Ce document est à destination du public peu initié aux statistiques et souhaitant savoir comment interpréter les passages comportant des écritures obscures comme « $X^2 = 5,702$; df = 1 ; p = 0,017 ; v de Cramer = 19,4% ; $BF_{10} = 3,776$ ». Ici plusieurs choses sont mentionnées :

- ➔ X^2 désigne le khi-carré, un test statistique qui consiste à étudier la relation entre deux variables nominales ou plus, et donc à comparer les participants concernant des caractéristiques qui s'expriment par un nom : par exemple l'appartenance au groupe « hommes » vs au groupe « femmes » au sein de la variable « genre ». Ce test est appliqué selon un certain nombre de degrés de liberté (df) sur lesquels nous revenons ci-dessous. Lorsque le test est appliqué, la valeur de X^2 reflète une distance entre ce qui a été observé sur notre échantillon et ce qu'il est attendu s'il n'y a pas de différence entre les résultats de nos groupes.
- ➔ df désigne « degree of freedom » en anglais, ce qu'on appelle les degrés de liberté (ddl) en français. Ceux-ci servent à comprendre le nombre de paramètres que le modèle statistique doit estimer pour représenter nos données avec nos variables. Or les statistiques visent à représenter le maximum de données via des lois simples. Ainsi, lors des tests t paramétriques le df rapporté vaut (la taille de notre échantillon de participants) – (le nombre de paramètres à estimer) et on cherche alors à ce que ce nombre soit le plus grand possible. En revanche lors des tests du khi-carré, les df rapportés ici correspondent au calcul : (nombre de lignes - 1) x (nombre de colonnes - 1) dans la table de contingence et on cherche alors à ce que ce nombre soit le plus petit possible.
- ➔ La valeur-p correspond en statistiques fréquentistes à un critère de décision. En effet, passé en-dessous d'un certain seuil fixé ici à 5% (donc p = 0,05) on parlera d'effet « significatif », pour dire que l'on a pu rejeter l'hypothèse selon laquelle, grossièrement, « nos variables n'expliquent pas nos données ». Cette hypothèse, on la note H_0 pour « hypothèse nulle ». Si elle est rejetée ($p < 0,05$) on acceptera alors son hypothèse alternative qui correspond grossièrement à « nos variables apportent une explication de nos données ». Mais attention, cette valeur-p qui estime la probabilité d'obtenir nos données sous l'hypothèse nulle n'implique pas que l'effet étudié soit fort, même si cette H_0 est rejetée via un $p < 0,001$! En effet la valeur-p est égale à $p(D \mid H_0)$, c'est-à-dire à la probabilité de nos données sous l'hypothèse nulle, et à rien d'autres.
- ➔ La taille d'effet : dans l'exemple ci-dessus nous rapportons un v de Cramér sous la forme d'un pourcentage qui doit s'interpréter comme à quel point le modèle H1 (qui prédit un effet de la variable que l'on teste) explique mieux nos résultats au questionnaire que le modèle H0 (qui émet toujours la même prédiction et reflète l'hypothèse d'absence d'effet de notre variable). Dans le cas du d de Cohen, rapporté pour les tests-t et leurs équivalents non paramétriques, s'il correspond lui aussi à une taille d'effet, il ne doit pas être interprété comme le pourcentage expliqué ci-dessus mais en regard de seuils qui ont été fixés par Jacob Cohen pour les sciences sociales. Dans le test-t, $d < 0.20$ indique un effet négligeable, $0.20 < d < 0.50$ indique un effet petit, $0.50 < d < 0.80$ indique un effet moyen, et $d > 0,80$ indique un effet fort. Le lecteur pourra se référer à la source mentionnée à la fin de ce document.

➔ Le facteur de Bayes. Cette valeur est un réel monstre des statistiques car c'est à la fois un critère de décision et une taille d'effet. Dans le cas des analyses statistiques les plus courantes en recherche, on dit qu'il est calculé à l'aide d'un a priori non informatif (voir le biais de l'oubli de la fréquence de base ci-dessous). Cet a priori est dit non informatif car on fait alors le choix d'analyser nos statistiques sans donner une crédence (importance, crédibilité) plus importante à H_0 ou à H_1 : on effectue nos statistiques sans préconcevoir le résultat, en d'autres termes nous sommes donc au départ dans une incertitude la plus totale. Or l'a priori pour ces deux hypothèses se décrit alors comme le rapport $p(H_1)/p(H_0)$, et il vaut donc 50%/50% = 1 (pour les puristes, nous appliquons ici le prior ultrawide de loi β(1 ; 1)). Le facteur de Bayes, lui, correspond un peu à une valeur-p augmentée. C'est le rapport $p(D \mid H_1)/p(D \mid H_0)$. Lui aussi sera donc incertain lorsque le rapport vaudra 1. En revanche, il peut permettre, au contraire de la valeur-p, de **tendre** à valider l'hypothèse nulle (quand la valeur-p ne peut que la rejeter) et cela car il effectue automatiquement une comparaison à la probabilité des données sous H_1. En statistiques, on dit que le facteur de Bayes rend les deux hypothèses symétriques, tandis que le monde fréquentiste ne fait que décider selon rejet ou non d'H_0. Mais mieux encore, dans le cas de la méthode statistique utilisée ici (avec un a priori non informatif valant 50%/50%) nous avons le gros avantage que l'a posteriori (le rapport $p(H_1 \mid D)/p(H_0 \mid D)$) vaut précisément le facteur de Bayes car :

$$\frac{p(H_1)}{p(H_0)} x \ \frac{p(D \mid H_1)}{p(D \mid H_0)} = \frac{p(H_1 \mid D)}{p(H_0 \mid D)}$$

Où :

- $p(H_1)$ désigne la probabilité qu'il existe un effet et $p(H_0)$ désigne la probabilité qu'il n'existe aucun effet de nos variables
- $p(D \mid H_1)$ correspond à la probabilité d'obtenir les données que l'on a obtenu lors du questionnaire s'il existe bien un effet de la variable et $p(H_1 \mid D)$ désigne la probabilité qu'il existe un effet de la variable en regard des données que l'on a obtenu, les deux sont à bien distinguer (idem pour les cas relatifs à H_0)
- le rapport $p(H_1)/p(H_0)$ désigne l'a priori (fixé à 1 pour nos analyses) ;
- le rapport $p(D \mid H_1)/p(D \mid H_0)$ correspond au facteur de Bayes ;
- $p(H_1 \mid D)/p(H_0 \mid D)$ désigne l'a posteriori, c'est-à-dire la croyance à avoir en chaque hypothèse après test.

Retenez qu'on peut dire que la croyance à avoir après les tests statistiques en H0 et H1 est donnée par le facteur de Bayes. Cette croyance en H1, en l'existence d'un effet, d'un lien entre nos variables, est ici en regard de nos données qui partent d'un a priori non informatif, neutre : à la prochaine expérience (ou réplication du questionnaire) on peut se servir du facteur de Bayes précédemment obtenu comme nouvel a priori (c'est ce que fait le courant du « bayésianisme subjectif »). De manière plus globale, il est adéquat d'établir notre adhésion plus ou moins importante à une hypothèse en prenant en compte toutes les études bien menées qui étudient l'élément qui nous intéresse, voire toutes les connaissances scientifiques cohérentes et incohérentes avec notre hypothèse.

- Nous pouvons donc avec le facteur de Bayes estimer la probabilité de nos deux hypothèses à la fois et confronter ces deux probabilités en un même et unique nombre. C'est une des raisons pour lesquelles on dit que le facteur de Bayes, en plus d'être un critère de décision, est une taille d'effet. Ainsi, un BF_{10} largement en dessous de 1 sera largement en faveur de l'hypothèse H_0 (c'est-à-dire l'absence d'effet de notre variable) tandis qu'un facteur de Bayes largement au-dessus de 1 sera largement en faveur de H_1 (c'est-à-dire l'existence d'un effet de notre variable, par exemple la variable familiarité à la zététique). On peut réinterpréter cela en mode « des paris » (e.g., faire un pari de 2 contre 1) en calculant la fraction associée à ce nombre, le numérateur correspondra alors à la croyance à accorder en H_1 et le dénominateur en la croyance à accorder à H_0. À noter ici que le terme de croyance est utilisé en regard du théorème (central pour la philosophie du bayésianisme) de Cox-Jaynes.
- Quant aux BF_{+0} rapportés, leurs principes ne changent pas, seule l'hypothèse H1 change de notation car elle n'est plus du type « l'apport de nos variables permet de mieux expliquer nos données » mais l'on précise par le « + » qu'il n'y a pas qu'une simple différence : nous donnons le sens positif de la différence (par exemple, une corrélation positive, ou un groupe 1 supérieur à un groupe 2). Il existe également des BF_{-0} qui là précisent un sens négatif dans notre jeu de données. Enfin, puisque le facteur de Bayes BF_{10} est un rapport, il est accepté de rapporter, dans la littérature, son inverse qui est le facteur de Bayes BF_{01}.

Ces facteurs de Bayes sont de plus en plus appréciés en sciences car ils facilitent une bonne pratique de la recherche en discutant les résultats en faveur de l'hypothèse nulle, ce qui évite la tentation de faire du « p-hacking » (voir l'article « Why most published research are false ») en statistiques fréquentistes. En tant que fraction le seuil du BF reste toujours le même contrairement à celui de la valeur-p (qu'on fixe le plus souvent 0,05 en sciences humaines), et le BF est moins sensible à la variation du nombre de sujets et à la multiplication des tests.

Référence :

Guyon, H. (2019). *L'inférence avec JASP (fréquentiste et bayésienne)*.
https://www.researchgate.net/publication/336284800_L'inference_avec_JASP_frequentiste_et_bayesienne