Eric Depiereux, Benoît DeHertogh, Grégoire Vincke

www.fundp.ac.be/biostats **Module 80**

80 DISTRIBUTION T DE STUDENT 2

80.1 UTILITE 2
80.2 PRINCIPE 2
80.3 TABLES ET GRAPHIQUES 3
80.4 EXEMPLE 5

Eric Depiereux, Benoît DeHertogh, Grégoire Vincke

80 Distribution t de Student[1]

80.1 Utilité

La distribution de Student est utilisée dans des circonstances semblables à celles de la variable normale réduite, chaque fois que la variance σ^2 n'est pas connue et doit être remplacée par son estimation S^2.

Cet outil est donc extrêmement utile, car il permet de calculer des probabilités relatives à des populations dont tout est inconnu au départ, en tenant compte de la variabilité de l'estimation de Mx et de celle de S^2.

80.2 Principe

Imaginons que nous testions une nouvelle méthode de dosage du cholestérol (X, mg/l), de précision (σ) inconnue, dans une solution de cholestérol de moyenne 100, et que nous obtenions les valeurs 75, 95, 101, 149 mg/l.
Quelle est la probabilité que dans un échantillon de 4 valeurs, la moyenne calculée soit supérieure à 110 mg/l?
L'opération de standardisation vue par ailleurs

$$Z = \frac{Mx - \mu}{\sigma / \sqrt{n}}$$

n'est pas praticable ici car nous ne connaissons pas σ.

Calculons un écart -type à partir des valeurs disponibles S_{n-1} = 31,4 et utilisons -le pour effectuer la standardisation :

$$z \approx \frac{Mx - \mu}{S_{n-1} / \sqrt{n}} = t_{n-1} = \frac{110 - 100}{31.4 / 2} = 0.63$$

[1] *Student est le pseudonyme de W.S. Gosset (1876 -1937) Anglais ou Irlandais, travaillant chez Guinness à Dublin puis à Londres.*

La publication à laquelle on se réfère le plus souvent date de 1908 : On the probable error of the mean. Biometrika 6, 1 -25.

Bibiliographie : E.S. Pearson en : "Student" as statistician. Biometrika 30, 210 -250. 1939
J.F. Box. Gosset, Fisher, and the t distribution. The American Statistician 35, 61 -66,1981.
J.F. Box. Guinness, Gosset, Fisher, and small samples. Statistical Science 2, 45 -52, 1987.

Nous obtenons ainsi une approximation de Z, dont la variabilité dépend de la v.a. Mx et de la v.a. S. Cette variable s'appelle la variable t de Student.

La valeur obtenue t_{n-1} sera d'autant plus proche de la valeur z correspondante que l'estimation S_{n-1} sera proche de la valeur réelle σ.
La valeur de S_{n-1} sera d'autant plus proche de la valeur réelle σ que n est élevé.

La variable t est caractérisée par n -1 degrés de liberté : plus le nombre de d.l. est élevé, plus la v.a . t de Student tend vers une v .a. Z(0 ;1).

80.3 Tables et graphiques

La dernière ligne de la table de t correspond à un nombre infini de degrés de liberté (σ^2 connu sans erreur). Elle représente un extrait de la table de Z.

La table de *t* est généralement présentée de la façon suivante :

	0,9	**0,95**	**0,975**	**0,99**
1	3,08	6,31	12,71	31,82
2	1,89	2,92	4,30	6,96
3	1,64	2,35	**3,18**	4,54
4	1,53	2,13	2,78	3,75
5	1,48	2,02	2,57	3,36
…	…	…	…	…
∞	1,28	1,64	1,96	2,32

Tableau 80 -80-1 Extrait de la table de t. La première ligne énumère des probabilités, la première colonne, des degrés de liberté. Chaque ligne comprend $P(t \leq t_\pi)$, π étant la probabilité reprise en tête de colonne.

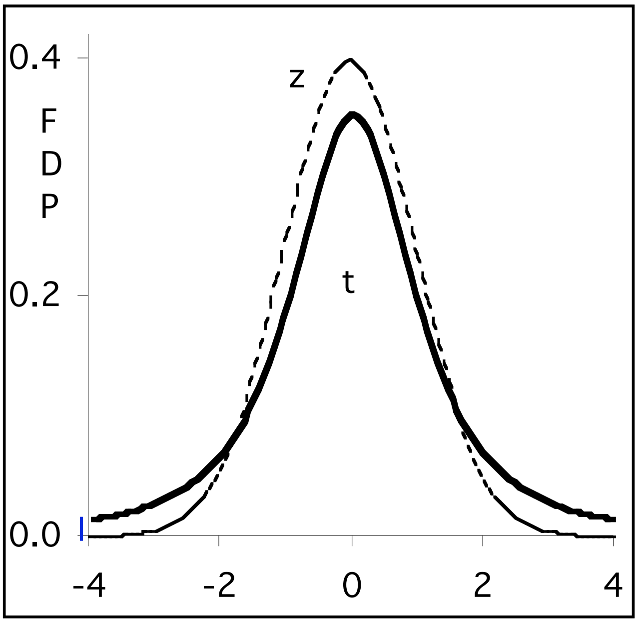

Figure 80 -1 Distribution de t avec 2 d.l. (trait continu) superposée à celle de Z(0 ;1) (trait pointillé). Observez que la distribution de t est nettement plus large à la base.

La distribution de t est symétrique et centrée sur zéro ce qui implique que
$t_\pi = - t(1 - \pi)$: $t_{3\,;0.975} = 3{,}18 \Leftrightarrow t_{3\,;0{,}025} = -3{,}18$

La distribution de t est nettement plus large à la base, ce qui signifie que pour une petite probabilité α, $t\,\alpha > z\,\alpha$. Par exemple : $t_{3\,;0{,}975} = 3{,}18$, $Z_{0{,}975} = 1{,}96$.

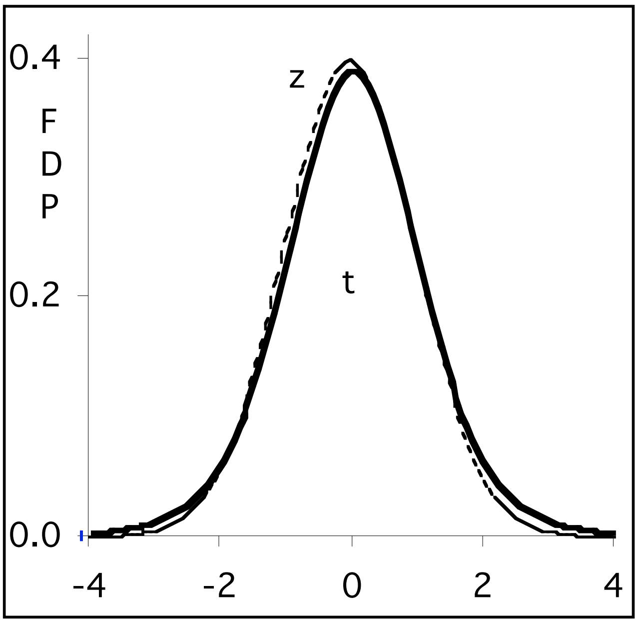

Figure 80 -2 Distribution de t avec 10 d.l. (trait continu) superposée à celle de Z(0 ;1) (trait pointillé). Observez que les distributions sont presque superposées.

La distinction entre Z et t est très importante lorsque l'on utilise de petits échantillons.

80.4 Exemple

Une espèce de spirifères a une longueur moyenne de 4cm.
Un échantillon de 4 individus permet de calculer un écart -type de 2cm.

Entre quelles valeurs a -t -on 95% de chances de trouver la moyenne d'un échantillon quelconque de 4 individus ?

$P(t_{3\,;0,025} \leq t_3 \leq t_{3;0,975}) = 0,95$
$P(-t_{3\,;0,975} \leq t_3 \leq t_{3;0,975}) = 0,95$
$P(-3,18 \leq t_3 \leq 3,18) = 0,95$

$$\frac{Mx-\mu}{\sqrt{S^2/n}} = t_{n-1} \Leftrightarrow \frac{Mx_{0.975}-4}{\sqrt{4/4}} = 3.18$$

$$Mx_{0.975} = 3.18 \times 1 + 4 = 7.18$$

$$Mx_{0.025} = -3.18 \times 1 + 4 = 0.82$$