Lien entre test et intervalle de confiance.
Cas particulier d'une moyenne, pour des échantillons gaussiens ou de grande taille.

Claudine Schwartz

Le texte ci-dessous a pour objectif de répondre à une question souvent posée : pourquoi ne dit-on pas que deux paramètres sont significativement différents au risque α si et seulement si leurs intervalles de confiance au niveau 1-α sont disjoints ?

Pour éclairer cette question, nous nous plaçons dans le cas simple de deux échantillons indépendants, de lois de Gauss de même variance connue σ^2, dont on veut comparer les espérances μ et μ'. Les tailles des échantillons sont notées n et n', leurs moyennes m et m'.

Deux procédures sont envisagées ci-dessous pour savoir si les moyennes empiriques m et m' diffèrent significativement ou non, à un risque α qui par défaut est choisi égal à 0,05. Le terme « différence significative » veut dire qu'on ne considérera pas que les lois des échantillons ont même espérance, alors qu'on le fera si la différence n'est pas jugée significative.

Procédure 1 : test d'hypothèses

On choisit un risque de première espèce α ; ici α=0,05.

La région de rejet du test est l'ensemble W_1 des couples (m,m') qui vérifient :

$$|m - m'| > 1{,}96\sigma \times \sqrt{\left(\frac{1}{n} + \frac{1}{n'}\right)}$$

-Si (m,m') est dans W_1 , on dit qu'il y a entre m et m'une différence significative au risque 0,05. La probabilité de se tromper est α=0,05.

-Si (m,m') n'est pas dans W_1, on dit que la différence entre m et m' n'est pas significative ; on accepte l'égalité des espérances des lois sous jacentes. La probabilité de se tromper est une fonction de la différence des espérances, fonction appelée risque de deuxième espèce.

Remarque : Si on change le risque α, le seuil 1,96 change.

Cas particulier : n=n'. La région de rejet s'écrit :

$$W_1 = \left\{(m,m'),\ |m - m'| > 1{,}96\sigma \times \sqrt{\frac{2}{n}}\ \right\}$$

Les moyennes empiriques m et m' sont des réalisations des variables aléatoires $\overline{X}_n$ et $\overline{X}'_{n'}$. Le calcul de la région de rejet se déduit du calcul de l'intervalle de dispersion de la variable $\overline{X}_n$ - $\overline{X}'_{n'}$ qui suit une loi de Gauss d'écart type $\sigma \times \sqrt{\left(\frac{1}{n} + \frac{1}{n'}\right)}$, et d'espérance nulle sous l'hypothèse μ=μ'.

Remarque :

Si les lois des échantillons sont inconnues, de même variance connue, et si n et n' sont « grands », le test précédent s'applique car le théorème central limite permet d'approximer les lois de $\overline{X}_n$ et $\overline{X}'_{n'}$ par des lois de Gauss.

Si les variances sont inconnues, le calcul de la région de rejet doit être fait à partir du calcul de l'intervalle de dispersion d'une loi de Student.

Procédure 2 : intersection d'intervalles de confiance[1]

On détermine les intervalles de confiance de μ et μ' au niveau 1-α=0.95, soit :

.

$$IC_{0.95}(\mu)=[m\text{-}1{,}96\frac{\sigma}{\sqrt{n}}\,,\, m\text{-}1{,}96\frac{\sigma}{\sqrt{n}}] \quad \text{et} \quad IC_{0.95}(\mu')=[m'\text{-}1{,}96\frac{\sigma}{\sqrt{n}}\,,\, m'\text{-}1{,}96\frac{\sigma}{\sqrt{n}}]$$

Si ces intervalles sont d'intersection vide, on dit que m et m' diffèrent significativement. Sinon, on dit qu'ils ne diffèrent pas significativement.

Remarque : si on change de niveau de confiance, il faut changer le seuil 1,96.

Cette procédure découle d'une rationalité aisée à comprendre : l'intervalle de confiance détermine, au vu des données, des valeurs « admissibles » pour les espérances μ et μ' des lois qui modélisent ces données. Si les intervalles sont disjoints, cela signifie qu'il n'existe aucune valeur admissible commune pour ces espérances.
Cette procédure ne quantifie pas l'erreur de première espèce, celle de rejeter à tort l'égalité des espérances : c'est par là qu'elle se différencie de la première.
La région de rejet, qui conduit à rejeter l'égalité des espérances, et à dire que les moyennes sont significativement différentes, s'écrit ici :

$$W_2=\left\{(m,m'),\ |m-m'|>1{,}96\sigma\times\left(\sqrt{\frac{1}{n}}+\sqrt{\frac{1}{n'}}\right)\right\}.$$

Dans le cas ou $n=n'$: $W_2=\left\{(m,m'),\ |m-m'|>1{,}96\sigma\times\frac{2}{\sqrt{n}}\right\}$.

Comparaison

On peut écrire :

$$\sqrt{\left(\frac{1}{n}+\frac{1}{n'}\right)}<\sqrt{\frac{1}{n}}+\sqrt{\frac{1}{n'}}.$$

D'où : $W_2 \subset W_1$.

Par conséquent, si on conclue à une différence significative avec la procédure 2, on conclue aussi à une différence significative avec la méthode 1.

Par contre, on peut conclure que la différence est significative avec la procédure 1 et pas avec la procédure 2 ; c'est le cas si :

[1] L'INSEE, dans les publications pour le grand public fait le choix de cette procédure.

$$1{,}96\times\sigma\sqrt{\left(\frac{1}{n}+\frac{1}{n'}\right)}<|m-m'|\leq 1{,}96\times\sigma\sqrt{\frac{1}{n}}+\sqrt{\frac{1}{n'}}\ ,$$

Soit, pour $n=n'$: $1{,}96\times\sigma\sqrt{\frac{2}{n}}<|m-m'|\leq 1{,}96\times\sigma\times 2\times\sqrt{\frac{1}{n}}$.

La procédure 2 conduit à accepter plus facilement que la procédure 1 l'égalité des espérances, c'est-à-dire a parler plus souvent de différences non significatives.
*La première procédure est plus élaborée, au sens où elle quantifie précisément la probabilité de se tromper si on rejette l'égalité des espérances. Mais elle est plus difficile à comprendre***.**

Cas $n=n'$:
La région de rejet de la procédure 2 est identique à celle de procédure 1 calculée avec un autre risque de première espèce α'. Ce risque α' doit être tel que dans le calcul de la région de rejet de la procédure 1, le seuil 1,96 soit remplacé par $1{,}96\times\sqrt{2}\approx 2{,}78$. On trouve un risque environ dix fois plus petit :

$$\alpha' = 0{,}0056$$

La procédure d'intersection des intervalles de confiance, au niveau 0,95 (α=0,05) est identique à la procédure 1 pour un risque α'=0,0056.

.
On peut aussi calculer le niveau de confiance à prendre dans la seconde procédure pour qu'elle soit équivalente à la procédure 1 avec un risque de première espèce 0,05. On trouve qu'il faudrait dans cette situation prendre un niveau de confiance 0,84 (les intervalles de confiance à ce niveau de confiance sont les mêmes que précédemment, en remplaçant 1,96 par $1{,}96/\sqrt{2}$).

La procédure d'intersection des intervalles de confiance au niveau 0,84 (α=0,16) est identique à la procédure 1 pour un risque α'=0,05.

Notons qu'il serait peu naturel de déterminer des intervalles de confiance au niveau 0,84 sauf à parler de risque de première espèce ; si on en parle, alors il est plus pertinent d'utiliser directement la procédure 1… sauf à considérer que la pratique de la statistique est l'application de recettes ésotériques.

Les deux procédures relèvent de conventions : en ce sens on ne peut pas dire que l'une est juste et l'autre non. Dans les deux cas, il s'agit, avant d'avoir eu connaissance des données, de définir une règle dont l'application conduira, au vu des données, à accepter ou à rejeter l'égalité des espérances.
Quand on connaît n et n', on peut adapter les calculs précédents : la procédure 2, avec des intervalles au niveau de confiance $1-\alpha$, est toujours identique à la procédure 1 pour un risque $\alpha' < \alpha$.