alimentaire touche 32,8 % des ménages.

Lire un intervalle face à un seuil — En Python

```
THRESHOLD = 0.15

if high < THRESHOLD:
    verdict = "below the threshold"
elif low > THRESHOLD:
    verdict = "above the threshold"
else:
    verdict = "cannot be distinguished from the threshold"
print(verdict)
```

Lire un intervalle face à un seuil — En R

```
dplyr::case_when(  
  high < 0.15 ~ "below the threshold",  
  low  > 0.15 ~ "above the threshold",  
  TRUE      ~ "cannot be distinguished from the threshold"  
)
```

- L'intervalle dit ce que votre échantillon entier permet.

Lire la leçon complète, avec le code exécutable

[https://data-analysis.cassion.dev/fr/cours/survey-analysis-sampling/
survey-07-confidence-intervals](https://data-analysis.cassion.dev/fr/cours/survey-analysis-sampling/survey-07-confidence-intervals)