Eric Depiereux, Benoît DeHertogh, Grégoire Vincke

www.fundp.ac.be/biostats **Module 100**

Eric Depiereux, Benoît DeHertogh, Grégoire Vincke

100 Distribution F de Fisher[1]

100.1 Utilité

La distribution de Fisher est à la base des modèles d'Analyse de la variance, technique qui a révolutionné les biostatistiques dans les années 1950.

Ces modèles permettent de puissantes comparaisons de moyennes par la comparaison d'une variance factorielle et d'une variance résiduelle.

100.2 Rapport de deux variances

Si deux variances S_1^2 et S_2^2 sont calculées sur des échantillons indépendants de taille n_1 et n_2, provenant de populations de même variance σ^2, le rapport S_1^2/S_2^2 est une v.a. F $n_1 -1 ;n_2 -1$ et S_2^2/S_1^2 est une v.a. F $n_2 -1 ;n_1 -1$. Cette variable est appelée distribution de Fisher.

On peut démontrer que la distribution de F est le rapport d'une $\chi 2$ avec n_1 -1 degrés de liberté et d'une $\chi 2$ avec n_2 -1 degrés de liberté.

Ce rapport est caractérisé par deux types de degrés de liberté. Par exemple le rapport S_1^2/S_2^2 est une distribution de F avec n_1 -1 degrés de liberté au numérateur[2] et n_2 -1 au dénominateur[3].

[1] Fisher, R.A. (1890 -1962) est un généticien anglais à qui l'on a attribué une avancée considérable dans la conciliation statistique des données de l'hérédité selon Mendel et la théorie de l'évolution de Darwin. Egalement statisticien, il révolutionna l'inférence en élaborant la théorie de l'analyse de la variance et de l'expérimentation factorielle.

[2] Numérateur : terme d'une fraction placé au -dessus de la barre horizontale et indiquant de combien de parties de l'unité se compose cette fraction.

[3] Dénominateur : diviseur, dans un quotient représenté par une fraction ; celui des deux termes d'une fraction qui indique en combien de parties l'unité a été divisée.

100.3 Tables et graphiques

On effectue le rapport de la plus grande variance sur la plus petite pour une question de facilité d'utilisation des tables.

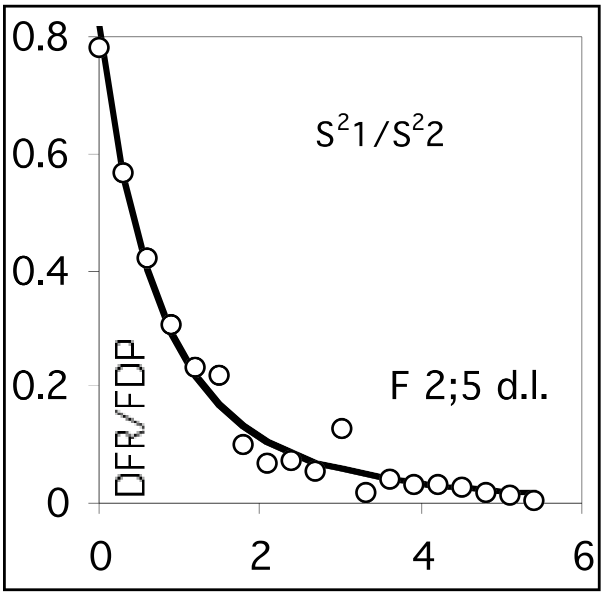

Figure 100 -1 Distribution de densité de fréquences relatives (DFR) observée du rapport S^2_{max}/S^2_{min} , S^2_{min} provenant d'un échantillon de 6 valeurs et S^2_{max} provenant d'un échantillon de 3 valeurs, superposée à la fonction de densité de probabilité (FDP) d'un v.a. $F_{2;5}$

Etant donné que la variable F est caractérisée par deux types de degrés de liberté, toute une table est nécessaire pour une seule valeur de probabilité.

0.95	1	2	3	4	5
1	161,4	199,5	215,7	224,6	230,2
2	18,5	19,0	19,2	19,2	**19,3**
3	10,1	9,6	9,3	9,1	9,01
4	7,7	6,9	6,6	6,4	6,26
5	6,6	**5,8**	5,4	5,2	5,05

Tableau 100 -100-1 Extrait d'une table éditée pour la probabilité 0,95. Les colonnes correspondent au nombre de d.l. du numérateur, les lignes au nombre de d.l. du dénominateur.

La table indique que la $P(F_{2;5} \leq 5,8) = 0,95$. Cela signifie qu'il est peu probable ($P < 0,05$) que le rapport de deux variances d'échantillons de 3 et 6 valeurs provenant de populations de même variance soit plus grand que 6.

La table indique que la $P(F_{5.2} \leq 19,3) = 0,95$. Cela signifie qu'il est probable (P > 0,95) que le rapport de deux variances d'échantillons de 3 et 6 valeurs provenant de populations de même variance soit inférieur à 19.

100.4 Fonctions Excel :

INVERSE.LOI.F(π;dl num; dl denom) retourne la valeur $F\pi$ = correspondant à la probabilité π, pour les degrés de liberté dl num; dl denom, sous la forme $P(F \geq F\pi) = \pi$.
LOI.F($F\pi$;dl num; dl denom) retourne la probabilité π , pour $F\pi$ correspondant aux degrés de liberté dl num; dl denom, sous la forme $P(F \geq F\pi) = \pi$.

Attention, le sens ≥ est contraire à la convention ≤ utilisée pour la plupart des autres fonctions de probabilité.

100.5 Exemples

A l'aide des tables : en dessous de quel rapport les variances de deux échantillons de 5 valeurs a -t -il 95% de chances de se situer ?

$F_{4\,;4\,;0.95} = 6.4$

Avec la fonction Excel : une pipette testée le mois dernier sur 100 pipetages a fourni une variance de 2μl^2.
Quelle est la probabilité qu'un test quelconque réalisé avec cette pipette sur 5 pipetages, fournisse une variance ≥ 6μl^2?

$P(F_{4\,;99} \geq 3) = 0,022$