Test de Student

Exemple d’application à l’enseignement

charlene.meyers@umons.ac.be
Service de traduction spécialisée et terminologie

Test de Student / T-test

But: Vérifier si 2 échantillons diffèrent l'un de l'autre.

Exemple 1:

Vérifier une différence de résultats entre des étudiants d'une groupe A et des étudiants d'un groupe B à un même test.

Exemple 2:

Vérifier le potentiel progrès des étudiants entre les résultats d'un test en septembre et les résultats du même test passé par les mêmes étudiants en juin.

2 types de test de Student

- Pour **échantillons indépendants** (two-tailed or independent test) : les groupes A et B sont constitués d'individus différents mais qui sont tous soumis au même test. Le chercheur veut démontrer une différence entre un groupe étudié et un groupe contrôle par exemple.

Exemple 1:

Vérifier une différence de résultats entre des étudiants d'un groupe A et des étudiants d'un groupe B à un même test.

2 types de test de Student

- Pour **échantillons appariés** (one-tailed or dependent test) : les groupes A et B sont constitués des mêmes individus mais ils ont pris part à un même test deux fois. Le chercheur veut démontrer s'il y a eu un changement entre le premier et le deuxième test.

Exemple 2:

Vérifier le potentiel progrès des étudiants entre les résultats d'un test en septembre et les résultats du même test passé par les mêmes étudiants en juin.

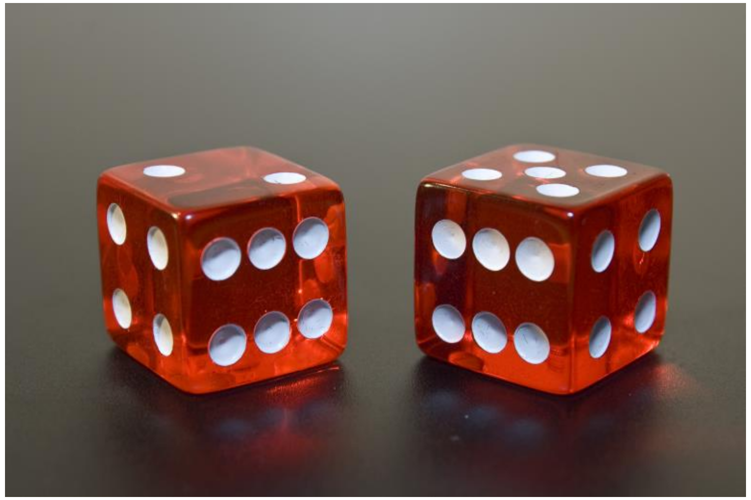
Conditions d’application

Dans le cas d’un test apparié :

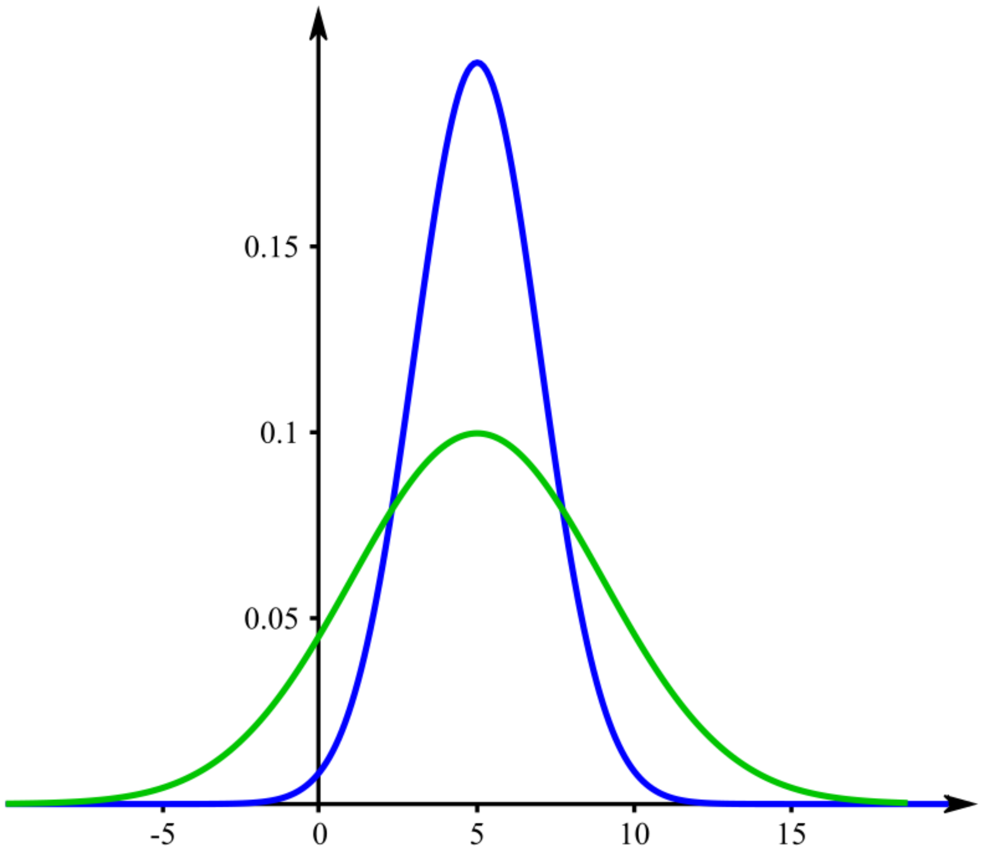
- Les données doivent suivre une distribution normale (courbe en cloche)
- Si la taille des échantillons est suffisamment grande ($n > 30$), il n’est pas nécessaire de vérifier si la loi normale est respectée.

Pourquoi appliquer un test de Student quand on peut facilement calculer la moyenne des deux échantillons et les comparer ?

Car la moyenne ne tient pas compte de la *dispersion* (variabilité) des données…



Par exemple, il est tout à fait plausible que la moyenne des résultats à d’un groupe A et d’un groupe B soit la même mais que les résultats individuels soient très **variés** au sein d’un groupe :



Le test de Student permet de comparer les moyennes des deux groupes tout en prenant en compte la **variabilité** des données et de conclure sur la significativité de cette différence :

> Roughly speaking, the t-test is a statistical procedure that compares the arithmetic means of two groups of data while taking their variability (i.e. their standard deviation or variance) into account. Hence, it allows us to draw conclusions whether or not there is a real difference between two groups. (Rasinger 2013: 192)

Statistiques inférentielles >< Statistiques descriptives

Calculer une moyenne, une médiane → Statistiques descriptives

Appliquer le test de Student, revient à vérifier une hypothèse → statistiques inférentielles

la formulation d'une **hypothèse** s'articule toujours sous la forme de deux propositions l'une annulant l'autre. L'**hypothèse alternative** est l'hypothèse selon laquelle une différence est perceptible. C'est l'hypothèse que le chercheur espère démontrer.

L'**hypothèse nulle**, selon laquelle aucune différence significative ne se manifeste entre les deux échantillons.

- **Hypothèse nulle** (H0) : il n’y a pas de différence *significative* entre les moyennes des deux groupes d’échantillons.
- **Hypothèse alternative** (H1) : il existe une différence *significative* entre les moyennes les deux groupes d’échantillons.

Vu qu’on s’attend à un amélioration des résultats entre septembre et juin, on peut formuler l’hypothèse suivante:

- **Hypothèse alternative** (H1): $\overline{X_A}$ (septembre) < $\overline{\mathrm{X_B}}$ (Juin)

L’hypothèse est vérifiée grâce à un test statistique (ici, test de Student). Les tests permettant de vérifier des hypothèses ont pour but de calculer la **valeur *p*.**

Cette valeur *p* permet de soit valider l’hypothèse alternative et de rejeter l’hypothèse nulle. Soit de valider l’hypothèse nulle et rejeter l’hypothèse alternative.

En règle générale, l’interprétation de la valeur *p* est la suivante:

Quand p < 0,05 l’hypothèse nulle est rejetée et l’hypothèse alternative est retenue.

Il ne faut pas se méprendre sur l'interprétation de la valeur *p.* La valeur *p* ne mesure pas *à quel point les résultats sont différents* mais bien *la confiance avec laquelle on peut affirmer que les résultats obtenus ne sont pas dus au hasard,* c'est ce que l'on appelle la significativité :

> Statistical significance is based on probability theory: how likely is something to happen or not. Statistical significance is denoted with *p*, with *p* fluctuating between zero and 1 inclusive, translating into zero to 100 per cent. The smaller *p*, the less likely our result is to be a fluke. (Rasinger 2013: 172)

la mesure de la différence entre deux groupes est exprimée par un indicateur de **dispersion** (ici, la **valeur *t*** – d'autres indicateurs de dispersion sont l'écart-type, la variance, l'écart interquartile, etc.):

The t score is a ratio between the **difference between two groups and the difference within the groups**. The larger the t score, the more difference there is between groups. The smaller the t score, the more similarity there is between groups. A t score of 3 means that the groups are three times as different *from* each other as they are within each other. When you run a t test, the bigger the t-value, the more likely it is that the results are repeatable.

- A large t-score tells you that the groups are different.
- A small t-score tells you that the groups are similar. (Aankul 2017)

Tout comme l'écart-type, il est difficile d'interpréter la valeur *t* isolément.

Règle empirique: Quand la valeur $t > 2$, la différence est généralement significative (si l'on ne dispose pas de la valeur *p* pour attester de la significativité).

Solution la plus simple → viser à créer une échelle ordinale

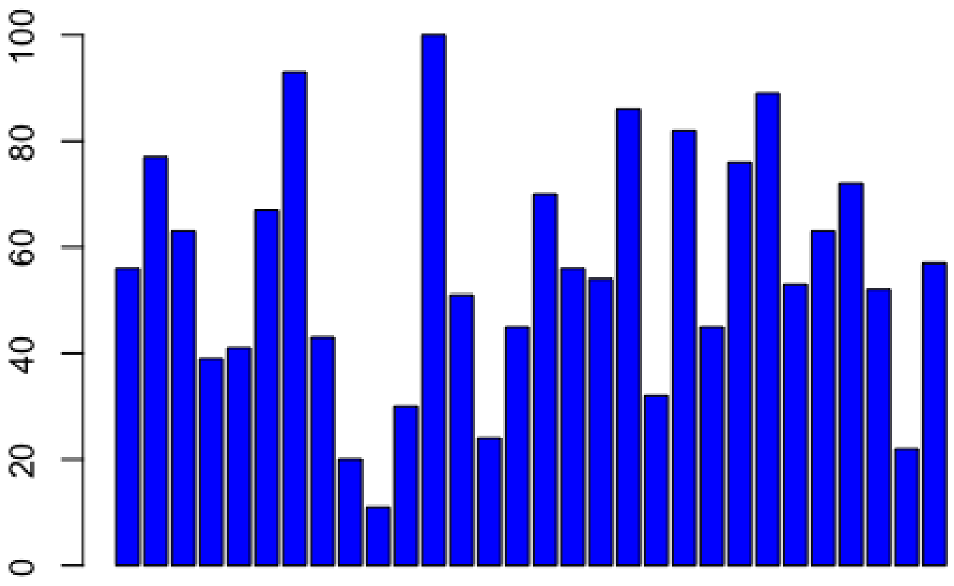
Exemple de comparaison d’une mesure de la variance (écart-type / standard deviation)

Mesure des interférences / 100 mots parmi différents groupes d’interprètes:

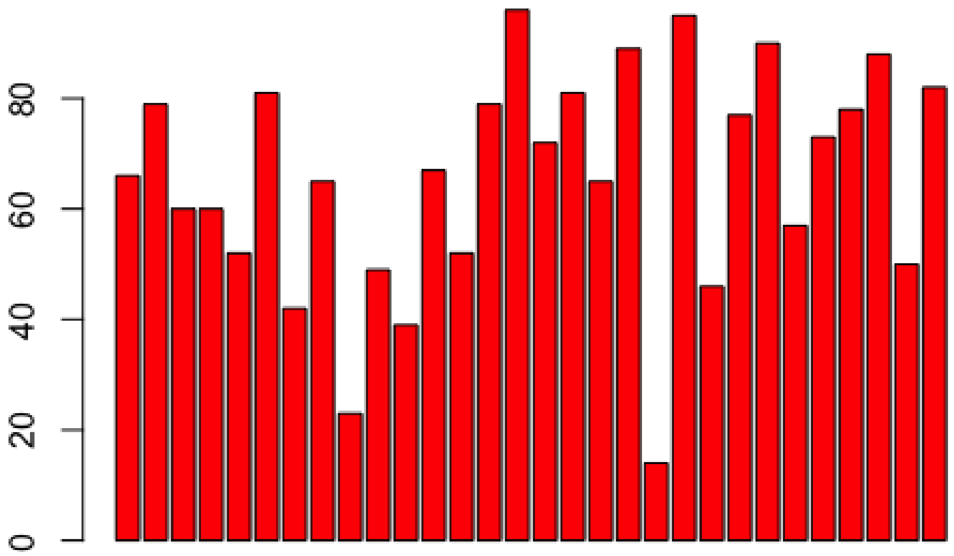
For interpreters working without text (O), standard deviation (SD) was +- 0.76 for an average of 2.29 INTs per 100 words/ST, for group PT SD was +- 0.75 for an average of 2.72 INTs per 100 words/ST, and for interpreters working with an unprepared text (T) SD was as high as +- 1.39 for an average of 2.28 INTs / 100 words /ST. (Lamberger-Felber & Schneider 2008: 223)

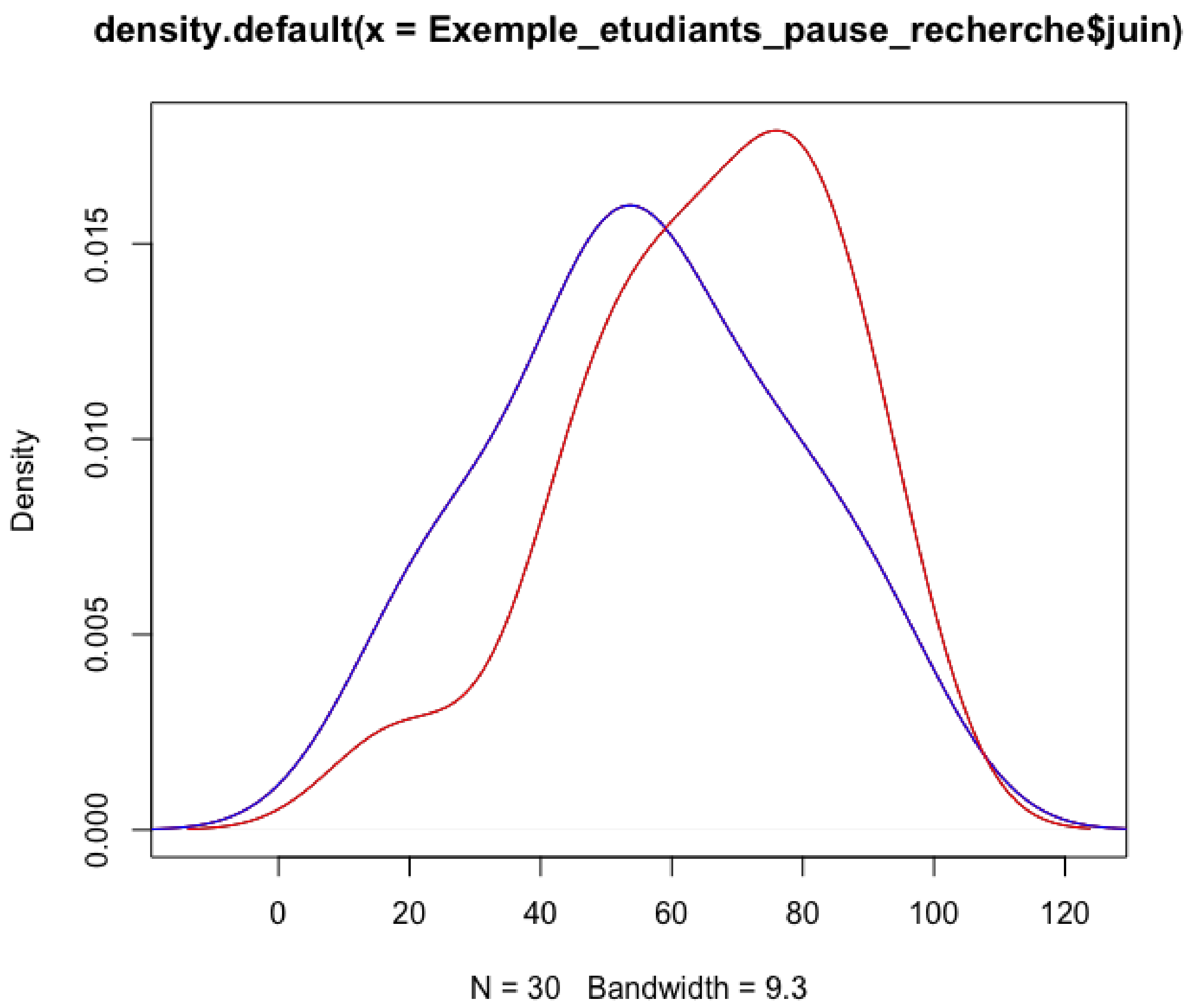
Méthode de travail	Écart-type
Sans texte	+- 0.76
Avec texte préparé	+- 0.75
Avec texte mais non préparé	+- 1.39

Histogramme résultats septembre



Histogramme résultats juin



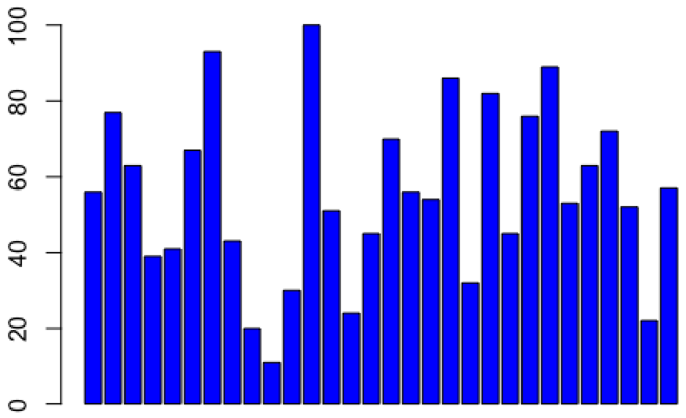

density.default(x = Exemple_etudiants_pause_recherche$juin)
Density
0.000
0.005
0.010
0.015
0
20
40
60
80
100
120
N = 30   Bandwidth = 9.3

Dans la pratique

- R Studio
- Import dataset → « Exemple_etudiants_pause_recherche »
- Résultats fictifs de 30 étudiants à un test (/100) de septembre et au même test en juin.
- Console : sert à interroger des données, coder des tests statistiques et générer des graphiques selon le langage informatique de R.

Visualiser ses données : histogramme

barplot(height= Exemple_etudiants_pause_recherche$septembre, col = "blue")

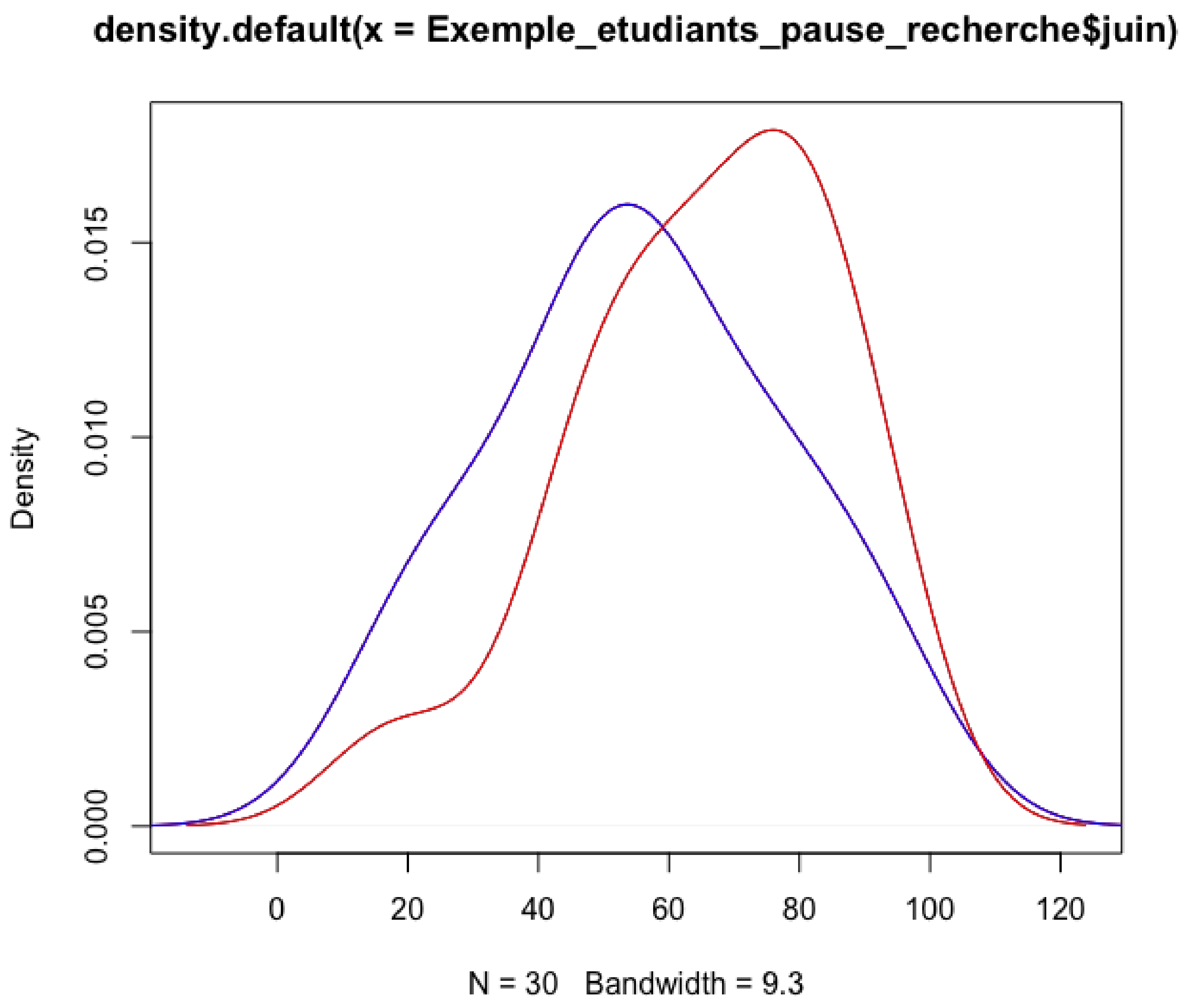


Visualiser ses données : courbe de densité

plot(density(x = Exemple_etudiants_pause_recherche$juin))

lines(density(x = Exemple_etudiants_pause_recherche$septembre), col = “blue”))

lines(density (x = Exemple_etudiants_pause_recherche$juin), col = “red”))


density.default(x = Exemple_etudiants_pause_recherche$juin)
Density
0.000
0.005
0.010
0.015
0
20
40
60
80
100
120
N = 30   Bandwidth = 9.3

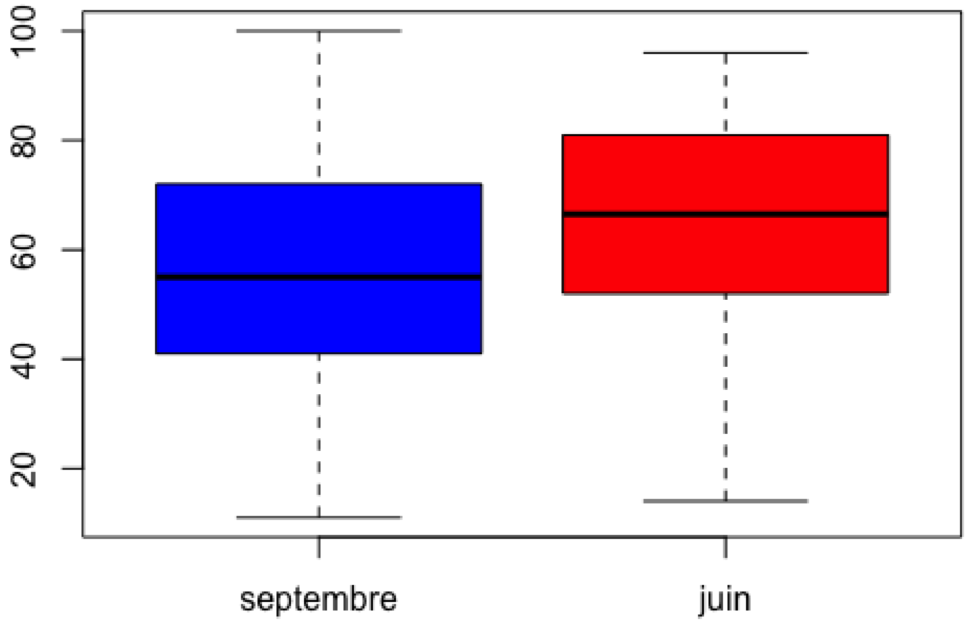
Décrire ses données: summary

```
summary(Exemple_etudiants_pause_recherche$septembre)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  11.00   41.50   55.00   55.63   71.50  100.00 


summary(Exemple_etudiants_pause_recherche$juin)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  14.00   52.00   66.50   65.57   80.50   96.00 
```

Décrire ses données: la boite à moustache

```
boxplot(x = Exemple_etudiants_pause_recherche, col=
c("blue","red"))
```



Tester ses données : test de Student

```
t.test(Exemple_etudiants_pause_recherche$septembre,
Exemple_etudiants_pause_recherche$juin, paired= TRUE, alternative =
"less")
```

Les deux groupes $septembre, $juin

Paired = TRUE car les deux groupes sont **appariés**

Alternative = « less » car cela signifie qu'on s'attend à ce que les résultats du premier groupe ($septembre) soient plus faibles que le deuxième groupe ($juin)

Tester ses données : test de Student

t = -2.535, df = 29, p-value = 0.008448

alternative hypothesis: true difference in means is less than 0

95 percent confidence interval:

-Inf -3.275472

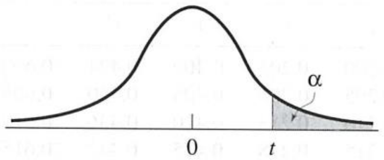
sample estimates:

mean of the differences

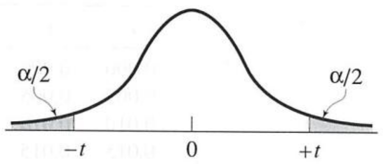
-9.933333

$p < 0.05$ → il y a une différence significative entre les résultats de septembre et de juin (t = -2.535)

Table t : points de pourcentage supérieurs de la distribution t



Test unilatéral



Test bilatéral

	Seuil de signification pour le test unilatéral								
	.25	.20	.15	.10	.05	.025	.01	.005	.0005
	Seuil de signification pour le test bilatéral								
dl	.50	.40	.30	.20	.10	.05	.02	.01	.001
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	636.620
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.599
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	3.496
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.390
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291

Bibliographie

Aankul, A.(2017). *T-test using python and Numpy.* *https://towardsdatascience.com/inferential-statistics-series-t-test-using-numpy-2718f8f9bf2f*

Lamberger-Fleber, H., & Schneider, J. (2008). Linguistic interference in simultaneous interpreting with text : A case study. In D. Gile, G. Hansen, A. Chesterman, & H. Gerzymisch-Arbogast (Éd.), *Efforts and models in interpreting and translation research : A tribute to Daniel Gile*. John Benjamins.

Levshina, N. (2015). *How to do linguistics with R : Data exploration and statistical analysis*. John Benjamins Publishing Company.

R foundation for Statistical Computing. (s. d.). *R: A language and environment for statistical computing*. http://www.R-project.org/

Rasinger, S. M. (2013). *Quantitative research in linguistics : An introduction* (Second Edition). Bloomsbury.

Xiao, R. (2013). *Making Statistic Claims*. Power Point presentation available online.