Introduction aux statistiques bayésiennes

Application dans le vieillissement avec JASP

GRÉGOIRE Coline

Directeur de thèse: Steve MAJERUS

Projet de thèse : « Vers une meilleure compréhension des capacités d'inhibition dans le vieillissement »

Tours, le 10/09/19

Sommaire

Partie A – Statistiques Bayésiennes

1. Postulats de base
2. Facteurs de Bayes
3. Exemples
4. Fréquentistes vs Bayésiennes

Partie B – JASP

Partie A

Introduction aux statistiques bayésiennes

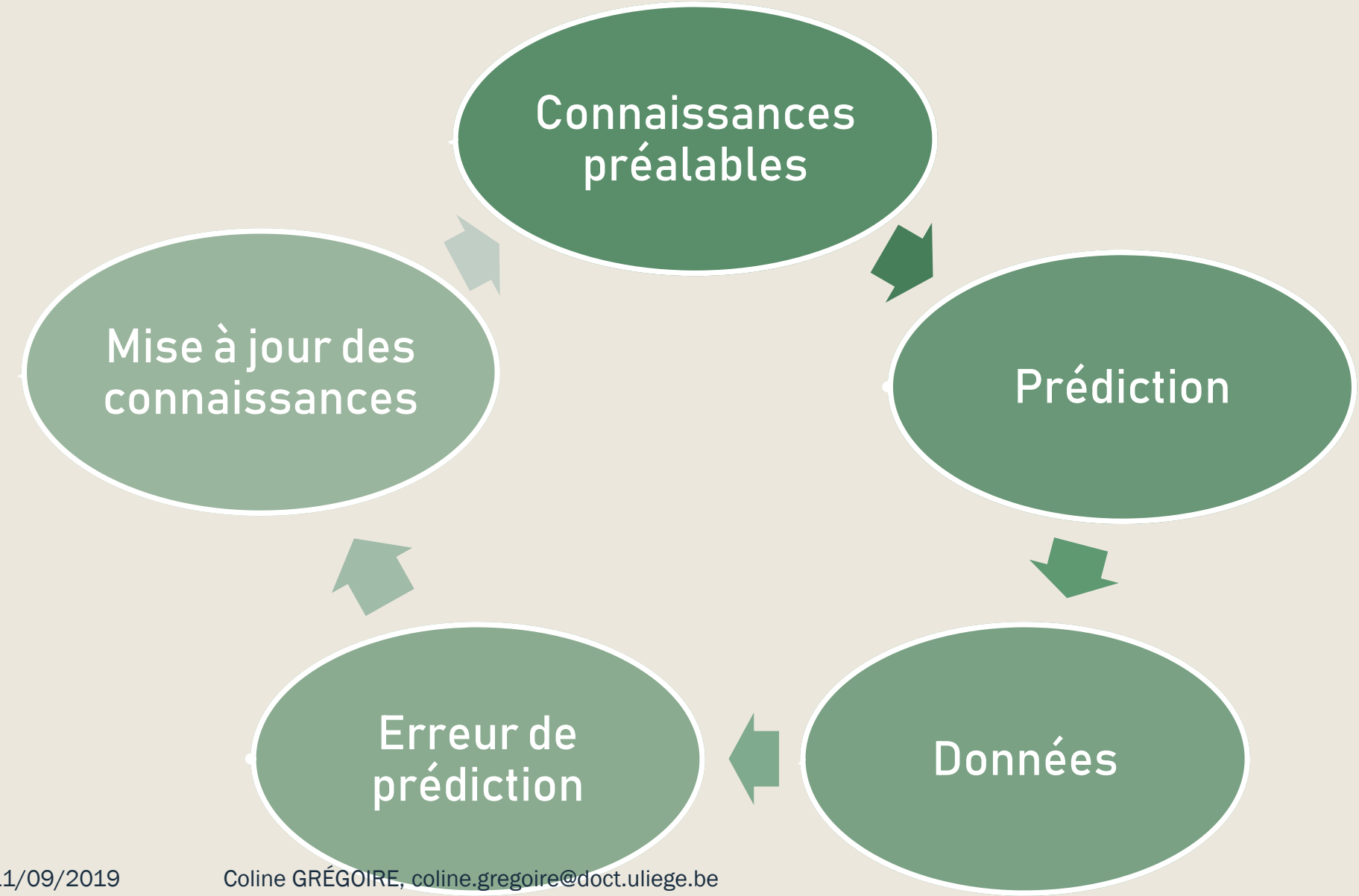
Postulat de base

Statistiques bayésiennes = Degré de **vraisemblance** (likelihood) d'un modèle

Données observées vs données a priori (prior)

Vraisemblance a posteriori (posterior)

Postulat de base



Postulat de base: comment ça fonctionne?

?

données *d*
data

Théorème de Bayes

Croyances postérieures
Posterior distribution

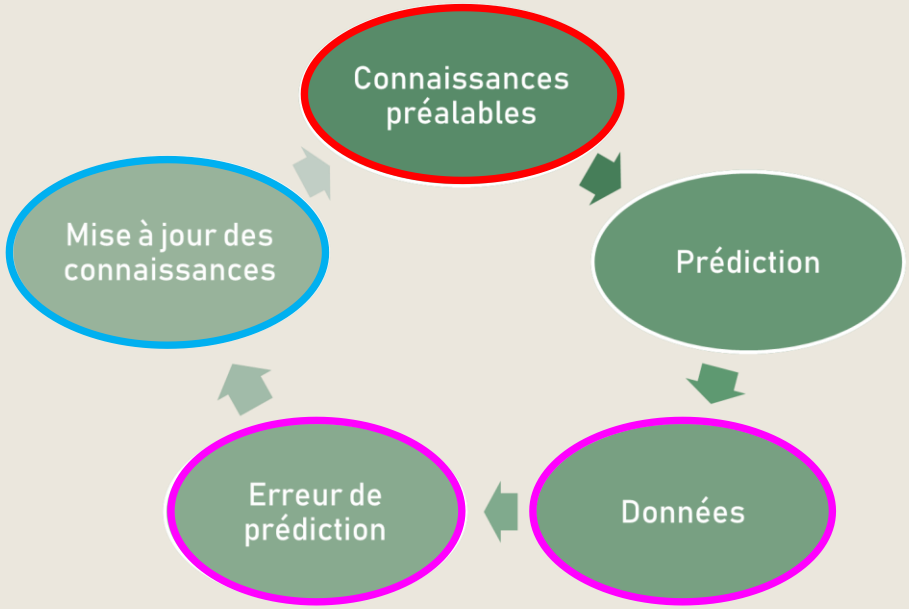
Croyances antérieures
Prior information

Jours de précipitations, 1 **mm/jour**[8]	10.7
Jours de précipitations, 10 **mm/jour**[8]	1.7

Stevens (2009)

Postulat de base

$$P(h|d) = \frac{P(d|h)}{P(d)} * P(h)$$



Les croyances antérieures sont mises à jour à l'aide des données afin de générer des croyances postérieures.

Postulat de base

①

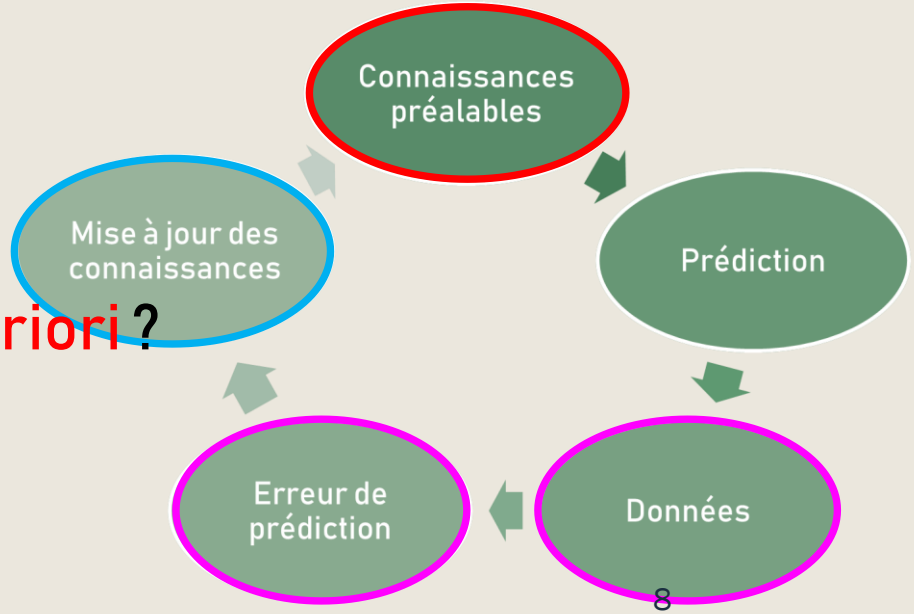
$$P(h|d) = \frac{P(d|h)}{P(d)} * P(h)$$

« Je dispose d'un parapluie. Pensez-vous qu'il va pleuvoir? »

h : hypothèse: il va pleuvoir

d : données : Je dispose d'un parapluie

↳ Quelles sont les croyances antérieures, a priori ?



Postulat de base

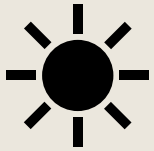
②

$$P(h|d) = \frac{P(d|h)}{P(d)} * P(h)$$

h : hypothèse: peut-être qu'il va pleuvoir, ou pas

d : données : Je dispose d'un parapluie

↳ Quelles sont les croyances antérieures, a priori ?

hypothèse h	Degré de croyance
	0,45
	0,55
	1

Jours de précipitations, 1 mm/jour[8]	10.7
Jours de précipitations, 10 mm/jour[8]	1.7

Postulat de base

③

$$P(h|d) = \frac{P(d|h)}{P(d)} * P(h)$$

↳ Quelles sont les probabilités concernant les données, leur vraisemblance?

	données d	
hypothèse h	☂	☂
☁	0,20	0,80
☀	0,05	0,95

Il est important de se rappeler que chaque cellule de ce tableau décrit vos opinions sur les données qui seront conservées, compte tenu de la véracité d'une hypothèse particulière $\boldsymbol{h}$.
Cette «probabilité conditionnelle» est écrite $P(d|h)$ que vous pouvez lire comme «la probabilité de d étant donné h». Dans les statistiques bayésiennes, on parle de probabilité de données $\boldsymbol{d}$ étant donné l'hypothèse $\boldsymbol{h}$.

Postulat de base

④

$$P(h|d) = \frac{P(d|h)}{P(d)} * P(h)$$

↳ Quelles sont les probabilités jointes des données $P(d, h)$?

hypothèse h	données d	
	☂	☂
☁	0,09	0,36
☀	0,0275	0,5225
	0,1175	0,8825

La probabilité qu'il pleuve aujourd'hui (c'est-à-dire que l'hypothèse $\boldsymbol{h}$ soit vraie) et que je porte un parapluie (c'est-à-dire que les données $\boldsymbol{d}$ sont observées).

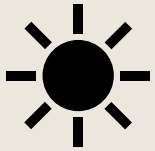
$\boldsymbol{P(d, h)} = P(d|h) * P(h)$
$= P(parapluie, pluie)$
$=$ probabilité que je me souvienne de mon parapluie un jour de pluie 20% $*$ prior de la probabilité d'un jour de pluie 45%
$= 0{,}20 * 0{,}45 = 0{,}09$

Postulat de base

5

$$P(h|d) = \frac{P(d|h)}{P(d)} * P(h)$$

↳ Comment met-on à jour les croyances selon la règle de Bayes ?

hypothèse h	données d	
	0,77	0
	0,33	0
	1	0

On vous dit que je porte vraiment un parapluie, ce qui n'avait que 11,75% de chance d'arriver, alors ma probabilité « parapluie » = 1

$$\text{plausibilité qu'il pleuve} = \frac{\text{probabilité jointe "pluie et parapluie"}}{\text{"chances totales de parapluie"}} = \frac{P(d,h)}{P(d)}$$

$$= \frac{\text{probabilété jointe de l'hypothèse et des données}}{\text{probabilité marginale des données "parapluie"}}$$

$$= \frac{0{,}09}{0{,}1175} = \mathbf{0,77}$$

Postulat de base

5

$$P(h|d) = \frac{P(d|h)}{P(d)} * P(h)$$

↳ Comment met-on à jour les croyances selon la règle de Bayes ?

	données d	
hypothèse h	☂	☂
🌧	0,77	0
☀	0,33	0
	1	0

La probabilité postérieure de pluie (sa plausibilité) à Liège en septembre étant donné que je porte un parapluie est de : $P(h|d)$ = 77%

Postulat de base

⑥

$$P(h|d) = \frac{P(d|h)}{P(d)} * P(h)$$

↳ Comment met-on à jour les croyances selon la règle de Bayes ?

	données d	
hypothèse h	☂	☂
🌧	0,77	0
☀	0,33	0
	1	0

$$P(h|d) = \frac{P(d,h)}{P(d)}$$

et

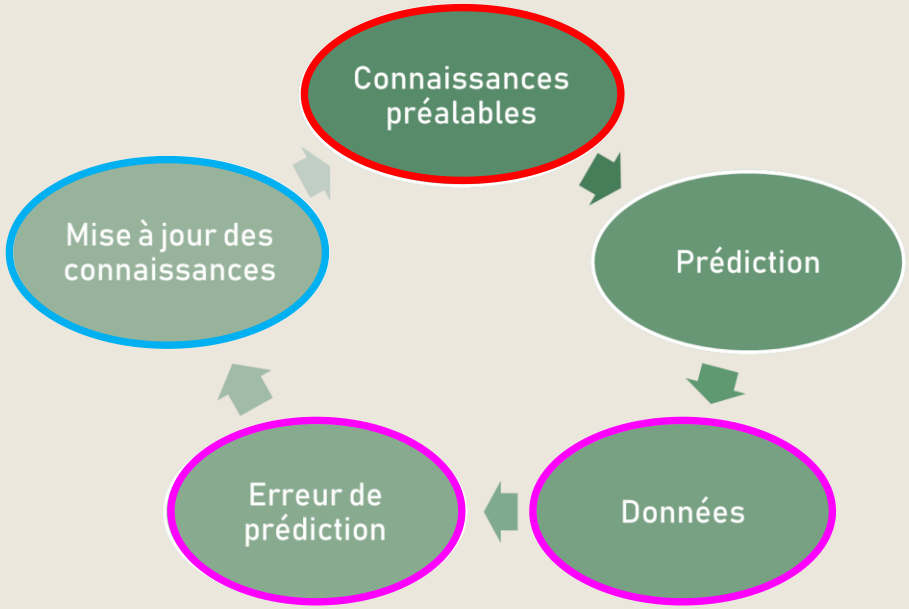
$$P(d,h) = P(d|h) * P(h)$$

alors

$$P(h|d) = \frac{P(d|h)}{P(d)} * P(h)$$

Postulat de base

$$P(h|d) = \frac{P(d|h)}{P(d)} * P(h)$$



Les adultes âgés n'ont peut-être pas systématiquement des scores d'inhibition (STROOP) moins bons que les adultes jeunes. → Implication théorique?

Les données tendent à ne montrer aucune différence entre les performances des jeunes et des âgés: variabilité dans les groupes.

Les adultes âgés ont des scores d'inhibition (STROOP) moins bons que les adultes jeunes « *et de toute façon, avec l'avancée en âge, les performances diminuent* ».

	Group	N	Mean	SD	SE
Indice d'inhibition 1	1	24	0.401	0.143	0.029
	2	24	0.488	0.137	0.028

BF_{10}
1.791

Test d'hypothèse bayésien

$$P(h|d) = \frac{P(d|h)}{P(d)} * P(h)$$

Les croyances antérieures sont mises à jour à l'aide des données afin de générer des croyances postérieures.

H_0: $$P(H_0|d) = \frac{P(d|H_0)}{P(d)} * P(H_0)$$

H_1: $$P(H_1|d) = \frac{P(d|H_1)}{P(d)} * P(H_1)$$

Test d'hypothèse bayésien

$$H_0: \quad P(H_0|d) = \frac{P(d|H_0)}{P(d)} * P(H_0)$$

$$H_1: \quad P(H_1|d) = \frac{P(d|H_1)}{P(d)} * P(H_1)$$

$$\frac{P(H_1|données)}{P(H_0|données)} = \frac{P(données|H_1)}{P(données|H_0)} * \frac{P(H_1)}{P(H_0)}$$

probabilités postérieures = facteur bayésien * probabilités antérieures

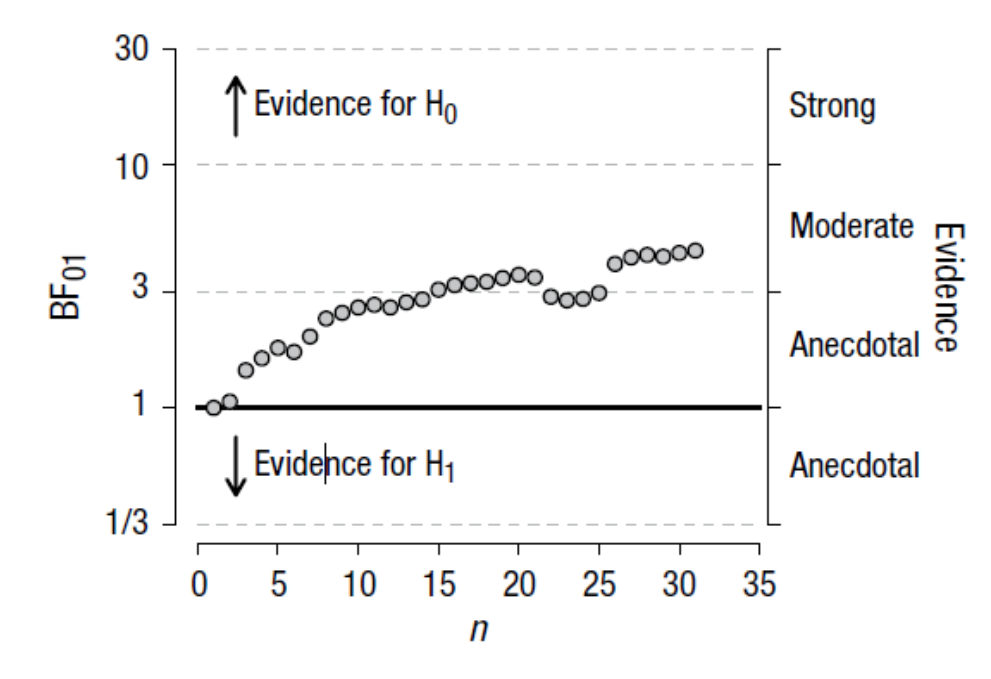
le Facteur Bayésien

- indice statistique qui quantifie l'évidence d'une hypothèse par rapport à une autre hypothèse.

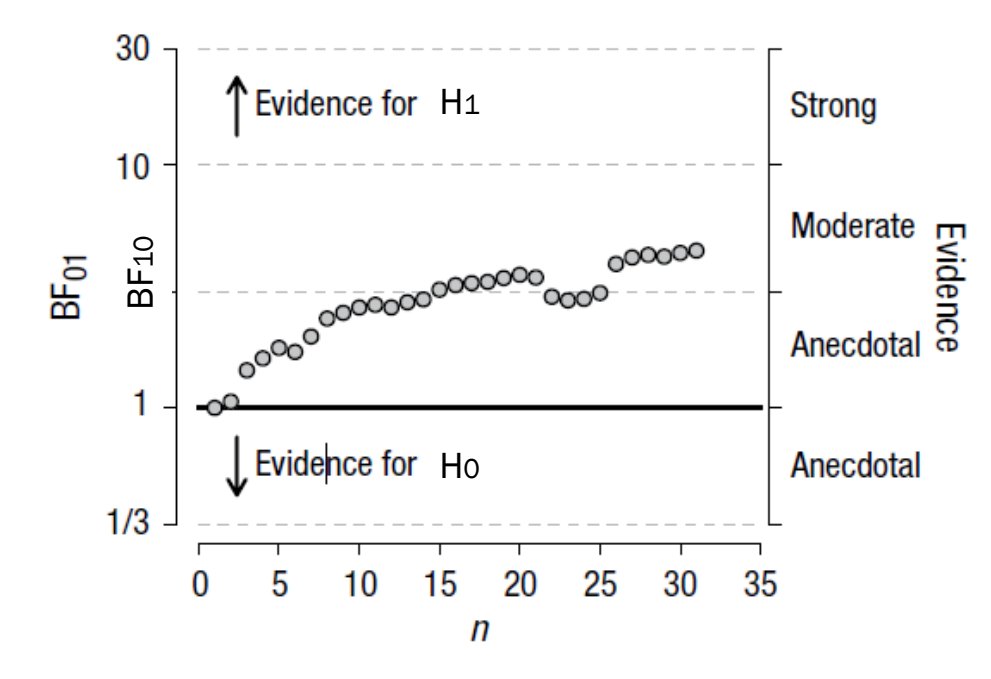
- occupe une place particulière dans les tests d'hypothèses bayésiennes, car il joue un rôle similaire à la valeur p dans les tests d'hypothèses orthodoxes.

what-does-a-bayes-factor-feel-like + wiki

BF_{01} ou BF_{10} ?



Preuve\Quantité d'évidence **en faveur de H_0**



Preuve\Quantité d'évidence **en faveur de H_1**

what-does-a-bayes-factor-feel-like + Wagenmakers et al. 2016

Interprétation : que nous dit le facteur bayes?

Les adultes âgés ont des scores d'inhibition (STROOP) moins bons que les adultes jeunes

	Group	N	Mean	SD	SE
Indice d'inhibition 1	1	24	0.401	0.143	0.029
	2	24	0.488	0.137	0.028

Independent Samples T-Test

	t	df	p
Indice d'inhibition 1	−2.144	46.000	0.037

Note. Student's t-test.

Bayesian Independent Samples T-Test ▼

	BF_{10}	error %
Indice d'inhibition 1	1.791	2.514e −4

Test de normalité ✓
Test d'égalité des variances ✓

Connaissances a posteriori ?

Interprétation : que nous dit le BF?

Expérience 1 : visant à évaluer l'impact des connaissances lexico-sémantiques sur une tâche MCTv à l'aide d'une procédure running-span rapide: manipulation de mots et non-mots.

Table 1
Bayesian Factor values as a function of serial position and stimulus condition – Experiment 1.

	List length					
	Item recall criterion			Strict serial recall criterion		
	6	9	12	6	9	12
Effects	*Bayesian Repeated Measure ANOVA*					
Lexicality	→ +∞	→ +∞	→ +∞	6.45e+9	→ +∞	→ +∞
Serial position	→ +∞	→ +∞	→ +∞	6.01e+15	→ +∞	→ +∞
Lexicality * Serial position	.95	2.50e+8	→ +∞	24.26	2.37e+10	→ +∞
Effects	*Bayesian Paired T-Test*					
Pre-recency	9617	218897.06	1218.263	4.40	44,085	7.461
Recency	2.90e+6	1.26e+9	1.96e+7	20241.11	184,736	256504.46

Note. For main effects and the interaction term (Bayesian ANOVA), the values represent $BF_{inclusion}$. For pre-recency and recency effects (Bayesian T-Test), values represent BF_{10}.

Kowialiewski, B., & Majerus, S. (2018). The non-strategic nature of linguistic long-term memory effects in verbal short-term memory.

Pour résumer

« La règle de Bayes permet la révision probabiliste, et est utilisée dans le cadre de probabilités conditionnelles binaires (malade / non malade sachant que je présente les symptômes).

[…] permet de réviser notre jugement dès lors que nous obtenons une nouvelle information.

Ceci équivaut à passer d'une probabilité a priori -précédent la prise en compte du nouvel élément- à une probabilité a posteriori -faisant suite à la prise de connaissance de la nouvelle information-. » (p.28)

statistiques bayésiennes vs statistiques fréquentistes

- Probabilité des évènements selon une théorie

- Probabilité des évènements selon certains évènements

Statistiques fréquentistes	Statistiques bayésiennes

• Test d'hypothèse H_0 vs H_1
• Probabilité d'observer un événement si H_0 est vraie
• alpha < .05
• Certains biais?
  • Calcul de probabilité?
  • Subjectivité
  • Que représente alpha ?
	• Test d'hypothèse H_0 vs H_1
	• Vraisemblance d'un événement compte tenu de données a priori
	• BF > 3 / >10 / >30
	• Vision moins biaisée?
	  • Preuve en faveur de H0 et H1
	  • Contrôle de la subjectivité
	  • Mesure plus continue

Tableau de M.Bouffier, présentation PsyNDoc

Partie B

Introduction aux statistiques bayésiennes avec JASP

Nécessite l'utilisation de JASP

Gratuit : https://jasp-stats.org/download/

Facile d'utilisation

Manuels JASP gratuits

- Statistical Analysis in JASP. A Guide for Students by Mark Goss-Sampson
- Learning Statistics with JASP: A Tutorial for Psychology Students and Other Beginners

→ Chapitre sur les statistiques bayésiennes

- Introduction à JASP et aux statistiques bayésiennes (diaporama)

- Vidéos et résumés d'utilisation: https://jasp-stats.org/how-to-use-jasp/

Sources

Téléchargement

https://jasp-stats.org/download/

JASP

Descriptives | T-Tests | ANOVA | Regression | Frequencies | Factor | Meta Analysis | SEM | Summary

- Open
- Save
- Save As
- Export Results
- Export Data
- Sync Data
- Close
- Preferences

Welcome to JASP

A Fresh Way to Do Statistics: Free, Friendly, and Flexible

- **Free:** JASP is an open-source project with structural support from the University of Amsterdam.
- **Friendly:** JASP has an intuitive interface that was designed with the user in mind.
- **Flexible:** JASP offers standard analysis procedures in both their classical and Bayesian manifestations.

So open a data file and take JASP for a spin!

Please keep in mind that this is a preview release and a number of features are still missing.

Taper ici pour rechercher
11/09/2019
Coline GRÉGOIRE coline.gregoire@ac-toulouse.fr
11:49
03/09/2019

JASP et ses utilisations

Descriptives • Descriptive stats • Reliability analysis*	**Regression** • Correlation matrix • Linear regression • Logistic regression
T-Tests • Independent • Paired • One sample	**Frequencies** • Binomial test • Multinomial test • Contingency tables • Log-linear regression*
ANOVA • Independent • Repeated measures • ANCOVA • MANOVA *	**Factor** • Principal Component Analysis (PCA)* • Exploratory Factor Analysis (EFA)* • Confirmatory Factor Analysis (CFA)*

Comment ouvrir un document dans JASP?

Formats supportés: .csv (comma separated value) sous excel et statistica

.txt (plain text) also can be saved in Excel

.sav (IBM SPSS data file)

.ods (Open Document spreadsheet)

Depuis Excel ou Open Office Calc

Ouvrir votre fichier dans Excel ou Open Office Calc.

Fichier → Enregistrer sous.

Dans le menu "Format du fichier" Sélectionner "CSV (Windows) (séparateur : point-virgule) (.csv)

Compléter les données nécessaires à vos analyses (noms dans la 1ère colonne ou 1ère ligne)

Depuis Statistica

Ouvrir votre fichier (.sta) dans Statistica

Fichier → Enregistrer sous

Compléter les données nécessaires à vos analyses (noms dans la 1ère colonne ou 1ère ligne)

Depuis statistica

Statistica - [Données : fichier statistica nouveau tableau de données.sta (115 var. et 52 obs.)]

Fichier | Accueil | Édition | Affichage | Format | Statistiques | Data Mining | Graphiques | Outils | Données | Recherche rapide | Options

Coller | Couper | Copier | Mise en Forme | Sélectionner Tout | Effacer | Supprimer | Déplacer | Capture d'Écran — Presse-Papiers/Données

Rechercher | Remplacer | Répéter | Atteindre — Rechercher/Remplacer

Valeurs Aléatoires | Recopier vers le Bas | Recopier vers la Droite — Remplissage Automatique

Variables | Observations — Centrer-Réduire

Variables | Observations | Objet OLE — Insertion

DDE | OLE | Objets — Liens

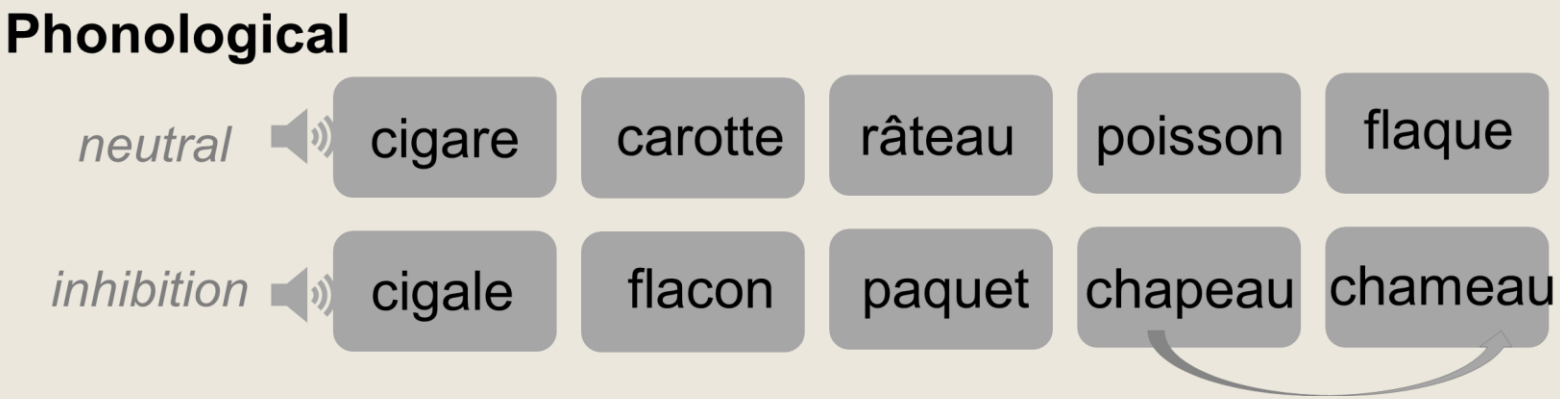
	1 Sujets	2 Groupe d'âge	3 âge réel	4 sexe	5 ContreB	6 niveau d'étude	7 MMSE	8 Mill Hill	9 N-Back	10 Anxiété	11 Dépression	12 Total	13 PS.mots	14 IP.mots	15 PS/IPmots	16 PS.nonmots	17 IP.nonmots	18 PS/IPnonmots	19 Stroop A	20 Stroop B	21 Stro
1	J01	1	40	1	1	17		23	25	9	7	16	61	61	1	29	29	1	114	66	3
2	J02	1	40	2	2	17		26	23	3	5	8	61	65	0,93846	41	42	0,97619	120	80	3
3	J03	1	23	2	1	13		22	26	5	11	16	48	51	0,94118	20	23	0,86957	83	62	3
4	J04	1	31	1	2	17		27	23	3	4	7	54	61	0,88525	37	48	0,77083	112	92	5
5	J05	1	29	2	1	17		27	24	5	8	13	66	68	0,97059	42	42	1	104	72	4
6	J06	1	25	1	2	17		20	24	2	5	7	63	65	0,96923	37	39	0,94872	107	87	7
7	J07	1	25	1	1	15		20	27	8	2	10	62	62	1	38	40	0,95	89	61	4
8	J08	1	22	2	2	15		28	27	13	5	18	61	63	0,96825	39	39	1	127	97	6
9	J09	1	28	1	1	17		27	26	3	4	7	61	63	0,96825	41	43	0,95349	133	107	6
10	J10	1	22	2	2	15		31	25	15	5	20	48	56	0,85714	32	36	0,88889	105	74	4
11	J11	1	24	2	1	15		28	27	9	3	12	35	67	0,52239	35	63	0,55556	94	76	6
12	J12	1	21	2	2	12		25	26	16	5	21	44	53	0,83019	38	41	0,92683	80	78	6
13	J13	1	22	2	1	12		19	24	5	10	15	55	57	0,96491	33	33	1	111	73	4
14	J14	1	37	2	2	15		27	17	11	5	16	44	47	0,93617	25	30	0,83333	79	62	2
15	J15	1	34	2	1	9		17	23	12	13	25	37	50	0,74	16	23	0,69565	75	56	2
16	J16	1	32	2	2	13		16	19	4	1	5	45	49	0,91837	16	16	1	90	66	4
17	J17	1	38	1	2	15		19	24	8	9	17	62	66	0,93939	40	43	0,93023	97	61	1
18	J18	1	22	1	1	15		25	27	20	13	33	68	73	0,93151	44	48	0,91667	99	77	3
19	J19	1	21	2	1	15		20	24	15	9	24	50	60	0,83333	35	38	0,92105	126	84	4
20	J20	1	21	2	2	15		25	26	9	5	14	54	63	0,85714	42	43	0,97674	118	103	5
21	J21	1	36	2	1	13		20	27	3	4	7	37	46	0,80435	20	23	0,86957	97	78	3
22	J22	1	22	1	2	15		30	22	2	6	8	60	64	0,9375	41	41	1	124	83	6
23	J23	1	21	2	2	15		23	22	1	3	4	49	54	0,90741	25	27	0,92593	99	78	4
24	J24	1	28	2	1	15		17	23	16	8	24	46	56	0,82143	24	36	0,66667	82	67	3
25	J25	1	23	1	1	12		20	26	3	2	5	49	57	0,85965	22	24	0,91667	119	91	5
26	J26	1	20	2	1	12		19	26	11	10	21	61	63	0,96825	39	39	1	110	71	5
27	A01	2	60	1	1	10	28	28	25	2	4	6	59	61	0,96721	27	28	0,96429	89	72	3

Pour l'Aide, appuyez sur la | fichier statistica nouveau table | C1,V1

Instant pratique

Comment faire un test-t ? Exemple

2 groupes
- 24 adultes jeunes
- 24 adultes âgés



Score d'interférence : neutre-inhibition

H1 → Les adultes âgés seront plus affectés par l'interférence phonologique que les adultes jeunes.

Descriptives | T-Tests | ANOVA | Regression | Frequencies | Factor | Meta Analysis | SEM | Summary St

Bayesian Independent Samples T-Test

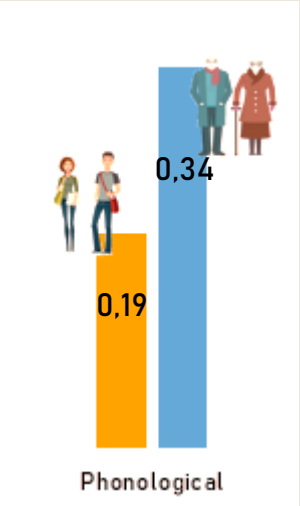
Bayesian Independent Samples T-Test

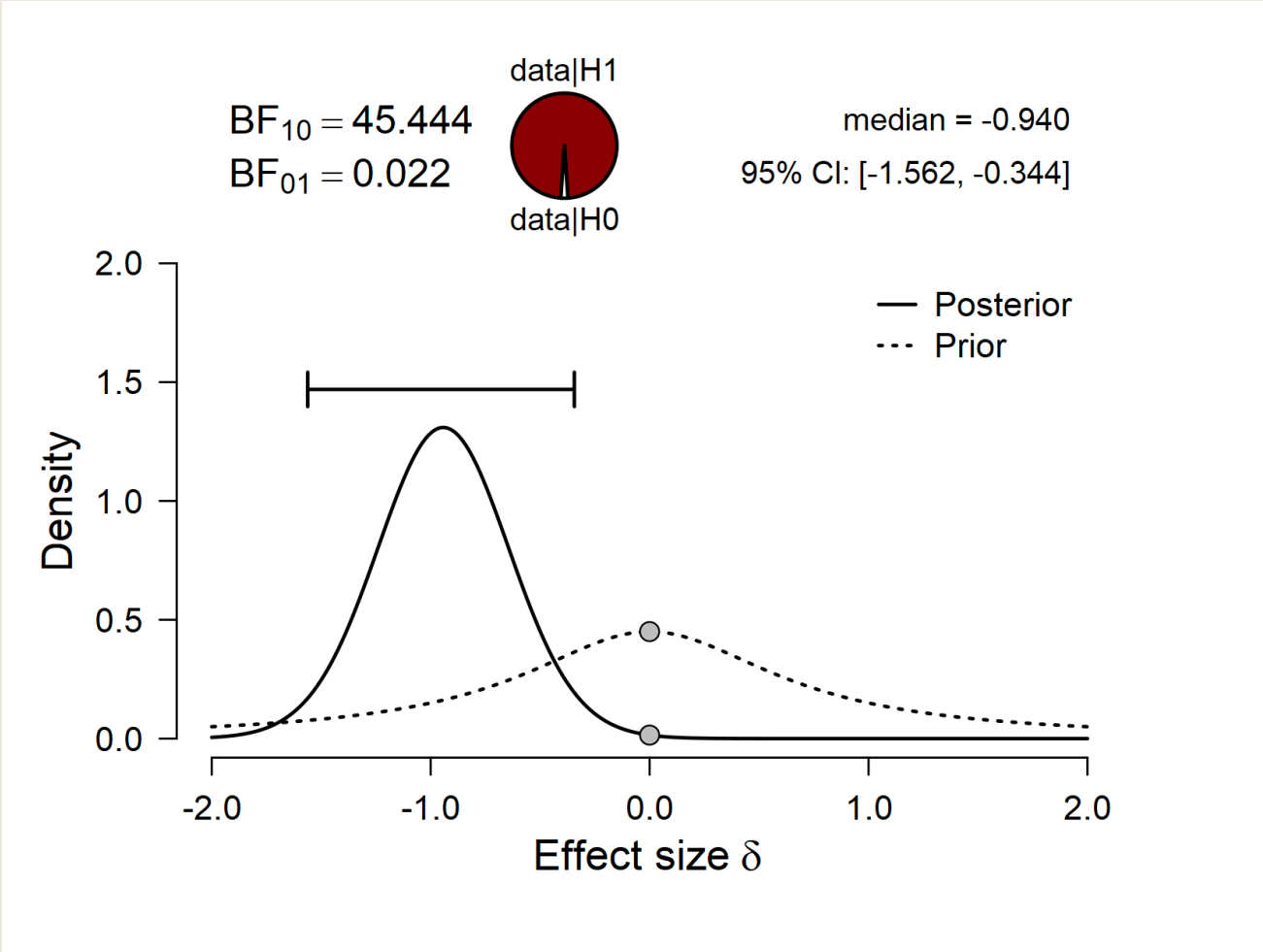
0.021

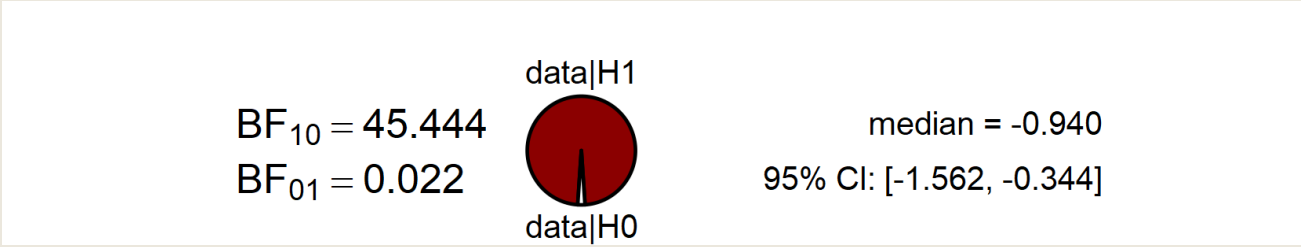
N	Mean	SD	S
24	14.625	2.039	0
24	12.250	2.786	0
24	23.208	4.384	0
24	28.375	3.487	0
24	0.401	0.143	0
24	0.488	0.137	0
24	24.250	2.524	0
23	23.652	1.991	0
24	1.042	2.016	0
24	0.750	2.703	0

Instant pratique

Bayesian Independant T-test





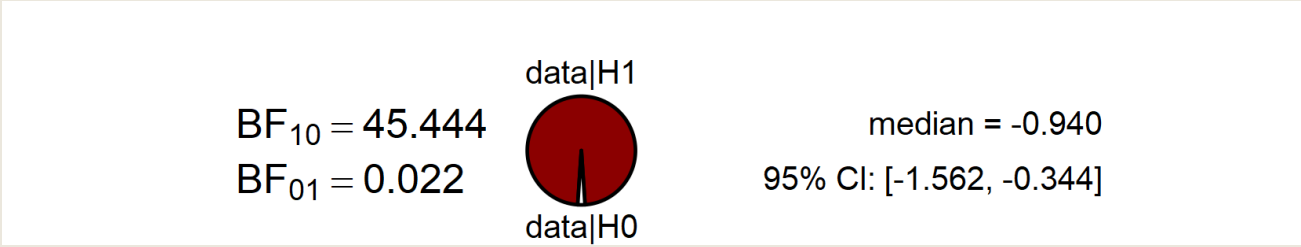


$BF_{10} = 45{,}444$ Odds de 45,444:1 en faveur de H_1

$BF_{01} = \frac{1}{45{,}444}$ Odds de 1:45,444 en faveur de H_0

↳ Quelle interprétation peut-on faire?

Les données sont 45,44 fois plus susceptibles de se produire sous H_1 que sous H_0.

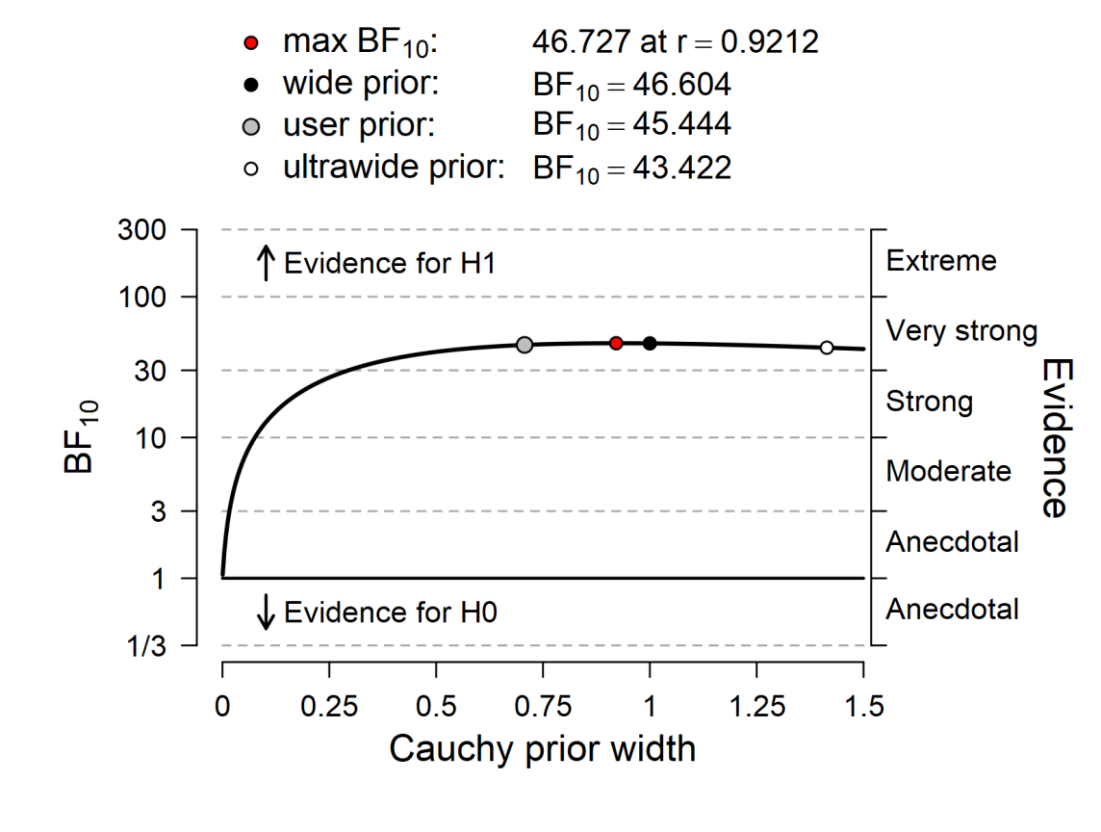


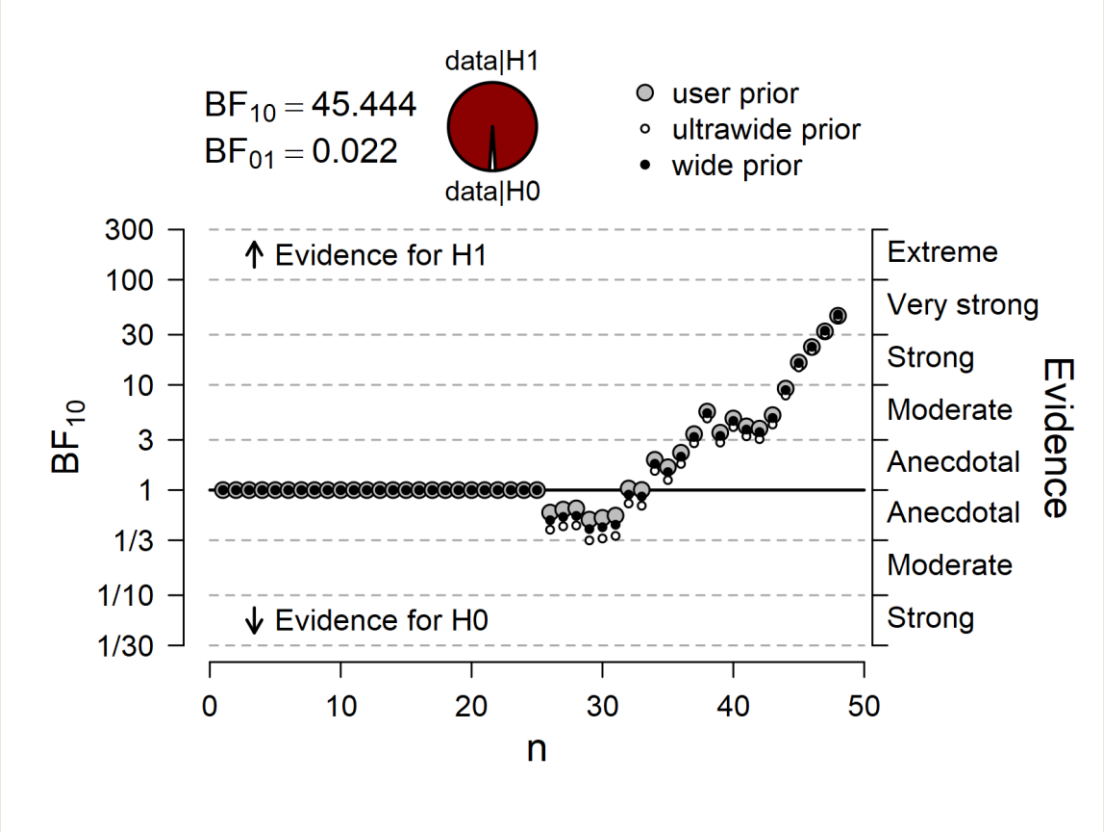
$BF_{10} = 45{,}444$ Odds de 45,444:1 en faveur de H_1

$BF_{01} = \frac{1}{45{,}444}$ Odds de 1:45,444 en faveur de H_0

↳ Quelle interprétation peut-on faire?

Il est 45 fois plus probable qu'il y ait une différence pour l'interférence phonologique entre âgés et jeunes , plutôt qu'il n'y en ait pas


max BF10: 46.727 at r = 0.9212
wide prior: BF10 = 46.604
user prior: BF10 = 45.444
ultrawide prior: BF10 = 43.422
Evidence for H1
Evidence for H0
BF10
300
100
30
10
3
1
1/3
0
0.25
0.5
0.75
1
1.25
1.5
Cauchy prior width
Extreme
Very strong
Strong
Moderate
Anecdotal
Anecdotal
Evidence


BF10 = 45.444
BF01 = 0.022
data|H1
data|H0
user prior
ultrawide prior
wide prior
Evidence for H1
Evidence for H0
BF10
300
100
30
10
3
1
1/3
1/10
1/30
0
10
20
30
40
50
n
Extreme
Very strong
Strong
Moderate
Anecdotal
Anecdotal
Moderate
Strong
Evidence

Instant pratique

Instant pratique
Régression Linéaire

Régression Linéaire

	adverts	sales	airplay	attract
1	10.256	330	43	10
2	985.685	120	28	7
3	1445.56	360	35	7
4	1188.19	270	33	7
5	574.513	220	44	5
6	568.954	170	19	5
7	471.814	70	20	1
8	537.352	210	22	9
9	514.068	200	21	7
10	174.093	300	40	7
11	1720.81	290	32	7
12	611.479	70	20	2
13	251.192	150	24	8
14	97.972	190	38	6
15	406.814	240	24	7
16	265.398	100	25	5
17	1323.29	250	35	5
18	196.65	210	36	8
19	1326.6	280	27	8
20	1380.69	230	33	8
21	792.345	210	33	7
22	957.167	230	28	6
23	1789.66	320	30	9
24	656.137	210	34	7

Régression Linéaire bayésienne: Big-V

	Neuroticism	Extraversion	Openness
1	2.47917	4.20833	3.9375
2	2.60417	3.1875	3.95833
3	2.8125	2.89583	3.41667
4	2.89583	3.5625	3.52083
5	3.02083	3.33333	4.02083
6	2.52083	3.29167	3.4375
7	2.35417	4.41667	4.58333
8	2.52083	3.5	2.89583
9	3.10417	3.8125	4.0625
10	2.6875	3.54708	3.78667
11	2.625	3.45833	2.89583
12	2.375	3.77083	3.16667
13	3.0625	3.41667	3.77083
14	3.125	2.52083	2.64583
15	2.58333	3.02083	3.5
16	2.97917	3.3125	3.35417
17	2.79167	3.79167	3.375
18	2.5625	3.54167	3.875
19	2.4375	3.72917	2.91188

Merci pour votre attention

MyORBI | PsyNCog | LinkedIn | GoogleScholar

✉ coline.gregoire@doct.uliege.be
✉ smajerus@uliege.be