

Un rapport de cotes de 0,43 pour un rapport de risques de 0,58

Les cas signalant un handicap aboutissent à 26,4 % contre 45,5 %. C'est 58 % du taux, et le modèle logistique imprime 0,43. Les deux nombres décrivent les mêmes données, un seul est ce que le lecteur croit lire, et l'écart se creuse précisément quand le résultat est fréquent.

Pierre Robentz Cassion

Leçon 3 sur 8

Régression appliquée aux données de programme — Des résultats en oui ou non

Ce que couvre cette leçon

- Ajustez-le, et lisez ce qu'il imprime
- Pourquoi ils diffèrent, et quand cela empire
- Le piège qui attrape les analystes, pas seulement les lecteurs
- Quand le rapport de cotes est le bon nombre
- La phrase à écrire, et trois à ne pas écrire
- Rapportez-le en entier
- La suite

Ajustez-le, et lisez ce qu'il imprime — En Python (suite)

```
import pandas as pd
import numpy as np
import statsmodels.formula.api as smf

referrals = pd.read_csv("protection-referrals-2024.v1.csv")
referrals["disability"] = referrals["disability_reported"].map(
    {"true": 1, "Yes": 1, "false": 0, "No": 0})

consenting = referrals[referrals["consent_to_refer"]].copy()
consenting["completed"] = (consenting["referral_accepted"]
                           & consenting["days_to_first_service"].notna()).astype(int)
d = consenting.dropna(subset=["disability", "case_category", "age_band",
                              "sex", "service_requested", "admin1"])

crude = smf.logit("completed ~ disability", data=d).fit(displ=0)
print(np.exp(crude.params).round(3))
```

Ajustez-le, et lisez ce qu'il imprime — En Python (suite)

```
print(np.exp(crude.conf_int()).round(3))
```

Ajustez-le, et lisez ce qu'il imprime — En R

```
library(dplyr)
```

```
crude <- glm(completed ~ disability, data = d, family = binomial())  
exp(cbind(OR = coef(crude), confint(crude)))
```

- Rapport de cotes 0,430, IC à 95 % 0,308 à 0,601

Ajustez-le, et lisez ce qu'il imprime — En Python

```
rates = d.groupby("disability")["completed"].agg(["sum", "size", "mean"])
print(rates.round(4))
p1, p0 = rates.loc[1, "mean"], rates.loc[0, "mean"]
print(f"risk ratio      {p1 / p0:.3f}")
print(f"risk difference {p1 - p0:+.4f}")
print(f"odds ratio      {(p1/(1-p1)) / (p0/(1-p0)):.3f}")
```

Ajustez-le, et lisez ce qu'il imprime — En R

```
d |> summarise(k = sum(completed), n = n(), rate = mean(completed), .by = disability)
```

Ajustez-le, et lisez ce qu'il imprime

Groupe	Aboutis	n	Taux
Handicap signalé	52	197	26,4 %
Non signalé	629	1 384	45,5 %

Ajustez-le, et lisez ce qu'il imprime

Mesure	Valeur	Se lit
Rapport de risques	0,581	58 % de chances d'aboutir
Différence de risques	−19,1 points	19 aboutissements de moins pour 100 cas
Rapport de cotes	0,430	—

Ajustez-le, et lisez ce qu'il imprime

- Le modèle a imprimé 0,430 et la réponse vaut 0,581 — Les deux sont corrects ; ce sont des réponses à des questions. . .

Pourquoi ils diffèrent, et quand cela empire — En Python

```
for p0 in (0.02, 0.10, 0.25, 0.45, 0.70):  
    p1 = 0.6 * p0                                # a true risk ratio of 0.6 throughout  
    odds = (p1 / (1 - p1)) / (p0 / (1 - p0))  
    print(f"baseline {p0:5.0%}    risk ratio 0.60    odds ratio {odds:.3f}")
```

Pourquoi ils diffèrent, et quand cela empire — En R

One true risk ratio, five baselines, five different odds ratios.

Pourquoi ils diffèrent, et quand cela empire

Risque de base	Vrai rapport de risques	Rapport de cotes
2 %	0,60	0,59
10 %	0,60	0,57
25 %	0,60	0,53
45 %	0,60	0,45
70 %	0,60	0,31

Pourquoi ils diffèrent, et quand cela empire

- Le rapport de cotes n'est pas une distorsion fixe du rapport de risques — il dépend du niveau de base — À 2 % de...
- Les résultats de programme ne sont pas rares — L'aboutissement des orientations vaut 43 %, la présence 88 %,...

Le piège qui attrape les analystes, pas seulement les lecteurs — En Python

```
adjusted = smf.logit(  
    "completed ~ disability + case_category + age_band + sex"  
    " + service_requested + admin1", data=d).fit(dis=0)  
print(f"crude OR      {np.exp(crude.params['disability']):.3f}")  
print(f"adjusted OR {np.exp(adjusted.params['disability']):.3f}")
```

Le piège qui attrape les analystes, pas seulement les lecteurs — En R

```
adjusted <- glm(completed ~ disability + case_category + age_band + sex +  
                service_requested + admin1, data = d, family = binomial())  
exp(coef(adjusted))["disability"]
```

Le piège qui attrape les analystes, pas seulement les lecteurs

	Rapport de cotes	Différence de risques
Brut	0,430	–19,05 points
Ajusté	0,388	–19,41 points

Le piège qui attrape les analystes, pas seulement les lecteurs

- Le rapport de cotes a bougé de 10 % et la différence de risques d'un tiers de point — La lecture habituelle — «... »
- Une mesure effondrable est égale à la moyenne des mesures de sous-groupes — Les différences et les rapports de risques. . .
- Un rapport de cotes ajusté et un rapport de cotes brut ne peuvent donc pas être comparés pour juger d'une confusion —. . .

Quand le rapport de cotes est le bon nombre

- Une étude cas-témoins — Cas et témoins sont échantillonnés séparément, si bien que les risques ne sont pas estimables. . .
- Un résultat rare — En dessous d'environ 10 %, le rapport de cotes approche le rapport de risques d'assez près pour que. . .
- Une comparaison avec une littérature publiée en rapports de cotes — Rapportez alors les deux, et menez avec la. . .

Quand le rapport de cotes est le bon nombre — En Python

```
def report(label, p1, p0):  
    return (f"{label}: {p1:.1%} vs {p0:.1%} -- "  
            f"risk difference {p1-p0:+.1%}, risk ratio {p1/p0:.2f}, "  
            f"odds ratio {(p1/(1-p1))/(p0/(1-p0)):.2f}")  
  
print(report("Referral completion, disability reported", p1, p0))
```

Quand le rapport de cotes est le bon nombre — En R

Print all three. The reader picks; you do not pick for them by omission.

La phrase à écrire, et trois à ne pas écrire

- Pas ceci — « Les cas signalant un handicap avaient 57 % de chances en moins d'aboutir (RC 0,43)
- Ni ceci — « Le handicap a divisé par deux les cotes d'aboutissement
- Ni ceci — « RC 0,43 (IC à 95 % 0,31–0,60, $p < 0,001$)
- Ceci — « Les cas signalant un handicap ont abouti à 26,4 % contre 45,5 % pour les cas sans handicap signalé — 19,1...
- La dernière proposition est celle qui travaille — Un rapport de cotes dans un rapport de programme sans traduction à...

Rapportez-le en entier — Exemple (suite)

Referral completion by disability status, 1,581 consenting cases

Disability reported	26.4%	52 of 197
Not reported	45.5%	629 of 1,384
Risk difference	-19.1 points	95% CI -25.7 to -12.4
Risk ratio	0.58	
Odds ratio	0.43 (crude), 0.39 (adjusted)	

The odds ratio is reported for comparability only. Completion is a common outcome (43% overall), so the odds ratio overstates the relative difference: 0.43 in odds is 0.58 in risk.

The adjusted odds ratio differs from the crude one partly because the odds ratio is not collapsible, not only because of confounding. The adjusted risk difference is -19.4 points, essentially unchanged from crude.

Rapportez-le en entier — Exemple (suite)

Fitted on the 1,581 consenting cases with complete covariates. The statistics course reported this gap on all 1,638 consenting cases and got -19.4 points; the 57-case difference is the complete-case restriction.

- Le dernier paragraphe fait deux lignes et prévient la surinterprétation la plus courante — celle qui veut que...

- Un modèle logistique imprime des coefficients sur une échelle où personne ne pense.

Lire la leçon complète, avec le code exécutable

```
https://data-analysis.cassion.dev/fr/cours/regression-for-programmes/  
reg-03-the-odds-ratio-your-reader-misreads
```